Convergence of probability measures
MASTER 2 MATH
Université Paris Dauphine

2024 - 2025

François Simenhaus*
Bureau B 640
simenhaus@ceremade.dauphine.fr

*Updated by Julien Poisat in 2023

Contents

The first part of the course (5*3 hours) is devoted to the study of convergence of probability measures on general (that is not necessarily $\mathbb{R}$ or $\mathbb{R}^n$) metric spaces or, equivalently, to the convergence in law of random variables taking values in general metric spaces. If this study has its own interest it is also useful to prove convergence of random objects in various random models that appear in probability theory. The main example we have to keep in mind is Donsker theorem that states that the path of a simple random walk on $\mathbb{Z}$ converges after proper renormalization to a Brownian motion. We will start this course with some properties of probability measures on metric spaces and in particular on $\mathscr{C}([0,1])$, the space of real continuous function on $[0,1]$. We will then study convergence of probability measures, having for aim Prohorov theorem that provides a useful characterization of relative compatctness via *tightness*. Finally we will gather everything to study convergence in law on $\mathscr{C}([0,1])$ and prove Donsker therorem. If there is still time we will consider other examples of convergence of random objects.

The main reference for this course is Billingsley [1]. I also used to write these notes many courses especially those by Jean Bertoin, Gilles Pages, Gregory Miermont and Zhan Shi. You should find easily corresponding lecture notes online. I have not been rigorous to cite them precisely each time I used their notes. I will try to fix this rapidly in future versions.

Code Teams : iov64dw

1 Introduction

A first but central example : *Donsker invariance principle.* For simplicity we consider $(\xi_k)_{k\geq 1}$ an i.i.d. family of random variables with value -1 or 1 with probability $1/2$. The sequence defined for $n \geq 1$ by

$$S_n = \sum_{k=1}^{n} \xi_k,$$

is the simple random walk on $\mathbb{Z}$. For large n, the central limit theorem tells us that S_n correctly renormalised, that is by $\sqrt{n}$, has law "close" from a gaussian $\mathcal{N}(0,1)$:

$$\frac{S_n}{\sqrt{n}} \overset{(law)}{\to} \mathcal{N}(0,1).$$

More generally, it is not difficult to derive from this result the convergence of the finite dimensional distributions: for all $k \geq 1$ and all $0 \leq t_0 \leq \cdots \leq t_k \leq 1$,

$$(\frac{S_{\lfloor t_0 N \rfloor}}{\sqrt{n}}, \cdots, \frac{S_{\lfloor t_k N \rfloor}}{\sqrt{N}}) \overset{(law)}{\to} (B_{t_0}, \cdots, B_{t_k}),$$

when N goes to $+\infty$ where $(B_t)_{0\leq t\leq 1}$ is a Brownian motion. If you do not know yet what is a Brownian motion, just consider that the random vector $(B_{t_0}, \cdots, B_{t_k})$ is a centred gaussian vector with variance $\mathrm{E}(B_{t_i} B_{t_j}) = t_i \wedge t_j$. Equivalently (exercise !), the increments variables $B_{t_0}, B_{t_1} - B_{t_0}, \cdots, B_{t_k} - B_{t_{k-1}}$ are gaussian centred independent with variance $\mathrm{Var}(B_{t_i} - B_{t_{i-1}}) = t_i - t_{i-1}$. Of course this implies that for all $0 \leq i \leq k$, B_{t_i} has law $\mathcal{N}(0, t_i)$.

We now want to go a step further and **describe the asymptotic law of the whole random function**, that is for large N the process,

$$S_t^{(N)} = \frac{1}{\sqrt{N}} S_{\lfloor Nt \rfloor} + (Nt - \lfloor Nt \rfloor) \frac{1}{\sqrt{N}} \xi_{\lfloor Nt \rfloor + 1} \qquad 0 \leq t \leq 1.$$

The process $(S_t^{(N)})_{0\leq t\leq 1}$ is just the renormalised path of the random walk (that is the linear interpolation of the S_n, $n \geq 1$). Our goal is to derive the convergence in law for this sequence of processes, when N goes to infinity, to the Wiener law that is the law of the Brownian motion. This is the aim of Donsker theorem that we will proof in the last section of this course:

Theorem 1 (Donsker Theorem)**.** *Assume that $(\xi_k)_{k\geq 1}$ is an i.i.d. family of square integrable real random variables with mean 0 and variance 1. Then*

$$(S_t^{(N)})_{0\leq t\leq 1} \overset{(law)}{\to} (B_t)_{0\leq t\leq 1} \quad \text{as } n \to +\infty,$$

where $(B_t)_{0\leq t\leq 1}$ is a standard Brownian motion and the convergence is relative to the uniform topology on $\mathscr{C}([0,1])$.

On our way to prove this theorem we will have to study probability measures on metric spaces and characterize weak convergence. We will finally apply our study to the space $(\mathscr{C}([0,1]), ||\cdot||_\infty)$ of continuous fonctions on $[0,1]$ endowed with the uniform topology.

You may see during the second semester similar study in the more delicate context of càdlàg functions space (see Julien Poisat's course about *Levy processes*), or also more convergence to nice continuum probabilist objects (see Jan Swart's course about *Brownian continuum objects*). I hope this course give you a good preparation for these forthcoming lectures !

2 Probabilty measures on metric spaces

2.1 Regularity and thigtness

The goal of this first section is to provide a description of probability measures on metric spaces. In the whole section we consider a metric space (E, d). We remind that a metric d is an application

$$d : E \times E \to \mathbb{R}^+$$

that satisfies for all $x, y, z \in E$

1. *Symmetry* : $d(x, y) = d(y, x)$
2. *Identity of indiscernibles* : $d(x, y) = 0 \Leftrightarrow \ x = y$
3. *Triangle inequality* : $d(x, z) \leq d(x, y) + d(y, z)$.

A *topology* on E is a collection $\mathcal{T}$ of subsets of E that satisfies

1. $\emptyset \in \mathcal{T}$ and $E \in \mathcal{T}$
2. Any union of elements of $\mathcal{T}$ is still an element of $\mathcal{T}$
3. Any intersection of a finite numbers of elements of $\mathcal{T}$ is still an element of $\mathcal{T}$.

In this course we will focus on the case where the topology is produced by the metric d. Any metric d defines indeed a topology $\mathcal{T}$ that is the collection of sets O satisfying,

$$\text{for all } x \in O \text{ there exists } \varepsilon > 0 \text{ such that } B(x, \varepsilon) \subset O,$$

where

$$B(x, \varepsilon) = \{y \in E, \ d(x, y) < \varepsilon\}$$

is the open ball centered in x and with radius ε. As usual elements of $\mathcal{T}$ are called *open sets*.

Exercise 1. *Check that the collection of subsets that satisfy the above property is indeed a topology.*

We denote by $\mathcal{F}$ the sigma-algebra generated by the topology $\mathcal{T}$,

$$\mathcal{F} = \sigma(O, \ O \in \mathcal{T}).$$

Often this sigma-algebra is called the *Borel sigma-algebra.* The space $(E, \mathcal{F})$ is our general framework for this course. This is a nice, and sometimes certainly difficult, chalenge to check what remains true if we work with a topological space instead of a metric space (still considering the Borel sigma-algebra).

We first prove that probability measures on $(E, \mathcal{F})$ are *regular* in the sense that the probability of any event A can be approximated correctly by the probability of open, closed or sometimes even compact sets.

Theorem 2. *Every probability measure* P *on* $(E, \mathcal{F})$ *is* ***regular****, that is: for every* $A \in \mathcal{F}$ *and every* $\varepsilon > 0$ *there exist a closed set* F *and an open one* O *such that*

$$F \subset A \subset O \quad \textit{and} \quad \mathrm{P}(O \setminus F) < \varepsilon.$$

Remark 1. *One can easily check that an equivalent definition for "*P *is regular" is: For all* $A \in \mathcal{F}$

$$\begin{aligned}\mathrm{P}(A) &= \sup\{\mathrm{P}(F), F \textit{ closed set included in } A\} \\ &= \inf\{\mathrm{P}(O), O \textit{ open set that contains } A\}.\end{aligned}$$

Proof of Theorem 2. Consider the collection $\mathcal{G}$ of sets $A \in \mathcal{F}$ that satisfy the property : *For every* $\varepsilon > 0$ *there exists a closed set* F *and an open one* O *such that*

$$F \subset A \subset O \quad \textit{and} \quad \mathrm{P}(O \setminus F) < \varepsilon.$$

The collection $\mathcal{G}$ is a sigma algebra :

1. $\emptyset \in \mathcal{G}$. This is trivial as $\emptyset$ is open and close.

2. If $A \in \mathcal{G}$ then $A^c \in \mathcal{G}$. Fix $\varepsilon > 0$ and a closed set F and an open one O such that $F \subset A \subset O$ and $\mathrm{P}(O \setminus F) < \varepsilon$. Then O^c is closed, F^c is open, $O^c \subset A^c \subset F^c$ and $\mathrm{P}(F^c \setminus O^c) = \mathrm{P}(O \setminus F) < \varepsilon$.

3. If $(A_n)_{n\geq 1}$ is a sequence in $\mathcal{G}$ then $\cup_{n\geq 1} A_n \in \mathcal{G}$. For all $n \geq 1$, we consider a closed set F_n and an open one O_n such that $F_n \subset A_n \subset O_n$ and $\mathrm{P}(O_n \setminus F_n) < \frac{\varepsilon}{2^{n+1}}$. Set $O = \cup_{n\geq 1} O_n$ and $F = \cup_{1\leq n\leq n_0} F_n$ where n_0 is large enough so that $\mathrm{P}(\cup_{n\geq 1} F_n \setminus F) < \varepsilon/2$. We obtain $F \subset \cup_{n\geq 1} A_n \subset O$ and

$$\mathrm{P}(O \setminus F) \leq \mathrm{P}(O \setminus (\cup_{n\geq 1} F_n)) + \mathrm{P}((\cup_{n\geq 1} F_n) \setminus F) \leq \sum_{n\geq 1} \mathrm{P}(O_n \setminus F_n) + \varepsilon/2 \leq \varepsilon,$$

as $\cup_{n\geq 1} O_n \setminus \cup_{n\geq 1} F_n \subset \cup_{n\geq 1} (O_n \setminus F_n)$.

Moreover $\mathcal{T} \subset \mathcal{G}$. Indeed, fix A a closed set and $\varepsilon > 0$. One defines for $\delta > 0$, the open δ-neighborhood of A,

$$A^\delta = \{x \in E,\ d(x, A) < \delta\}.$$

As A is closed $\cap_{\delta>0} A^\delta = A$ and one can choose δ small enough so that $\mathrm{P}(A^\delta) < \mathrm{P}(A) + \varepsilon$ (see Exercise 5 for basic properties of the distance to some subset of E). We set $F = A$ and $O = A^\delta$ so that $F \subset A \subset O$ and $\mathrm{P}(O \setminus F) < \varepsilon$.

Finally $\mathcal{G}$ is a sigma algebra and contains $\mathcal{T}$ thus it contains $\mathcal{F}$.

□

Exercise 2. *Prove that in general the role of "open" and "closed" can not be reversed in the definition of regular. When is it the case ?*

We have proven that on a metric space, any probability P on the borel sigma algebra is regular. This implies that it is **completely determined by its values on closed sets or, on open sets.**

Exercise 3. *Is it still the case on a topological space (not necessarily induced by a metric) ?*

Another usefull way to characterise probability measures on metric spaces is to make use of integrals for fucntions that live in a a large enough class of functions:

Proposition 1. *Let* P *and* Q *be two probability measures on* $(E, \mathcal{F})$. *They coincide if and only if* ***for all bounded and uniformly continuous real functions*** f***,*** $\mathrm{E}_\mathrm{P}(f) = \mathrm{E}_\mathrm{Q}(f)$***.***

Remark 2. *One can safely replace "uniformly continuous" by "Lipschitz" in the above result.*

Proof. The key point is to approximate the indicator of a closed set by a sequence of bounded and uniformly continuous real functions.

Lemma 1. 1. *Let F be a closed set. The sequence of Lipschitz bounded functions*

$$x \in E \to (1 - kd(x, F))_+ \qquad k \geq 1$$

decreases to 1_F when k goes to infinity.

2. *Let O be an open set. The sequence of Lipschitz bounded functions*

$$x \in E \to \min(kd(x, O^c), 1), \qquad k \geq 1$$

increases to 1_O when k goes to infinity.

For all $\varepsilon > 0$, choosing k large enough, this proves that there exists a Lipschitz bounded function f that approximates 1_F in the sense that

$$1_F \leq f \leq 1_{F^\varepsilon}, \tag{1}$$

where F^ε is the ε-open neighbourhood of F, that is $F^\varepsilon = \{x \in E,\ d(x, F) < \varepsilon\}$. This implies that

$$\mathrm{P}(F) \leq \mathrm{E}_\mathrm{P}(f) = \mathrm{E}_\mathrm{Q}(f) \leq \mathrm{Q}(F^\varepsilon).$$

As F is closed $F = \cap_{\varepsilon>0} \downarrow F^\varepsilon$ and we obtain $\lim_{\varepsilon\to 0} \mathrm{Q}(F^\varepsilon) = \mathrm{Q}(F)$. It leads to $\mathrm{P}(F) \leq \mathrm{Q}(F)$ and by symmetry this concludes the proof.

□

A key notion in the following is the notion of **tightness** :

Définition 1. *A probability measure* P *on* $(E, \mathcal{F})$ *is said to be tight if for all* $\varepsilon > 0$ *the exists a compact set* K *such that* $\mathrm{P}(K) > 1 - \varepsilon$*.*

We remind that a **Polish space** is a metrisable complete separable topological space. This definition is useful as we will see throughout the course that Polish spaces are the good framework for many properties in probability theory. Here is a first example:

Theorem 3 (Ulam Theorem)**.** *We suppose that* (E, d) *is a Polish space. Then any probability measure on* $(E, \mathcal{F})$ *is tight.*

Proof. We consider a dense countable family $(x_n)_{n\geq 1}$. For all $k \geq 1$, $E = \cup_n \overline{B(x_n, 1/k)}$ so that there exists n_k such that $\mathrm{P}(\cup_{n\leq n_k} \overline{B(x_n, 1/k)}) \geq 1 - \varepsilon/2^k$. We consider the set

$$K = \bigcap_{k\geq 1} \bigcup_{n\leq n_k} \overline{B(x_n, 1/k)}.$$

The set K is closed in the complete space E so that it is complete. Moreover K is clearly totally bounded (that is for all $\varepsilon > 0$ one can cover K with a finite union of ball of radius ε). The set K is thus compact (see Exercise 6) and

$$\mathrm{P}(K^c) \leq \sum_{k\geq 1} \mathrm{P}\left(\left(\bigcup_{n\leq n_k} \overline{B(x_n, 1/k)}\right)^c\right) \leq \varepsilon.$$

□

Exercise 4 (An example of a non tight probability measure)**.** *Consider* $\mathcal{T}$ *the* lower limit topology *on* $\mathbb{R}$ *that is the topology generated by the basis of all half-open intervals* $[a, b)$ *where* a *and* b *are real numbers.*

1. *Prove that $\mathcal{T}$ is finer than the usual topology. One can actually prove that however $\sigma(\mathcal{T}) = \mathcal{B}(\mathbb{R})$ (this is quite difficult but you will easily find help online ! The key is to prove that any open set (for $\mathcal{T}$) is a countable union of basic open subsets $[a, b)$ even if $\mathcal{T}$ has no countable base. We say that $(\mathbb{R}, \mathcal{T})$ is hereditarily Lindelöf).*

2. *Prove that sets that writes $[a, b)$, $]-\infty, a)$ or $[a, +\infty)$ are both closed and open sets.*

3. *Prove that any compact set (for $\mathcal{T}$) is a countable set.*

4. *Deduce from the previous questions that any probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that has a density is non tight (relatively to the topology $\mathcal{T}$ considered in this exercise of course !).*

This topology is not metrisable. If one wants to construct a non tight probability measure on a metric space (that could not be a Polish space !) it is a much more challenging exercise !

From Theorem 2 and Remark 1, we know that any probability measure P on $(E, \mathcal{F})$ satisfies for all event A

$$\mathrm{P}(A) = \sup\{\mathrm{P}(F), F \text{ closed set included in } A\}.$$

When we assume moreover that E is a Polish space, one can improve this approximation by restraining the supremum over compact sets

Proposition 2. *Assume that E is a Polish space. Then for all $A \in \mathcal{F}$*

$$\mathrm{P}(A) = \sup\{\mathrm{P}(K), K \text{ compact set included in } A\}.$$

Proof. Fix $\varepsilon > 0$. From Theorem 2 there exists a closed set $F \subset A$ such that $\mathrm{P}(F) \geq \mathrm{P}(A) - \varepsilon$. From Theorem 3 there exists a compact set L such that $\mathrm{P}(L) > 1 - \varepsilon$. The set $K = F \cap L$ is compact (because F is closed and L is compact) and satisfies $K \subset F \subset A$ and

$$\mathrm{P}(K) \geq \mathrm{P}(A) - 2\varepsilon,$$

as

$$\begin{aligned} \mathrm{P}(A) &\leq \mathrm{P}(A \cap K) + \mathrm{P}(A \cap F^c) + \mathrm{P}(A \cap L^c) \\ &\leq \mathrm{P}(K) + \mathrm{P}(A \setminus F) + \mathrm{P}(L^c) \\ &\leq \mathrm{P}(K) + 2\varepsilon. \end{aligned}$$

□

2.2 An important example : $\mathscr{C}([0,1])$

. We start here the study of $\mathscr{C}([0,1])$ viewed as a metric space once endowed with the uniform metric that is the metric induced by the uniform norm :

$$||f||_\infty = \sup\{|f(x)|,\ x \in [0,1]\}.$$

Theorem 4. *The set $(\mathscr{C}([0,1]), ||\cdot||_\infty)$ is a Polish space.*

It is a particular case of the more general

Theorem 5. *If (E,d) is a compact metric space then $(\mathscr{C}(E), ||\cdot||_\infty)$, the space of real-valued continuous functions on E endowed with the uniform norm, is a complete separable normed vector space (viewed as metric space it is a Polish space).*

Proof. First observe that E is separable. Indeed for all $n \geq 1$, using the Borel property, one can extract a finite family of open balls $(B(x_p^n, 1/n))_{p\in I_p}$ that covers E. The family $(x_n^p)_{p\geq 1, n\in I_p}$ is a countable dense family. For convenience, we rename it $(y_n)_{n\geq 1}$ and introduce for all $n \geq 1$

$$f_n : x \in E \to d(x, y_n).$$

We also set $f_0 = 1$. We consider $\mathcal{A}$ to be the subalgebra generated by all linear combinaisons with rational coefficients of f_n, $n \geq 0$. It satisfies hypothesis of Stone-Weierstrass theorem :

1. $\mathcal{A}$ contains a non zero constant function.
2. $\mathcal{A}$ separates points of E. Indeed if $u \neq v$, consider a subsequence $(y_{\phi(n)})_{n\geq 1}$ that converges to u. Then $(f_{\phi(n)}(u))_{n\geq 1}$ converges to 0 while this is not the case for $(f_{\phi(n)}(v))_{n\geq 1}$. This implies that for some n, $f_{\phi(n)}(u) \neq f_{\phi(n)}(v)$.

Using Stone-Weierstrass theorem we obtain that $\mathcal{A}$ is dense in $\mathscr{C}(E)$. As it is also countable (it is actually the set of all linear combinaisons of finite product of f_n, $n \geq 1$) this achieves the proof.

The fact that $\mathscr{C}(E)$ is complete is a classical result. Consider $(f_n)_{n\geq 1}$ a Cauchy sequence in $(\mathscr{C}(E), ||\cdot||_\infty)$. Clearly for all $x \in E$, $(f_n(x))_{n\geq 1}$ is a Cauchy sequence in $\mathbb{R}$ so that it converges to some $f(x) \in \mathbb{R}$. Let us prove that the convergence to f is actually uniform. Fix $\varepsilon > 0$ and $N \geq 1$ large enough so that for all $n, p \geq N$, $||f_n - f_p||_\infty < \varepsilon$. With $p = N$ and letting n going to infinity, this implies that for all $x \in E$, $|f(x) - f_N(x)| < \varepsilon$. Finally for $n \geq N$,

$$||f_n - f||_\infty \leq ||f_n - f_N||_\infty + ||f_N - f||_\infty < 2\varepsilon.$$

It remains to prove that f is continuous. Fix $x \in E$. As f_N is continuous there exists $\delta > 0$ so that for all $y \in B(x,\delta)$, $|f(y) - f(x)| < \varepsilon$. We obtain that for all $y \in B(x,\delta)$

$$\begin{aligned}|f(y) - f(x)| &\leq |f(y) - f_N(y)| + |f_N(y) - f_N(x)| + |f_N(x) - f(x)| \\ &\leq 2||f_N - f||_\infty + \varepsilon \\ &\leq 3\varepsilon.\end{aligned}$$

□

There are two natural sigma algebras one may consider on the $\mathscr{C}([0,1])$.

Définition 2 (Cylinder σ-algebra)**.** *We call cylinder set any subset of $\mathscr{C}([0,1])$ that writes*

$$\{f \in \mathscr{C}([0,1],\mathbb{R}) \text{ such that } f(t_1) \in B_1, \cdots, f(t_n) \in B_n\}$$

where $n \geq 1$ is an integer, $t_1, \cdots, t_n$ are in $[0,1]$ and $B_1, \cdots, B_n$ are in $\mathcal{B}(\mathbb{R})$. We call cylinder sigma algebra, and note $\mathcal{E}$, the sigma algebra generated by the cylinder sets :

$$\mathcal{E} = \sigma(C,\ C \text{ cylinder set of } \mathscr{C}([0,1])),$$

that is the smallest sigma algebra containing all cylinder sets.

One can check that we define actually the same sigma algebra by replacing the Borel sets in the definition of cylinder by open intervals. There is also another useful definition of $\mathcal{E}$

Proposition 3. *The sigma algebra $\mathcal{E}$ is also the smallest sigma algebra that makes all the coordinate applications measurable.*

We remind that the coordinate applications are the functions π_t, $t \geq 0$, defined by :

$$\begin{aligned}\pi_t : \mathscr{C}([0,1]) &\to \mathbb{R} \\ f &\mapsto f(t).\end{aligned}$$

Proof that both definitions are equivalent. For all $t \geq 0$, π_t is $\mathcal{E} - \mathcal{B}(\mathbb{R})$-measurable as for all $B \in \mathcal{B}(\mathbb{R})$,

$$\{\pi_t \in B\} = \{f \in \mathscr{C}([0,1]),\ f(t) \in B\}$$

is a cylinder set and thus in $\mathcal{E}$.

For the other inclusion, we consider $\tilde{\mathcal{E}}$ a sigma algebra that makes all coordinate applications measurable. We have to check that it contains all cylinder set. We consider such a set

$$C = \{f \text{ s.t. } f(t_1) \in B_1, \cdots, f(t_n) \in B_n\}.$$

We can rewrite

$$C = \bigcap_{i=1}^{n} \{\pi_{t_i} \in B_i\}$$

and this implies that $C \in \tilde{\mathcal{E}}$. We have proven that $\mathcal{E} \subset \tilde{\mathcal{E}}$. □

Remark that all properties stated so far are also true on the space $\mathcal{F}([0,1])$ of functions from $[0,1]$ with values in $\mathbb{R}$. They have nothing to do with continuity. However in the continuous case there is another natural way to build a sigma algebra. One can consider $\mathscr{C}([0,1])$ as a metric space for the distance induced by the norm $||\cdot||_\infty$ and then consider the Borel sigma algebra $\mathcal{F}$ on $\mathscr{C}([0,1],\mathbb{R})$ that is the smallest sigma algebra that contains all open sets of the topology of uniform convergence.

Proposition 4. *The cylinder sigma algebra and the borel sigma algebra coincide:*

$$\mathcal{F} = \mathcal{E}.$$

Proof. $\underline{\mathcal{F} \supset \mathcal{E}}$. For all $t \geq 0$,

$$\pi_t : (\mathscr{C}(\mathbb{R}^+,\mathbb{R}), ||\cdot||_\infty) \to (\mathbb{R}, |\cdot|)$$

is continuous and thus measurable. This implies that $\mathcal{F}$ makes all coordinate applications measurable and this enough for this first inclusion.

$\underline{\mathcal{F} \subset \mathcal{E}}$. From Theorem 4, we know that $\mathscr{C}([0,1])$ is separable. This implies that $\mathcal{F}$ is also generated by open (or closed) balls (see Exercice 8) and we are reduced to prove that $\mathcal{F}$ contains all closed balls. Let f be in $\mathscr{C}([0,1])$ and $\varepsilon > 0$. Using that f is continuous we obtain

$$\begin{aligned} \overline{B(f,\varepsilon)} &= \{g \in \mathscr{C}([0,1],\mathbb{R}) \text{ such that } ||g-f||_\infty \leq \varepsilon\} \\ &= \cap_{t\in[0,1]} \{g \in \mathscr{C}([0,1],\mathbb{R}) \text{ such that } g(t) \in [f(t)-\varepsilon, f(t)+\varepsilon]\} \\ &= \cap_{t\in[0,1]\cap\mathbb{Q}} \{\pi_t \in [f(t)-\varepsilon, f(t)+\varepsilon]\}. \end{aligned}$$

From this we deduce that $\overline{B(f,\varepsilon)} \in \mathcal{E}$ and this concludes the proof. □

Proposition 5. *A probability measure on* $(\mathscr{C}([0,1]),\mathcal{F})$ *is characterised by its finite dimensional marginals that is if* P *and* Q *are two probability measures on* $(\mathscr{C}([0,1]),\mathcal{F})$ *such that for all cylinder sets* C *it holds that* $\mathrm{P}(C) = \mathrm{Q}(C)$ *then* $\mathrm{P} = \mathrm{Q}$.

Proof. Indeed, $\mathcal{F} = \mathcal{E} = \sigma(C,\ C$ cylinder of $\mathscr{C}([0,1]))$ and moreover the class of cylinder sets is stable by finite intersection. The result comes next from the monotone class lemma (=Dynkin's lemma). □

2.3 More exercices

(including topology stuff quite far from the topic !)

Exercise 5. *Let $A \subset E$. We remind that for all $x \in E$, $d(x, A) = \inf\{d(x, y),\ y \in A\}$. Prove that the function $x \in E \to d(x, A)$ is 1-Lipschitz and that $\bar{A} = \{x \in E$ tel que $d(x, A) = 0\}$.*

Exercise 6. *Prove that the three following properties are equivalent :*

1. *(E, d) satisfies the Borel property*
2. *(E, d) is sequentially compact*
3. *(E, d) is a totally bounded and complete metric space*

"totally bounded" = "precompact" in french.

Exercise 7. *Prove that the two following definitions of "$(E, \mathcal{T})$ is separable" (where E is a metric space) are equivalent*

1. *it exists a countable dense subset in E,*
2. *$\mathcal{T}$ is generated by a countable family of open sets;*

Note that this is not the case anymore if $(E, \mathcal{T})$ is not a metric space, see Exercise 11.

Exercise 8. *Compare the sigma-algebra generated by open balls and $\mathcal{F}$. Prove that they coincide when E is a separable set and that it is however not the case in the general setting.*

Exercise 9. *Prove that a sequence $(x_n)_{n\geq 1}$ with values in E converges to $x \in E$ if and only if for any subsequence of $(x_n)_{n\geq 1}$ one can extract a further subsequence (a "subsubsequence") converging to x.*

Exercise 10. *Let $\mathcal{A}$ be a collection of subsets of E. Prove that the topology generated by $\mathcal{A}$ (that is the smallest topology that contains $\mathcal{A}$) is the collection of sets that are union of sets B that writes*

$$B = \cap_{i\in I} A_i,$$

where I is finite and A_i, $i \in I$ are in $\mathcal{A}$. The collection of sets of this form is called a base of the topology. Prove that a sequence $(x_n)_{n\geq 1}$ converges to x if and only if for all B in the base all x_n are in B for n large enough.

Exercise 11. *Any topology generated by a metric is clearly Hausdorff and this simple observation provides easy to find examples of topologies that are not metrizable (in the sense that it does not coincide with any topology associated to a metric on E). In this exercise we prove that the topology of the pointwise convergence on the space $\mathcal{A}([0,1])$ of applications from $[0,1]$ in $\mathbb{R}$ is separated but not metrizable.*

1. *We consider the topology $\mathcal{T}$ generated by the sets*

$$U^{\varepsilon}_{x,z} = \{f \in \mathcal{A}([0,1]),\ |f(x) - z| < \varepsilon\}, \qquad x \in [0,1], z \in \mathbb{R}, \varepsilon > 0.$$

   Prove that a sequence of function converges pointwise if and only if it converges for the topology $\mathcal{T}$.

2. *Prove that $\mathcal{T}$ is separated.*

3. *A function is said to be simple if it takes value 0 except for a finite number of points. Check that the set of simple functions is dense in $\mathcal{A}([0,1])$.*

4. *Check that the function identically equals to 1 is not limit for the pointwise convergence of a sequence of simple functions.*

5. *Deduce that there exists no metric that generates the topology $\mathcal{T}$.*

Exercise 12. *We consider the set $E = [0,1]$ endowed with the topology*

$$\mathcal{T} = \{A \subset E \text{ s.t. } A^c \text{ is a countable set}\} \cup \{\emptyset\}.$$

1. *Prove that $\mathcal{T}$ is a topology and that it is not separated and thus non metrizable.*

2. *Describe $\mathcal{F} = \sigma(\mathcal{T})$.*

3. *We consider the restriction of the uniform probability to $\mathcal{F}$. Prove that it is not regular.*

3 Weak convergence

3.1 Definition and Portmanteau Theorem

We consider in this section a metric space (E, d) and the probability space $(E, \mathcal{F})$ where $\mathcal{F}$ is the Borel sigma algebra generated by the topology $\mathcal{T}$ associated to d.

Définition 3. *A sequence of probability measures* $(\mathrm{P}_n)_{n\geq 1}$ *is said to converge weakly to a probability measure* P *if* ***for all continuous and bounded real function*** f*, the sequence* $(\int f d\mathrm{P}_n)_{n\geq 1}$ *converges to* $\int f d\mathrm{P}$*.*

When such convergence holds we use the notation $\mathrm{P}_n \implies \mathrm{P}$. Note that the limit is unique because if $\mathrm{P}_n \implies \mathrm{P}$ and $\mathrm{P}_n \implies \mathrm{P}'$, this implies that $\int f d\mathrm{P} = \int f d\mathrm{P}'$ for all continuous bounded functions and thus, from Proposition 1, $\mathrm{P} = \mathrm{P}'$.

Définition 4. *A sequence of random variables* $(X_n)_{n\geq 1}$ *built on a probability space* $(\Omega, \mathcal{G}, \mathrm{Q})$ *with value in* $(E, \mathcal{F})$ *is said to converge in law to some variable* X *if* ***for all continuous and bounded real function*** f*, the sequence* $(\mathrm{E}(f(X_n)))_{n\geq 1}$ *converges to* $\mathrm{E}(f(X))$*.*

Denoting by P_n the law of X_n and P the law of X, we can reformulate this last definition: $(X_n)_{n\geq 1}$ is said to converge in law to some variable X if $(\mathrm{P}_n)_{n\geq 1}$ converges weakly to P on $(E, \mathcal{F})$. This is just due to the definition of the law of a random variable that implies that for all continuous and bounded real function f, and all $n \geq 1$

$$\mathrm{E}_{\mathrm{Q}}(f(X_n)) = \mathrm{E}_{\mathrm{P}_n}(f) \quad \text{and} \quad \mathrm{E}_{\mathrm{Q}}(f(X)) = \mathrm{E}_{\mathrm{P}}(f).$$

The notion of convergence in law depends thus of the random variables only through their laws.

Let see some examples to make this definition more familiar.

1. If $(x_n)_{n\geq 1}$ is a sequence with values in E that converges to some $x \in E$ then $\delta_{x_n} \implies \delta_x$.

2. The sequence $1/n \sum_{i=1}^{n} \delta_{i/n}$ converges weakly on $([0,1], \mathcal{B}([0,1]))$ to the uniform probability measure. Indeed for all bounded continuous function
$$\int f \, \mathrm{d}(\frac{1}{n}\sum_{i=1}^{n} \delta_{i/n}) = \frac{1}{n}\sum_{i=1}^{n} f(i/n) = \int_{[0,1]} f_n(x)\, dx,$$
where f_n is the piecewise constant function $f_n = \sum_{i=1}^{n} f(i/n) 1_{[(i-1)/n, i/n[}$. We conclude easily using Lebesgue convergence theorem. This proves

that a sequence of discrete probability measures may converge to a diffusive one.

3. The sequence $(\mathcal{N}(0,1/n))_{n\geq 1}$ converges weakly to δ_0 on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$.

4. **Central Limit Theorem.** Consider a sequence $(\xi_n)_{n\geq 1}$ of i.i.d. centred square integrable random variables such that $\mathrm{E}(\xi^2)=1$. Then $\sum_{k=1}^n \xi_k/\sqrt{n} \implies \mathcal{N}(0,1)$.

   The proof of this well-known result is postponed to the section relative to characteristic function.

5.

   Exercise 13. ***Scheffé Lemma.*** *Consider* $(\mathrm{P}_n)_{n\geq 1}$ *a sequence of probability measures on* $(E,\mathcal{F})$ *defined by their densities* $(f_n)_{n\geq 1}$ *with respect to a reference measure measure* Q. *We assume moreover that* $(f_n)_{n\geq 1}$ *converges* Q*−almost surely to some density function* f. *Then the convergence holds in* $\mathrm{L}^1(\mathrm{Q})$ *and moreover*

$$\mathrm{P}_n \implies f\,\mathrm{Q}.$$

The following theorem, known as Portmanteau's theorem (even if no Portmanteau seems to have ever existed !) gives a useful characterisation of weak convergence :

Theorem 6 (Portmanteau)**.** *The following properties are equivalent*

1. $(\mathrm{P}_n)_{n\geq 1}$ *converges weakly to* P

2. $(\mathrm{E}_{\mathrm{P}_n}(f))_{n\geq 1}$ *converges to* $\mathrm{E}_{\mathrm{P}}(f)$ *for all Lipschitz and bounded real function*

3. $\limsup \mathrm{P}_n(F) \leq \mathrm{P}(F)$ *for all closed set* F

4. $\liminf \mathrm{P}_n(O) \geq \mathrm{P}(O)$ *for all open set* O

5. $\lim \mathrm{P}_n(A) = \mathrm{P}(A)$ *for all event* $A\in\mathcal{F}$ *such that* $\mathrm{P}(\partial A)=0$.

Proof. $\underline{1\to 2}$. This is a straightforward consequence of the definition.

$\underline{2\to 3}$. Fix $\varepsilon>0$ and consider the Lipschitz bounded function f that appears in (1). It holds that

$$\limsup \mathrm{P}_n(F) = \limsup \mathrm{E}_{\mathrm{P}_n}(1_F) \leq \limsup \mathrm{E}_{\mathrm{P}_n}(f) \overset{(2.)}{=} \mathrm{E}_{\mathrm{P}}(f) \leq \mathrm{P}(F^\varepsilon).$$

As F is closed, $F=\cap_{n\geq 1}\downarrow F^{1/n}$ and we obtain *3*.

3 ↔ 4. This is just matter of complementation !

3 & 4 → 5. Let A be an event in $\mathcal{F}$ such that $\mathrm{P}(\partial A) = 0$. Using points *3* and *4*,

$$\mathrm{P}(A^\circ) \le \liminf \mathrm{P}_n(A^\circ) \le \liminf \mathrm{P}_n(A) \le \limsup \mathrm{P}_n(A) \le \limsup \mathrm{P}_n(\bar{A}) \le \mathrm{P}(\bar{A}).$$

As $\mathrm{P}(\partial A) = 0$, $\mathrm{P}(A^\circ) = \mathrm{P}(\bar{A}) = \mathrm{P}(A)$ and one can deduce that $\liminf \mathrm{P}_n(A) = \limsup \mathrm{P}_n(A) = \mathrm{P}(A)$.

5 → 1. Consider f a bounded continuous function. Let say that M is a bound ($\sup |f| \le M$). Considering $f + M$ instead of f, one can consider that f is non negative so that by Fubini theorem

$$\mathrm{E}_{\mathrm{P}}(f) = \int_0^{+\infty} \mathrm{P}(f > t)\ dt = \int_0^M \mathrm{P}(f > t)\ dt,$$

and the same holds if we replace P by P_n for $n \ge 1$. As f is continuous $\partial\{f > t\} \subset \{f = t\}$ (as $\overline{\{f > t\}} \subset \{f \ge t\}$ and $(\{f > t\}^\circ)^c = \{f > t\}^c = \{f \le t\}$). If $\mathrm{P}(f = t) = 0$, one can thus use *5.* and $\mathrm{P}_n(f > t) \to \mathrm{P}(f > t)$. As the set of t such that $\mathrm{P}(f = t) > 0$ is at most countable (for all $k \ge 1$ there are only a finite number of t such that $\mathrm{P}(f = t) > 1/k$) it is negligible and we use the dominated convergence theorem to conclude.

□

The Portmanteau theorem has obviously a counterpart characterising convergence in law of a sequence of random variables :

Theorem 7 (Portmanteau again)**.** *Let* $(X_n)_{n\ge1}$ *and* X *be random variables on* $(\Omega, \mathcal{G}, \mathrm{Q})$*. The following properties are equivalent*

1. $(X_n)_{n\ge1}$ *converges in law to* X
2. $(\mathrm{E}(f(X_n)))_{n\ge1}$ *converges to* $\mathrm{E}(f(X))$ *for all Lipschitz and bounded real function*
3. $\limsup \mathrm{Q}(X_n \in F) \le \mathrm{Q}(X \in F)$ *for all closed set* F
4. $\liminf \mathrm{Q}(X_n \in O) \ge \mathrm{Q}(X \in O)$ *for all open set* O
5. $\lim \mathrm{Q}(X_n \in A) = \mathrm{Q}(X \in A)$ *for all event* $A \in \mathcal{F}$ *such that* $\mathrm{Q}(X \in \partial A) = 0$.

Note that weak convergence is at the lowest place in the hierarchy of random variables convergence.

Exercise 14. *Prove that*

1. *for all* $1 \leq p \leq q \leq \infty$, *convergence in* L^q *implies convergence in* L^p,

2. *convergence in* L^1 *implies convergence in probability,*

3. *almost sure convergence implies convergence in probability,*

4. *convergence in probability implies convergence almost surely for a subsequence,*

5. *convergence in probability implies convergence in law.*

The Portmanteau theorem has the nice and useful following consequence

Theorem 8 (Mapping Theorem)**.** *Consider* (E, d) *and* (E', d') *two metric spaces and* $(\mathrm{P}_n)_{n\geq 1}$ *a sequence of probability measures on* (E, d) *weakly converging to some probability measure* P. *Consider also* $f : E \mapsto E'$ *a measurable function such that* $Disc(f)$, *the set of all discontinuity points of* f, *satisfies* $Disc(f) \in \mathcal{F}$ *and* $\mathrm{P}(Disc(f)) = 0$. *Then*

$$\mathrm{P}_n\, f^{-1} \implies \mathrm{P}\, f^{-1}.$$

Proof. First remark that when f is continuous everywhere the result is easy to prove as for any bounded continuous function ϕ, using that $\phi \circ f$ is also bounded and continuous,

$$\int \phi \mathrm{d}(\mathrm{P}_n\, f^{-1}) = \int \phi \circ f \mathrm{d}\mathrm{P}_n \overset{(n\to+\infty)}{\to} \int \phi \circ f \mathrm{d}\mathrm{P} = \int \phi \mathrm{d}(\mathrm{P}\, f^{-1}).$$

For the general case, we use the third condition of Theorem 6. We consider F a closed set of E'. We have to check that

$$\limsup \mathrm{P}_n(f \in F) \leq \mathrm{P}(f \in F).$$

Note that this is straightforward if f is continuous as in this case $\{f \in F\}$ is a closed set and this provides another proof in this easy case. To manage with the general case we deal with the closure of $\{f \in F\}$ and use that $\mathrm{P}(Disc(f)^c) = 1$,

$$\limsup \mathrm{P}_n(f \in F) \leq \limsup \mathrm{P}_n(\overline{f \in F}) \leq \mathrm{P}(\overline{f \in F}) = \mathrm{P}(\overline{f \in F}, Disc(f)^c).$$

If $x \in \overline{\{f \in F\}}$, there exists a sequence $(x_n)_{n\geq 1}$ with value in $\{f \in F\}$ converging to x. If moreover $x \in Disc(f)^c$, as F is closed, $f(x) \in F$. This implies that $\overline{\{f \in F\}} \cap Disc(f)^c \subset \{f \in F\}$ and this concludes the proof. $\square$

Remark 3. *For readers familiar with functional analysis and that may be surprised by the terminology used there, let us try to make clear the link with the present course. Given* (E, d) *a metric space, we consider the space* $(\mathcal{C}_b(E), ||\cdot||_\infty)$ *that is a Banach space and we denote by* $\mathcal{C}_b(E)'$ *its topological dual dual that is the space of all continuous linear forms on* $\mathcal{C}_b(E)$*. As you know, there is on* $\mathcal{C}_b(E)$ *a notion of weak convergence but we do not care about it here. On* $\mathcal{C}_b(E)'$ *there are several topologies of interest. We are interested here in the weak* $\star$ *topology that is the smallest topology that makes all applications (that is for all* $f \in \mathcal{C}_b(E)$*)*

$$\begin{aligned} \mathcal{C}_b(E)' &\to \mathbb{R} \\ l &\to l(f) \end{aligned}$$

continuous. A sequence $(l_n)_{n\geq 1}$ *in* $\mathcal{C}_b(E)'$ *converges weakly* $\star$ *if and only if for all* $f \in \mathcal{C}_b(E)$ *the sequence* $(l_n(f))_{n\geq 1}$ *converges to* $l(f)$*.*

Finally one defines the weak $\star$ *topology on* $\mathcal{M}_f(E)$ *the set of all finite measure on* $(E, \mathcal{B})$ *as the smallest topology that makes*

$$\begin{aligned} \mathcal{M}_f(E) &\to \mathcal{C}_b(E)' \\ \nu &\to \int \cdot \ \nu \end{aligned}$$

continuous (where the topology considered on $C_b(E)'$ *is the weak* $\star$ *topology). Equivalently the weak* $\star$ *topology on* $\mathcal{M}_f(E)$ *may be defined as the smallest topology that makes all applications (that is for all* $f \in \mathcal{C}_b(E)$*)*

$$\begin{aligned} \mathcal{M}_f(E) &\to \mathbb{R} \\ \nu &\to \int f \ \nu \end{aligned}$$

continuous. The convergence of a sequence $(\nu_n)_{n\geq 1}$ *is characterised by : for all* $f \in \mathcal{C}_b(E)$*,* $\int f \ d\nu_n \to \int f \ d\nu$*.*

We find back the definition of the weak convergence (or convergence in law), we are dealing with in this course. Weak convergence is the usual terminology in probability theory to refer to the weak $\star$ *topology on* $\mathcal{M}_f(E)$*. Any sequence of finite measure* $(\nu_n)_{n\geq 1}$ *converges weakly to* ν *(*$\nu_n \implies \nu$ *with our notations) if and only if* $\nu_n \overset{\star weak}{\to} \nu$*.*

3.2 A metric for the weak convergence ?

In this section we study if the weak convergence can also be defined via a metric ρ on the space $\mathcal{M}$ of all probability measures on $(E, \mathcal{F})$. When this is the case, what kind of properties of $(\mathcal{M}, \rho)$ can we deduced from the properties of (E, d) ?

Définition 5 (Prohorov metric)**.** *For* P *and* Q *in* $\mathcal{M}$*, one defines*

$$\rho(P,Q) = \inf\{\varepsilon > 0,\ \mathrm{P}(A) \leq \mathrm{Q}(A^{\varepsilon})+\varepsilon \ \textit{and}\ \mathrm{Q}(A) \leq \mathrm{P}(A^{\varepsilon})+\varepsilon,\ \textit{for all}\ A \in \mathcal{F}\}.$$

We first have to prove that the function ρ defines just above is indeed a metric on $\mathcal{M}$. It is clear from the definition that it is symmetric. Consider now P and Q such $\rho(\mathrm{P}, \mathrm{Q}) = 0$. We obtain that for all $\varepsilon > 0$ and all closed set F,

$$\mathrm{P}(F) \leq \mathrm{Q}(F^{\varepsilon}) + \varepsilon.$$

As F is closed $F = \cap_{\varepsilon>0} F^{\varepsilon}$ so that letting ε going to 0 we obtain $\mathrm{P}(F) \leq \mathrm{Q}(F)$. By symmetry the reverse inequality holds and we deduce that $\mathrm{P} = \mathrm{Q}$. It remains to prove the triangle inequality. Let P, Q and R be probability measures in $\mathcal{M}$ such that $\rho(\mathrm{P}, \mathrm{Q}) < \varepsilon_1$ and $\rho(\mathrm{Q}, \mathrm{R}) < \varepsilon_2$. It holds that for all $A \in \mathcal{F}$

$$\mathrm{P}(A) \leq \mathrm{Q}(A^{\varepsilon_1}) + \varepsilon_1 \leq \mathrm{R}((A^{\varepsilon_1})^{\varepsilon_2}) + \varepsilon_1 + \varepsilon_2 \leq \mathrm{R}(A^{\varepsilon_1+\varepsilon_2}) + \varepsilon_1 + \varepsilon_2.$$

This implies that $\rho(\mathrm{P}, \mathrm{R}) \leq \varepsilon_1 + \varepsilon_2$ and leads to the triangle inequality.

Before studying the properties of ρ and in particular it links with the weak convergence, we formulate the useful simplified version of its expression:

Lemma 2. *One could equivalently define* ρ *by: For* P *and* Q *in* $\mathcal{M}$*,*

$$\rho(P,Q) = \inf\{\varepsilon > 0,\ \mathrm{P}(A) \leq \mathrm{Q}(A^{\varepsilon}) + \varepsilon,\ \textit{for all}\ A \in \mathcal{F}\}.$$

Proof. Suppose that for some $\varepsilon > 0$ it holds that for all $A \in \mathcal{F}$, $\mathrm{P}(A) \leq \mathrm{Q}(A^{\varepsilon}) + \varepsilon$. We have to prove that for all $A \in \mathcal{F}$, $\mathrm{Q}(A) \leq \mathrm{P}(A^{\varepsilon}) + \varepsilon$. For this we apply the first inequality to $B = (A^{\varepsilon})^c$. This easy to check that $A = (B^{\varepsilon})^c$ as both equalities are actually equivalent to

$$d(x,y) \geq \varepsilon \qquad \text{for all } x \in B,\ y \in A.$$

We thus obtain

$$1 - \mathrm{P}(A^{\varepsilon}) = \mathrm{P}(B) \leq \mathrm{Q}(B^{\varepsilon}) + \varepsilon = 1 - \mathrm{Q}(A) + \varepsilon.$$

□

Proposition 6. *Let* $(\mathrm{P}_n)_{n\geq 1}$ *and* P *be probability measures.*

1. *If* $(\mathrm{P}_n)_{n\geq 1}$ *converges for the metric* ρ *to* P *(that is if* $(\rho(\mathrm{P}_n, \mathrm{P}))_{n\geq 1}$ *converges to* 0*) then* $(\mathrm{P}_n)_{n\geq 1}$ *converges weakly to* P*.*

2. *Assume moreover that* (E, d) *is* ***separable****. If* $(\mathrm{P}_n)_{n\geq 1}$ *converges weakly to* P *then* $(\rho(\mathrm{P}_n, \mathrm{P}))_{n\geq 1}$ *converges to* 0

Proof. 1. Let F be a closed subset of E. We consider a sequence $(\varepsilon_n)_{n\geq 1}$ of positive real numbers converging to 0 and such that for all $n \geq 1$ $\varepsilon_n > \rho(\mathrm{P}_n, \mathrm{P})$ (for example $\varepsilon_n = 2\rho(\mathrm{P}_n, \mathrm{P}) \vee \frac{1}{n}$. By definition of ρ,

$$\limsup \mathrm{P}_n(F) \leq \limsup \mathrm{P}(F^{\varepsilon_n}) + \varepsilon_n.$$

As F is closed, $F = \cap_{n\geq 1} \downarrow F^{\varepsilon_n}$ and $\mathrm{P}(F^{\varepsilon_n}) + \varepsilon_n$ goes to $\mathrm{P}(F)$ when n goes to infinity. We conclude with the Portmanteau theorem (Theorem 6).

2. We now assume that (E, d) is separable. Consider $(x_n)_{n\geq 1}$ a dense countable subset of (E, d). Fix $\varepsilon > 0$. Set $D_1 = B(x_1, \varepsilon)$, $D_2 = B(x_2, \varepsilon) \setminus D_1$ and for $k \geq 3$, $D_k = B(x_k, \varepsilon) \setminus \cup_{i\leq k-1} D_i$. The sets D_i, $i \geq 1$ provides a countable partition of E such that each element has diameter at most 2ε. We fix K large enough so that $\mathrm{P}(\cup_{i\geq K+1} D_i) < \varepsilon$. We consider $\mathcal{G}$ the family of ε-neighbourhood of finite union of D_i, $i \leq K$, that is the sets D that write

$$D = \Big(\bigcup_{i_1,\cdots,i_j \leq K} D_{i_j}\Big)^{\varepsilon}.$$

For each $D \in \mathcal{G}$, as it is an open set we can use Theorem 6 that provides n_0 such that for all $n \geq n_0$,

$$\mathrm{P}_n(D) \geq \mathrm{P}(D) - \varepsilon.$$

As $\mathcal{G}$ is finite one can actually find n_0 such that this inequality holds for all $D \in \mathcal{G}$. For $A \in \mathcal{F}$ we define I to be the set of indexes $i \leq K$ such that D_i intersects A and consider the set $\tilde{A} = \bigcup_{i\in I} D_i$. As $\tilde{A}^{\varepsilon}$ belongs to $\mathcal{G}$, for n larger than n_0,

$$\mathrm{P}(A) \leq \mathrm{P}(\tilde{A}) + \mathrm{P}\Big(\bigcup_{i\geq K+1} D_i\Big) \leq \mathrm{P}(\tilde{A}^{\varepsilon}) + \varepsilon \leq \mathrm{P}_n(\tilde{A}^{\varepsilon}) + 2\varepsilon + \leq \mathrm{P}_n(A^{3\varepsilon}) + 2\varepsilon,$$

as $\tilde{A}^{\varepsilon} \subset A^{3\varepsilon}$.

This is enough to conclude using Lemma 2. □

Theorem 9. *If* (E, d) *is compact then* $(\mathcal{M}, \rho)$ *is also compact.*

Remind Theorem 5 and also the famous

Theorem 10 (Riesz representation theorem)**.** *Let E be a compact set and $\ell : \mathscr{C}(E) \mapsto \mathbb{R}$ be a linear form that is positive ($f \geq 0 \Rightarrow \ell(f) \geq 0$) and satisfies $\ell(1) = 1$. Then there exists a unique probability measure on $(E, \mathcal{F})$ such that for all $f \in \mathscr{C}(E)$,*

$$\ell(f) = \int f \; d\mathrm{P}.$$

For a proof of this well-known result, see the Riesz-Markov-Kakutani theorem in [4, Theorem 2.14].

Sketch of the proof.

1. Define for all open set O

$$\mu(O) = \sup\{l(f), f \in \mathscr{C}(E) \; s.t. \; ||f||_\infty \leq 1, support(f) \subset O\}$$

   and then for all $A \in \mathcal{F}$

$$\mu(A) = \sup\{\mu(O), O \text{ open set, } O \subset A\}$$

2. Prove that μ is a probability measure

3. Prove that $l(f) = \int f d\mu$ for all $f \in \mathscr{C}(E)$

Proof of Theorem 9. We consider a sequence $(\mathrm{P}_n)_{n\geq 1}$ of probability measures on (E, d). We have to prove that there exists a converging subsequence in the sense of the metric ρ. From Proposition 6, this is equivalent to prove that there is a weakly converging subsequence. This proves that $(\mathrm{P}_n)_{n\geq 1}$ is sequentially compact and, as $(\mathcal{M}, \rho)$ is a metric space (see Proposition 6), this proves that it is compact.

We consider again the algebra $\mathcal{A} = \{g_k, \; k \geq 1\}$ introduced in Theorem 5. It is dense in $(\mathscr{C}(E), || \cdot ||_\infty)$. We also define $g_0 = 1$. For all $k \geq 0$, $(\int g_k \; d\mathrm{P}_n)_{n\geq 1}$ is a bounded sequence of real numbers so that we can extract from it a converging subsequence. Using a diagonal extraction procedure [Details ?] one can actually build a subsequence $(\mathrm{P}_{\psi(n)})_{n\geq 1}$ so that all these sequences converge together : for $k \geq 0$,

$$\int g_k \; \mathrm{dP}_{\psi_n} \to \ell(g_k) \quad \text{when } n \text{ goes to infinity.}$$

It is easy to check that ℓ is $1-$Lipschitz on $\mathcal{A}$ and we can thus consider the continuous continuation of ℓ to $(\mathscr{C}(E), || \cdot ||_\infty)$. We now prove that ℓ is linear and positive. Consider g_i, g_j in $(g_k)_{k\geq 1}$ and a, b two positive rational numbers.

As $(g_k)_{k\geq 1}$ is an algebra, there exists g_m in $\mathcal{A}$ such that $ag_i + bg_j = g_m$. We then easily obtain

$$\ell(g_m) = \lim_{n\to+\infty} \int g_m \mathrm{dP}_{\psi(n)} = a\ell(g_i) + b\ell(g_j).$$

As ℓ is continuous, it is actually linear on $\mathscr{C}(E)$. We now prove that it is positive. Let $g \in \mathscr{C}(E)$ such that $g \geq 0$. As $(g_k)_{k\geq 1}$ is dense one can extract a subsequence $(g_{\phi(k)})_{k\geq 1}$ that converges uniformly to g. For all $\varepsilon > 0$, $g_{\phi(k)} \geq -\varepsilon$ for k large enough so that $\ell(g_{\phi(k)}) \geq -\varepsilon$. As ℓ is continuous this is enough to conclude that $\ell(g) \geq -\varepsilon$, and, as ε is arbitrary small, that $\ell(g) \geq 0$. Finally note that $\ell(1) = 1$.

From Riesz representation theorem, there existe a probability measure $\mathrm{P} \in \mathcal{M}(E)$ such that for all $g \in \mathscr{C}(E)$,

$$l(g) = \int g \ \mathrm{dP}.$$

It remains to prove that P is the weak limit of $(\mathrm{P}_{\psi_n})_{n\geq 1}$. For that consider again $g \in \mathscr{C}(E)$ and $(g_{\phi(k)})_{k\geq 1}$ that converges uniformly to g. It holds that

$$\begin{aligned}&\left|\int g \ \mathrm{dP}_{\psi(n)} - \int g \ \mathrm{dP}\right| \\ &\leq \left|\int g \ \mathrm{dP}_{\psi(n)} - \int g_{\phi(k)} \ \mathrm{dP}_{\psi(n)}\right| + \left|\int g_{\phi(k)} \ \mathrm{dP}_{\psi(n)} - \int g_{\phi(k)} \ \mathrm{dP}\right| + \left|\int g_{\phi(k)} \ \mathrm{dP} - \int g \ \mathrm{dP}\right|.\end{aligned}$$

Both first and last term are smaller than $||g - g_{\phi(k)}||_\infty$ (that can be made arbitrary small choosing k large enough) while the second term goes to 0 with n going to infinity by definition of ℓ and P. [Try to find a more probabilistic proof] □

We present now a last result as it is easy to memorise together with the previous one but we postpone the proof to the next section as it uses Prohorov Theorem. This result is interesting specially when one wants to study random mesures that are random variables with values in a space of measures. It guarantees then that we are in the confortable framework of Polish spaces.

Theorem 11. *If (E, d) is a Polish space then $(\mathcal{M}, \rho)$ is also a Polish space.*

Exercise 15. *Let us turn back to Exercise 9. Prove that the result is still valid if we consider the weak convergence of a sequence of probability measure instead of convergence in a metric space. Of course the results are completely equivalent by Proposition 6 when (E, d) is separable.*

3.3 Tightness and Prohorov Theorem

Prohorov Theorem is a decisive tool to prove some convergences in law on the space $\mathscr{C}([0,1])$ as we will see later. It provides a characterisation of *compactness* via the notion of *tightness*. We first make precise what we mean by these two notions.

By *(relatively) compact*, we mean here, sequentially compact: that is a set Π of probability measures on (E,d) is compact if from any sequence $(\mathrm{P}_n)_{n\geq 1}$ of elements of Π one can extract a weakly convergent subsequence. Of course when weak convergence can be defined by a metric (see 3.2 for a discussion about this point) one can use equivalently any other definition of compactness using Borel Theorem (see Exercise 6).

We turn to the definition of *tight* that can be viewed as an extension of Definition 1 .

Définition 6. *A family $(\mathrm{P}_n)_{n\geq 0}$ of probability measures on (E,d) is said to be tight if for every $\varepsilon > 0$ there exists a compact set K such that for all $n \geq 0$*

$$\mathrm{P}_n(K) > 1 - \varepsilon.$$

To understand better this definition let us check if the following family are tight or not :

1. Due to Theorem 3, if (E,d) is a Polish space then any singleton $\{\mathrm{P}\}$ is tight.

   Exercise 16. *Prove that any finite family of probability measures is tight.*

2. Consider the sequence of probability measures δ_n, $n \geq 1$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ then it is easy to check that it is not tight. This a first example to keep in mind : the mass escapes to infinity.

3. Consider the sequence of probability measures P_n, $n \geq 1$ defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by
$$\mathrm{dP}_n = \frac{1}{\sqrt{2\pi n}} e^{-x^2/2n}\ \mathrm{dx}.$$
   Here again it is easy to check that this family is not tight and this a second example to keep in mind : the mass spreads on $\mathbb{R}$.

Exercise 17. *Let f be a continuous function from (E,d) with values in (E',d') and $\Pi \subset \mathcal{M}(E)$ be a tight family of probability measures. Prove that the pushforward probability measures of elements of Π by f are also a tight family of probability measures on E'.*

Theorem 12 (Prohorov). *Let Π be a familly of probability measures on (E, d).*

1. *If Π is tight then it is relatively compact that is: From any sequence with values in Π one can extract a weakly converging subsequence (not necessarily in Π).*

2. *Assume moreover that (E, d) is a Polish space. Then the converse is true that is: If Π is relatively compact then it is tight.*

Proof. For applications, we are actually mainly interested by the direct part of the theorem that is however the more difficult to prove. We thus start with the *direct sense.*

Step 1. Let see first how we build the **extraction.** As $(\mathrm{P}_n)_{n\geq 0}$ is tight, for all $p \geq 1$ there exists a compact set K_p such that for all $n \geq 1$, $\mathrm{P}_n(K_p^c) < 1/p$. We may suppose that $(K_p)_{p\geq 1}$ is an increasing sequence just replacing (without changing the name !) K_p by $K_1 \cup \cdots \cup K_p$. For all $p \geq 1$ we consider the restriction $\mathrm{P}_n^{(p)}$ of P_n to K_p that is $\mathrm{P}_n^{(p)}(A) = \mathrm{P}_n(A \cap K_p)$ for all $A \in \mathcal{B}(K_p)$. Note that $\mathrm{P}_n^{(p)}$ is not a probability measure as its total mass may be less than 1. As $\mathcal{M}_{\leq 1}(K_p)$, the set of measures on K_p with mass less than 1 is compact (this is an improvement of Theorem 9) one can extract a weakly converging subsequence from $(\mathrm{P}_n^{(p)})_{n\geq 0}$. Using a diagonal extraction procedure, one can actually consider an extraction ϕ that works for all $p \geq 1$ that is

$$\mathrm{P}_{\phi(n)}^{(p)} \implies \bar{\mathrm{Q}}_p,$$

where $\bar{\mathrm{Q}}_p$ is a measure in $\mathcal{M}_{\leq 1}(K_p)$. We extend $\bar{\mathrm{Q}}_p$ to a measure Q_p on E setting for all $A \in \mathcal{F}$,

$$\mathrm{Q}_p(A) := \bar{\mathrm{Q}}_p(A \cap K_p),$$

that makes sense as $A \cap K_p \in \mathcal{B}(K_p)$.

Step 2. We are now ready to **define the limit.** For this we prove that for all $A \in \mathcal{F}$, $(\mathrm{Q}_p(A))_{p\geq 1}$ is non decreasing. Note that it is easy if we assume $\bar{Q}_p(\partial A) = \bar{Q}_{p+1}(\partial A) = 0$ using Portmanteau theorem. If it is not the case, we fix $p \geq 1$ and first consider the case where $A = F$ is a closed set in K_p. We define, for $\delta > 0$, F_δ to be the closed δ-neighborhood of F in K_p :

$$F_\delta = \{x \in K_p,\ d(x, F) \leq \delta\}.$$

Remark that $F \subset F_\delta \subset K_p \subset K_{p+1}$. From the Portmanteau theorem we get for all $\delta > 0$

$$\mathrm{Q}_{p+1}(F_\delta) \overset{(def.)}{=} \bar{\mathrm{Q}}_{p+1}(F_\delta \cap K_{p+1}) = \bar{\mathrm{Q}}_{p+1}(F_\delta) \geq \limsup_n \mathrm{P}_{\phi(n)}^{(p+1)}(F_\delta). \tag{2}$$

As $F_\delta \subset K_p$, $\mathrm{P}^{(p+1)}_{\phi(n)}(F_\delta) = \mathrm{P}^{(p)}_{\phi(n)}(F_\delta)$. Moreover as $\bar{\mathrm{Q}}_p$ is a finite measure the set of δ such that the boundary of F_δ is of non zero $\bar{\mathrm{Q}}_p$ measure is at most countable (for all $i \geq 1$, there are at most i distincts δ such that $\bar{\mathrm{Q}}_p(\partial F_\delta) \geq 1/i$) and one can define a sequence $(\delta_k)_{k\geq 1}$ decreasing to 0 such that for all $k \geq 1$, $\bar{\mathrm{Q}}_p(\partial F_{\delta_k}) = 0$. Using the Portemanteau theorem we obtain $\lim_{n\geq 1} \mathrm{P}^{(p)}_{\phi(n)}(F_{\delta_k}) = \bar{\mathrm{Q}}_p(F_{\delta_k})$ and thus from (2),

$$\mathrm{Q}_{p+1}(F_{\delta_k}) \geq \bar{\mathrm{Q}}_p(F_{\delta_k}) \geq \bar{\mathrm{Q}}_p(F) = \mathrm{Q}_p(F).$$

As $F = \cap_k \geq 1 \downarrow F_{\delta_k}$ one deduce from this last inequality that

$$\mathrm{Q}_{p+1}(F) \geq \mathrm{Q}_p(F). \tag{3}$$

Using Theorem 2, for all $A \in \mathcal{F}$,

$$\mathrm{Q}_{p+1}(A \cap K_p) = \sup\{\mathrm{Q}_{p+1}(F),\ F \text{ closed in } E \text{ and } F \subset A \cap K_p\}.$$

If $F \subset A \cap K_p$ is closed in E then F is closed in K_p and one can use (3) so that

$$\begin{aligned}\mathrm{Q}_{p+1}(A \cap K_p) &\overset{(3)}{\geq} \sup\{\mathrm{Q}_p(F),\ F \text{ closed in } E \text{ and } F \subset A \cap K_p\} \\ &= \mathrm{Q}_p(A \cap K_p).\end{aligned}$$

From this we deduce that $\mathrm{Q}_{p+1}(A) \geq \mathrm{Q}_{p+1}(A \cap K_p) \geq \mathrm{Q}_p(A \cap K_p) = \mathrm{Q}_p(A)$. And one can finally define for all $A \in \mathcal{F}$,

$$\mathrm{Q}(A) = \lim_{p\geq 1} \uparrow \mathrm{Q}_p(A).$$

Step 3. **We now prove that** Q **is a probability measure.** We use the following lemma that is a non difficult exercise left to the reader.

Lemma 3. *Suppose that $m : \mathcal{F} \to \mathbb{R}^+$ satisfies*

1. *for all $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$,*

$$m(A \cup B) = m(A) + m(B);$$

2. *For all increasing sequence $(A_n)_{n\geq 1}$ of events in $\mathcal{F}$,*

$$m(\cup_{n\geq 1} A_n) = \lim_{n\geq 1} \uparrow m(A_n).$$

Then m is a measure on $(E, \mathcal{F})$.

We prove now that Q satisfies both assumption of Lemma 3. For the first point, just note that for all $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$,

$$\mathrm{Q}(A \cup B) = \lim_{p \geq 1} \uparrow \mathrm{Q}_p(A \cup B) = \lim_{p \geq 1} \uparrow [\mathrm{Q}_p(A) + \mathrm{Q}_p(B)] = \mathrm{Q}(A) + \mathrm{Q}(B).$$

While for the seconde one,

$$\mathrm{Q}(\cup_{n \geq 1} \uparrow A_n) = \lim_{p \geq 1} \uparrow \mathrm{Q}_p(\cup_{n \geq 1} \uparrow A_n) = \sup_{p \geq 1} \sup_{n \geq 1} \mathrm{Q}_p(A_n) = \sup_{n \geq 1} \sup_{p \geq 1} \mathrm{Q}_p(A_n) = \sup_{n \geq 1} \mathrm{Q}(A_n).$$

We stress that this last argument turns wrong if we do not suppose the sequence $(A_n)_{n \geq 1}$ to be increasing and that is why we use Lemma 3. Moreover

$$\mathrm{Q}(E) = \lim_{p \geq 1} \uparrow \mathrm{Q}_p(E) = \sup_{p \geq 1} \bar{\mathrm{Q}}_p(K_p) = \sup_{p \geq 1} \lim_{n \geq 1} \mathrm{P}^{(p)}_{\phi(n)}(K_p) = \sup_{p \geq 1} \lim_{n \geq 1} \mathrm{P}_{\phi(n)}(K_p) = 1$$

so that Q is a probability measure.

Step 4. Finally we prove that Q **is the weak limit of** $(\mathrm{P}_{\phi(n)})_{n \geq 1}$ using once again the Portmanteau theorem. For any open set O, as $O \cap K_p$ is an open set of K_p,

$$\mathrm{Q}(O) = \lim_{p \geq 1} \mathrm{Q}_p(O) = \lim_{p \geq 1} \bar{\mathrm{Q}}_p(O \cap K_p) \leq \lim_{p \geq 1} \liminf_n \mathrm{P}_{\phi(n)}(O \cap K_p).$$

For all $p \geq 1$, $\mathrm{P}_{\phi(n)}(O \cap K_p) \leq \mathrm{P}_{\phi(n)}(O)$ so that the last inequality rewrites

$$\mathrm{Q}(O) \leq \liminf_n \mathrm{P}_{\phi(n)}(O).$$

This implies that $(\mathrm{P}_{\phi(n)})_{n \geq 1}$ converges weakly to Q.

We turn to the *converse proposition.*

Step 1. We prove here the following result: If $(O_n)_{n \geq 1}$ is an increasing sequence of open sets such that $\cup_{n \geq 1} O_n = E$ then for all $\varepsilon > 0$ there exists n such that for all $\mathrm{P} \in \Pi$, $\mathrm{P}(O_n) > 1 - \varepsilon$. Suppose indeed that it is not the case: then for some $\varepsilon > 0$ one can build a sequence $(\mathrm{P}_n)_{n \geq 1}$ with values in Π such that for all $n \geq 1$, $\mathrm{P}_n(O_n) \leq 1 - \varepsilon$. As Π is relatively compact one can extract a weakly converging subsequence $(\mathrm{P}_{\phi(n)})_{n \geq 1}$. We denote by Q its limit. This implies that for all $k \geq 1$

$$\mathrm{Q}(O_k) \leq \liminf_n \mathrm{P}_{\phi(n)}(O_k).$$

For n large enough $\phi(n)$ is of course larger than k so that $O_k \subset O_{\phi(n)}$ and finally

$$\mathrm{Q}(O_k) \leq \liminf_n \mathrm{P}_{\phi(n)}(O_{\phi(n)}) \leq 1 - \varepsilon.$$

This is of course a contradiction as $Q(O_k)$ increases to 1 when k goes to infinity.

Step 2. As E is separable there exists a countable family $(x_i)_{i\geq 1}$ that is dense. Fix $k \geq 1$ and consider $O_n^k = \cup_{i\leq n} B(x_i, 1/k)$. Using the first step, there exists n_k so that for all $P \in \Pi$

$$P(O_{n_k}^k) > 1 - \frac{\varepsilon}{2^k}.$$

We finally define

$$K = \overline{\cap_{k\geq 1} O_{n_k}^k}.$$

Clearly K is totally bounded ($K \subset \cap_{k\geq 1}\overline{O_{n_k}^k}$, as this last set is closed and contains $\cap_{k\geq 1} O_{n_k}^k$, so that that for all $k \geq 1$, $K \subset \overline{\cup_{i\leq n_k} B(x_i, 1/k)} \subset \cup_{i\leq n_k} B(x_i, 2/k)$) and complete (it is closed in a complete space) so that it is compact. Moreover for all $P \in \Pi$

$$P(K^c) \leq \sum_{k\geq 1} P\left((O_{n_k}^k)^c\right) \leq \varepsilon.$$

□

3.4 (E, d) is a Polish space implies $(\mathcal{M}, \rho)$ is also a Polish space

As a first nice consequence of Prohorov theorem we can now prove Theorem 11.

Proof of Theorem 11. In view of Proposition 6, it is clear that the weak convergence is metrizable. It remains to prove that $(\mathcal{M}, \rho)$ is separable and complete.

1. We prove first that if (E, d) is **separable** (and this is of course the case when (E, d) is a Polish space), **then it is the same for** $(\mathcal{M}, \rho)$. Fix $\varepsilon > 0$ and consider again the partition that we introduce in the proof of Proposition 6. For each $i \geq 1$ if D_i is non empty, consider a point $y_i \in D_i$ (note that one can not always choose x_i !). Define Π_ε the set of probability measures on E that writes

$$\sum_{i=1}^{k} r_i \delta_{y_i},$$

where $k \geq 1$ is a natural integer and the r_i, $i \geq 1$ are positive rational numbers. Clearly Π_ε is countable. Let us prove that it intersects any open ball of radius 2ε: Fix $\mathrm{P} \in \mathcal{M}$ and prove that $B(\mathrm{P}, 2\varepsilon) \cap \Pi_\varepsilon \neq \emptyset$ (the ball here is with respect to the metric ρ). Consider K large enough so that $\mathrm{P}(\bigcup_{i>K} D_i) < \varepsilon$. Consider also a family of rational numbers r_i, $i \leq K$ that approximates correctly $\mathrm{P}(D_i)$, $i \leq K$ in the sense that

$$\sum_{i=1}^{K} r_i = 1 \text{ and } \sum_{i=1}^{K} |r_i - \mathrm{P}(D_i)| < \varepsilon.$$

This is possible as $\sum_{i=1}^{K} \mathrm{P}(D_i) > 1 - \varepsilon$. It remains to check that $\mathrm{Q} := \sum_{i=1}^{K} r_i \delta_{y_i}$ belongs to $B(\mathrm{P}, 2\varepsilon)$. Consider $A \in \mathcal{F}$ and $\tilde{A} = \bigcup_{i \in I} D_i$ as in the proof of Proposition 6. One obtains

$$\mathrm{P}(A) \leq \mathrm{P}(\tilde{A}) + \varepsilon = \sum_{i \in I} \mathrm{P}(D_i) + \varepsilon \leq \sum_{i \in I} r_i + 2\varepsilon = \mathrm{Q}(\tilde{A}) + 2\varepsilon \leq \mathrm{Q}(A^{2\varepsilon}) + 2\varepsilon.$$

Using Lemma 2, this proves that $\mathrm{Q} \in B(\mathrm{P}, 2\varepsilon)$. The set $\bigcup_{n \geq 1} \Pi_{1/n}$ is thus countable and dense in $(\mathcal{M}, \rho)$.

2. Our last point is that **if (E, d) is separable and complete then $(\mathcal{M}, \rho)$ is also complete**. We consider a Cauchy sequence $(\mathrm{P}_n)_{n \geq 1}$. To prove that it converges we only need to prove that it is relatively compact. For this, using Prohorov Theorem, we prove that it is tight.

   Step 1. We prove that for all $\varepsilon, \delta > 0$ there exists finitely many $\delta-$balls $(B_i)_{1 \leq i \leq M}$ such that for all $n \geq 1$,

   $$\mathrm{P}_n(\cup_{1 \leq i \leq M} B_i) \geq 1 - \varepsilon.$$

   For this we choose η such that $0 < \eta < \frac{\varepsilon}{2} \wedge \frac{\delta}{2}$. As $(\mathrm{P}_n)_{n \geq 1}$ is a Cauchy sequence one can fix N such that for all $n \geq N$, $\rho(\mathrm{P}_n, \mathrm{P}_N) < \eta$. We consider a dense sequence $(x_i)_{i \geq 1}$ so that the balls $B(x_i, \eta)$, $i \geq 1$ cover E. For all $n \leq N$ one can thus find M large enough so that $\mathrm{P}_n(\cup_{n \leq M} B(x_i, \eta)) \geq 1 - \eta$. As $(\mathrm{P}_n)_{n \leq N}$ is a finite family one can actually choose M that works simultaneously for all $n \leq N$. Set for $i \geq 1$, $B_i = B(x_i, 2\eta)$. If $n \geq N$, from the definition of ρ,

   $$\mathrm{P}_n(\cup_{i \leq M} B_i) \geq \mathrm{P}_n((\cup_{i \leq M} B(x_i, \eta))^\eta) \geq \mathrm{P}_N(\cup_{i \leq M} B(x_i, \eta)) - \eta \geq 1 - 2\eta.$$

   If instead $n \leq N$,

   $$\mathrm{P}_n(\cup_{i \leq M} B_i) \geq \mathrm{P}_n(\cup_{i \leq M} B(x_i, \eta)) \geq 1 - \eta.$$

   Step 2. In order to conclude we mimic the argument used in Step 2 in the proof of the converse half of Theorem 12. We fix $\varepsilon > 0$. For all

$k \geq 1$, using the first step, one can find finitly many $1/k$-balls $(B_i^k)_{i \leq M_k}$ such that for all $n \geq 1$

$$\mathrm{P}_n(\cup_{1 \leq i \leq M_k} B_i^k) \geq 1 - \frac{\varepsilon}{2^k}.$$

We finally define

$$K = \overline{\cap_{k \geq 1} \cup_{1 \leq i \leq M_k} B_i^k}.$$

The set K is compact (because totally bounded and complete) and for all $n \geq 1$

$$\mathrm{P}_n(K^c) \leq \sum_{k \geq 1} \mathrm{P}\left((\cup_{1 \leq i \leq M_k} B_i^k)^c\right) \leq \varepsilon.$$

□

There are actually **other notions of convergence** on $\mathcal{M}(E)$: Fortet, Wasserstein, Total variation [COMPLETE ?]

3.5 Characteristic functions

We consider a probability measure P on $\mathbb{R}^d$ and define its *characteristic function* $\hat{\mathrm{P}}$ by

$$\hat{\mathrm{P}}(\xi) = \int_{\mathbb{R}^d} e^{i\,\xi \cdot u}\ \mathrm{P}(du) \qquad \xi \in \mathbb{R}^d.$$

It is well-known (see [2] for example) that the characteristic function characterises the law in the sense that if $\hat{\mathrm{P}} = \hat{\mathrm{Q}}$ it implies that $\mathrm{P} = \mathrm{Q}$.

The following theorem makes a deep link between weak convergence and convergence of the characteristic functions :

Theorem 13. *Consider* $(\mathrm{P}_n)_{n \geq 1}$ *and* P *probability measures on* $\mathbb{R}^d$. *The two following assertions are equivalent:*

1. $\mathrm{P}_n \implies \mathrm{P}$
2. *for all* $\xi \in \mathbb{R}^d$, $\lim_{n \to +\infty} \hat{\mathrm{P}}_n(\xi) = \hat{\mathrm{P}}(\xi)$.

However when one wants to use this theorem we have to check that the pointwise limit $(\hat{\mathrm{P}}_n(\cdot))_{n \geq 1}$ is indeed the characteristic function of some probability measure P. That is why next result is stronger and useful

Theorem 14. *Assume that* $(\mathrm{P}_n)_{n \geq 1}$ *is a sequence of probability measures on* $\mathbb{R}^d$ *such that*

1. $(\hat{\mathrm{P}}_n(\cdot))_{n \geq 1}$ *converges pointwise to some* ϕ,

2. ϕ *is continuous at* 0.

Then there exists a probability measure P *such that*

1. $\hat{\mathrm{P}} = \phi$*,*

2. $\mathrm{P}_n \implies \mathrm{P}$*.*

Proof. We only manage with the case $d = 1$.

We first prove that for all probability measure P on $\mathbb{R}$ and all $\varepsilon > 0$,

$$\mathrm{P}(|u| > 2/\varepsilon) \leq \frac{1}{\varepsilon}\int_{-\varepsilon}^{+\varepsilon}(1 - \hat{\mathrm{P}}(\xi))\,d\xi. \tag{4}$$

Indeed, from Fubini theorem

$$\frac{1}{\varepsilon}\int_{-\varepsilon}^{+\varepsilon}(1 - \hat{\mathrm{P}}(\xi))\,d\xi = 2\int\left(1 - \frac{\sin(\varepsilon u)}{\varepsilon u}\right)\,\mathrm{P}(du).$$

As $u \to 1 - \frac{\sin(\varepsilon u)}{\varepsilon u}$ is non negative and larger that $1/2$ when $|\varepsilon u| > 2$ we obtain

$$2\int\left(1 - \frac{\sin(\varepsilon u)}{\varepsilon u}\right)\,\mathrm{P}(du) \geq 2\int_{|\varepsilon u|>2}\left(1 - \frac{\sin(\varepsilon u)}{\varepsilon u}\right)\,\mathrm{P}(du) \geq \mathrm{P}(|\varepsilon u| > 2).$$

This inequality provides a control on the tail of the distribution considering how fast its characteristic function converges to 1 at 0.

We now use (4) to prove that $(\mathrm{P}_n)_{n\geq 1}$ is tight. We fix $\eta > 0$ and prove that there exists K large enough so that for all n larger than some n_0 (see Exercice 22),

$$\mathrm{P}_n(|x| > K) < \eta. \tag{5}$$

As ϕ is continuous at 0 we can choose $\varepsilon > 0$ small enough so that

$$\frac{1}{\varepsilon}\int_{-\varepsilon}^{+\varepsilon}(1 - \phi(\xi))\,d\xi < \eta.$$

As $(\hat{\mathrm{P}}_n)_{n\geq 1}$ converges pointwise to ϕ one deduces from Lebesgue theorem that for n larger that some n_0

$$\left|\frac{1}{\varepsilon}\int_{-\varepsilon}^{+\varepsilon}(1 - \hat{\mathrm{P}}_n(\xi))\,d\xi - \frac{1}{\varepsilon}\int_{-\varepsilon}^{+\varepsilon}(1 - \phi(\xi))\,d\xi\right| < \eta.$$

Thus for $n \geq n_0$

$$\left|\frac{1}{\varepsilon}\int_{-\varepsilon}^{+\varepsilon}(1 - \hat{\mathrm{P}}_n(\xi))\,d\xi\right| < 2\eta.$$

One can thus choose $K = 2/\varepsilon$ to establish (5).

From Prohorov theorem one deduces that any subsequence of $(\mathrm{P}_n)_{n\geq 1}$ admits a subsequence converging to a probability measure. All limits are however the same has they admit ϕ for characteristic function. Wee call P this limit and it clearly satisfies the conclusion of the theorem. □

Exercise 18. ***Central limit theorem.*** *Consider a family of i.i.d. centred and square integrable random variables* $(\xi_i)_{i\geq 1}$ *such that* $\mathrm{E}(\xi^2) = 1$. *Define for all* $n \geq 1$, *the variable* $S_n = \sum_{i=1}^n \xi_i$. *Then* $(S_n/\sqrt{n})$ *converges weakly to a* $\mathcal{N}(0,1)$.

3.6 Skorohod's Representation Theorem

Theorem 15. *Let* (E,d) *be a Polish space and* $(\mathrm{P}_n)_{n\geq 1}$ *be a sequence of probability measure converging to* P. *Then there exist andom variables* $(X_n)_{n\geq 1}$ *and* X *all defined on the same probability space* $(\Omega, \mathcal{G}, P)$ *such that*

1. $X_n \overset{P-p.s.}{\to} X$ *when* n *goes to infinity*

2. *for all* $n \geq 1$, X_n *has law* P_n *and* X *has law* P.

Voir Miermont par exemple

Proof. (See [3]) You should first do Exercise 21 that takes care of the case $E = \mathbb{R}$.

First step. Let us first build a single random variable on $(]0,1], \mathcal{B}(]0,1]), Leb)$ with law P. We consider for that a sequence of even finer partitions of E. We consider thus a collection $A_{\mathbf{i}}^m$ of elements in $\mathcal{F}$, where $m \geq 1$ and $\mathbf{i} = (i_1, \cdots, i_m) \in \mathbb{N}^m$ is a multi-index that is very convenient to work with as each new coordinate index (that is when we go from m to $m+1$) encodes a partition of the previous cell :

- $A_\emptyset^0 = E$,
- for all $m \geq 0$ and $\mathbf{i}$, $(A_{\mathbf{i},j}^{m+1})_{j\geq 1}$ is a partition of $A_{\mathbf{i}}^m$,
- for all $m \geq 1$ and $\mathbf{i}$, $\mathrm{diam}(A_{\mathbf{i}}^m) \leq 2^{-m}$.

We let the reader check that, as (E,d) is a Polish space it is possible to construct such a sequence of partition. In each cell $A_{\mathbf{i}}^m$ we fix a point $x_{\mathbf{i}}^m$.

For each $m \geq 1$, we define a corresponding partition of $]0,1]$ denoted by $(B_{\mathbf{i}}^m)$ where $\mathbf{i} = (i_1, \cdots, i_m)$ is again a multi-indexes. Each $B_{\mathbf{i}}^m$ is defined as an interval $]\alpha_{\mathbf{i}}^m, \beta_{\mathbf{i}}^m]$ with

$$\alpha_{\mathbf{i}}^m = \sum_{\mathbf{j}<\mathbf{i}} \mathrm{P}(A_{\mathbf{j}}^m) \quad \text{and } \beta_{\mathbf{i}}^m = \sum_{\mathbf{j}\leq\mathbf{i}} \mathrm{P}(A_{\mathbf{j}}^m),$$

where the order here is the lexicographic order. This implies (check again !) that

- for all $m \geq 0$ and $\mathbf{i}$, $Leb(B^m_{\mathbf{i}}) = \mathrm{P}(A^m_{\mathbf{i}})$,
- for all $m \geq 0$ and $\mathbf{i}$, $B^{m+1}_{\mathbf{i},j} \subset B^m_{\mathbf{i}}$.

We now define for all $m \geq 1$, the random variable Z^m on $(]0,1], \mathcal{B}(]0,1]))$ with values in E by

$$Z^m : u \in]0,1] \to Z^m(u) = \sum_{\mathbf{i}} x^m_{\mathbf{i}} 1_{B^m_{\mathbf{i}}}.$$

One can easily check that for all $u \in]0,1]$, $(Z^m(u))_{m\geq 1}$ is a Cauchy sequence in E so that it converges to some $Z(u)$. Note that the convergence is actually uniform on $]0,1]$ as for all $u \in]0,1]$

$$d(Z^m(u), Z(u)) < 2^{-m}. \tag{6}$$

Let us check now that Z has law P. For any f that is $1-$Lipschitz and $m \geq 1$,

$$\begin{aligned} &\left| \int_{]0,1]} f(Z(u))\,du - \int f(x)\,\mathrm{P}(dx) \right| \\ &\leq \left| \int_{]0,1]} f(Z(u))\,du - \int_{]0,1]} f(Z^m(u))\,du \right| + \left| \int_{]0,1]} f(Z^m(u))\,du - \int f(z)\,\mathrm{P}(dx) \right| \end{aligned}$$

Using that $(Z^m)_{m\geq 1}$ converges uniformly to Z, one easily conclude that the first term goes to 0 with m going to infinity. The second one rewrites (check again)

$$\sum_{\mathbf{i}} \int_{A^m_{\mathbf{i}}} (f(x) - f(x^m_{\mathbf{i}}))\,\mathrm{P}(dx).$$

so that, as f is 1-Lipschitz and $diam(A^m_{\mathbf{i}}) \leq 2^{-m}$, it is bounded by

$$\sum_{\mathbf{i}} \mathrm{P}(A^m_{\mathbf{i}}) 2^{-m} = 2^{-m},$$

and goes to 0 when m goes to infinity.

Second step. Let us see now how to use this to manage with a family of law. We define the sequence of partition exactly in the same way except that we ask for the following additional condition :

- for all $m \geq 1$ and $\mathbf{i}$, $\mathrm{P}(\partial A^m_{\mathbf{i}}) = 0$.

Once again we let the reader check why we can add this constraint. For all $n \geq 1$ we consider the partition $(B_{\mathbf{i}}^{m,n})$ induced by the intervals $]\alpha_{\mathbf{i}}^{m,n}, \beta_{\mathbf{i}}^{m,n}]$ defined analogously to $\alpha_{\mathbf{i}}^{m,n}$ and $\beta_{\mathbf{i}}^{m,n}$ but with P_n instead of P. We then construct exactly in the same way a random variable Z_n on on $(]0,1], \mathcal{B}(]0,1]), Leb)$ with law P_n. It remains to prove that $Z_n \overset{p.s.}{\to} Z$.

Let us first admit that for all $m \geq 1$ and $\mathbf{i}$

$$\lim_{n\to+\infty} \alpha_{\mathbf{i}}^{m,n} = \alpha_{\mathbf{i}}^{m} \quad \text{ and } \lim_{n\to+\infty} \beta_{\mathbf{i}}^{m,n} = \beta_{\mathbf{i}}^{m}. \tag{7}$$

We prove now the convergence on the set

$$D =]0,1] \setminus \{\alpha_{\mathbf{i}}^{m}, \beta_{\mathbf{i}}^{m}, m \geq 1, \mathbf{i} \in \mathbb{N}^m\}$$

that is of Lebesgue measure 1. We fix $u \in D$ and $\varepsilon > 0$ and m such that $2^{-m} < \varepsilon$. For all $n \geq 1$,

$$d(Z_n(u), Z(u)) \leq d(Z_n(u), Z_n^m(u)) + d(Z_n^m(u), Z(u)).$$

For the first term: there exists $\mathbf{i}$ such that $u \in]\alpha_{\mathbf{i}}^{m}, \beta_{\mathbf{i}}^{m}[$ and, using (7) this implies that for n large enough $u \in]\alpha_{\mathbf{i}}^{m,n}, \beta_{\mathbf{i}}^{m,n}[$ and, in consequence, $Z_n^m(u) = Z^m(u)$. The first term coincides with $d(Z_n(u), Z^m(u))$ that is less than 2^{-m} (thus less than $\varepsilon/2$) due to (6) with Z_n instead of Z. For the second term, we use again (6).

The proof will be complete once (7) proven via an iteration. Suppose that it is true for some $m \geq 0$ and consider a $m+1$-multi index $(\mathbf{i}, k)$. We observe that for all $n \geq 1$,

$$\alpha_{\mathbf{i},k}^{m+1,n} = \alpha_{\mathbf{i}}^{m,n} + \sum_{j \leq k} \mathrm{P}_n(A_{\mathbf{i},j}^{m+1}).$$

As $\mathrm{P}(\partial A_{\mathbf{i},j}^{m+1}) = 0$ for all $j \leq k$, using Theorem 6, we obtain that $\mathrm{P}_n(A_{\mathbf{i},j}^{m+1})$ converges to $\mathrm{P}(A_{\mathbf{i},j}^{m+1})$ when n goes to infinity. Moreover by iteration hypothesis, $\alpha_{\mathbf{i}}^{m,n}$ goes to $\alpha_{\mathbf{i}}^{m}$ when n goes to infinity. This gives the desired convergence for $\alpha_{\mathbf{i},k}^{m+1,n}$.

□

3.7 More exercices

Exercise 19. *Consider probability measure* $(\mathrm{P}_n)_{n\geq 1}$ *and* P *on* $(E, \mathcal{F})$ *such that for all continuous function* f ***with bounded support***

$$\int f \, d\mathrm{P}_n \overset{n\to+\infty}{\to} \int f \, d\mathrm{P}.$$

Prove that $\mathrm{P}_n \implies \mathrm{P}$*. One could first prove for* $x \in E$ *and* $r > 0$, $\phi_r \mathrm{P}_n \implies \phi_r \mathrm{P}$ *where* ϕ_r *is a non negative continuous with bounded support function such that* $\phi_r(x) = 1$ *if* $d(y, x) \leq r$*. Use Portmanteau theorem to conclude.*

Exercise 20. *True or False ?*

1. *If* $\mathrm{P}_n \implies \mathrm{P}$ *and* P *is atomic then* P_n *is also atomic for* n *large enough.*
2. *If* $\mathrm{P}_n \implies \mathrm{P}$ *(where they are probability measures on* $\mathbb{R}^d$, $d \geq 1$*) and* P *is absolutely continuous with respect to Lebesgue measure, then it is the same for* P_n *for* n *large enough.*
3. *Same context as prévious question : we suppose this time that* P_n *is absolutely continuous with respect to Lebesgue measure. Does it imply that* P *is also absolutely continuous ?*

Exercise 21 (Proof of Skorohod's theorem when $E = \mathbb{R}$)**.** *Let* P *be a probability measure on* $\mathbb{R}$ *and* F *be its cumulative distribution function. As it is non decreasing on can consider it generalised inverse*

$$F^{-1} : u \in]0,1[\to \inf\{x \in \mathbb{R},\ F(x) \geq u\}.$$

1. *Prove that* F^{-1} *is also a càdlàg non decreasing function.*
2. *Prove that if* U *is an uniform on* $[0,1]$ *random variable then* $F^{-1}(U)$ *has law as* P*.*
3. *We now turn to the proof of Skorohod's Representation Theorem in the case* $E = \mathbb{R}$*. With same notation as in Theorem 15, we denote by* F_n *the cumulative distribution function of* P_n*. Prove that for all* $u \in]0,1[$*,*

$$F^{-1}(u) \leq \liminf F_n^{-1}(u) \leq \limsup F_n^{-1}(u) \leq F^{-1}(u+).$$

4. *Conclude*

Theorem 16. *Let* $(\mathrm{P}_n)_{n \geq 1}$ *and* P *be probability measures on* $\mathbb{R}$ *and* $(F_n)_{n \geq 1}$ *and* F *be their cumulative distribution function. Then the two following assertions are equivalent.*

(a) $\mathrm{P}_n \implies \mathrm{P}$

(b) For all x *that is a continuity point of* F, $\lim_{n \to +\infty} F_n(x) = F(x)$*.*

Proof. <u>*1* $\Longrightarrow$ *2*</u>. If x is continuity point of F the $\mathrm{P}(\partial(-\infty, x]) = \mathrm{P}(\{x\}) = 0$ so that $F_n(x) = \mathrm{P}_n((-\infty, x]) \to \mathrm{P}((-\infty, x]) = F(x)$ when n goes to infinity.

<u>*2* $\Longrightarrow$ *1*</u>. One can complete Exercise 21 or alternatively prove that $(\mathrm{P}_n)_{n\geq 1}$ is tight. Indeed for all $\varepsilon > 0$, one can fix K large enough so that $F(K) - F(-K) \geq 1-\varepsilon$ and neither K neither $-K$ are discontinuity point of F. Using that $(F_n(K))_{n\geq 1}$ converges to $F(K)$ and $(F_n(-K))_{n\geq 1}$ converges to $F(-K)$ we obtain that for all n larger than some n_0, $|F_n(K) - F_n(-K)| \geq 1 - 2\varepsilon$. From Prohorov theorem we know that any subsequence of $(\mathrm{P}_n)_{n\geq 1}$ admits a weakly converging subsequence. Moreover the limit is unique as it is admits, using the direct implication of the theorem, F as cumulative distribution function. □

Exercise 22. *1. Prove that any finite family of probability measures on a Polish space (E, d) is tight.*

2. Prove that a family of probability measures $(\mathrm{P}_n)_{n\geq 1}$ on a Polish space (E, d) is tight if and only if for some $n_0 \geq 1$ the family $(\mathrm{P}_n)_{n\geq n_0}$ is tight.

Exercise 23. *Let (E, d) be a metric space.*

1. Consider a sequence $(x_n)_{n\geq 1}$ that converges to x. Prove that

$$\delta_{x_n} \Longrightarrow \delta_x.$$

2. Assume now that $(\delta_{x_n})_{n\geq 1}$ converges weakly to some probability measure P. Prove that P is a Dirac probability measure at some point $x \in E$ and that $(x_n)_{n\geq 1}$ converges to x.

<u>**Correction of point 2.**</u> Consider for all $N \geq 1$ the closed set $F_N = \overline{\cup_{n\geq N}\{x_n\}}$. From Portmanteau theorem

$$\mathrm{P}(F_N) \geq \limsup \delta_{x_n}(F_N) = 1.$$

As this sets are decreasing with N we obtain $\mathrm{P}(\cap_{N\geq 1} F_N) = 1$. The set $\cap_{N\geq 1} F_N$ is the set of all accumulation point for the sequence $(x_n)_{n\geq 1}$. As it has probability 1 it contains at least one point x. From question 1, $\delta_{x_{\phi_n}}$ converges weakly to δ_x and P that are thus equal.

Finally if (x_n) does not converge to x there is $\varepsilon > 0$ so that infinitely many x_k lie outside $B(x, \varepsilon)$. That implies that $\liminf \delta_{x_n}(B(x, \varepsilon)) = 0$ that is a contradiction with Portmanteau theorem as $\delta_x(B(x, \varepsilon)) = 1$.

Exercise 24. *We consider* $(X_n)_{n\geq 1}$*,* $(Y_n)_{n\geq 1}$*,* X*,* Y *random variables that take values in a Polish space* (E,d)*. We assume that* $X_n \implies X$ *and* $Y_n \implies Y$*. We assume moreover that for all* $n \geq 1$*,* X_n *and* Y_n *are independent and that* X *and* Y *are independent. Prove that*

$$(X_n, Y_n) \implies (X, Y).$$

Exercise 25. *(Slutzky theorem) We consider* $(X_n)_{n\geq 1}$ *and* $(Y_n)_{n\geq 1}$ *two sequences of random variables that take values in a Polish space* (E,d)*.*

1. *Prove that if* $(X_n)_{n\geq 1}$ *converges in probability, it also converges in law.*
2. *Prove that if* $(X_n)_{n\geq 1}$ *converges in law to some constant* c*, it also converges in probability to* c*. Prove also that this property is wrong if we do not suppose anymore the limit to be a constant random variable.*
3. *Prove Slutzky theorem: We consider* $(X_n)_{n\geq 1}$*,* $(Y_n)_{n\geq 1}$*,* X *random variables that take values in a Polish space* (E,d) *and* $c \in E$*. We assume that*

   (a) $X_n \implies X$*,*

   (b) $(Y_n)_{n\geq 1}$ *converges in probability to* c*.*

   Then

   $$(X_n, Y_n) \implies (X, c).$$

4 Weak convergence in $\mathscr{C}([0,1])$

4.1 Processes as random functions

A process is a collection $(X_t)_{t\in[0,1]}$ of real random variables built on the same probability space $(\Omega, \mathcal{G}, \mathrm{Q})$. For $\omega \in \Omega$, the process defines a function :

$$\begin{aligned} X(\omega) : \mathbb{R}^+ &\to \mathbb{R} \\ t &\mapsto X_t(\omega). \end{aligned}$$

We say that the process $(X_t)_{t\in[0,1]}$ is a continuous process if for all $\omega \in \Omega$ the function $X(\omega)$ is continuous (that is lies in $\mathscr{C}([0,1])$). We wonder if a process can be viewed as a random variable that is if one can find a sigma algebra $\mathcal{F}$ such that

$$X : (\Omega, \mathcal{G}, \mathrm{Q}) \to (\mathscr{C}([0,1]), \mathcal{F})$$

is measurable. This is actually the case if one choose $\mathcal{F} = \mathcal{B}$ the Borel (for the uniform norm) sigma algebra. Indeed, from Proposition 4, we know that $\mathcal{B}$ coincides with $\mathcal{E}$ the sigma fiel generated by cylinder sets. We consider

$$C = \bigcap_{i=1}^{n} \{\pi_{t_i} \in B_i\}$$

a cylinder set and check that

$$\{X \in C\} = \{\omega \text{ s.t. } \big(t \mapsto \; X_t(\omega)\big) \in C\} = \bigcap_{i=1}^{n} \{X_{t_i} \in B_i\}$$

that clearly belongs to $\mathcal{G}$.

The process $(X_t)_{t\in[0,1]}$ can thus be viewed as random function that takes values in $(\mathscr{C}([0,1]), ||\cdot||_\infty)$ endowed with its Borel sigma algebra. Its law is a probability measure on this space and we are in the confortable framework of Polish spaces that we have studied in the previous chapters.

4.2 Convergence of finite dimensional marginals

Définition 7. *The finite dimensional distribution of a sequence of processes $(X^{(N)})_{N\geq 1}$ converges to X if for all $n \geq 1$ and all $0 \leq t_0 \leq \cdots \leq t_n \leq 1$, $(X^{(N)}_{t_0}, \cdots, X^{(N)}_{t_n})$ converges weakly in $\mathbb{R}^n$ to $(X_{t_0}, \cdots, X_{t_n})$ when N goes to $+\infty$. When this is the case we denote this convergence by*

$$X^{(N)} \overset{(D_f)}{\Longrightarrow} X.$$

We note that weak convergence of a sequence $(\mathrm{P}_n)_{n\geq 0}$ implies the convergence of the finite dimensional marginal. Indeed, for all $n \geq 1$ and all $0 \leq t_0 \leq \cdots \leq t_n \leq 1$, the application

$$f \in (\mathscr{C}([0,1]), ||\cdot||_\infty) \to (f(t_0), \cdots, f(t_n)) \in \mathbb{R}^n$$

is continuous and we conclude easily with Theorem 8.

However, the converse is false in general as we can see from the following easy example : consider for $N \geq 1$ the process $(X^{(N)})_{0\leq t\leq 1}$ with law the Dirac measure on

$$f_N : t \in [0,1] \to Nt1_{[0,1/N]} + N(2/N - t)1_{(1/N,2/N]},$$

that is just a tent function. For all $n \geq 1$ and $0 \leq t_0 \leq \cdots \leq t_n$, the laws of $(X^{(N)}_{t_0}, \cdots, X^{(N)}_{t_n})$ are the same for N large enough so that

$$X^{(N)} \overset{(D_f)}{\Longrightarrow} X,$$

where X has law δ_0. However $(f_N)_{n\geq 1}$ does not converge for the uniform norm to the zero function and this implies, see Exercise 23, that

$$X^{(N)} \not\Rightarrow X,$$

for the uniform topolgy. Thus, to establish weak convergence in $\mathscr{C}([0,1])$ we need another ingredient that is *tightness*.

4.3 A general strategy to prove convergence in law of continuous processes

We have now all the ingredients we need to study the weak convergence on the functional space $\mathscr{C}([0,1])$. The reader may reread, if necessary, Theorem 12 (Prohorov), Exercise 15 and Proposition 5. Here is a general strategy to establish that a sequence $(\mathrm{P}_n)_{n\geq 0}$ of probability measures on $\mathscr{C}([0,1])$ converges weakly to P :

1. We first prove **convergence of the finite dimensional marginals** of $(\mathrm{P}_n)_{n\geq 0}$ to those of P. We recall the reader that this is not enough to conclude as explained at the end of the previous section.

2. We then prove that the sequence $(\mathrm{P}_n)_{n\geq 0}$ is **tight**.

This is enough to conclude: Using Exercice 15 we just have to prove that from any subsequence of $(\mathrm{P}_n)_{n\geq 0}$ one can extract a further subsequence weakly converging to P. Let us thus consider a subsequence $(\mathrm{P}_{\phi(n)})_{n\geq 0}$. As a subsequence of $(\mathrm{P}_n)_{n\geq 0}$ it is also tight and using Theorem 12 one deduce that it is relatively compact, that is it admits a converging further subsequence. Using the first ingredient, the limit has same finite dimensional marginals as P. From Proposition 5 this implies that the limit is actually P. This concludes the proof.

We sumarize what we have just proven in

Theorem 17. *Let* $(\mathrm{P}_n)_{n\geq 0}$ *and* P *be probability measures on* $(\mathscr{C}([0,1]), \mathcal{F})$. *Assume that*

- *finite-dimensional marginals of* $(\mathrm{P}_n)_{n\geq 0}$ *converges to those of* P
- $(\mathrm{P}_n)_{n\geq 0}$ *is tight.*

Then $(\mathrm{P}_n)_{n\geq 0}$ *converges weakly to* P.

In consequence we have to give criteria for tightness in $\mathscr{C}([0,1])$.

Remark 4. *The converse proposition of Theorem 17 is also true as convergence in law on* $\mathscr{C}([0,1])$ *implies both finite dimensional marginals convergence (see the remark after Definition 7) and tightness via Prohorov's theorem.*

4.4 Tightness in $\mathscr{C}([0,1])$

We first recall Ascoli theorem that provides a characterisation of compact sets in $\mathscr{C}([0,1])$. From Heine theorem, any continuous function in $\mathscr{C}([0,1])$ is actually uniformly continuous. For $f \in \mathscr{C}([0,1])$ one can define the *modulus of continuity* by

$$w(f,\delta) = \sup\{|f(x) - f(y)|,\ x, y \in [0,1], |x-y| < \delta\}, \qquad \delta > 0.$$

Uniform continuity for a function f defined on $[0,1]$ is equivalent to $\lim_{\delta\to 0} w(f,\delta) = 0$.

Theorem 18 (Ascoli)**.** *Let* $\mathcal{A}$ *be a subset of* $\mathscr{C}([0,1])$. *Then* $\mathcal{A}$ *is relatively compact if and only if*

1. $\sup_{f\in\mathcal{A}} |f(0)| < +\infty$
2. $\lim_{\delta\to 0} \sup_{f\in\mathcal{A}} w(f,\delta) = 0$.

Remark 5. *We formulate here Ascoli's theorem on $\mathscr{C}([0,1])$ but it can be generalised with the same ideas to the space $\mathscr{C}(K)$ of continuous functions defined on some compact K.*

Proof. Direct implication. We suppose first that $\mathcal{A}$ is relatively compact. It implies that it is bounded as subset of the norm vector space $(\mathscr{C}([0,1]), ||\cdot||_\infty)$ and in particular $\sup_{f\in\mathcal{A}} f(0) < +\infty$. As K is a compact set, from Heine Theorem, any continuous function f on K is uniformly continuous that is $w(f,\delta)$ converges to 0 when δ goes to 0. Moreover $w(f,\delta)$ seen as a function of f is continuous and $w(f,\delta)$ seen as a function of δ is non decreasing. Dini Theorem thus implies that the pointwise convergence is actually uniform. Indeed for any sequence $(\delta_k)_{k\geq 1}$ decreasing to 0 and $\varepsilon > 0$ we introduce the sets

$$V(\delta_k) = \{f \in \mathcal{A},\ w(f,\delta_k) < \varepsilon\}.$$

These sets are open because of the continuity of w in f. Moreover due to the pointwise convergence $\mathcal{A} = \cup_{k\geq 1} V(\delta_k)$. As $\mathcal{A}$ is relatively compact we can extract from this covering a finite covering :

$$\mathcal{A} = \cup_{k\leq K} V(\delta_k) = V(\delta_K),$$

using for the last equality the monotonicity with δ. This is enough to conclude as for $k \geq K$ and any $f \in \mathcal{A}$

$$w(f,\delta_k) \leq w(f,\delta_K) < \varepsilon.$$

We turn to the *converse implication.* We have to prove that $\bar{\mathcal{A}}$ is totally bounded and complete. Actually as $\mathscr{C}([0,1])$ is a Polish space (see Theorem 4) it is complete and as $\bar{\mathcal{A}}$ is closed it implies that it is complete. We thus just have to prove that $\bar{\mathcal{A}}$ is totally bounded.

We first prove that $\mathcal{A}$ is bounded. Fix k large enough so that $\sup_{f\in\mathcal{A}} w(f,1/k) \leq 1$. This implies that for all $f \in \mathcal{A}$ and $x \in [0,1]$

$$|f(x)| \leq |f(0)| + \sum_{i=1}^{k} |f(ix/k) - f((i-1)x/k)| \leq \sup_{f\in\mathcal{A}} |f(0)| + k < +\infty.$$

This implies that $\mathcal{A}$ is bounded for the uniform norm by some M.

We now prove that $\mathcal{A}$ is totally bounded. Fix $\varepsilon > 0$. We consider in $[-M,M]$, $N = \lceil 2M/\varepsilon \rceil$ points $Y = \{y_i,\ i = 1,\cdots,N\}$ such that any $t \in [-M,M]$ is at distance at most ε of Y. We choose now k large enough so that $\sup_{f\in\mathcal{A}} w(f,1/k) < \varepsilon$. We define the set G of functions g in $\mathscr{C}([0,1])$ that are linear on each interval $[j/k,(j+1)/k]$ and such $g(j/k) \in Y$, $j = 0,\cdots k-1$.

We prove now that the open balls of radius 2ε and centred in the finite set G cover $\mathcal{A}$.

Fix $f \in \mathcal{A}$. By definition of Y there exists for all $j = 0, \cdots k-1$ a point $y(j) \in Y$ such that $|f(j/k) - y(j)| < \varepsilon$. One can thus consider the function g in G such that for all $j = 0, \cdots k-1$, $g(j/k) = y(j)$. This implies of course that for all $j = 0, \cdots k-1$, $|f(j/k) - g(j/k)| < \varepsilon$. If t lies in $[j/k, (j+1)/k]$ for some j then $|f(t) - f(j/k)| < \varepsilon$ and $|f(t) - f((j+1)/k)| < \varepsilon$ because $\sup_{f\in\mathcal{A}} w(f, 1/k) < \varepsilon$. It also holds that $|g(j/k) - f(j/k)| < \varepsilon$ and $|g((j+1)/k) - f((j+1)/k)| < \varepsilon$ by definition of g. As $g(t)$ lies between $g(j/k)$ and $g((j+1)/k)$ whatever how they are ordered, this implies $|f(t) - g(t)| < 2\varepsilon$. $\square$

With this theorem in hands one can now formulate a characterization of tightness. In the next theorem we consider a sequence of continuous processes that have to be thought as random continuous functions. Each of them defines thus a law on $(\mathscr{C}([0,1]), \mathcal{F})$ and our goal is to provide a characterisation for tightness of this sequence.

Theorem 19. *Let $(X^N)_{N\geq 1}$ be a sequence of continuous processes. The three following assertions are equivalent.*

1. *The family of laws on $\mathscr{C}([0,1])$, $(\mathcal{L}(X^N))_{N\geq 1}$ is tight*
2. (a) *The family of laws on $\mathbb{R}$, $(\mathcal{L}(X_0^N))_{N\geq 1}$ is tight*
   (b) *For all $\eta > 0$, $\lim_{\delta\to 0} \sup_N \mathrm{P}(w(X^N, \delta) > \eta) = 0$*
3. (a) *The family of laws on $\mathbb{R}$, $(\mathcal{L}(X_0^N))_{N\geq 1}$ is tight*
   (b) *For all $\eta > 0$, $\lim_{\delta\to 0} \limsup_N \mathrm{P}(w(X^N, \delta) > \eta) = 0$*

Proof. **1. $\Longrightarrow$ 2.** To prove (a), we use that the coordinate application in 0: $\pi_0 : f \in \mathscr{C}([0,1]) \to f(0) \in \mathbb{R}$ is continuous. This implies (see Exercise 17) that if $(\mathcal{L}(X^N))_{N\geq 1}$ is tight then its image by π_0, $(\mathcal{L}(X_0^N))_{N\geq 1}$ is also tight.

We turn to (b). Fix $\eta, \varepsilon > 0$. As the sequence $(\mathcal{L}(X^N))_{N\geq 1}$ is tight there exists a compact set $K_\varepsilon \subset \mathscr{C}([0,1])$ such that for all $N \geq 1$,

$$\mathrm{P}(X^N \in K_\varepsilon) \geq 1 - \varepsilon.$$

From Ascoli theorem, $\lim_{\delta\to 0} \sup_{f\in K_\varepsilon} w(f, \delta) = 0$ so that there exists δ_0 such that for $\delta \leq \delta_0$, $\sup_{f\in K_\varepsilon} w(f, \delta) \leq \eta$. This implies that for $\delta \leq \delta_0$ and for all $N \geq 1$

$$\mathrm{P}(w(X^N, \delta) > \eta) \leq \mathrm{P}(X^N \notin K_\varepsilon) < \varepsilon.$$

2. $\Longrightarrow$ 3. There is here nothing to prove !

3. $\implies$ 2. We only have to prove that 3.(b) implies 2.(b). Fix $\eta, \varepsilon > 0$. From 3.(b) there exists δ_0 so that for $\delta \leq \delta_0$ there exists $N_0(\delta)$ so that for $N \geq N_0(\delta)$, $\mathrm{P}(w(X^N, \delta) > \eta) < \varepsilon$. Actually as $\delta \to w(f, \delta)$ is non decreasing one can use $N_0(\delta_0)$ for all $\delta \leq \delta_0$. For $i \leq N_0(\delta_0)$, $\mathcal{L}(X^i)$ is tight using Theorem 3 so that the finite family $(\mathcal{L}(X^i))_{i \leq N_0(\delta_0)}$ is also tight. Arguing as in the proof of "1. $\implies$ 2." there exists $\delta_1 > 0$ so that for $\delta < \delta_1$

$$\sup_{i \leq N_0(\delta_0)} \mathrm{P}(w(X^i, \delta) > \eta) < \varepsilon.$$

We define $\bar{\delta} = \min(\delta_0, \delta_1)$ so that for $\delta < \bar{\delta}$, agains as $\delta \to w(f, \delta)$ is non decreasing, it holds that for all $N \geq 1$, $\mathrm{P}(w(X^N, \delta) > \eta) < \varepsilon$.

2. $\implies$ 1. We use here again Ascoli theorem. Fix $\varepsilon > 0$. From 2.(a) there exists a compact $\bar{K} \subset \mathbb{R}$ such that for all $N \geq 1$

$$\mathrm{P}(X_0^N \notin \bar{K}) < \varepsilon.$$

This $\bar{K}$ is included in $[-M, M]$ for some lare enough M. From 2.(b), for all $k \geq 1$ there exists δ_k such that for all $N \geq 1$,

$$\mathrm{P}(w(X^N, \delta_k) > \frac{1}{k}) < \frac{\varepsilon}{2^{k+1}}.$$

With a basic union bound we obtain for all $N \geq 1$

$$\mathrm{P}(X_0^N \in [-M, M], \cap_{k \geq 1} w(X^N, \delta_k) \leq \frac{1}{k}) \geq 1 - 2\varepsilon. \tag{8}$$

We set

$$K_\varepsilon = \{f \in \mathscr{C}([0,1]) \text{ such that } |f(0)| \leq M,\ \cap_{k \geq 1} w(f, \delta_k) \leq \frac{1}{k}\}.$$

Using Ascoli theorem this a relatively compact set and (8) rewrites

$$\mathrm{P}(X^N \in K_\varepsilon) \geq 1 - 2\varepsilon.$$

□

This last theorem is actually a simple reformulation of the definition of tightness using Ascoli theorem as can be seen from the proof. It is however difficult to establish criterium 2.(b) and 3.(b). We now see two (only) sufficient conditions that are more convenient to work with.

Proposition 7. *Let $(X^N)_{N \geq 1}$ be a sequence of continuous processes. Suppose that*

1. *The family of laws on* $\mathbb{R}$, $(\mathcal{L}(X_0^N))_{N\geq 1}$ *is tight*

2. *For all* $\eta, \varepsilon > 0$, *there exists* $\delta \in]0,1[$ *such that*

$$\limsup_{N} \sup_{t\in[0,1]} \frac{1}{\delta} \mathrm{P}(\sup_{s\in[t,t+\delta]} |X_s^N - X_t^N| > \eta) \leq \varepsilon.$$

Then the family of laws on $\mathscr{C}([0,1])$, $(\mathcal{L}(X^N))_{N\geq 1}$ *is tight.*

Proof. We prove that condition 2 above implies condition 3.(b) in Theorem 19. Of course condition 2 rewrites: for all $\eta > 0$

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \sup_{t\in[0,1]} \frac{1}{\delta} \mathrm{P}(\sup_{s\in[t,t+\delta]} |X_s^N - X_t^N| > \eta) = 0,$$

so that the question is how we can enter the "sup" in the probability.

We first observe that for all $f \in \mathscr{C}([0,1])$, all $\delta > 0$ (let us admit that $1/\delta$ is an integer) and all $0 \leq s, t \leq 1$ such that $|t - s| \leq \delta$

$$|f(s) - f(t)| \leq 3 \max_{k\leq 1/\delta} \sup_{s\in[k\delta,(k+1)\delta]} |f(s) - f(k\delta)|.$$

Indeed if s and t are in the same interval $[k\delta, (k+1)\delta]$ for some $k \geq 0$ then $|f(s) - f(t)| \leq |f(s) - f(k\delta)| + |f(t) - f(k\delta)|$. If they are not in the same interval they are in neighbour intervals that is for some $k \geq 0$, $k\delta \leq s \leq (k+1)\delta \leq t \leq (k+2)\delta$ and this time

$$|f(s) - f(t)| \leq |f(s) - f(k\delta)| + |f(k\delta) - f((k+1)\delta)| + |f(t) - f((k+2)\delta)|.$$

This implies that

$$\{w(X^N, \delta) > \eta\} \subset \bigcup_{k=0\cdots,1/\delta} \left\{ \sup_{s\in[k\delta,(k+1)\delta]} |X_s^N - X_{k\delta}^N| \geq \frac{\eta}{3} \right\}$$

and with an union bound

$$\begin{aligned} \mathrm{P}(w(X^N, \delta) > \eta) &\leq \sum_{k=0}^{1/\delta} \mathrm{P}(\sup_{s\in[k\delta,(k+1)\delta]} |X_s^N - X_{k\delta}^N| \geq \frac{\eta}{3}) \\ &\leq \frac{1}{\delta} \sup_{t} \mathrm{P}(\sup_{s\in[t,t+\delta]} |X_s^N - X_t^N| > \frac{\eta}{3}), \end{aligned}$$

and this concludes the proof. □

Another classic characterisation of tightness is provided by this criterium due to Kolmogorov

Proposition 8. *Let $(X^N)_{N\geq 1}$ be a sequence of continuous processes. Suppose that*

1. *The family of laws on $\mathbb{R}$, $(\mathcal{L}(X_0^N))_{n\geq 1}$ is tight*
2. *There exists $\alpha, \beta, C > 0$ such that for all $0 \leq s, t \leq 1$ and all $N \geq 1$*

$$\mathrm{E}(|X_s^N - X_t^N|^\alpha) \leq C|t-s|^{1+\beta}.$$

Then the family of laws on $\mathscr{C}([0,1])$, $(\mathcal{L}(X^N))_{N\geq 1}$ is tight.

Proof. It is a consequence of Kolmogorov's continuity criterium. □

4.5 Brownian motion and Wiener measure

We remind three classical, and of course equivalent, definitions of the Brownian motion:

Définition 8 (B1)**.** *We call **Brownian motion** any **continuous** process $(B_t)_{t\geq 0}$ with **stationary and independent increments** and such that $B_0 = 0$ a.s. and for all $0 \leq s \leq t$*

$$B_t - B_s \rightsquigarrow \mathcal{N}(0, t-s).$$

Définition 9 (B2)**.** *We call **Brownian motion** any **continuous** process $(B_t)_{t\geq 0}$ that is **gaussian centred** with variance defined for all $0 \leq s \leq t$ by*

$$R(s,t) = s \wedge t.$$

Définition 10 (B3)**.** *We call **Brownian motion** any **continuous** process $(B_t)_{t\geq 0}$ such that $B_0 = 0$ a.s. and for all $0 \leq s \leq t$*

$$\begin{aligned} &B_t - B_s \perp\!\!\!\perp \sigma(B_r;\, 0 \leq r \leq s), \\ &B_t - B_s \rightsquigarrow \mathcal{N}(0, t-s). \end{aligned}$$

[Canonical process and Wiener measure]
[Discussion about the existence of the Wiener measure]

4.6 Donsker Theorem

[Motivation. Example of continuous function on $\mathscr{C}([0,1])$: sup,...]

The goal of Donsker theorem is to provide a description of the rescaled sum of i.i.d. square integrable variables viewed as a random function. Let us consider $(\xi_k)_{k\geq 1}$ an i.i.d. family of square integrable real random variables with mean 0 and variance 1 (of course if the mean is not 0 or the variance not 1 one can still work with a renormalised version of ξ). We consider for $n \geq 1$,

$$S_n = \sum_{k=1}^{n} \xi_k,$$

and for all $N \geq 1$, we the define the random function, that is the process,

$$S_t^N = \frac{1}{\sqrt{N}} S_{\lfloor Nt \rfloor} + (Nt - \lfloor Nt \rfloor) \frac{1}{\sqrt{N}} \xi_{\lfloor Nt \rfloor + 1} \qquad 0 \leq t \leq 1.$$

[Add a picture]

Theorem 20 (Donsker Theorem)**.** *Assume that $(\xi_k)_{k\geq 1}$ is an i.i.d. family of square integrable real random variables with mean* 0 *and variance* 1*. Then*

$$(S_t^N)_{0\leq t\leq 1} \overset{(law)}{\to} (B_t)_{0\leq t\leq 1},$$

where $(B_t)_{0\leq t\leq 1}$ is a standard Brownian motion and the convergence is relative to the uniform topology on $\mathscr{C}([0,1])$.

Proof. From Theorem 17 we have to prove that

1. For all $n \geq 1$ and all $0 \leq t_0 \leq \cdots \leq t_n \leq 1$, $(S_{t_0}^N, \cdots, S_{t_n}^N)$ converges weakly in $\mathbb{R}^n$ to $(B_{t_0}, \cdots, B_{t_n})$ when N goes to $+\infty$.
2. The family of laws $(\mathcal{L}(S^N))_{N\geq 1}$ on $\mathscr{C}([0,1])$ is tight.

Proof of 1. This is mainly an application of the central limit theorem. Fix $n \geq 1$ and $0 \leq t_0 \leq \cdots \leq t_n \leq 1$. For all $1 \leq i \leq n$

$$\frac{S_{\lfloor Nt_{i+1} \rfloor} - S_{\lfloor Nt_i \rfloor}}{\sqrt{N}} = \frac{S_{\lfloor Nt_{i+1} \rfloor} - S_{\lfloor Nt_i \rfloor}}{\sqrt{\lfloor Nt_{i+1} \rfloor - \lfloor Nt_i \rfloor}} \frac{\sqrt{\lfloor Nt_{i+1} \rfloor - \lfloor Nt_i \rfloor}}{\sqrt{N}}.$$

Using the central limit theorem, the first part converges in law to a gaussian $\mathcal{N}(0,1)$ while the deterministic second one converges to $\sqrt{t_{i+1} - t_i}$. From Slutsky theorem one deduces that $(\frac{S_{\lfloor Nt_{i+1} \rfloor} - S_{\lfloor Nt_i \rfloor}}{\sqrt{N}})_{N\geq 1}$ converges to a gaussian $\mathcal{N}(0, t_{i+1} - t_i)$ that is also the law of $B_{t_{i+1}} - B_{t_i}$. One can say actually more as for all $N \geq 1$ the variables $\frac{S_{\lfloor Nt_{i+1}\rfloor} - S_{\lfloor Nt_i\rfloor}}{\sqrt{N}}$, $i = 0, \cdots n$ are independent. This implies that

$$\left(\frac{S_{\lfloor Nt_0\rfloor}}{\sqrt{N}}, \frac{S_{\lfloor Nt_1\rfloor} - S_{\lfloor Nt_0\rfloor}}{\sqrt{N}}, \cdots, \frac{S_{\lfloor Nt_n\rfloor} - S_{\lfloor Nt_{n-1}\rfloor}}{\sqrt{N}}\right) \overset{N\to+\infty}{\Longrightarrow} (B_{t_0}, B_{t_1} - B_{t_0} \cdots, B_{t_n} - B_{t_{n-1}}),$$

as, from the definition of the Brownian motion this last vector has law $\mathcal{N}(0, K)$ with

$$K = \begin{pmatrix} t_0 & & & & \\ & \ddots & & & \\ & & t_{i+1} - t_i & & \\ & & & \ddots & \\ & & & & t_n - t_{n-1} \end{pmatrix}.$$

From this one can easily prove that

$$\left(\frac{S_{\lfloor Nt_0\rfloor}}{\sqrt{N}}, \cdots, \frac{S_{\lfloor Nt_n\rfloor}}{\sqrt{N}}\right) \overset{N\to+\infty}{\Longrightarrow} (B_{t_0}, \cdots, B_{t_n}).$$

Moreover for all $t \in [0,1]$ and $\varepsilon > 0$

$$\mathrm{P}\left(\left|(Nt - \lfloor Nt\rfloor)\frac{1}{\sqrt{N}}\xi_{\lfloor Nt\rfloor+1}\right| > \varepsilon\right) = \mathrm{P}\left(\xi > \varepsilon\frac{\sqrt{N}}{Nt - \lfloor Nt\rfloor}\right)$$

goes to 0 with N going to $+\infty$ and more generally

$$\left((Nt_0 - \lfloor Nt_0\rfloor)\frac{1}{\sqrt{N}}\xi_{\lfloor Nt_0\rfloor+1}, \cdots, (Nt_n - \lfloor Nt_n\rfloor)\frac{1}{\sqrt{N}}\xi_{\lfloor Nt_n\rfloor+1}\right) \overset{Proba.}{\longrightarrow} 0.$$

Using again Slutsky theorem we obtain that $(S^N_{t_0}, \cdots, S^N_{t_n})$ converges weakly in $\mathbb{R}^n$ to $(B_{t_0}, \cdots, B_{t_n})$. We turn to the proof of tightness that is more intricated.

Proof of 2. We use for that the criterium stated in Proposition 7. The first point is obviously verified and we are left to prove the second one. Fix $\varepsilon, \eta > 0$. We want to prove that there exists $\delta > 0$ such that

$$\limsup_{N\to\infty} \sup_{t\in[0,1]} \frac{1}{\delta}\mathrm{P}(\sup_{s\in[t,t+\delta]} |S^N_s - S^N_t| > \eta) \leq \varepsilon.$$

For all $0 \leq t \leq 1$ and $\delta > 0$, for all $t \leq s \leq t + \delta$,

$$\begin{aligned} |S^N_s - S^N_t| &\leq \frac{1}{\sqrt{N}}\left(\left|\sum_{k=\lfloor Nt\rfloor+1}^{\lfloor Ns\rfloor} \xi_k\right| + |\xi_{\lfloor Nt\rfloor+1}| + \sup_{t\leq s\leq t+\delta} |\xi_{\lfloor Ns\rfloor+1}|\right) \\ &\leq \frac{1}{\sqrt{N}}\left|\sum_{k=\lfloor Nt\rfloor+1}^{\lfloor Ns\rfloor} \xi_k\right| + \frac{2}{\sqrt{N}} \sup_{t\leq s\leq t+\delta} |\xi_{\lfloor Ns\rfloor+1}| \end{aligned}$$

so that

$$\frac{1}{\delta}\mathrm{P}(\sup_{s\in[t,t+\delta]}|S_s^N - S_t^N| > \eta)$$
$$\leq \frac{1}{\delta}\mathrm{P}\left(\sup_{s\in[t,t+\delta]}\left|\sum_{k=\lfloor Nt\rfloor+1}^{\lfloor Ns\rfloor}\xi_k\right| > \frac{\eta\sqrt{N}}{2}\right) + \frac{1}{\delta}\mathrm{P}\left(\sup_{t\leq s\leq t+\delta}|\xi_{\lfloor Ns\rfloor+1}| > \frac{\eta\sqrt{N}}{4}\right).$$

We start with the second term and observe that for all $N \geq 1$ and $t \in [0,1]$ with the notation $I = \{\lfloor Ns\rfloor + 1, t \leq s \leq t+\delta\}$ (that satisfies $Card(I) \leq \lfloor N\delta\rfloor + 2$)

$$\begin{aligned}\mathrm{P}\left(\sup_{t\leq s\leq t+\delta}|\xi_{\lfloor Ns\rfloor+1}| > \frac{\eta\sqrt{N}}{4}\right) &\leq \mathrm{P}\left(\exists i \in I \text{ s.t. } |\xi_i| > \frac{\eta\sqrt{N}}{4}\right)\\ &\leq (\lfloor N\delta\rfloor + 2)\mathrm{P}\left(|\xi_1| > \frac{\eta\sqrt{N}}{4}\right)\\ &\leq (\lfloor N\delta\rfloor + 2)\frac{16}{N\eta^2}\mathrm{E}\left(\xi_1^2 1_{|\xi_1|>\frac{\eta\sqrt{N}}{4}}\right),\end{aligned}$$

(where we used that for all non negative random variable

$$\mathrm{P}(X > a) \leq \mathrm{E}(X^2 1_{X>a})$$

as $X^2 1_{X>a} > a^2 1_{X>a}$) and this last inequality is uniform in $t \in [0,1]$. The second term is thus bounded, for all $\delta > 0$, by

$$\frac{\lfloor N\delta\rfloor + 2}{\delta}\frac{16}{N\eta^2}\mathrm{E}\left(\xi_1^2 1_{|\xi_1|>\frac{\eta\sqrt{N}}{4}}\right)$$

that goes to 0 when N goes to infinity because ξ is square integrable.

To manage with the first one we need the following lemma

Lemma 4. *For all $\lambda > 0$ and all $N \geq 1$,*

$$\mathrm{P}(\max_{n\leq N}|S_n| > \lambda\sqrt{N}) \leq 2\mathrm{P}(|S_N| > (\lambda - \sqrt{2})\sqrt{N}).$$

Proof. Note that it is not an usual version of reflexion principle as our variables $(\xi_k)_{k\geq 1}$ are not supposed to be symmetric. We call $(\mathcal{F}_n^\xi)_{n\geq 1}$ the filtration generated by the process $(\xi_k)_{k\geq 1}$ (that is for all $n \geq 1$, $\mathcal{F}_n^\xi = \sigma(\xi_1, \cdots, \xi_n)$) and τ the hitting time of $\lambda\sqrt{N}$ by $(|S_n|)_{n\geq 1}$,

$$\tau = \inf\{n \geq 1,\ |S_n| \geq \lambda\sqrt{N}\}.$$

Note that τ is a hitting time for the filtration $\mathcal{F}^\xi$ and that

$$\{\max_{n\le N}|S_n|\ge \lambda\sqrt{N}\}=\{\tau\le N\} \tag{9}$$

so that,

$$\begin{aligned}
\mathrm{P}(\max_{n\le N}|S_n|\ge \lambda\sqrt{N}) &= \mathrm{P}(\tau\le N)\\
&= \mathrm{P}(|S_N|\ge(\lambda-\sqrt{2})\sqrt{N})+\sum_{k=1}^{N}\mathrm{P}(|S_N|<(\lambda-\sqrt{2})\sqrt{N},\tau=k).
\end{aligned}$$

We easily observe that for all $1\le k\le N$, $\{|S_N|<(\lambda-\sqrt{2})\sqrt{N},\tau=k\}\subset\{|S_N-S_k|\ge\sqrt{2N}\}$ so that for all $1\le k\le n$

$$\begin{aligned}
\mathrm{P}(|S_N|<(\lambda-\sqrt{2})\sqrt{N},\tau=k) &\le \mathrm{P}(|S_N-S_k|\ge\sqrt{2N},\tau=k)\\
&= \mathrm{P}(|S_N-S_k|\ge\sqrt{2N})\mathrm{P}(\tau=k),
\end{aligned}$$

as $\{|S_N-S_k|\ge\sqrt{2N}\}$ is independent from $\mathcal{F}^\xi_k$ while $\{\tau=k\}\in\mathcal{F}^\xi_k$. Using that

$$\mathrm{P}(|S_N|<(\lambda-\sqrt{2})\sqrt{N},\tau=k)\le\frac{\mathrm{E}(|S_N-S_k|^2)}{2N}\mathrm{P}(\tau=k)=\frac{N-k+1}{2N}\mathrm{P}(\tau=k)\le\frac{1}{2}\mathrm{P}(\tau=k),$$

and we finally obtain

$$\mathrm{P}(\max_{n\le N}|S_n|\ge\lambda\sqrt{N})\le\mathrm{P}(|S_N|\ge(\lambda-\sqrt{2})\sqrt{N})+\frac{1}{2}\mathrm{P}(\tau\le N).$$

This enough to conclude using (9). Note that $\sqrt{2}$ is somehow arbitrary in this lemma and one could write an analogous version with $\sqrt{\theta}$ instead at the cost of replacing the factor 2 in the right hand term by $\theta/(\theta-1)$. □

Using Lemma 4, the fact that $\lfloor N(t+\delta)\rfloor-\lfloor Nt\rfloor\le\lceil N\delta\rceil$ and also that $(\xi_k)_{k\ge1}$ are i.i.d. we obtain

$$\begin{aligned}
\frac{1}{\delta}\mathrm{P}\left(\sup_{s\in[t,t+\delta]}\left|\sum_{k=\lfloor Nt\rfloor+1}^{\lfloor Ns\rfloor}\xi_k\right|>\frac{\eta\sqrt{N}}{2}\right) &\le \frac{1}{\delta}\mathrm{P}\left(\max_{i\le\lceil N\delta\rceil}\left|\sum_{k=1}^{i}\xi_k\right|>\frac{\eta\sqrt{N}}{2}\right)\\
&\le\frac{2}{\delta}\mathrm{P}\left(\left|\frac{S_{\lceil N\delta\rceil}}{\sqrt{\lceil N\delta\rceil}}\right|>\frac{\eta\sqrt{N}}{2\sqrt{\lceil N\delta\rceil}}-\sqrt{2}\right)\\
&\le\frac{2}{\delta}\mathrm{P}\left(\left|\frac{S_{\lceil N\delta\rceil}}{\sqrt{\lceil N\delta\rceil}}\right|>\frac{\eta}{2\sqrt{\delta+1/N}}-\sqrt{2}\right)\\
&\le\frac{2}{\delta}\mathrm{P}\left(\left|\frac{S_{\lceil N\delta\rceil}}{\sqrt{\lceil N\delta\rceil}}\right|>\frac{\eta}{4\sqrt{\delta}}-\sqrt{2}\right),
\end{aligned}$$

for N large enough.

From the central limit theorem $\frac{S_{\lceil N\delta\rceil}}{\sqrt{\lceil N\delta\rceil}}$ converges to a gaussian $\mathcal{N}(0,1)$ and we obtain for Z a $\mathcal{N}(0,1)$ random variable

$$\begin{aligned}\limsup_{N\to\infty}\sup_{t\in[0,1]}\frac{1}{\delta}\mathrm{P}\left(\sup_{s\in[t,t+\delta]}\left|\sum_{k=\lfloor Nt\rfloor+1}^{\lfloor Ns\rfloor}\xi_k\right|>\frac{\eta\sqrt{N}}{2}\right)&\\ \leq \limsup_{N\to\infty}\frac{2}{\delta}\mathrm{P}\left(\left|\frac{S_{\lceil N\delta\rceil}}{\sqrt{\lceil N\delta\rceil}}\right|>\frac{\eta}{4\sqrt{\delta}}-\sqrt{2}\right)&\\ \leq \frac{2}{\delta}\mathrm{P}\left(|Z|>\frac{\eta}{4\sqrt{\delta}}-\sqrt{2}\right)&\\ \leq \frac{2}{\delta(\frac{\eta}{4\sqrt{\delta}}-\sqrt{2})^3}\mathrm{E}\left(|Z|^3\right)&\end{aligned}$$

and this last quantity goes to 0 with δ going to 0. This concludes the proof.

□

4.7 More exercises

Exercise 26. *Which of the following functional are continuous on $(\mathscr{C}([0,1]),||\cdot||_\infty)$? In the case where they are not continuous everywhere precise where their continiuty point.*

1. $\phi : f \in \mathscr{C}([0,1]) \to ||f||_\infty$
2. $\phi : f \in \mathscr{C}([0,1]) \to \int_{[0,1]} f(t)\, dt$
3. *For* $a \geq 0$, $\phi : f \in \mathscr{C}([0,1]) \to T_a(f) = \inf\{t \geq 0,\ |f(t) - f(0)| \geq a\}$.

Exercise 27. *We use the same notations as for Donsker theorem. Prove that*

$$\frac{1}{\sqrt{n}}\max_{k\leq n} S_k \implies \sup_{t\in[0,1]} B_t.$$

Prove that $\sup_{t\in[0,1]} B_t \overset{(law)}{=} |Z|$ *where* Z *has law* $\mathcal{N}(0,1)$.

4.8 Some applications

[COMPLETE]

5 Weak convergence in $\mathcal{D}([0,1])$

This section is based on the Julien Poisat's notes written 2023 - 2024. It also used an old course at Paris 6 by Jean Bertoin. And off course the indispensable [1].

We denote by $\mathcal{D}([0,1]$ the space of real-valued *càdlàg* (right-continuous with left limits) functions defined on $[0,1]$ that is function $f:[0,1]\to\mathbb{R}$ that satisfies for all $t\in[0,1]$

$$f(t)=\lim_{\substack{s\to t\\ s>t}} f(s)$$

and $\lim_{\substack{s\to t\\ s<t}} f(s)$ exists in $\mathbb{R}$ we note

$$f(t^-)=\lim_{\substack{s\to t\\ s<t}} f(s).$$

Motivations : martingales admits càdlàg versions ; càdlàg is the good space for modeling paths with jumps ; càdlàg processes appears often as limit processes in limit theorems (especially when summing variables with heavy tails)

We need a *good* topology on $\mathcal{D}([0,1])$: it has to be large enough so that many functionals (as supremum, integrals,...) are continuous but not to large in order to keep the confortable framework of a separable and complete space.

5.1 Skorhokhod topology

Let us start with basic properties of càdlàg functions.

Proposition 9. *For every $f\in\mathcal{D}([0,1])$,*

1. $\|f\|_\infty := \sup_{0\le t\le 1}|f(t)| < +\infty$ *but the supremum may not be achieved.*
2. $\forall\delta>0$, *the set* $\{t\in(0,1]\colon |f(t)-f(t^-)|\ge\delta\}$ *is finite.*
3. *The set of discontinuity points* $\mathrm{Disc}(f)$ *is at most countable.*
4. $\lim_{\delta\to 0} w'(f,\delta)=0$ *(non-increasing limit), where*

$$w'(f,\delta) := \inf_{\substack{0=t_0<t_1<\ldots<t_k=1\\ \min t_i-t_{i-1}>\delta}} \max_{1\le i\le k} \Delta(f,[t_{i-1},t_i))$$

with the notation

$$\Delta(f,I) := \sup\{|f(s)-f(t)|\colon s,t\in I\}, \qquad I\subseteq[0,1].$$

Proof of Proposition 9. 1. Assume by contradiction that there exists a sequence $(x_n)_{n\geq 0}$ in $[0,1]$ such that for all $n \geq 0$ $|f(x_n)| > n$. One can extract a subsequence converging to some $x \in [0,1]$ and actually a further subsequence that is monotonous. This leads to a contradiction as f is càdlàg.

2. Assume by contradiction that for some $\delta > 0$ there exists a sequence $(x_n)_{n\geq 0}$ in $[0,1]$ such that for all $n \geq 0$ $|f(x_n) - f(x_n^-)| \geq \delta$. One can extract a subsequence converging to some $x \in [0,1]$ and actually a further subsequence $(x_{\phi_n})_{n\geq 0}$ that is monotonous say non-increasing but the other case can be treated in the same way. For all $n \geq 1$ there exists $y_n \in [x, x_{\phi_n}]$ such that $|f(x_{\phi_n}) - f(y_n)| > \delta/2$ and we obtain easily a contradiction with the fact that f is càd.

3. This is a direct consequence of previous point as

$$\mathrm{Disc}(f) = \cup_{n\geq 1}\{t \in (0,1]\colon |f(t) - f(t^-)| \geq \frac{1}{n}\}$$

4. The proof relies on the following lemme

   Lemma 5. *Let $f \in \mathcal{D}([0,1])$ and $\varepsilon > 0$. There exists $k \in \mathbb{N}$ and a subdivision $0 = t_0 < t_1 < \ldots < t_k = 1$ such that*

   $$\max_{1\leq i\leq k} \Delta(f, [t_{i-1}, t_i)) < \varepsilon.$$

   Proof. Fix $f \in \mathcal{D}([0,1])$ and $\varepsilon > 0$.

   Define $\mathcal{J}_\varepsilon$ as all t's between 0 and 1 such that $[0,t)$ may be decomposed into finitely many such subintervals. As f is right continuous at 0 we easly obtain that $\mathcal{J}_\varepsilon$ is non empty and contains $[0,\alpha]$ for α small enough. One can defined $\mathcal{T}_\varepsilon = \sup \mathcal{J}_\varepsilon$ and it is clear that $\mathcal{T}_\varepsilon > 0$. One can check that $\mathcal{T}_\varepsilon \in \mathcal{J}_\varepsilon$ (with a similar argument to the left this time) and that $\mathcal{T}_\varepsilon < 1$ leads to a contradiction (still with the same argument). □

   The function w' may be seen as a càdlàg version of the modulus of continuity w in $\mathscr{C}([0,1])$.

□

We now introduce a new metric, called Skorokhod metric, and investigate how projection maps interact with the induced topology and Borel σ-field. Let Λ be the space of increasing continuous functions $\Lambda\colon [0,1] \to [0,1]$ such that $\Lambda(0) = 0$ and $\Lambda(1) = 1$ (the time parametrizations). One can check that the following map $d_S\colon D \times D \mapsto \mathbb{R}_+$:

$$d_S(f,g) := \inf_{\Lambda \in \Lambda} \max(\|\Lambda - \mathrm{Id}\|_\infty, \|f - g \circ \Lambda\|_\infty)$$

defines a metric that is called the Skorokhod metric [ADD DETAILS ?]. We use $\mathcal{T}$ to denote the topology induced by d_S.

This distance is quite natural as we would like two functions having similar jumps at close times to be close. It is indeed the case as d_S makes possible a transformation of time. For example one can check that if $f = 1_{[0,1/2)}$ and $g = 1_{[0,1/2+\varepsilon)}$ then $d_S(f,g) = \varepsilon$ goes to 0 with ε going to 0 while $\|f-g\|_\infty = 1$ for all $0 < \varepsilon < 1/2$.

One can check that both

$$\sup : f \in \mathcal{D}([0,1]) \to \|f\|_\infty \in \mathbb{R}$$

and

$$f \in \mathcal{D}([0,1]) \to \int_a^b f(t)\ dt \in \mathbb{R}$$

are continuous functionals for Skorokhod topology. [COMPLETE]

Warning about Skorhokod topology : It is not a vectorial space topology as $+$ is not continuous. For example if $f_n = 1_{[0,1/2+1/n[}$ and $f = 1_{[0,1/2[}$ then $f_n \overset{d_S}{\to} f$ but $f_n + f \overset{d_S}{\not\to} 2f$. [DETAILS]

Exercise 28. *Let f and g be functions in $\mathcal{D}([0,1])$. It is clear that that $d_S(f,g) \le \|f-g\|_\infty$. Prove the following property :*

assume that $(f_n)_{n\ge1}$ in $\mathcal{D}([0,1])$ and $f \in \mathscr{C}([0,1])$. Then $f_n \overset{d_S}{\to} f$ implies $f_n \overset{\|\cdot\|_\infty}{\to} f$

Theorem 21. *The space $(\mathcal{D}([0,1]), \mathcal{T})$ is a Polish space in the sense that one can find a metric that generates $\mathcal{T}$ and for which $\mathcal{D}([0,1])$ is complete. However note that $(\mathcal{D}([0,1]), d_S)$ is not complete.*

The following exercise provides a good counter-example to prove that $(\mathcal{D}([0,1]), d_S)$ is not complete.

Exercise 29. *Consider $f_n := 1_{[0,1/2^n)}$.*

1. *Show that (f_n) is a Cauchy sequence for d_S.*

2. *Assume that $d_S(f_n, f) \overset{n\to\infty}{\longrightarrow} 0$ for some $f \in D$. Show that $f(t) = 0$, $\forall t \in [0,1]$.*

3. *Conclude.*

Lemma 6. *Let $(f_n), f$ in $\mathcal{D}([0,1])$ such that (f_n) converges to f w.r.t. the Skorokhod metric. Then,*

1. *$f_n(t) \to f(t)$ for all $t \notin \mathrm{Disc}(f)$ and thus Lebesgue-a.e. on $[0,1]$.*

2. *(f_n) is uniformly bounded (for the uniform metric).*

Proof. 1. Show that $f_n(t)$ converges to $f(t)$ for every $t \notin \mathrm{Disc}(f)$.

2. Use that $\|\Lambda_n - \mathrm{Id}\|_\infty \to 0$ and $\|f \circ \Lambda_n - f_n\|_\infty \to 0$ for some sequence (Λ_n) in Λ along with Item (3) from Corollary 9.

□

In the following we denote by $\mathcal{B}$ the Borel σ-algebra on $\mathcal{D}([0,1])$ induced by the Skorokhod metric (generated by the Skorokhod-open sets). Let us now investigate the continuity and measurability properties of the projection maps.

As in the continuous case we understand a càdlàg random process $(X_t)_{t\in[0,1]}$ (= a collection of random variables such that each path is càdlàg) as a random function that is a random variable with values in $(\mathcal{D}([0,1]), \mathcal{B})$. It is possible as

1. as in the continuous case, the borel $\sigma-$algebra (for $\mathcal{T}$) coincides with the cylinder $\sigma-$algebra that is

$$\mathcal{B} = \sigma(\pi_t,\ 0 \le t \le 1)$$

where π_t, $t \in [0,1]$ the coordinate application from $\mathcal{D}([0,1])$ in $\mathbb{R}$.

2. for every $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in [0,1]$, $\pi_{t_1,\ldots t_k} : (\mathcal{D}([0,1]), \mathcal{B}) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is measurable.

Both points above are not trivial. They are actually consequence of the following proposition that say more and will be useful later.

Proposition 10. *The following properties hold:*

1. *π_0 and π_1 are continuous from $(\mathcal{D}([0,1]), d_S)$ to $(\mathbb{R}, |\cdot|)$.*

2. *For every $t \in (0,1)$, π_t is continuous at $f \in \mathcal{D}([0,1])$ if and only if f is continuous at t.*

3. *For every $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in [0,1]$, $\pi_{t_1,\ldots t_k} : (\mathcal{D}([0,1]), \mathcal{B}) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is measurable.*

4. *For every $T \subseteq [0,1]$ that satisfies $1 \in T$ and $\bar{T} = [0,1]$, $\mathcal{B} = \sigma(\pi_t, t \in T)$.*

Proof. 1. This follows from $\Lambda(0) = 0$ and $\Lambda(1) = 1$ when $\Lambda \in \Lambda$.

2. If f is continuous at t, decompose $f(t) - f_n(t)$ as $[f(t) - f(\Lambda_n(t))] + [f(\Lambda_n(t)) - f_n(t)]$ for some appropriate sequence (Λ_n) and use the triangular inequality. If f is not continuous at t, construct a sequence (f_n) as a counterexample [shift the jump at x to $x + \varepsilon$ and let $\varepsilon \to 0$].

3. It is enough to consider $k = 1$ and $t < 1$. Show that for every $\varepsilon > 0$, the map $f \in (D, d_S) \mapsto h_\varepsilon(f) := \frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s) \mathrm{d}s \in (\mathbb{R}, |\cdot|)$ is continuous. Conclude by noting that π_t is the pointwise limit of $h_{1/m}$ as $m \to \infty$.

4. The proof is divided into several steps:

   - Assume that $0 \in T$ (w.l.o.g.).
   - Show that for every $m \in \mathbb{N}$, there exists $k \in \mathbb{N}$, $s_0, \ldots, s_k \in T$ such that $0 = s_0 < \ldots < s_k = 1$ and $\max_{1 \le i \le k}(s_i - s_{i-1}) < 1/m$. Let $\sigma_m := \{s_i\}_{0 \le i \le k}$ in the sequel.
   - For all $f \in \mathcal{D}([0,1])$, let $A_m f \in \mathcal{D}([0,1])$ be the step function such that (i) f and $A_m f$ coincide on σ_m and (ii) $A_m f$ is constant on the intervals $[s_{i-1}, s_i)$ for every $1 \le i \le k$. Check that

$$d_S(f, A_m f) \le \max\left(\tfrac{1}{m}, w'(f, \tfrac{1}{m})\right).$$

   - Deduce thereof that $(A_m f)$ converges to f in the Skorokhod topology.
   - Show that A_m is measurable from $(\mathcal{D}([0,1]), \sigma(\pi_t,\ t \in T))$ to $(\mathcal{D}([0,1]), \mathcal{B})$.
   - Conclude.

□

Corollary 1. *The last item implies using Dynkin's lemma that finite dimensional marginal laws characterises laws on $(\mathcal{D}([0,1], \mathcal{B})$ and actually even only finite dimensional marginal laws with with coordinates in a subset $T \subseteq [0,1]$ that satisfies $1 \in T$ and $\bar{T} = [0,1]$.*

5.2 A general stragegy to prove weak convergence in $\mathcal{D}([0,1])$

In this section we explain the general mechanism for proving weak convergence of càdlàg processes. We use actually almost the same strategy as in $\mathscr{C}([0,1])$:

1. as $(\mathcal{D}([0,1],\mathcal{B})$ is a Polish space Prohorov's theorem takes care of the characterisation of relatively compact set. Note by the way that we often only use the direct implication in Prohorov's theorem for which there is not need of assumptions of completeness nor separability. We still have however to produce a criterium for tigthness (see below section 5.3).

2. In $\mathscr{C}([0,1])$ it was possible to prove the unicity of the accumulation points using that the projections are continuous and that finite dimensional marginal laws characterise the law. In $\mathcal{D}([0,1])$ it is not true anymore that projections are continuous (see Proposition 10) and thus weak convergence does not imply anymore convergence of finite dimensional marginal laws.

In this section we manage to fix this second point. In the next one we give a criterium for tightness in $\mathcal{D}([0,1])$. We start with the following proposition:

Proposition 11. *Let* $\mathrm{P} \in \mathcal{M}_1(\mathcal{D}([0,1]),\mathcal{B})$ *and define*

$$T_{\mathrm{P}} := \{t \in [0,1] \colon \mathrm{P}(\mathrm{Disc}(\pi_t)) = 0\}.$$

Then

1. T_{P} *contains* 0 *and* 1.

2. $[0,1] \setminus T_{\mathrm{P}}$ *is at most countable.*

Proof. The first item is a direct consequence of Item (1) in Proposition 10 so we only need to prove the second one. Let $t \in (0,1)$ and

$$J_t = \{f \in D \colon f(t) \neq f(t^-)\}.$$

Then $t \in T_{\mathrm{P}}$ if and only if $\mathrm{P}(J_t) = 0$, by Proposition 10. For every $\varepsilon > 0$, let

$$J_t(\varepsilon) := \{f \in D \colon |f(t) - f(t^-)| > \varepsilon\},$$

so that $\mathrm{P}(J_t)$ is the nondecreasing limit of $\mathrm{P}(J_t(\varepsilon))$ as $\varepsilon \to 0$. Therefore,

$$\mathrm{P}(J_t) > 0 \Rightarrow \exists \delta, \varepsilon \in \mathbb{Q} \cap (0,1) \colon \mathrm{P}(J_t(\varepsilon)) \geq \delta.$$

Assume that the assertation on the right-hand side holds for some fixed pair $\delta, \varepsilon \in \mathbb{Q} \cap (0,1)$ and infinitely many t_n's, $n \geq 1$. Then, for that pair (ε, δ), one has

$$\mathrm{P}\Big(\limsup_{n\to\infty} J_{t_n}(\varepsilon)\Big) \geq \delta.$$

This is absurd, since any $f \in \mathcal{D}([0,1])$ must have only finitely many jumps larger than ε in absolute value. This means that the set

$$\bigcup_{\delta,\varepsilon\in\mathbb{Q}\cap(0,1)} \{t \in [0,1] \colon \mathrm{P}(J_t(\varepsilon)) \geq \delta\}$$

is at most countable, which concludes the proof. □

As a consequence of this Lemma and of the mapping theorem (Theorem 8) if $t_1, \cdots, t_k$ are in T_{P} and $\mathrm{P}_n \implies \mathrm{P}$ then

$$\mathrm{P}_n \pi^{-1}_{t_1,\cdots,t_k} \implies \mathrm{P}\pi^{-1}_{t_1,\cdots,t_k}.$$

Theorem 22. *Let* (P_n) *and* P *be probability measures on* $(\mathcal{D}([0,1]), \mathcal{B})$. *Suppose that the two following conditions hold:*

1. (P_n) *is tight;*
2. $(\mathrm{P}_n\pi^{-1}_{t_1,\ldots,t_k})$ *converges weakly to* $\mathrm{P}\pi^{-1}_{t_1,\ldots,t_k}$ *for every* $t_1, \ldots t_k \in T_{\mathrm{P}}$.

Then, (P_n) *converges weakly to* P *w.r.t. the Skorokhod topology.*

Proof. The sequence (P_n) is tight, hence relatively compact by Prokhorov's theorem. Then, it is enough to show that P is the only possible limit for any subsequence that converges weakly. Let $\varphi(n)$ be such a subsequence and Q the limit. By the continuous mapping theorem, for all $t_1, \ldots, t_k \in T_{\mathrm{Q}}$,

$$\mathrm{P}_{\varphi(n)}\pi^{-1}_{t_1,\ldots,t_k} \to \mathrm{Q}\pi^{-1}_{t_1,\ldots,t_k}, \qquad n \to \infty, \quad \text{weakly.}$$

From our second assumption, we get for all $t_1, \ldots, t_k \in T_{\mathrm{P}} \cap T_{\mathrm{Q}}$,

$$\mathrm{P}\pi^{-1}_{t_1,\ldots,t_k} = \mathrm{Q}\pi^{-1}_{t_1,\ldots,t_k}.$$

Moreover, by Proposition 11, $T_{\mathrm{P}} \cap T_{\mathrm{Q}}$ contains $\{0,1\}$ and is dense in $[0,1]$ (as its complement is countable). We conclude that $\mathrm{P} = \mathrm{Q}$ by using Corollary 1. □

5.3 Tightness in $\mathcal{D}([0,1])$

We close this section by stating without proof a tightness criterion for càdlàg processes.

Theorem 23. *Let $(X^N)_{N\geq 1}$ be a sequence of càdlàg processes. The two following assertions are equivalent.*

1. *The family of laws on $\mathcal{D}([0,1])$, $(\mathcal{L}(X^N))_{N\geq 1}$ is tight*
2. (a) *The family of laws on $\mathbb{R}$, $(\mathcal{L}(\|X^N\|_\infty))_{N\geq 1}$ is tight*
   (b) *For all $\eta > 0$, $\lim_{\delta\to 0} \limsup_N \mathrm{P}(w'(X^N,\delta) > \eta) = 0$*

This criterion relies on the analogue of the Ascoli theorem for càdlàg functions :

Theorem 24. *A subset $A \subset \mathcal{D}([0,1])$ is relatively compact if and only if*

1. $\sup_{f\in A} \|f\|_\infty < +\infty$
2. $\lim_{\delta\to 0} \sup_{f\in A} w'(f,\delta) = 0$

Note that this criterium is not easy to establish as we need to compute the càdlàg modulus of continuity. There are other criterium much easier to manage with but it is the end of this course !

References

[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] W. Feller. *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons, Inc., New York-London-Sydney, second edition, 1971.

[3] G. Miermont. *Théorèmes limites et processus de Poisson, Notes de cours de M2.* 2011-2012.

[4] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, third edition, 1987.