Convergence of probability measures and limit theorems

Work in progress

Michel Pain

Largely based on lecture notes by François Chapon

KMAS9AG1

Université de Toulouse

M2RI 2026–2027

Contents

Preamble

This notes correspond to the first part of the course *Convergence of probability measures, limit theorems and optimal transport*. I am deeply grateful to François Chapon for sending me his lecture notes, on which these notes are largely based. The main reference for this first part of the course is the book [1] by Billingsley, more precisely Sections 1 through 8 (except for Section 4). The other references [2, 3] are used for background on topology and measure theory.

References

[1] Patrick 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] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

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

1 Introduction

1.1 A motivation: scaling limit of random walks

Let $(\xi_i)_{i\geq 1}$ be i.i.d. random variables, with $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[\xi_1^2] = 1$, and consider the random walk $(S_n)_{n\geq 0}$ defined by $S_0 = 0$ and for $n \geq 1$,

$$S_n \coloneqq \xi_1 + \cdots + \xi_n.$$

The central limit theorem gives the asymptotic position of the rescaled random walk:

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{(d)}} N,$$

where N is a normal random variable with distribution $\mathcal{N}(0, 1)$, and where $\xrightarrow{\text{(d)}}$ means convergence in distribution.

But one can wonder what happens for the *whole* trajectory of the walk, and if one has a *functional* central limit theorem. More precisely, for any $n \geq 1$, we can consider the interpolated rescaled random walk up to time n, defined as a function of $t \in [0, 1]$ by

$$X_t^{(n)} \coloneqq \frac{1}{\sqrt{n}}\Big(S_{\lfloor nt\rfloor} + (nt - \lfloor nt\rfloor)\xi_{\lfloor nt\rfloor+1}\Big),$$

where $\lfloor\cdot\rfloor$ denotes the floor function, i.e. $\lfloor u\rfloor$ is the largest integer such that $\lfloor u\rfloor \leq u < \lfloor u\rfloor + 1$. Thus, for any $t \in [\frac{k}{n}, \frac{k+1}{n})$,

$$X_t^{(n)} = \frac{1}{\sqrt{n}}(S_k + (nt - k)\xi_{k+1}),$$

so $X_t^{(n)}$ is the affine interpolation between $\frac{1}{\sqrt{n}}S_k$ and $\frac{1}{\sqrt{n}}S_{k+1}$, see Figure 1.

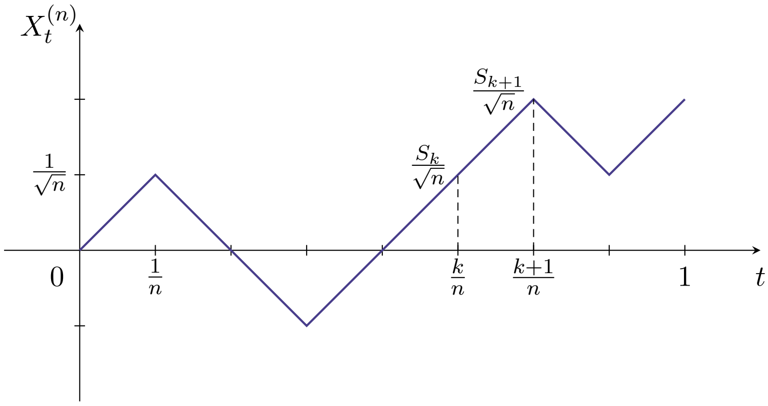


Figure 1: A realisation of interpolated random walk in the case where ξ_1 follows the Rademacher distribution.

Now we can consider the random walk at different times. Fix some times $t_0 = 0 < t_1 < \cdots < t_k \leq 1$. Our goal is to study the joint convergence of $(X_{t_1}^{(n)}, X_{t_2}^{(n)}, \ldots, X_{t_k}^{(n)})$. To this end, it is more convenient to start by working at times which are of the form m/n, so that the affine interpolation plays no role. For $i \in [\![0, k]\!]$, let $m_i = \lfloor nt_i\rfloor$. Then, for each $i \in [\![1, k]\!]$, we have

$$X_{m_i/n}^{(n)} - X_{m_{i-1}/n}^{(n)} = \frac{S_{m_i} - S_{m_{i-1}}}{\sqrt{n}} = \frac{\sqrt{m_i - m_{i-1}}}{\sqrt{n}} \cdot \frac{S_{m_i} - S_{m_{i-1}}}{\sqrt{m_i - m_{i-1}}} \xrightarrow[n\to\infty]{\text{(d)}} N_i,$$

where $N_i \sim \mathcal{N}(0, t_i - t_{i-1})$, using the central limit theorem again, the fact that $\sqrt{m_i - m_{i-1}}/\sqrt{n} \to \sqrt{t_i - t_{i-1}}$, Slutsky's theorem[1] and the scaling property of the normal distribution. The increments

[1]It states that if $(X_n)_{n\geq 1}$ converges in distribution to some random variable X and $(Y_n)_{n\geq 1}$ converges in probability to some *constant* y, then $(X_n, Y_n)_{n\geq 1}$ converges in distribution to (X, y). In particular, combined with the continuous mapping theorem, it shows that $(Y_n X_n)_{n\geq 1}$ converges in distribution to yX.

of the random walk being independent, we deduce[2] that

$$\left(X^{(n)}_{m_1/n}, X^{(n)}_{m_2/n} - X^{(n)}_{m_1/n}, \dots, X^{(n)}_{m_k/n} - X^{(n)}_{m_{k-1}/n}\right) \xrightarrow[n\to\infty]{\text{(d)}} (N_1, N_2, \dots, N_k),$$

where we now assume that the random variables $N_1, N_2, \dots, N_k$ are independent. By the continuous mapping theorem[3], we get that

$$\left(X^{(n)}_{m_1/n}, X^{(n)}_{m_2/n}, \dots, X^{(n)}_{m_k/n}\right) \xrightarrow[n\to\infty]{\text{(d)}} (N_1, N_1 + N_2, \dots, N_1 + N_2 + \dots + N_k).$$

Furthermore, for all $i \in [\![1, k]\!]$, recalling that $m_i = \lfloor nt_i \rfloor$,

$$X^{(n)}_{t_i} - X^{(n)}_{m_i/n} = \frac{nt_i - \lfloor nt_i \rfloor}{\sqrt{n}} \xi_{m_i+1} \xrightarrow[n\to\infty]{(\mathbb{P})} 0.$$

Combining this with the previous convergence (using Slutsky's theorem) shows

$$\left(X^{(n)}_{t_1}, X^{(n)}_{t_2}, \dots, X^{(n)}_{t_k}\right) \xrightarrow[n\to\infty]{\text{(d)}} (N_1, N_1 + N_2, \dots, N_1 + N_2 + \dots + N_k).$$

Hence, for any fixed number of times, we have convergence of the *finite-dimensional distributions* of the sequence of random function $(X^{(n)}_t, t \in [0, 1])$.

But we would like to have the convergence in distribution of the whole random function $(X^{(n)}_t, t \in [0, 1])$, which does not follow from the finite-dimensional distribution convergence. This is the content of Donsker theorem or Donsker invariance principle, which states that the rescaled random walk converges in distribution to the *Brownian motion*:

$$\left(X^{(n)}_t, t \in [0, 1]\right) \xrightarrow[n\to\infty]{\text{(d)}} (B_t, t \in [0, 1]),$$

where $(B_t, t \in [0, 1])$ is a stochastic process that satisfies:

- $B_0 = 0$,
- for all $k \geq 1$, for all $t_0 = 0 < t_1 < \dots < t_k \leq 1$, the random variables

  $$B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}}$$

  are independent with $B_{t_j} - B_{t_{j-1}} \sim \mathcal{N}(0, t_j - t_{j-1})$ (one says that the increments are independent and stationary),
- $t \mapsto B_t$ is almost surely continuous.

See Figure 2 for an illustration of Brownian motion.

The stochastic process $(B_t, t \in [0, 1])$ is thus a random variable with values in $\mathcal{C}([0, 1])$, the space of continuous functions from $[0, 1]$ to $\mathbb{R}$, whose distribution has some prescribed properties: for any $k \geq 1$, and all $0 \leq t_1 < \dots < t_k \leq 1$, the random vector $(B_{t_1}, \dots, B_{t_k})$ is a centered Gaussian vector in $\mathbb{R}^k$ with covariance given by $\mathbb{E}[B_s B_t] = \min\{s, t\}$. It is not obvious that such a process exists! And to prove Donsker's theorem, we will need criteria to show convergence in distribution for random variables with values in $\mathcal{C}([0, 1])$, i.e. random continuous functions.

[2] We use here that if $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ converge in distribution to X and Y respectively, and if, for any $n \geq 1$, X_n and Y_n are independent, then $(X_n, Y_n)_{n\geq 1}$ converges in distribution to (X, Y'), where Y' has the same distribution as Y and is independent of X.

[3] It states that if a sequence $(X_n)_{n\geq 1}$ of r.v. with values in E converges in distribution to X and $f \colon E \to F$ is continuous, then $(f(X_n))_{n\geq 1}$ of r.v. with values in E converges in distribution to $f(X)$.

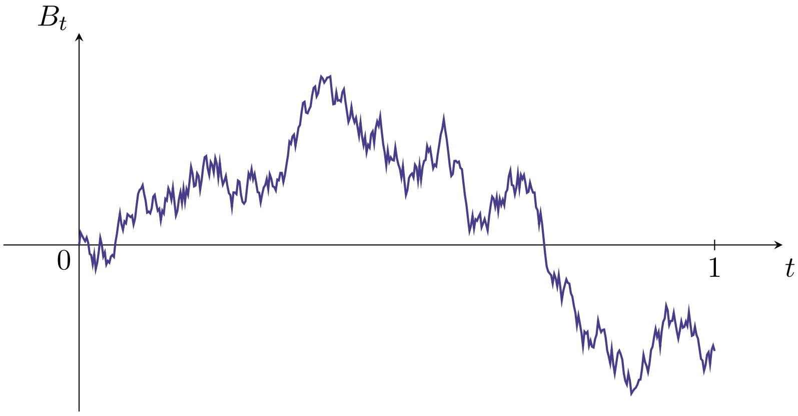


Figure 2: A Brownian trajectory.

1.2 Content of these notes

To prove convergence in distribution of a sequence $(X_n)_{n\geq 1}$ of random variables in $\mathbb{R}^d$ towards a limit X, one rarely uses the definition, which requires to prove the convergence of $\mathbb{E}[f(X_n)]$ towards $\mathbb{E}[f(X)]$ for any continuous bounded $f\colon \mathbb{R}^d \to \mathbb{R}$. Instead, it is sufficient to work with simpler families of test functions f: for example, it is enough to prove the convergence of the characteristic function (i.e. $f(x) = e^{i\langle x,\xi\rangle}$ for $\xi \in \mathbb{R}^d$) or of the cumulative distribution function (i.e. $f(x) = \prod_{i=1}^d \mathbb{1}_{x_i \leq a_i}$ for $a \in \mathbb{R}^d$). No similar criteria exists for random variables in $\mathcal{C}([0,1])$, but relying directly on the definition of convergence in distribution in that case seems hopeless as well. We present in these notes tools to prove convergence in distribution for random variables in general metric spaces (and which are even useful in $\mathbb{R}$). However, we mainly use the equivalent point of view of weak convergence of probability measures on those metric space.

Therefore, we start by studying topological properties of probability measures on metric spaces and more precisely on *Polish spaces*, which is a convenient type of space to work with. Then, we introduce weak convergence of probability measures on those spaces and describe the associated topology on the set of probability measures. One of the most important results (if not the most!) is Prokhorov's theorem, which provides a necessary and sufficient condition for the relative compactness of a family of probability measures, in terms of a notion called *tightness*. This yields a general method to prove weak convergence of probability measures. Finally, we show how this method applies to the particular case of the space $\mathcal{C}([0,1])$ and use it to prove Donsker's theorem.

2 Probability measures on metric spaces

2.1 Reminder: Metric spaces and compactness

We recall here the definition of metric spaces, compactness and related notions, as well as some basic facts.

Definition 2.1. *Let E be a set. A function $d\colon E \times E \to \mathbb{R}$ is called a* metric *(or a* distance*) if it satisfies the following properties:*

- *For any $x \in E$, $d(x,x) = 0$;*
- *Positivity: for any $x, y \in E$, if $x \neq y$ then $d(x,y) > 0$;*
- *Symmetry: for any $x, y \in E$, $d(x,y) = d(y,x)$;*
- *Triangle inequality: for any $x, y, z \in E$, $d(x,z) \leq d(x,y) + d(y,z)$.*

In that case, (E,d) is called a metric space*.*

For $x \in E$ and $r > 0$, we denote the ball centered at x of radius r by

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

In the sequel, all the balls are meant to be open, except if it is explicitly said they are closed.

Definition 2.2. *Let X be a topological space and $K \subset X$.*

- *K is called* compact *if every open cover of K has a finite subcover.*
- *K is called* sequentially compact *if any sequence in K has a subsequence converging to some point in K.*

Remark 2.3. This is the Anglo-Saxon definition of compactness. The French definition (used e.g. by Bourbaki) additionally requires that K is Hausdorff, i.e. that any two distinct points in K have disjoint neighborhoods. However, in this course we mainly work with metric spaces, which are always Hausdorff: if (E,d) is a metric space and $x, y \in E$ are distinct, then the balls $B(x,r)$ and $B(y,r)$ with $r = d(x,y)/2$ are disjoint neighborhoods of x and y.

Definition 2.4. *A metric space (E,d) is called* totally bounded *if for all $\varepsilon > 0$, there exists a finite cover of E by balls of radius (at most) ε whose centers lie in E.*

This is a generalization of compactness: a compact metric space (E,d) is totally bounded, since from the cover

$$E \subset \bigcup_{x \in E} B(x,\varepsilon),$$

one can extract a finite subcover by compactness. The converse statement is not true, for instance $(0,1)$ is totally bounded, but not compact.

Note also that any totally bounded space is bounded (as a finite union of bounded sets is bounded), but the reverse is not true. For instance, consider an infinite set E equipped with the discrete metric $d(x,y) = 1$ if $x \neq y$ and 0 otherwise. Then it is bounded as $\sup_{x,y \in E} d(x,y) = 1$, but not totally bounded if X is not finite, as any ball of radius $\varepsilon = \frac{1}{2}$ is a singleton, and no finite union of singletons can cover E.

Exercise 1. Let (E,d) be a metric space and let $A \subset E$ a totally bounded subset. Show that any subset B of A is totally bounded.

Proposition 2.5. *Let (E,d) be a metric space and let $A \subset E$. Then A is totally bounded if and only if $\overline{A}$ is totally bounded.*

Proof. ($\Leftarrow$) This is easy, since $A \subset \overline{A}$, so if $\overline{A}$ is totally bounded, so is A.

($\Rightarrow$) Assume A is totally bounded. Let $\varepsilon > 0$, and consider the finite cover

$$A \subset \bigcup_{i=1}^{n} B(x_i, \varepsilon/2),$$

for some $x_1, \ldots, x_n$ in A. Since the closure of a *finite* union is the union of the closure, and since the closure of any open ball with radius $\varepsilon/2$ is included in an open ball with the same center and radius ε, we get

$$\overline{A} \subset \bigcup_{i=1}^{n} \overline{B(x_i, \varepsilon/2)} \subset \bigcup_{i=1}^{n} B(x_i, \varepsilon).$$

Hence $\overline{A}$ is totally bounded. $\square$

Before stating the main result of this subsection, which links the different notions introduced above, we recall the definition of completeness.

Definition 2.6. *A metric space (E, d) is said to be complete if any Cauchy sequence in E is converging in E.*

Theorem 2.7. *Let (E, d) be a metric space. The following assertions are equivalent:*

(i) *E is compact.*

(ii) *E is sequentially compact.*

(iii) *E is complete and totally bounded.*

Proof. The equivalence between (i) and (ii) is well-known and sometimes referred to as Bolzano–Weierstrass theorem, although this is not historically accurate. It is left as an exercise below. We prove here the equivalence between (ii) and (iii).

(ii) $\Rightarrow$ (iii): Let $(x_n)_{n\geq 1}$ be a Cauchy sequence. By (ii), $(x_n)_{n\geq 1}$ admits a converging subsequence, hence the whole sequence converges. Thus, E is complete. Suppose by contradiction that, for some $\varepsilon > 0$, E does not admit a finite subcover by balls of radius ε. Let $x_1 \in E$. We build a sequence $(x_n)_{n\geq 1}$ by choosing inductively $x_n \notin \bigcup_{i=1}^{n-1} B(x_i, \varepsilon)$ for all $n \geq 2$ (this is possible because $\bigcup_{i=1}^{n-1} B(x_i, \varepsilon)$ does not cover E). Then, for any $k \neq l$, $d(x_k, x_l) \geq \varepsilon$, and thus the sequence $(x_n)_{n\geq 1}$ cannot admit a converging subsequence. This contradicts (ii) and E is totally bounded.

(iii) $\Rightarrow$ (ii): Let $(x_n)_{n\geq 1}$ in E. We first construct successive subsequences which are contained in smaller and smaller balls using the following fact: for all $k \geq 1$, by total boundedness, there exists a finite cover of E by balls of radius $\frac{1}{k}$ and so from any sequence one can extract a subsequence staying in the same ball of radius $\frac{1}{k}$. Thus, for $k = 1$, there exists a subsequence $(x_{\varphi_1(n)})_{n\geq 1}$ staying in the same ball of radius 1, so that, for all $m, n \geq 1$,

$$d(x_{\varphi_1(m)}, x_{\varphi_1(n)}) \leq 1 + 1 = 2.$$

For $k = 2$, there exists a subsequence of the previous subsequence $(x_{\varphi_1\circ\varphi_2(n)})_{n\geq 1}$ such that for all $m, n \geq 1$,

$$d(x_{\varphi_1\circ\varphi_2(m)}, x_{\varphi_1\circ\varphi_2(n)}) \leq \frac{1}{2} + \frac{1}{2} = 1.$$

Proceeding inductively on k, we obtain that, for any $k \geq 1$, there exists $(x_{\varphi_1\circ\cdots\circ\varphi_k(n)})_{n\geq 1}$ such that for all $m, n \geq 1$,

$$d(x_{\varphi_1\circ\cdots\circ\varphi_k(m)}, x_{\varphi_1\circ\cdots\circ\varphi_k(n)}) \leq \frac{2}{k}. \tag{2.1}$$

Now define φ by

$$\varphi(k) = \varphi_1 \circ \cdots \circ \varphi_k(k),$$

for all $k \geq 1$. Then, for all $m \geq 1$ and all $n \geq m$,

$$d(x_{\varphi(m)}, x_{\varphi(n)}) = d(x_{\varphi_1 \circ \cdots \circ \varphi_m(m)}, x_{\varphi_1 \circ \cdots \circ \varphi_m(\varphi_{m+1} \circ \cdots \circ \varphi_n(n))}) \leq \frac{2}{m},$$

using (2.1) with $k = m$ and n replaced by $\varphi_{m+1} \circ \cdots \circ \varphi_n(n)$. Hence, $(x_{\varphi(n)})_{n \geq 1}$ is a Cauchy sequence, and since E is complete, it converges. This proves (ii). ☐

Exercise 2. Let (E, d) be a metric space. We prove here the equivalence between compactness and sequential compactness.

1. Assume that E is compact.
   (a) Prove that if $(F_n)_{n \geq 0}$ is a non-increasing sequence of non-empty closed subsets of E, then $\bigcap_{n \geq 0} F_n$ is non-empty.
   (b) Let $(x_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$. Letting F_n be the closure of $\{x_m : m \geq n\}$, prove that $(x_n)_{n \in \mathbb{N}}$ has a convergent subsequence, hence showing that E is sequentially compact.
2. Assume that E is sequentially compact. Let $(U_i)_{i \in I}$ be an open cover of E.
   (a) Prove by contradiction that there exists $r > 0$ such that, for any $x \in E$, there exists $i(x) \in I$ such that $B(x, r) \subset U_{i(x)}$. Such a number r is called a *Lebesgue's number* of the cover $(U_i)_{i \in I}$.
   (b) Using total boundedness of E with the radius r (this makes use of the implication (ii) $\Rightarrow$ (iii) proved above), deduce that there exists a finite subcover of $(U_i)_{i \in I}$, hence showing that E is compact.

Exercise 3 (Tychonoff theorem). Using diagonal extraction, prove that $[0, 1]^{\mathbb{N}}$ is sequentially compact for the product topology.

Exercise 4. Let (E, d) be a metric space. Show that

1. If E is compact, then any continuous function is bounded.
2. If E is totally bounded, then any uniformly continuous function is bounded.
3. If E is bounded, then any Lipschitz function is bounded.

Definition 2.8. *Let (E, d) be a metric space. A subset $A \subset E$ is* relatively compact *if its closure $\overline{A}$ is compact.*

In terms of sequences, a subset $A \subset E$ is relatively compact if and only if for any sequence $(x_n)_{n \geq 1}$ in A, there exists a convergent subsequence $(x_{\varphi(n)})_{n \geq 1}$, but the limit may not lie in A.

On a *complete* metric space, we have the following.

Proposition 2.9. *Let (E, d) be a complete metric space and let $A \subset E$. Then A is relatively compact if and only if A is totally bounded.*

Proof. If A is relatively compact, then $\overline{A}$ is compact, hence $\overline{A}$ is totally bounded by Theorem 2.7 and so A is totally bounded by Proposition 2.5 (no completude needed for this implication).

Now, if A is totally bounded, then $\overline{A}$ is also totally bounded (it is included in the same finite cover by balls as A by definition of the closure). Moreover, it is closed in the complete metric space E, so it is also complete. Hence, $\overline{A}$ is complete and totally bounded, hence compact by Theorem 2.7. ☐

2.2 Polish spaces

Definition 2.10. *Let E be a topological space.*

- *We say that E is* metrizable *if there exists a metric $d \colon E \times E \to [0, +\infty)$ which induces the same topology than the one of E.*

- *We say that E is separable if it contains a countable dense subset.*

Definition 2.11. *A topological space E is said to be a Polish space if it is separable, metrizable and such that the metric d makes (E, d) complete.*

Remark 2.12. It is useful to know that a space is metrizable even if the distance is unknown! For instance, one can use sequential compactness to prove compactness, etc...

Note that completeness is a metric property and not a topological one: two metrics d_1 and d_2 can induce the same topology on E and E can be complete for d_1 and not for d_2. For instance, on $\mathbb{R}$, the metric

$$d'(x, y) := |\arctan x - \arctan y|$$

induces the same topology than the one given by the usual metric $d(x, y) = |x - y|$. But $(\mathbb{R}, d')$ is not complete, see the exercise below.

In the sequel, "(E, d) is a Polish space" means that E is a Polish space whose topology is induced by the metric d and such that (E, d) is complete.

Exercise 5. Prove that $(\mathbb{R}, d')$ is not complete.

Example 2.13.
- $\mathbb{R}$ and more generally $\mathbb{R}^n$, equipped with the Euclidean distance, are Polish spaces.
- A compact metric space K is a Polish space. Indeed, compactness implies completeness (recall Theorem 2.7) so it remains to prove separability. For any $n \geq 1$,

  $$K \subset \bigcup_{x \in K} B\Big(x, \frac{1}{n}\Big),$$

  so by compactness, one can extract a finite subcover: there exists k_n and $x_{n,1}, \dots, x_{n,k_n}$ in K such that

  $$K \subset \bigcup_{j=1}^{k_n} B\Big(x_{n,j}, \frac{1}{n}\Big).$$

  The subset $\{x_{n,j} : 1 \leq j \leq k_n, n \geq 1\}$ is then countable and dense in K. Hence, K is separable.
- The space

  $$\mathcal{C}([0,1]) := \{f : [0,1] \to \mathbb{R} \mid f \text{ continuous}\}$$

  is a separable complete metric space, with the metric given by the uniform norm

  $$d(f, g) = \|f - g\|_\infty = \sup_{t \in [0,1]} |f(t) - g(t)|,$$

  so that the sequence $(f_n)_{n \geq 1}$ converges to f iff $(f_n)_{n \geq 1}$ converges *uniformly* to f. Indeed, by the Stone-Weierstrass theorem, the set of polynomials on $[0,1]$ is dense in $\mathcal{C}([0,1])$, and thus the set of polynomials with rational coefficients, which is countable, is dense in $\mathcal{C}([0,1])$, so $\mathcal{C}([0,1])$ is separable. To prove completeness, let $(f_n)_{n \geq 1}$ be a Cauchy sequence in $\mathcal{C}([0,1])$, i.e. for all $\varepsilon > 0$, there exists N, such that, for all $m, n \geq N$, for all $t \in [0,1]$,

  $$|f_n(t) - f_m(t)| < \varepsilon.$$

  Hence, for all $t \in [0,1]$, the sequence of real numbers $(f_n(t))_{n \geq 1}$ is a Cauchy sequence in $\mathbb{R}$ which is complete, hence, it converges to some limit $f(t) \in \mathbb{R}$. This defines a function f on $[0,1]$. Letting $m \to \infty$ in the bound above, one gets that for all $n \geq N$, for all $t \in [0,1]$,

  $$|f_n(t) - f(t)| < \varepsilon,$$

  i.e. $(f_n)_{n \geq 1}$ converges uniformly to f. It is well known that the uniform limit of a sequence of continuous functions is continuous. Thus, f is continuous and $\mathcal{C}([0,1])$ is complete.

- Note that the above is still true when $[0,1]$ is replaced by a compact metric space K, i.e. $\mathcal{C}(K)$, the space of continuous functions from K to $\mathbb{R}$, is a Polish space. One has to use the general statement of the Stone–Weierstrass theorem (see [2]):

  Theorem 2.14 (Stone–Weierstrass theorem)**.** *Let K be a compact metric space. Let $\mathcal{A}$ be a subalgebra of $\mathcal{C}(K)$ such that $\mathcal{A}$ contains a non-zero constant function. Then $\mathcal{A}$ is dense in $\mathcal{C}(K)$ if and only if it separates points: for any $x, y \in X$, there exists a function $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.*

  Let $\{x_n : n \geq 1\}$ be a countable dense subset of K (recall that K is separable since it is compact). Define $f_0 \equiv 1$, and for $n \geq 1$, $f_n(x) = d(x, x_n)$, for $x \in K$. Let $\mathcal{A}$ be the subalgebra of $\mathcal{C}(K)$ generated by linear combinations with rationals coefficients of the f_n, $n \geq 0$. Then $\mathcal{A}$ is countable and satisfies the assumptions of Stone–Weierstrass theorem, hence it is dense in $\mathcal{C}(K)$.

- Note that the space $\mathcal{C}(\mathbb{R}_+)$, the space of continuous function from $\mathbb{R}_+$ to $\mathbb{R}$, equipped with the topology of uniform convergence is not separable. But if one endows $\mathcal{C}(\mathbb{R}_+)$ with the compact-open topology, i.e. the topology of uniform convergence *on compact sets* given by the metric
$$d(f,g) = \sum_{k \geq 1} \sup_{t \in [0,k]} |f(t) - g(t)| \wedge \frac{1}{2^k},$$
  then one can prove that $\mathcal{C}(\mathbb{R}_+)$ is a Polish space.

- Recall a Banach space is a normed vector space that is complete. Thus, any separable Banach space is a Polish space. For example, for any $p \in [1, +\infty)$, $L^p(\mathbb{R})$ is a Polish space. The completeness of $L^p(\mathbb{R})$ is usually proved in a measure theory course, so we rather give some details on how to show separability.

  First, one checks that integrable simple functions are dense in $L^p(\mathbb{R})$: this follows from the fact that any non-negative measurable function is the limit of an increasing sequence of non-negative simple functions. Then, using exterior regularity of Lebesgue measure (see Proposition 2.18, which can be applied to the Lebesgue measure restricted to finite intervals), one can deduce that integrable simple functions built only with indicator functions of open sets are dense in $L^p(\mathbb{R})$. Furthermore, using that open sets of $\mathbb{R}$ are a countable union of disjoint open intervals shows that integrable simple functions built only with indicator functions of intervals are also dense. Finally, considering only such simple functions with rational coefficients and rational endpoints for the intervals, one gets a countable set which is dense in $L^p(\mathbb{R})$.

- A finite product of Polish spaces is Polish.

- As a counterexample, $L^\infty([0,1])$, the space of essentially bounded functions on $[0,1]$, is complete, but not separable.

  To prove the non-separability, denote for $t \in [0,1]$, $f_t := \mathbb{1}_{[0,t]}$. Then, for any $t \neq s$,
$$\|f_t - f_s\|_\infty = 1,$$
  where $\|\cdot\|_\infty$ denotes the essential supremum, i.e. $\|f\|_\infty = \inf\{a > 0 \mid \lambda(|f| > a) = 0\}$ where λ is the Lebesgue measure. Thus, the open balls $B(f_t, \frac{1}{2})$, for $t \in [0,1]$, are pairwise distincts. Now suppose that $L^\infty([0,1])$ is separable, and let $A \subset L^\infty([0,1])$ be a countable dense subset. By density, any non-empty open set intersects A, hence each balls $B(f_t, \frac{1}{2})$, for $t \in [0,1]$ contains an element of A. But since they are disjoint and there are uncountably many of them, this contradicts the fact that A is countable.

- Note also that the space

$$\mathcal{C}_b(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ continuous and bounded}\}$$

equipped with the uniform norm $\|\cdot\|_\infty$ is a complete normed space, but is not separable.

Exercise 6. Prove that $\mathcal{C}_b(\mathbb{R})$ is not separable (mimic the proof for $L^\infty([0,1])$, i.e. find an uncountable family $\{f_i\}_{i\in I}$ in $\mathcal{C}_b(\mathbb{R})$ such that $\|f_i - f_j\|_\infty \geq 1$ for $i \neq j$).

2.3 Regularity of probability measures on metric spaces

Let E be a topological space. We will always, unless specified, equip E with the Borel σ-algebra, denoted by $\mathcal{B}(E)$, which is the σ-algebra generated by the class of open sets of E. We will denote by $\mathcal{M}_1(E)$ the space of probability measures on $(E, \mathcal{B}(E))$.

Definition 2.15. *Let (E, d) be a metric space. Let $x \in E$ and $A \subset E$. We define the* distance between x and A *by*

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

Exercise 7. Let $A \subset E$.

1. Show that $x \mapsto d(x, A)$ is 1-Lipschitz on E.
2. Let $x \in E$. Show that $x \in \overline{A}$ if and only if $d(x, A) = 0$.
3. Let $x \in E$. Show that $d(x, A) = d(x, \overline{A})$.

We will need the following useful lemma. We use the standard notations: for $a, b \in \mathbb{R}$,

$$a \wedge b := \min\{a, b\} \qquad \text{and} \qquad a \vee b := \max\{a, b\}.$$

Lemme 2.16. *Let (E, d) be a metric space.*

(i) *Let $F \subset E$ be a closed set. Define, for all $n \geq 0$ and $x \in E$, $f_n(x) := (1 - nd(x, F)) \vee 0$.*

(ii) *Let $G \subset E$ be an open set. Define, for all $n \geq 0$ and $x \in E$, $g_n(x) := nd(x, G^c) \wedge 1$.*

Then, for all $n \geq 0$, f_n and g_n are bounded Lipschitz functions, and, as $n \to \infty$,

$$f_n \searrow \mathbb{1}_F \quad \text{and} \quad g_n \nearrow \mathbb{1}_G.$$

Proof. Let $n \geq 0$. We have $0 \leq f_n \leq 1$ so f_n is bounded. Moreover, since $x \mapsto \max(x, 0)$ is 1-Lipschitz and $x \mapsto nd(x, F)$ is n-Lipschitz by Exercise 7, one obtains that f_n is n-Lipschitz. The fact that $(f_n)_{n\geq 0}$ is non-increasing follows from the fact that $d(x, F) \geq 0$ for any $x \in E$ It remains to prove that, for any $x \in E$, $f_n(x) \to \mathbb{1}_F(x)$:

- If $x \in F$, then $d(x, F) = 0$, hence $f_n(x) = 1$ for all $n \geq 0$ so $f_n(x) \to \mathbb{1}_F(x)$;
- If $x \notin F$, then $d(x, F) > 0$ because F is closed (by Exercise 7), hence there exists n_0 such that $d(x, F) > \frac{1}{n_0}$, so for all $n \geq n_0$, $nd(x, F) > 1$, that is $f_n(x) = 0$ and thus $f_n(x) \to \mathbb{1}_F(x)$.

This proves the result for f_n. The result for g_n then follows by noting that, if f_n denotes the function associated to the closed set $F = G^c$, then $g_n = 1 - f_n$. □

A first consequence of this lemma is the following: a probability measure on a metric space is uniquely characterized by the integrals of bounded continuous functions.

Proposition 2.17. *Let (E, d) a metric space and let $\mu, \nu \in \mathcal{M}_1(E)$. If for all bounded and continuous (or Lipschitz) function f,*

$$\int_E f \, \mathrm{d}\mu = \int_E f \, \mathrm{d}\nu,$$

then $\mu = \nu$.

Proof. Let G be an open set. By Lemma 2.16, there exists a nondecreasing sequence $(g_n)_{n\geq 0}$ of bounded continuous functions such that $g_n \nearrow \mathbb{1}_G$ as $n \to \infty$. By assumption, we have

$$\int_E g_n \,\mathrm{d}\mu = \int_E g_n \,\mathrm{d}\nu,$$

for all n, and by monotone convergence theorem, we get

$$\mu(G) = \int_E \mathbb{1}_G \,\mathrm{d}\mu = \int_E \mathbb{1}_G \,\mathrm{d}\nu = \nu(G).$$

Since probability measures on the Borel σ-algebra $\mathcal{B}(E)$ are characterized by their values on open sets (by Dynkin π–λ theorem[4]), we get $\mu = \nu$. □

Proposition 2.18. *Let (E, d) be a metric space and let $\mu \in \mathcal{M}_1(E)$. Then, for any Borel set A, one has*

$$\mu(A) = \inf\{\mu(U) \mid U \text{ open such that } A \subset U\}.$$

One says that μ is exterior regular. We also have,

$$\mu(A) = \sup\{\mu(F) \mid F \text{ closed such that } F \subset A\}.$$

In other words, for any Borel set A and for every $\varepsilon > 0$, there exists an open set U and a closed set F such that $F \subset A \subset U$ and

$$\mu(U \setminus F) \leq \varepsilon.$$

Remark 2.19. One may wonder why we want to approximate a Borel set from the outside by an open set, and from the inside by a closed set, and not the other way around. But in that case, the best possible choice would be $\mathring{A} \subset A \subset \overline{A}$, which obviously does not work: consider for instance the Lebesgue measure on $[0, 1]$ and the set of rationals $\mathbb{Q}$.

Proof. Define

$$\mathcal{A} = \{A \in \mathcal{B}(E) \mid \forall\varepsilon > 0,\ \exists U \text{ open and } F \text{ closed such that } F \subset A \subset U \text{ and } \mu(U \setminus F) \leq \varepsilon\}.$$

We will show that $\mathcal{A}$ is a σ-algebra that contains the open sets, hence, we will have $\mathcal{A} = \mathcal{B}(E)$.

We start by showing that $\mathcal{A}$ contains the class of open sets. Let A be an open set. Let $\varepsilon > 0$, we want to find an open set U and a closed set F with $F \subset A \subset U$ and such that $\mu(U \setminus F) \leq \varepsilon$. Set $U \coloneqq A$. Then, for any $n \geq 1$, define

$$F_n \coloneqq \left\{x \in A \;\middle|\; d(x, A^c) \geq \frac{1}{n}\right\}.$$

The sequence of sets F_n is clearly increasing, i.e. $F_n \subset F_{n+1}$, for all $n \geq 1$. Since the function $x \mapsto d(x, A^c)$ is continuous (see Exercise 7), each F_n is closed. Moreover, since A^c is closed, we have $d(x, A^c) > 0$ for every $x \in A$ (by Exercise 7 again), and thus there exists n large enough such that $d(x, A^c) \geq \frac{1}{n}$. Therefore,

$$A = \bigcup_{n\geq 1} F_n,$$

and, since the union is increasing,

$$\lim_{n\to\infty} \mu(F_n) = \mu\left(\bigcup_{n\geq 1} F_n\right) = \mu(A).$$

[4]More precisely, we use the following important consequence of Dynkin π–λ theorem: if $\mathcal{A}$ is a π-system, i.e. is closed under finite intersections, and if two probability measures on $\sigma(\mathcal{A})$ are equal on $\mathcal{A}$, then they are equal on $\sigma(\mathcal{A})$. Here the π-system is the class of open sets.

Hence, for n large enough, we have

$$\mu(F_n) \geq \mu(A) - \varepsilon,$$

so we choose F of the form F_n for some sufficiently large n. Hence, $A \in \mathcal{A}$, and so $\mathcal{A}$ contains the class of open sets.

Now let us show that $\mathcal{A}$ is a σ-algebra:

(i) It clearly contains E (take $U = F = E$).

(ii) It is closed under complementation: if $A \in \mathcal{A}$, then there exist an open set U and a closed set F such that $F \subset A \subset U$ and $\mu(U \setminus F) \leq \varepsilon$. Set $U' = F^c$, which is open, and $F' = U^c$, which is closed. We have $F' \subset A^c \subset U'$ and since $U' \setminus F' = U \setminus F$, it follows that $\mu(U' \setminus F') \leq \varepsilon$.

(iii) It is closed under countable unions: let, for each n, $A_n \in \mathcal{A}$ and set $A := \bigcup_{n\geq 1} A_n$. Let $\varepsilon > 0$. Then, there exist open sets U_n and closed sets F_n such that $F_n \subset A_n \subset U_n$ and

$$\mu(U_n \setminus F_n) \leq \frac{\varepsilon}{2^{n+1}}.$$

Then

$$\bigcup_{n\geq 1} F_n \subset A \subset \bigcup_{n\geq 1} U_n$$

and

$$\left(\bigcup_{n\geq 1} U_n\right) \setminus \left(\bigcup_{n\geq 1} F_n\right) \subset \bigcup_{n\geq 1} (U_n \setminus F_n)$$

(since if x belongs to one of the U_n and to none of the F_n's, there exists n such that $x \in U_n \setminus F_n$). Therefore,

$$\mu\left(\left(\bigcup_{n\geq 1} U_n\right) \setminus \left(\bigcup_{n\geq 1} F_n\right)\right) \leq \mu\left(\bigcup_{n\geq 1} (U_n \setminus F_n)\right) \leq \sum_{n\geq 1} \mu(U_n \setminus F_n) \leq \frac{\varepsilon}{2},$$

where we used σ-subadditivity.

Set $U' = \bigcup_{n\geq 1} U_n$, which is open as a union of open sets. We can't do the same for F, since an arbitrary union of closed sets isn't necessarily closed. So define $F'_n = \bigcup_{k\leq n} F_k$, which is closed as a *finite* union of closed sets. Since $\bigcup_n F_n = \bigcup_n \bigcup_{k\leq n} F_k$, and the sequence is increasing, and since $\mu(\bigcup_n F_n) < \infty$, there exists N such that

$$\mu\left(\bigcup_{n\geq 1} F_n\right) \leq \mu(F'_N) + \frac{\varepsilon}{2}.$$

Then $F'_N \subset A \subset U'$, and

$$\mu(U' \setminus F'_N) = \mu(U') - \mu(F'_N) = \mu(U') - \mu\left(\bigcup_{n\geq 1} F_n\right) + \mu\left(\bigcup_{n\geq 1} F_n\right) - \mu(F'_N) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus, $A \in \mathcal{A}$.

Therefore, $\mathcal{A}$ is a σ-algebra that contains the open sets, and hence $\mathcal{A}$ is the Borel σ-algebra $\mathcal{B}(E)$. $\square$

2.4 Probability measures on Polish spaces

One of the most important notion of this course is the notion of *tightness*. We first introduce this notion for a single probability measure on a Polish space.

Proposition 2.20. *Let E be a Polish space and let $\mu \in \mathcal{M}_1(E)$. Then, μ is tight, i.e. for all $\varepsilon > 0$, there exists a compact subset $K_\varepsilon \subset E$, such that*

$$\mu(K_\varepsilon^c) \leq \varepsilon.$$

It says that, on a Polish space, almost all the mass of a probability measure is concentrated on a compact set. This would be obvious on a metric space E where balls are relatively compact (such as $\mathbb{R}^n$): writing, for some $x \in E$,

$$E = \bigcup_{n \geq 1} B(x, n),$$

one gets that there exists $n \geq 1$ such that $\mu(B(x,n)) \geq 1 - \varepsilon$ and then one can choose $K = \overline{B(x,n)}$ which is compact. However, our goal is to work with metric spaces where balls are not necessarily relatively compact, such as infinite-dimensional normed vector spaces (for example $\mathcal{C}([0,1])$). In that case, the result above is more surprising: in such a space, a compact set is quite "small" because it has an empty interior.

Proof of Proposition 2.20. Let $\varepsilon > 0$. Since E is separable, there exists a dense sequence $(x_n)_n$ in E, hence for all $k \geq 1$, we have,

$$E = \bigcup_{n \geq 1} B\left(x_n, \frac{1}{k}\right).$$

Therefore,

$$1 = \mu(E) = \lim_{N \to \infty} \mu\left(\bigcup_{n=1}^{N} B\left(x_n, \frac{1}{k}\right)\right),$$

so there exists $N_k \geq 1$ such that

$$\mu\left(\bigcup_{n=1}^{N_k} B\left(x_n, \frac{1}{k}\right)\right) \geq 1 - \frac{\varepsilon}{2^k}.$$

Now define

$$A := \bigcap_{k \geq 1} \bigcup_{n=1}^{N_k} B\left(x_n, \frac{1}{k}\right).$$

By construction A is totally bounded, hence, by Theorem 2.7, $K := \overline{A}$ is a compact set since E is complete. Then,

$$\mu(K^c) \leq \mu(A^c) = \mu\left(\bigcup_{k \geq 1} \left[\bigcup_{n=1}^{N_k} B\left(x_n, \frac{1}{k}\right)\right]^c\right) \leq \sum_{k \geq 1} \mu\left(\left[\bigcup_{n=1}^{N_k} B\left(x_n, \frac{1}{k}\right)\right]^c\right) \leq \sum_{k \geq 1} \frac{\varepsilon}{2^k} = \varepsilon,$$

which concludes the proof. □

On a Polish space, we can then improve the regularity of probability measures:

Proposition 2.21. *Let E be a Polish space. Then every $\mu \in \mathcal{M}_1(E)$ is inner regular, i.e. for all $A \in \mathcal{B}(E)$,*

$$\mu(A) = \sup\{\mu(K) \mid K \subset A \text{ compact}\}.$$

Proof. Let $A \in \mathcal{B}(E)$. By Proposition 2.18,

$$\mu(A) = \sup\{\mu(F) \mid F \subset A \text{ closed}\},$$

hence, for all $\varepsilon > 0$, there exists a closed set $F \subset A$ such that

$$\mu(A) - \mu(F) \leq \varepsilon.$$

By tightness, there exists a compact set K_ε such that $\mu(K_\varepsilon^c) \leq \varepsilon$. Put $K = F \cap K_\varepsilon$. Then K is a compact set included in F, hence also in A, and we have

$$\begin{aligned}\mu(A \setminus K) = \mu(A \setminus (F \cap K_\varepsilon)) &= \mu((A \setminus F) \cup (A \setminus K_\varepsilon)) \\ &\leq \mu(A \setminus F) + \mu(K_\varepsilon^c) \\ &\leq 2\varepsilon,\end{aligned}$$

so $\mu(A) \leq \mu(K)+2\varepsilon$. This proves $\mu(A) \leq \sup\{\mu(K) \mid K \subset A \text{ compact}\}$ and the reversed inequality is direct. $\square$

3 Weak convergence and topology on probability measures

3.1 Weak convergence

Recall $\mathcal{C}_b(E)$ denotes the set of bounded continuous functions from E to $\mathbb{R}$.

Definition 3.1. *Let E be a metric space. Let $(\mu_n)_{n\geq 0} \in \mathcal{M}_1(E)^{\mathbb{N}}$ and $\mu \in \mathcal{M}_1(E)$. We say that $(\mu_n)_{n\geq 0}$ converges weakly to μ if*

$$\forall f \in \mathcal{C}_b(E), \quad \int_E f \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int_E f \,\mathrm{d}\mu.$$

We denote this convergence by $\mu_n \xrightarrow[n\to\infty]{(\mathrm{w})} \mu$ (another common notation is $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$).

Remark 3.2. Note that since integrals $\int_E f \,\mathrm{d}\mu$ for all $f \in \mathcal{C}_b(E)$ characterize the measure μ by Proposition 2.17, the limit is unique.

In probabilistic term, if X_n has distribution μ_n and X has distribution μ, then weak convergence corresponds to convergence in distribution for random variables:

$$X_n \xrightarrow[n\to\infty]{(\mathrm{d})} X \quad \Leftrightarrow \quad \mu_n \xrightarrow[n\to\infty]{(\mathrm{w})} \mu.$$

So all results stated in these notes can be translated to the context of convergence in distribution for random variables.

Example 3.3. Let (E, d) be a metric space. Let $(x_n)_{n\geq 0} \in E^{\mathbb{N}}$ and $x \in E$. Then $(\delta_{x_n})_{n\geq 0}$ converges weakly to δ_x if and only if $(x_n)_{n\geq 0}$ converges to x.

Indeed, assume that $x_n \to x$. Then, for any continuous function f on E, we have

$$\int_E f \,\mathrm{d}\delta_{x_n} = f(x_n) \xrightarrow[n\to\infty]{} f(x) = \int_E f \,\mathrm{d}\delta_x,$$

by continuity of f. Hence, $(\delta_{x_n})_{n\geq 0}$ converges weakly to δ_x. Now suppose that $x_n \not\to x$. Then, there exists some $\varepsilon > 0$, such that $d(x_n, x) > \varepsilon$ for infinitely many indices n. Consider the bounded continuous function

$$f(y) = \max\left\{1 - \frac{1}{\varepsilon} d(y, x), 0\right\}.$$

Then, $f(x) = 1$ while $f(x_n) = 0$ for infinitely many indices n. Hence, $(f(x_n))_{n\geq 0}$ does not converge to $f(x)$ and so $(\delta_{x_n})_{n\geq 0}$ does not converge weakly to δ_x.

Let $\mu \in \mathcal{M}_1(E)$ and let $g \colon E \to F$ be a measurable function between two metric spaces. We denote by $g_{\#}\mu$ the *pushforward measure* (or image measure) of μ by g, i.e.

$$g_{\#}\mu(A) := \mu\Big(g^{-1}(A)\Big) = \mu(\{g \in A\}),$$

for all Borel set $A \in \mathcal{B}(F)$. It is a probability measure on $(F, \mathcal{B}(F))$ and we recall that by the change of variables formula, one has

$$\int_F f \,\mathrm{d}g_{\#}\mu = \int_E f \circ g \,\mathrm{d}\mu,$$

for any integrable measurable function $f \colon F \to \mathbb{R}$.

The following result is well known for convergence in distribution of random variables, in term of weak convergence, it states:

Theorem 3.4 (Continuous mapping theorem)**.** *Let E and F be metric spaces. Let $(\mu_n)_{n\geq 0} \in \mathcal{M}_1(E)^{\mathbb{N}}$ be a sequence converging weakly to $\mu \in \mathcal{M}_1(E)$. Let $g\colon E \to F$ be a continuous function. Then,*

$$g_{\#}\mu_n \xrightarrow[n\to\infty]{(\mathrm{w})} g_{\#}\mu.$$

Proof. Let f be a bounded continuous function on F. Then, by the change of variable formula,

$$\int_F f \, \mathrm{d}g_{\#}\mu_n = \int_E f \circ g \, \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int_E f \circ g \, \mathrm{d}\mu = \int_F f \, \mathrm{d}g_{\#}\mu,$$

since $f \circ g$ is a bounded continuous function on E. □

3.2 Portmanteau theorem

We now give a list of criteria for weak convergence.

Theorem 3.5 (Portmanteau theorem)**.** *Let E be a metric space. Let $(\mu_n)_{n\geq 0} \in \mathcal{M}_1(E)^{\mathbb{N}}$ and $\mu \in \mathcal{M}_1(E)$. The following assertions are equivalent:*

(i) *The sequence $(\mu_n)_n$ converges weakly to μ.*

(ii) *For all bounded and uniformly continuous function f, $\int f \, \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int f \, \mathrm{d}\mu$.*

(iii) *For all bounded and Lipschitz function f, $\int f \, \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int f \, \mathrm{d}\mu$.*

(iv) *For all open set $G \subset E$, $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$.*

(v) *For all closed set $F \subset E$, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.*

(vi) *For all Borel set $A \in \mathcal{B}(E)$ of μ-continuity, i.e. such that $\mu(\partial A) = 0$, $\mu_n(A) \xrightarrow[n\to\infty]{} \mu(A)$.*

(vii) *For all bounded measurable function f which is continuous μ-a.e., $\int f \, \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int f \, \mathrm{d}\mu$.*

Remark 3.6. The name of this theorem does not come from a mathematician named Portmanteau, but from an invention of Billingsley in his book [1]. He is the first one to name this theorem this way and he adds in his bibliography this reference: Jean-Pierre Portmanteau, Espoir pour l'ensemble vide?, *Annales de l'Université de Felletin*, CXLI(1915)322–325. But this mathematician does not exist, there is no university in Felletin and the title is obviously a joke.

Proof. For any function $f\colon E \to \mathbb{R}$, we have

$$f \text{ Lipschitz} \Rightarrow f \text{ uniformly continuous} \Rightarrow f \text{ continuous} \Rightarrow f \text{ continuous } \mu\text{-a.e.},$$

so we directly obtain (vii) $\Rightarrow$ (i) $\Rightarrow$(ii) $\Rightarrow$ (iii).

We now prove (iii) $\Rightarrow$ (iv). Let G be an open set. From Lemma 2.16, there exists a sequence $(g_k)_{k\geq 0}$ such that g_k is bounded and Lispschitz for all $k \geq 0$ and $g_k \nearrow \mathbb{1}_G$ as $k \to \infty$. Hence,

$$\mu_n(G) = \int \mathbb{1}_G \, \mathrm{d}\mu_n \geq \int g_k \, \mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int g_k \, \mathrm{d}\mu.$$

Using monotone convergence theorem, one obtains,

$$\liminf_{n\to\infty} \mu_n(G) \geq \int g_k \, \mathrm{d}\mu \xrightarrow[k\to\infty]{} \int \mathbb{1}_G \, \mathrm{d}\mu = \mu(G),$$

which concludes the proof.

The equivalence (iv) $\Leftrightarrow$ (v) is direct by taking the complement.

Let us prove that (iv) + (v) $\Rightarrow$ (vi). Let $A \in \mathcal{B}(E)$ such that $\mu(\partial A) = 0$, then we have

$$\mu(\mathring{A}) = \mu(A) = \mu(\overline{A}).$$

Since for all $n \geq 0$, $\mu_n(\mathring{A}) \leq \mu_n(A) \leq \mu_n(\overline{A})$, we have

$$\limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\overline{A}) \overset{\text{(v)}}{\leq} \mu(\overline{A}) = \mu(A) = \mu(\mathring{A}) \overset{\text{(iv)}}{\leq} \liminf_{n\to\infty} \mu_n(\mathring{A}) \leq \liminf_{n\to\infty} \mu_n(A),$$

hence we get $\liminf_n \mu_n(A) = \limsup_n \mu_n(A) = \mu(A)$, i.e. $\mu_n(A) \to \mu(A)$ as $n \to \infty$.

It remains to prove that (vi) $\Rightarrow$ (vii). Let f be continuous μ-a.e. and bounded. By linearity, one can suppose that f is bounded by 1. We can also suppose that $f \geq 0$ (if not, just decompose $f = f_+ - f_-$). Then, for any $\nu \in \mathcal{M}_1(E)$, using Fubini's theorem, we have

$$\int_E f(x)\,\mathrm{d}\nu(x) = \int_E \int_0^1 \mathbb{1}_{\{t\leq f(x)\}}\,\mathrm{d}t\,\mathrm{d}\nu(x) = \int_0^1 \int_E \mathbb{1}_{\{f(x)\geq t\}}\,\mathrm{d}\nu(x)\,\mathrm{d}t = \int_0^1 \nu(\{x \mid f(x) \geq t\})\,\mathrm{d}t.$$

In particular, for all $n \geq 0$,

$$\int_E f\,\mathrm{d}\mu_n = \int_0^1 \mu_n(f \geq t)\,\mathrm{d}t.$$

Our goal is to apply (vi) to the set $\{f \geq t\}$ for Lebesgue-a.e. t. Denote by D the set of discontinuities of f. First, we claim that for all $t \in \mathbb{R}$,

$$\partial\{f \geq t\} \subset \{f = t\} \cup D.$$

Indeed, let $x \in \partial\{f \geq t\} = \overline{\{f \geq t\}} \setminus \operatorname{Int}\{f \geq t\}$. If $x \in D^c$, i.e. f is continuous at x, then $x \in \{f = t\}$. Indeed, let $x_n \in \{f \geq t\}$ such that $x_n \to x$. Then, $f(x_n) \to f(x)$ by continuity at x, and thus $x \in \{f \geq t\}$. Moreover, if $f(x) > t$, then by continuity $\{f > t\}$ would contain a neighborhood of x, hence $x \in \operatorname{Int}\{f > t\} \subset \operatorname{Int}\{f \geq t\}$, which contradicts the fact that $x \in \partial\{f \geq t\}$. Hence, $x \in \{f = t\} \cup D$ proving the inclusion claimed above. It follows that

$$\mu(\partial\{f \geq t\}) \leq \mu(\{f = t\}) + \mu(D) = \mu(\{f = t\}),$$

because $\mu(D) = 0$ by assumption. Now remark that the set $\{t \in \mathbb{R} \,|\, \mu(f = t) > 0\}$ is the set of the atoms of the probability measure $f_{\#}\mu$, and as such is at most countable[5] and therefore has Lebesgue measure equal to zero. Hence, for Lebesgue-a.e. $t \in \mathbb{R}$, $\mu(\{f = t\}) = 0$ and thus $\mu(\partial\{f \geq t\}) = 0$. So, we get that:

- $\mu_n(f \geq t) \xrightarrow[n\to\infty]{} \mu(f \geq t)$ Lebesgue-a.e., by assumption (vi),
- $\mu_n(f \geq t) \leq 1$, which is integrable on $[0,1]$ with respect to Lebesgue measure,

so by dominated convergence theorem,

$$\int_E f\,\mathrm{d}\mu_n = \int_0^1 \mu_n(f \geq t)\,\mathrm{d}t \xrightarrow[n\to\infty]{} \int_0^1 \mu(f \geq t)\,\mathrm{d}t = \int_E f\,\mathrm{d}\mu.$$

This proves (vii). □

Remark 3.7. It can be surpising that Criterion (iv) characterizes weak convergence despite the fact that it provides only lower bounds of the measures μ_n in terms of μ. This is due to the fact that we work only with probability measure, so we have one inequality in the other direction, $\mu_n(E) \leq \mu(E)$, and this is enough to characterize weak convergence. For weak convergence on the larger set of finite measures on E, Criterion (iv) does not characterize weak convergence, one needs to add the condition $\limsup_n \mu_n(E) \leq \mu(E)$. A similar remark hols for Criterion (v).

Example 3.8. To understand the direction of the inequalities in Conditions (iv) and (v), it is helpful to take as example Dirac measures. Consider $(x_n)_{n\geq 0} \in E^{\mathbb{N}}$ converging to some $x \in E$, so that $(\delta_{x_n})_{n\geq 0}$ converges weakly to δ_x by Example 3.3. Let $G \subset E$ be an open set.

[5] A probability measure ν can have at most countably many atoms. Indeed, for each $n \geq 1$, the set A_n of atoms of mass at least $1/n$ has at most n elements (otherwise ν would have a total mass greater than 1) and, since the set of all atoms can be written as $\bigcup_{n\geq 1} A_n$, it is at most countable.

- If $x \in G$, then, for n large enough, $x_n \in G$ and so $\delta_{x_n}(G) = 1 = \delta_x(G)$.
- If $x \notin G$, then different things can happen. One can have $x_n \notin G$ for all n and thus $\delta_{x_n}(G) = 0 = \delta_x(G)$. On the other hand, if $x \in \partial G$, then one could have $x_n \to x$ with $x_n \in G$ for all n and thus $\delta_{x_n}(G) = 1 \not\to 0 = \delta_x(G)$. Finally, a mixture of both cases can occur with $\liminf_n \delta_{x_n}(G) = 0$ and $\limsup_n \delta_{x_n}(G) = 1$.

In words, one can describe Condition (iv) by the fact that, in a weak convergence, mass can escape from an open set in the limit but no new mass can appear. A similar discussion can be done for closed sets.

Example 3.9. Recall that if X is a real random variable, its cumulative distribution function is the function

$$\begin{aligned} F_X \colon \mathbb{R} &\to [0,1] \\ t &\mapsto \mathbb{P}(X \leq t). \end{aligned}$$

In other words, if μ denotes the distribution of X (i.e. the image measure of $\mathbb{P}$ by X), then $F_X(t) = \mu((-\infty, t])$, for all $t \in \mathbb{R}$. Our goal here is to use Portmanteau theorem to prove the following well-known characterization:

$$X_n \xrightarrow[n\to\infty]{\text{(d)}} X \qquad \Longleftrightarrow \qquad \forall t \in \mathbb{R} \text{ continuity point of } F_X, \quad F_{X_n}(t) \xrightarrow[n\to\infty]{} F_X(t).$$

Recall that F_X is a càdlàg ("continue à droite avec limite à gauche", French for "right continuous with left limits") function and that the points of discontinuity of F_X are exactly the atoms of the distribution of X. Hence, F_X continuous at t is equivalent to $\mu(\{t\}) = 0$. Moreover, $\partial(-\infty, t] = \{t\}$, hence F_X continuous at t is equivalent to say that $(-\infty, t]$ is a set of μ-continuity. Hence, Condition (vi) of Portmanteau theorem implies that if $X_n \to X$ in distribution, i.e. $\mu_n \to \mu$ weakly, then, $F_{X_n}(t) \to F_X(t)$, for all t continuity point of F_X.

Conversely, assume $F_{X_n}(t) \to F_X(t)$, for all t continuity point of F_X. Let $a < b$, we first prove convergence of $\mathbb{P}(X_n \in (a,b))$. Let D be the set of discontinuity points of F_X. Since D is at most countable (see Footnote 5), there exists $a_k \searrow a$, and $b_k \nearrow b$ such that for all k, a_k and b_k are continuity points of F_X (since D is countable, D^c is dense in $\mathbb{R}$). Hence,

$$\begin{aligned} \liminf_{n\to\infty} \mathbb{P}(X_n \in (a,b)) &\geq \liminf_{n\to\infty} \mathbb{P}(X_n \in (a_k, b_k]) \\ &\geq \liminf_{n\to\infty} (F_{X_n}(b_k) - F_{X_n}(a_k)) \\ &= F_X(b_k) - F_X(a_k) \qquad\qquad \text{(by assumption)} \\ &= \mathbb{P}(X \in (a_k, b_k]). \end{aligned}$$

Letting $k \to \infty$, we get

$$\liminf_{n\to\infty} \mathbb{P}(X_n \in (a,b)) \geq \mathbb{P}(X \in (a,b)). \tag{3.1}$$

Now, we aim at proving Condition (iv) of Portmanteau theorem. For this, let U be an open set and $\varepsilon > 0$. Since any open set is the countable union of disjoint open intervals, we write $U = \bigcup_{i\geq 1}(a_i, b_i)$, for some a_i and b_i. Then, there exists N large enough such that

$$V_N := \bigcup_{i=1}^{N} (a_i, b_i)$$

satisfies $\mathbb{P}(X \in V_N) > \mathbb{P}(X \in U) - \varepsilon$. Eventually, we get

$$\liminf_{n\to\infty} \mathbb{P}(X_n \in U) \geq \liminf_{n\to\infty} \mathbb{P}(X_n \in V_N) \overset{(3.1)}{=} \mathbb{P}(X \in V_N) > \mathbb{P}(X \in U) - \varepsilon.$$

Since ε is arbitrary, we obtain $\liminf_{n\to\infty} \mathbb{P}(X_n \in U) \geq \mathbb{P}(X \in U)$, and we conclude, using Portmanteau theorem, that X_n converges in distribution to X.

Exercise 8. Let (E, d) be a separable metric space. Let $(X_n)_{n\geq 0}$ be a sequence of random variables in E and X be r.v. in E. We say that the sequence of random variables $(X_n)_{n\geq 0}$ *converges in probability* to X if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(d(X_n, X) > \varepsilon) = 0.$$

We denote this convergence by $X_n \xrightarrow[n\to\infty]{(\mathbb{P})} X$. Note that the separability assumption ensures that $d(X, Y)$ is a random variable for any random variables X and Y, so that the definition above makes sense: more details in the solution.

1. Show that if $X_n \xrightarrow[n\to\infty]{(\mathbb{P})} X$, then $X_n \xrightarrow[n\to\infty]{(\mathrm{d})} X$.
2. Show that if $X_n \xrightarrow[n\to\infty]{(\mathrm{d})} x$, where $x \in E$ is deterministic, then $X_n \xrightarrow[n\to\infty]{(\mathbb{P})} x$.
3. Show that if $X_n \xrightarrow[n\to\infty]{(\mathrm{d})} X$ and $d(X_n, Y_n) \xrightarrow[n\to\infty]{(\mathbb{P})} 0$, then $Y_n \xrightarrow[n\to\infty]{(\mathrm{d})} X$.
4. *Slutsky's theorem:* Show that if $X_n \xrightarrow[n\to\infty]{(\mathrm{d})} X$ and $Y_n \xrightarrow[n\to\infty]{(\mathrm{d})} y$, where $y \in E$ is deterministic, then $(X_n, Y_n) \xrightarrow[n\to\infty]{(\mathrm{d})} (X, y)$.

   Hint: First prove that $(X_n, y) \xrightarrow[n\to\infty]{(\mathrm{d})} (X, y)$ and then use Question 3.

3.3 Weak topology on $\mathcal{M}_1(E)$ (bonus)

Recall that the topology generated by a subbase is the smallest topology containing this subbase. Equivalently, every open set of the topology, except possibly the full space, can be written as the union of finite intersections of elements of the subbase.

Definition 3.10. *The* weak topology *on $\mathcal{M}_1(E)$ is the topology generated by the subbase consisting of the sets*

$$U(f, a, b) := \left\{\mu \in \mathcal{M}_1(E) : \int_E f \,\mathrm{d}\mu \in (a, b)\right\},$$

for $f \in \mathcal{C}_b(E)$ and real numbers $a < b$.

Exercise 9. Prove that the weak topology is the smallest topology making the map $\mu \in \mathcal{M}_1(E) \mapsto \int_E f \,\mathrm{d}\mu$ continuous for any $f \in \mathcal{C}_b(E)$.

Proposition 3.11. *For the weak topology, a basis of neighborhoods of $\mu \in \mathcal{M}_1(E)$ is given by the sets*

$$V_\mu(f_1, \dots, f_k, \varepsilon) := \bigcap_{i=1}^{k} \left\{\nu \in \mathcal{M}_1(E) : \left|\int_E f_i \,\mathrm{d}\nu - \int_E f_i \,\mathrm{d}\mu\right| < \varepsilon\right\},$$

for $k \geq 1$, $f_1, \dots, f_k \in \mathcal{C}_b(E)$ and $\varepsilon > 0$.

Proof. First note that the sets $V_\mu(f_1, \dots, f_k, \varepsilon)$ are open as finite intersection of sets of the subbase and contain μ so they are neighborhoods of μ. Now consider an open set U containing μ, we want to prove that it contains one of the $V_\mu(f_1, \dots, f_k, \varepsilon)$. The set U is a union of finite intersections of sets of the subbase (even if U is the full space with our subbase), so μ is necessarily in one of those finite intersections: there exists $k \geq 1$, $f_1, \dots, f_k \in \mathcal{C}_b(E)$ and $a_1 < b_1, \dots, a_k < b_k$ such that

$$\mu \in \bigcap_{i=1}^{k} U(f_i, a_i, b_i).$$

For each $i \in [\![1, k]\!]$, $\int_E f_i \,\mathrm{d}\mu \in (a_i, b_i)$, so one can choose $\delta > 0$ such that, for each i,

$$a_i < \int_E f_i \,\mathrm{d}\mu - \delta < \int_E f_i \,\mathrm{d}\mu + \delta < b_i.$$

Then, it follows that

$$V_\mu(f_1,\dots,f_k,\delta) \subset \bigcap_{i=1}^k U(f_i,a_i,b_i) \subset U$$

and this concludes the proof. □

Proposition 3.12. *Sequential convergence for the weak topology is equivalent to the weak convergence.*

Proof. Let $(\mu_n)_{n\geq 0} \in \mathcal{M}_1(E)^{\mathbb{N}}$ and $\mu \in \mathcal{M}_1(E)$.

$$\begin{aligned}
\mu_n \xrightarrow[n\to\infty]{(\mathrm{w})} \mu \quad &\Leftrightarrow \quad \forall f \in \mathcal{C}_b(E), \int_E f\,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int_E f\,\mathrm{d}\mu \\
&\Leftrightarrow \quad \forall k\geq 1, \forall f_1,\dots,f_k \in \mathcal{C}_b(E), \max_{1\leq i\leq k}\left|\int_E f_i\,\mathrm{d}\mu_n - \int_E f_i\,\mathrm{d}\mu\right| \xrightarrow[n\to\infty]{} 0 \\
&\Leftrightarrow \quad \forall k\geq 1, \forall f_1,\dots,f_k \in \mathcal{C}_b(E), \forall\varepsilon>0, \exists n_0\geq 0, \forall n\geq n_0, \mu_n \in V_\mu(f_1,\dots,f_k,\varepsilon) \\
&\Leftrightarrow \quad \mu_n \xrightarrow[n\to\infty]{} \mu \text{ for the weak topology,}
\end{aligned}$$

where the last equivalence follows from the fact that the sets $V_\mu(f_1,\dots,f_k,\varepsilon)$ form a basis of neighborhoods of μ. □

Remark 3.13. The facts above are a particular case of results on the weak* topology. Let Y be a real normed space. Denote by Y^* the topological dual of Y, i.e. the space of continuous real linear forms on Y. The *weak* topology* is the weakest topology on Y^* making the evaluation maps

$$\begin{array}{rccc} \mathrm{ev}_x : & Y^* & \longrightarrow & \mathbb{R} \\ & \phi & \longmapsto & \phi(x) \end{array}$$

continuous for all $x \in Y$. Equivalently, it is the topology generated by the subbase consisting of the sets $\{\phi \in Y^* : \phi(x) \in (a,b)\}$, for $x \in Y$ and $a < b$. As above, a basis of neighborhood of $\phi \in Y^*$ for the weak* topology is given by finite intersections of sets $\{\psi \in Y^* : |\psi(x)-\phi(x)| < \varepsilon\}$ for $x \in Y$ and $\varepsilon > 0$. Furthermore, a sequence $(\phi_n)_{n\geq 0}$ in Y^* converges to $\phi \in Y^*$ for the weak* topology, if and only if

$$\forall x \in Y, \quad \phi_n(x) \xrightarrow[n\to\infty]{} \phi(x),$$

i.e. if ϕ_n converges pointwise to ϕ.

Our particular case corresponds to $Y = \mathcal{C}_b(E)$. For any $\mu \in \mathcal{M}_1(E)$, $f \mapsto \int_E f\,\mathrm{d}\mu$ defines a continuous linear map on $\mathcal{C}_b(E)$. Furthermore, by Proposition 2.17, the map

$$\begin{array}{ccc} \mathcal{M}_1(E) & \longrightarrow & \mathcal{C}_b(E)^* \\ \mu & \longmapsto & (f \mapsto \int_E f\,\mathrm{d}\mu) \end{array}$$

is injective, so $\mathcal{M}_1(E)$ can be identified to a subset of $\mathcal{C}_b(E)^*$. The weak topology on $\mathcal{M}_1(E)$ is exactly the restriction of the weak* topology of $\mathcal{C}_b(E)^*$ to $\mathcal{M}_1(E)$.

Exercise 10. Consider the topology on $\mathcal{M}_1(E)$ generated by the subbase consisting of the sets $\{\nu \in \mathcal{M}_1(E) : \nu(U) > a\}$, for $U \subset E$ open and $a > 0$.

1. Prove that, for this topology, a basis of neighborhoods of $\mu \in \mathcal{M}_1(E)$ is given by the sets

$$\bigcap_{i=1}^k \{\nu \in \mathcal{M}_1(E) : \nu(U_i) > \mu(U_i) - \varepsilon\},$$

for $k \geq 1$, open sets $U_1,\dots,U_k \subset E$ and $\varepsilon > 0$.

2. Prove that sequential convergence for this topology is equivalent to the weak convergence.

3. Prove that the subbase consisting of the sets $\{\nu \in \mathcal{M}_1(E) : \nu(F) < a\}$, for $F \subset E$ closed and $a > 0$ generates the same topology.

Remark 3.14. Convergence of sequences does not characterize the topology in general: there can be two different topologies on the same space inducing the same notion of sequential convergence. It can however be proved, for any metric space E, that the topology introduced in Exercise 10 is the same as the weak topology.

In the next subsection, we are going to see that, if E is separable, then there is a metric on $\mathcal{M}_1(E)$ for which sequential convergence is equivalent to the weak convergence. The topology of such a metric is unique: there cannot be two different *metrizable* topologies on the same space inducing the same notion of sequential convergence. Furthermore, one can prove that the topology of such a metric is the weak topology (this is not clear a priori because we do not know if the weak topology is metrizable).

Proofs of the facts mentioned in this remark can be found in Appendix III of the *1st edition* of the book [1] by Billingsley (this has been removed in the 2nd edition).

3.4 The Lévy–Prokhorov distance

We will now see that the weak convergence is metrizable when E is separable, and there are several metrics of interest on $\mathcal{M}_1(E)$. We first introduce the Lévy–Prokhorov distance.

Let (E, d) be a metric space. Define the (open) *ε-neighborhood* of a subset $A \subset E$ by

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

It is useful to note that, for any $A \subset E$,

$$\bigcap_{\varepsilon>0} A^\varepsilon = \{x \in E \mid d(x, A) = 0\} = \overline{A},$$

as a direct consequence of Exercise 7.

Definition 3.15. *Let (E, d) be a metric space and let $\mu, \nu \in \mathcal{M}_1(E)$. The* Lévy–Prokhorov distance *between μ and ν is defined by*

$$\rho(\mu, \nu) := \inf\{\varepsilon > 0 \mid \forall B \in \mathcal{B}(E), \mu(B) \le \nu(B^\varepsilon) + \varepsilon\}.$$

To manipulate this definition, it is useful to note that the set

$$I := \{\varepsilon > 0 \mid \forall B \in \mathcal{B}(E), \mu(B) \le \nu(B^\varepsilon) + \varepsilon\}$$

is an interval of the form (a, ∞) or $[a, \infty)$ because, if $\varepsilon \in I$, then any $\varepsilon' > \varepsilon$ is in I as well. Therefore, here are the methods to prove upper and lower bounds on $\rho(\mu, \nu)$ using its definition:

- If for all Borel set B, $\mu(B) \le \nu(B^\varepsilon) + \varepsilon$, then $\rho(\mu, \nu) \le \varepsilon$.
- If there exists a Borel set B such that $\mu(B) > \nu(B^\varepsilon) + \varepsilon$, then $\rho(\mu, \nu) \ge \varepsilon$.

Also note that, for all $B \in \mathcal{B}(E)$, $\mu(B) \le \nu(B^1) + 1$, hence $\rho(\mu, \nu) \le 1$.

Proposition 3.16. *Let (E, d) be a metric space. The Lévy–Prokhorov distance ρ is a metric on $\mathcal{M}_1(E)$.*

Proof. Let $\mu \in \mathcal{M}_1(E)$. Let $\varepsilon > 0$. Then, for all $B \in \mathcal{B}(E)$, since $B \subset B^\varepsilon$, one has $\mu(B) \le \mu(B^\varepsilon) + \varepsilon$. This proves $\rho(\mu, \mu) \le \varepsilon$ and by letting $\varepsilon \downarrow 0$, we get $\rho(\mu, \mu) = 0$.

Let us show that ρ is symmetric. Let $\mu, \nu \in \mathcal{M}_1(E)$ and $\varepsilon \in (0, \rho(\mu, \nu))$. Then, there exists a Borel set B such that $\mu(B) > \nu(B^\varepsilon) + \varepsilon$. Hence, we have

$$\nu((B^\varepsilon)^c) > \mu(B^c) + \varepsilon \ge \mu(((B^\varepsilon)^c)^\varepsilon) + \varepsilon,$$

using that $((B^\varepsilon)^c)^\varepsilon \subset B^c$ (check it!). Thus there exists $A = (B^\varepsilon)^c$ (which is a Borel set) such that $\nu(A) > \mu(A^\varepsilon) + \varepsilon$. Hence, $\rho(\nu, \mu) \geq \varepsilon$. Letting $\varepsilon \uparrow \rho(\mu, \nu)$ gives that $\rho(\nu, \mu) \geq \rho(\mu, \nu)$. Exchanging the roles of μ and ν gives the equality.

Now we prove that $\mu \neq \nu$ implies $\rho(\mu, \nu) > 0$. Since $\rho \geq 0$ by definition, it is enough to prove that $\rho(\mu, \nu) = 0$ implies $\mu = \nu$. Let $\mu, \nu \in \mathcal{M}_1(E)$ such that $\rho(\mu, \nu) = 0$. For any $\varepsilon > 0$, since $\rho(\mu, \nu) < \varepsilon$, one has, for any $B \in \mathcal{B}(E)$, $\mu(B) \leq \nu(B^\varepsilon) + \varepsilon$. Taking $\varepsilon = 1/k$ and using that $(B^{1/k})_{k \geq 1}$ is non-increasing in k, we get

$$\mu(B) \leq \lim_{k \to \infty} \downarrow \nu\Big(B^{1/k}\Big) = \nu\left(\bigcap_{k \geq 1} B^{1/k}\right) = \nu(\overline{B}),$$

because $\overline{B} = \{x \in E \mid d(x, B) = 0\}$ by Exercise 7. Therefore, if B is closed, $\mu(B) \leq \nu(B)$ and, by symmetry, $\mu(B) = \nu(B)$. Since probability measures on $(E, \mathcal{B}(E))$ are characterized by their values on closed sets (recall Footnote 4), we get $\mu = \nu$.

It remains to prove the triangle inequality. Let $\mu, \nu, \gamma \in \mathcal{M}_1(E)$. Let $\varepsilon > 0$ and $\delta > 0$ such that $\rho(\mu, \nu) < \varepsilon$ and $\rho(\nu, \gamma) < \delta$. Then, for all Borel set B,

$$\mu(B) \leq \nu(B^\varepsilon) + \varepsilon \quad \text{and} \quad \nu(B) \leq \gamma(B^\delta) + \delta.$$

Hence, noting that $(B^\varepsilon)^\delta \subset B^{\varepsilon + \delta}$,

$$\mu(B) \leq \nu(B^\varepsilon) + \varepsilon \leq \gamma\Big((B^\varepsilon)^\delta\Big) + \delta + \varepsilon \leq \gamma(B^{\varepsilon+\delta}) + \delta + \varepsilon.$$

Since this holds for any $B \in \mathcal{B}(E)$, we get $\rho(\mu, \gamma) \leq \varepsilon + \delta$. Letting $\varepsilon \downarrow \rho(\mu, \nu)$ and $\delta \downarrow \rho(\nu, \gamma)$ gives the triangle inequality. □

Remark 3.17.

- Note that since $B^\varepsilon = \overline{B}^\varepsilon$ (since $d(x, B) = d(x, \overline{B})$), one can replace in the definition of the Lévy–Prokhorov distance $B \in \mathcal{B}(E)$ by $B \subset E$ closed. This is sometimes taken as the definition.
- Since the symmetry of ρ is not obvious in its definition, it is sometimes useful to use the definition of $\rho(\nu, \mu)$ to bound $\rho(\mu, \nu)$. Furthermore, the Lévy–Prokhorov distance between μ and ν is sometimes defined as

$$\inf\{\varepsilon > 0 \mid \forall B \in \mathcal{B}(E), \mu(B) \leq \nu(B^\varepsilon) + \varepsilon \text{ and } \nu(B) \leq \mu(B^\varepsilon) + \varepsilon\}.$$

  This makes the definition clearly symmetric, but it does not change it: with our notation, this quantity equals $\rho(\mu, \nu) \wedge \rho(\nu, \mu) = \rho(\mu, \nu)$ by symmetry of ρ.
- The Lévy–Prokhorov distance has been introduced by Prokhorov as a generalization of the Lévy distance on $\mathcal{M}_1(\mathbb{R})$: for $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ with cumulative distribution functions F_μ and F_ν, define

$$\mathcal{L}(\mu, \nu) = \inf\{\varepsilon > 0 \mid \forall x \in \mathbb{R},\ F_\mu(x - \varepsilon) - \varepsilon \leq F_\nu(x) \leq F_\mu(x + \varepsilon) + \varepsilon\}.$$

  This is a metric on $\mathcal{M}_1(\mathbb{R})$ which induces the weak convergence.

Example 3.18. Let us compute the Lévy–Prokhorov distance between two Dirac masses δ_x and δ_y, with $x \neq y$, by proving that

$$\rho(\delta_x, \delta_y) = d(x, y) \wedge 1.$$

Let $\varepsilon > d(x, y)$. Let B be a Borel set. We distinguish case:

- If $x \notin B$, then $\delta_x(B) = 0 \leq \delta_y(B^\varepsilon) + \varepsilon$.
- If $x \in B$, then $y \in B^\varepsilon$ because $d(x, y) < \varepsilon$, so $\delta_x(B) = 1 \leq 1 + \varepsilon = \delta_y(B^\varepsilon) + \varepsilon$.

Hence, $\rho(\delta_x, \delta_y) \le \varepsilon$ and letting $\varepsilon \downarrow d(x,y)$ gives $\rho(\delta_x, \delta_y) \le d(x,y) \wedge 1$, by recalling that $\rho \le 1$. Now, let $0 < \varepsilon < d(x,y) \wedge 1$. Then, $y \notin \{x\}^\varepsilon$, hence

$$\delta_x(\{x\}) = 1 > \varepsilon = \delta_y(\{x\}^\varepsilon) + \varepsilon,$$

hence, we obtain $\rho(\delta_x, \delta_y) \ge d(x,y) \wedge 1$, concluding the proof.

Exercise 11. Let $\mu, \nu \in \mathcal{M}_1(E)$. The *total variation distance* between μ and ν is defined by

$$\delta(\mu, \nu) := \sup_{A \in \mathcal{B}(E)} |\mu(A) - \nu(A)|.$$

1. Prove that δ is a metric on $\mathcal{M}_1(E)$.
2. Prove that $\rho(\mu, \nu) \le \delta(\mu, \nu)$.
3. Find $(\mu_n)_{n \ge 0} \in \mathcal{M}_1(\mathbb{R})^{\mathbb{N}}$ and $\mu \in \mathcal{M}_1(\mathbb{R})$ such that $\rho(\mu_n, \mu) \to 0$ as $n \to \infty$ but $\delta(\mu_n, \mu) = 1$ for all $n \ge 1$.

When E is separable, the weak convergence is metrizable by the Lévy–Prokhorov distance:

Theorem 3.19. *Let (E, d) be a separable metric space. Let $(\mu_n)_{n \ge 0} \in \mathcal{M}_1(E)^{\mathbb{N}}$ and $\mu \in \mathcal{M}_1(E)$. Then,*

$$\mu_n \xrightarrow[n \to \infty]{(\mathrm{w})} \mu \qquad \Longleftrightarrow \qquad \rho(\mu_n, \mu) \xrightarrow[n \to \infty]{} 0.$$

Remark 3.20. Separability is only used for the implication $\Rightarrow$, the other one is true on any metric space E.

Proof. Suppose that $\rho(\mu_n, \mu) \to 0$. Let $\varepsilon > 0$. For n large enough, $\rho(\mu_n, \mu) < \varepsilon$ and so, for all closed set F, $\mu_n(F) \le \mu(F^\varepsilon) + \varepsilon$. Thus, for any $k \ge 1$, taking $\varepsilon = 1/k$, we get

$$\limsup_{n \to \infty} \mu_n(F) \le \mu\left(F^{1/k}\right) + \frac{1}{k}.$$

Noticing that $F = \bigcap_{k \ge 1} F^{1/k}$ because F is closed and that the intersection is non-increasing, we get, letting $k \to \infty$, that

$$\limsup_{n \to \infty} \mu_n(F) \le \mu(F),$$

which proves that $(\mu_n)_{n \ge 0}$ converges weakly to μ by Portmanteau theorem.

Conversely, assume that $\mu_n \to \mu$ weakly. Let $\varepsilon > 0$. Since E is separable, there is a dense sequence $(x_n)_{n \ge 1}$. Then

$$E = \bigcup_{n \ge 1} B(x_n, \varepsilon).$$

Then, define $B_1 = B(x_1, \varepsilon)$, and for $n \ge 1$, $B_{n+1} = B(x_{n+1}, \varepsilon) \setminus (B_1 \cup \cdots \cup B_n)$. The B_n's are pairwise disjoints, they have diameters smaller than 2ε, and

$$E = \bigcup_{n \ge 1} B_n.$$

Now let N be large enough, such that

$$\mu\left(\bigcup_{n=1}^{N} B_n\right) > 1 - \varepsilon.$$

Denote by $\mathcal{G}$ the family of open sets of the form

$$(B_{i_1} \cup \cdots \cup B_{i_k})^\varepsilon, \quad \text{with } \{i_1, \ldots, i_k\} \subset \{1, \ldots, N\}.$$

Hence, for any $G \in \mathcal{G}$, by the Portmantean theorem, we have

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G),$$

thus there exists n_0 such that for all $n \geq n_0$,

$$\mu_n(G) \geq \mu(G) - \varepsilon.$$

Moreover, as $\mathcal{G}$ is a finite family, this holds uniformly over $\mathcal{G}$: there exists n_0 such that for all $n \geq n_0$, for all $G \in \mathcal{G}$,

$$\mu_n(G) \geq \mu(G) - \varepsilon.$$

Now, let $A \in \mathcal{B}(E)$ be a Borel set. Define, $I_A = \{i \in \{1, \dots, N\} \mid B_i \cap A \neq \varnothing\}$, and let

$$G(A) = \bigcup_{i \in I_A} B_i.$$

It follows from this definition that, for $x \in \bigcup_{n=1}^N B_n$, if $x \in A$ then $x \in G(A)$, so $A \cap G(A)^c \subset (\bigcup_{n=1}^N B_n)^c$. Therefore,

$$\mu(A) = \mu(A \cap G(A)) + \mu(A \cap G(A)^c) \leq \mu(A \cap G(A)) + \varepsilon.$$

Then, for $n \geq n_0$, $\mu(A \cap G(A)) \leq \mu(G(A)^\varepsilon) \leq \mu_n(G(A)^\varepsilon) + \varepsilon$ because $G(A)^\varepsilon \in \mathcal{G}$. Furthermore, using that the B_k's have diameters less than 2ε, one gets that $G(A) \subset A^{2\varepsilon}$ and so $G(A)^\varepsilon \subset (A^{2\varepsilon})^\varepsilon \subset A^{3\varepsilon}$. Hence, for $n \geq n_0$,

$$\mu(A) \leq \mu_n(G(A)^\varepsilon) + 2\varepsilon \leq \mu_n\left(A^{3\varepsilon}\right) + 3\varepsilon.$$

Hence, for all $n \geq n_0$, $\rho(\mu_n, \mu) \leq 3\varepsilon$, which proves that $\rho(\mu_n, \mu) \to 0$. □

Exercise 12. Let (E, d) be a separable metric space. The goal of this exercise is to prove that $(\mathcal{M}_1(E), \rho)$ is separable. Let $(x_n)_{n\geq 1}$ be a dense sequence in E.

1. Let $\varepsilon > 0$ and $\mu \in \mathcal{M}_1(E)$. Define the sets B_n as in the proof of Theorem 3.19 and fix $N \geq 1$ such that $\mu(\bigcup_{n=1}^N B_n) > 1 - \varepsilon$. Prove that there exists $q_1, \dots, q_N \in \mathbb{Q} \cap [0, 1]$ such that

$$\sum_{k=1}^N |q_k - \mu(B_k)| < 3\varepsilon \quad \text{and} \quad \sum_{k=1}^N q_k = 1.$$

2. Set $\pi = \sum_{k=1}^N q_k \delta_{x_k}$. Using the sets $G(A)$ introduced in the proof of Theorem 3.19, prove that $\rho(\mu, \pi) \leq 4\varepsilon$.
3. Conclude.

3.5 The Kantorovich–Rubinstein distance

We now introduce another metric on $\mathcal{M}_1(E)$. Define, for all $f : E \to \mathbb{R}$,

$$\|f\|_{\mathrm{Lip}} := \sup_{\substack{x,y\in E \\ x\neq y}} \frac{|f(x) - f(y)|}{d(x, y)},$$

and

$$\|f\|_{\mathrm{BL}} := \|f\|_\infty + \|f\|_{\mathrm{Lip}}.$$

This defines a norm on the space of bounded Lipschitz functions on E.

Definition 3.21. *Let (E, d) be a metric space and let $\mu, \nu \in \mathcal{M}_1(E)$. The* Kantorovich–Rubinstein (or Fortet–Mourier) distance *between μ and ν is defined by*

$$\beta(\mu, \nu) := \sup_{\|f\|_{\mathrm{BL}} \leq 1} \left| \int_E f \, \mathrm{d}\mu - \int_E f \, \mathrm{d}\nu \right|.$$

Note that $\beta(\mu, \nu) \leq 2$ using that $\|f\|_{\mathrm{BL}} \leq 1$ implies $\|f\|_\infty \leq 1$. Furthermore, for any bounded Lipschitz function f on E,

$$\left| \int_E f \, \mathrm{d}\mu - \int_E f \, \mathrm{d}\nu \right| \leq \|f\|_{\mathrm{BL}} \beta(\mu, \nu),$$

by noting that $g = f / \|f\|_{\mathrm{BL}}$ satisfies $\|g\|_{\mathrm{BL}} \leq 1$.

Exercise 13. Prove that β is a metric on $\mathcal{M}_1(E)$.

Exercise 14. Show that

$$\beta(\delta_x, \delta_y) = \frac{2d(x, y)}{2 + d(x, y)}.$$

Hint: Use the function $f(z) = \frac{d(z,y) - d(z,x)}{2 + d(x,y)}$.

We now prove that the Lévy–Prokhorov distance and the Kantorovich–Rubinstein distance are equivalent, i.e. they induce the same topology, as a consequence of the following explicit inequalities. Note that this does not require separability.

Proposition 3.22. *Let (E, d) be a metric space For all $\mu, \nu \in \mathcal{M}_1(E)$,*

$$\frac{1}{4} \beta(\mu, \nu) \leq \rho(\mu, \nu) \leq \sqrt{2\beta(\mu, \nu)}.$$

Proof. We start with the second inequality. Assume that $\mu \neq \nu$, so $\beta(\mu, \nu) > 0$ (if not the inequality is trivial). Let $\varepsilon > 0$ and let $B \in \mathcal{B}(E)$. Consider the functions

$$f_\varepsilon(x) := \left(1 - \frac{1}{\varepsilon} d(x, B)\right) \vee 0,$$

as in Lemma 2.16 (but here B is not necessarily closed). Then, $\mathbb{1}_B \leq f_\varepsilon \leq \mathbb{1}_{B^\varepsilon}$ and $\|f_\varepsilon\|_{\mathrm{BL}} \leq 1 + \frac{1}{\varepsilon}$, so

$$\mu(B) \leq \int_E f_\varepsilon \, \mathrm{d}\mu \leq \int_E f_\varepsilon \, \mathrm{d}\nu + \left(1 + \frac{1}{\varepsilon}\right) \beta(\mu, \nu) \leq \nu(B^\varepsilon) + \left(1 + \frac{1}{\varepsilon}\right) \beta(\mu, \nu).$$

Thus, for any $\varepsilon > 0$,

$$\rho(\mu, \nu) \leq \max\left\{\varepsilon, \left(1 + \frac{1}{\varepsilon}\right) \beta(\mu, \nu)\right\}.$$

Choosing $\varepsilon = \sqrt{2\beta(\mu, \nu)}$, we get

$$\rho(\mu, \nu) \leq \sqrt{\beta(\mu, \nu)} \cdot \max\left\{\sqrt{2}, \sqrt{\beta(\mu, \nu)} + \frac{1}{\sqrt{2}}\right\},$$

If $\beta(\mu, \nu) \leq 1/2$, then the second argument in the maximum is at most $\sqrt{2}$ and so the result follows. If $\beta(\mu, \nu) > 1/2$, then the inequality can be directly obtained by using that $\rho(\mu, \nu) \leq 1$.

Now we prove the first inequality. Suppose without loss of generality that $\rho(\mu, \nu) > 0$ and let $\varepsilon > \rho(\mu, \nu)$. Let f be such that $\|f\|_{\mathrm{BL}} \leq 1$. Hence, $\|f\|_\infty \leq 1$ and $\|f\|_{\mathrm{Lip}} \leq 1$. Since f takes value in $[-1, 1]$, we have, using Fubini's theorem,

$$\int_E f \, \mathrm{d}\mu = \int_E \left(\int_{-1}^1 \mathbb{1}_{\{t < f(x)\}} \, \mathrm{d}t - 1 \right) \mathrm{d}\mu(x) = \int_{-1}^1 \mu(\{f > t\}) \, \mathrm{d}t - 1.$$

Using the same formula for ν, we get

$$\begin{aligned}\left|\int_E f\,\mathrm{d}\mu - \int_E f\,\mathrm{d}\nu\right| &= \left|\int_{-1}^1 \mu(\{f>t\})\,\mathrm{d}t - \int_{-1}^1 \nu(\{f>t\})\,\mathrm{d}t\right| \\ &\le \int_{-1}^1 |\mu(\{f>t\}) - \nu(\{f>t\})|\,\mathrm{d}t.\end{aligned}$$

Now remark that since f is 1-Lipschitz, we have $\{f>t\}^\varepsilon \subset \{f > t-\varepsilon\}$. Then, since $\rho(\mu,\nu) < \varepsilon$, we have

$$\begin{aligned}\mu(\{f>t\}) &\le \nu(\{f>t\}^\varepsilon) + \varepsilon \\ &\le \nu(\{f>t-\varepsilon\}) + \varepsilon.\end{aligned}$$

Hence,

$$\begin{aligned}\mu(\{f>t\}) - \nu(\{f>t\}) &\le \nu(\{f>t-\varepsilon\}) - \nu(\{f>t\}) + \varepsilon \\ &= \nu(\{t-\varepsilon < f \le t\}) + \varepsilon.\end{aligned}$$

Exchanging the role of μ and ν, we get by symmetry,

$$|\mu(\{f>t\}) - \nu(\{f>t\})| \le \mu(\{t-\varepsilon < f \le t\}) + \nu(\{t-\varepsilon < f \le t\}) + \varepsilon.$$

Using Fubini's theorem again, one obtains

$$\int_{-1}^1 \mu(\{t-\varepsilon < f \le t\})\,\mathrm{d}t = \int_E \int_{-1}^1 \mathbb{1}_{\{f \le t < f+\varepsilon\}}\,\mathrm{d}t\,\mathrm{d}\mu \le \varepsilon.$$

The same holds for the measure ν, so we get

$$\left|\int_E f\,\mathrm{d}\mu - \int_E f\,\mathrm{d}\nu\right| \le \int_{-1}^1 \Big(\mu(\{t-\varepsilon < f \le t\}) + \nu(\{t-\varepsilon < f \le t\}) + \varepsilon\Big)\,\mathrm{d}t \le 4\varepsilon.$$

Since this holds for any function f such that $\|f\|_{\mathrm{BL}} \le 1$, we deduce $\beta(\mu,\nu) \le 4\varepsilon$. Letting $\varepsilon \downarrow \rho(\mu,\nu)$ gives $\beta(\mu,\nu) \le 4\rho(\mu,\nu)$. □

As a consequence of the previous proposition together with Theorem 3.19, we obtain that the Kantorovich–Rubinstein distance induces the weak convergence.

Corollaire 3.23. *Let (E,d) be a separable metric space. Let $(\mu_n)_{n\ge 0} \in \mathcal{M}_1(E)^{\mathbb{N}}$ and $\mu \in \mathcal{M}_1(E)$. Then,*

$$\mu_n \xrightarrow[n\to\infty]{(\mathrm{w})} \mu \qquad \Longleftrightarrow \qquad \beta(\mu_n,\mu) \xrightarrow[n\to\infty]{} 0.$$

4 Tightness and Prokhorov's theorem

In this section, we establish compactness results for sets of probability measures. More precisely, all proofs establish *sequential compactness*, and compactness can be deduced as soon as weak convergence can be induced by a metric (i.e. as soon as E is separable by the previous section). However, in most applications, it is the sequential compactness which is useful.

4.1 Compactness of probability measures on a compact set

We start with a first compactness result when the underlying space is compact. The main ingredient for the proof is Riesz representation theorem (or Riesz-Markov-Kakutani theorem):

Theorem 4.1 (Riesz representation theorem)**.** *Let X be a compact metric space. Let $\Lambda\colon \mathcal{C}(X) \to \mathbb{R}$ be a continuous linear form such that:*

(i) *Λ is positive: for all $f \geq 0$, $\Lambda(f) \geq 0$,*

(ii) *$\Lambda(1) = 1$.*

Then, there exists a unique $\mu \in \mathcal{M}_1(X)$ such that, for all $f \in \mathcal{C}(X)$,

$$\Lambda(f) = \int_X f \, \mathrm{d}\mu.$$

Hence, the Riesz representation theorem asserts that the topological dual of $\mathcal{C}(X)$ corresponds to the set of bounded signed measures on X (recall that a finite signed measure is a measure taking values in $\mathbb{R}$).

In fact, the above theorem holds if X is only assumed to be Hausdorff and locally compact (any point of X admits a compact neighborhood) and separable and with $\mathcal{C}(X)$ replaced by $\mathcal{C}_c(X)$, the space of continuous functions with compact support, or $\mathcal{C}_0(X)$, the space of continuous functions vanishing at infinity. Of course, if X is compact, then $\mathcal{C}_c(X) = \mathcal{C}_0(X) = \mathcal{C}(X)$. This theorem is not easy and we refer for instance to Rudin [3] for a proof. Note that Riesz theorem is a way to construct Lebesgue measure on $\mathbb{R}$ starting with the Riemann integral. Note also that, when X is not compact, the topological dual of $\mathcal{C}_b(X)$ consists of objects more general that measures (in fact bounded finitely additive measures).

Theorem 4.2. *Let K be a compact metric space. Then, $\mathcal{M}_1(K)$ is sequentially compact for the weak convergence. Equivalently, $(\mathcal{M}_1(K), \rho)$ is compact.*

Proof. First recall that a compact metric space K is separable (see Example 2.13), so the Lévy–Prokhorov distance ρ induces the weak convergence on $\mathcal{M}_1(K)$. Therefore, sequential compactness of $\mathcal{M}_1(K)$ for the weak convergence is equivalent to sequential compactness of $(\mathcal{M}_1(K), \rho)$, which is itself equivalent to compactness of $(\mathcal{M}_1(K), \rho)$ because it is a metric space (see Theorem 2.7).

We now aim at proving that $\mathcal{M}_1(K)$ is sequentially compact for the weak convergence. Let $(\mu_n)_{n\geq 0} \in \mathcal{M}_1(K)^{\mathbb{N}}$. Recall that the space of continuous functions $\mathcal{C}(K)$ is separable (see Example 2.13). Let $\mathcal{A} = \{f_k : k \geq 0\}$ be a dense subset of $\mathcal{C}(K)$. Since K is compact, each f_k is bounded and so, for any $k, n \geq 0$,

$$\left| \int_K f_k \, \mathrm{d}\mu_n \right| \leq \|f_k\|_\infty < \infty.$$

Thus, since $[-\|f_k\|_\infty, \|f_k\|_\infty]$ is compact, any subsequence of $(\int f_k \, \mathrm{d}\mu_n)_{n\geq 0}$ has a convergent sub-subsequence. Using a diagonal extraction argument, one can construct an extraction φ such that, for all $k \geq 0$,

$$\int_K f_k \, \mathrm{d}\mu_{\varphi(n)}$$

converges, as $n \to \infty$, to some $\Lambda(f_k) \in \mathbb{R}$.

We now prove that this convergence holds on the whole set $\mathcal{C}(K)$. Let $\varepsilon > 0$ and let $f \in \mathcal{C}(K)$. By density of $\mathcal{A}$, there exists $k \geq 0$ such that $\|f - f_k\|_\infty < \varepsilon$. Then,

$$\begin{aligned}
&\left|\int f \, \mathrm{d}\mu_{\varphi(n)} - \int f \, \mathrm{d}\mu_{\varphi(m)}\right| \\
&\leq \left|\int f \, \mathrm{d}\mu_{\varphi(n)} - \int f_k \, \mathrm{d}\mu_{\varphi(n)}\right| + \left|\int f_k \, \mathrm{d}\mu_{\varphi(n)} - \int f_k \, \mathrm{d}\mu_{\varphi(m)}\right| + \left|\int f_k \, \mathrm{d}\mu_{\varphi(m)} - \int f \, \mathrm{d}\mu_{\varphi(m)}\right| \\
&\leq 3\varepsilon,
\end{aligned}$$

for n, m large enough, using $\|f - f_k\|_\infty < \varepsilon$ and that $(\int f_k \, \mathrm{d}\mu_{\varphi(n)})_{n\geq 0}$ converges. Hence, the sequence $(\int f \, \mathrm{d}\mu_{\varphi(n)})_{n\geq 0}$ in $\mathbb{R}$ is Cauchy and thus converges to some $\Lambda(f)$. Define $\Lambda \colon f \in \mathcal{C}(K) \mapsto \Lambda(f)$. Then Λ is a linear form on $\mathcal{C}(K)$ as the pointwise limit of a sequence of linear forms. Moreover, as

$$\left|\int f \, \mathrm{d}\mu_{\varphi(n)}\right| \leq \|f\|_\infty,$$

letting $n \to \infty$, one has that

$$|\Lambda(f)| \leq \|f\|_\infty.$$

As such, Λ is a continuous linear form on $\mathcal{C}(K)$. Moreover, $\Lambda(1) = 1$ and Λ is clearly positive, i.e. if $f \geq 0$, then $\Lambda(f) \geq 0$. By the Riesz representation theorem, there exists a probability measure $\mu \in \mathcal{M}_1(K)$ such that

$$\Lambda(f) = \int_K f \, \mathrm{d}\mu,$$

for all $f \in \mathcal{C}(K)$. Then, for all $f \in \mathcal{C}(K)$,

$$\int_K f \, \mathrm{d}\mu_{\varphi(n)} \underset{n\to\infty}{\longrightarrow} \int_K f \, \mathrm{d}\mu,$$

i.e. $(\mu_{\varphi(n)})_{n\geq 0}$ converges weakly to μ, hence proving that $\mathcal{M}_1(K)$ is sequentially compact for the weak convergence. □

Remark 4.3. Let Y be a normed vector space, recall that its topological dual Y^* can be equipped with the norm

$$\|\phi\| := \sup_{\|x\|=1} |\phi(x)|.$$

This norm induces a topology which is finer (i.e. with more open sets) than the weak* topology. More precisely, it is strictly finer iff Y is infinite dimensional. For example, the Banach–Alaoglu theorem states that the closed (for the norm) unit ball of Y^* is always compact for the weak* topology, whereas it is compact for the norm iff Y^* (or equivalently Y) is finite dimensional.

Theorem 4.2 can be seen as a consequence of the Banach–Alaoglu theorem. Indeed, by Riesz representation theorem and definition of the weak topology, $\mathcal{M}_1(K)$ is homeomorph to $\{\Lambda \in \mathcal{C}(K)^* : \Lambda \text{ positive and } \Lambda(1) = 1\}$, which is a closed subset of the norm-closed unit ball of $\mathcal{C}(K)^*$ and therefore compact by the Banach–Alaoglu theorem.

Note also that if Y is not separable, the closed unit ball of Y^* is not metrizable for the weak* topology. In particular, compactness and sequential compactness are not equivalent.

4.2 Prokhorov's theorem

For general metric spaces E, $\mathcal{M}_1(E)$ is not necessarily compact, see Exercise 15 below.

Exercise 15. Prove that the sequence $(\delta_n)_{n\geq 0}$ of Dirac measures on $\mathbb{R}$ does not have any convergent subsequence.

So we need a criterion that will allow us to still extract weakly converging subsequences. This is the content of Prokhorov's theorem, whose criterion is given by the following notion of tightness for sequences of probability measures.

Definition 4.4. *Let E be a metric space. Let $\mathcal{A} \subset \mathcal{M}_1(E)$. We say that $\mathcal{A}$ is* tight*, if for all $\varepsilon > 0$, there exists a compact subset $K \subset E$, such that*

$$\forall \mu \in \mathcal{A}, \quad \mu(K^c) \leq \varepsilon.$$

We have already seen that a single measure is tight on a Polish space. As a consequence, it follows that any *finite* family $\mathcal{A}$ of probability measures on a Polish space is tight.

Definition 4.5. *Let X be a topological space and $A \subset X$. We say that A is* relatively sequentially compact *if any sequence in A has a convergent subsequence in X.*

Exercise 16. Let E be a metric space and $A \subset E$. Prove that A is relatively sequentially compact iff it is relatively compact.

The most important result in this course is the following theorem.

Theorem 4.6 (Prokhorov's theorem)**.** *Let E be a metric space and let $\mathcal{A} \subset \mathcal{M}_1(E)$.*

(i) *If $\mathcal{A}$ is tight, then $\mathcal{A}$ is relatively sequentially compact for the weak convergence.*

(ii) *If E is Polish and $\mathcal{A}$ is relatively sequentially compact for the weak convergence, then $\mathcal{A}$ is tight.*

Remark 4.7. Part (i) is the part which is useful in applications (see Section 4.3), and it does not require further assumptions on the space E. It is also the most difficult part to prove. The proof presented here is based on Theorem 4.2, which itself relies on Riesz representation theorem, which we have not proved here. Billingsley follows a more direct approach in his book [1] constructing the limiting measure for a subsequence via an outer measure.

Remark 4.8. If E is separable, then the Lévy–Prokhorov distance ρ metrizes the weak convergence and so the assertion "$\mathcal{A}$ is relatively sequentially compact for the weak convergence" is equivalent to "$\mathcal{A}$ is relatively compact for ρ" by Exercise 16.

Proof. Part (ii). Assume E is Polish and $\mathcal{A}$ is relatively sequentially compact for the weak convergence. We first prove the following claim: If $(G_n)_{n\geq 1}$ is a non-decreasing sequence of open sets such that $E = \bigcup_{n\geq 1} G_n$, then, for all $\varepsilon > 0$, there exists $n \geq 1$ such that for all $\mu \in \mathcal{A}$,

$$\mu(G_n) > 1 - \varepsilon.$$

By contradiction, assume there exists $\varepsilon > 0$, such that for all $n \geq 1$, there exists $\mu_n \in \mathcal{A}$, such that

$$\mu_n(G_n) \leq 1 - \varepsilon.$$

Since $\mathcal{A}$ is relatively sequentially compact for the weak convergence, there exists a subsequence $(\mu_{\varphi(n)})_{n\geq 1}$ weakly converging to some probability measure μ. Then, for all $k \geq 1$, by the Portmanteau theorem and then the fact that $(G_n)_{n\geq 1}$ is non-decreasing,

$$\mu(G_k) \leq \liminf_{n\to\infty} \mu_{\varphi(n)}(G_k) \leq \liminf_{n\to\infty} \mu_{\varphi(n)}(G_{\varphi(n)}) \leq 1 - \varepsilon,$$

which contradicts the facts that $E = \bigcup_{n\geq 1} \uparrow G_n$ and $\mu(E) = 1$. Hence, the claim is proved.

We now prove tightness of $\mathcal{A}$ with ideas similar to the proof of Proposition 2.20. Let $\varepsilon > 0$. Let $(x_n)_{n\geq 1}$ be a dense sequence in E. Then, for any $k \geq 1$, one has

$$E = \bigcup_{n\geq 1} B\Big(x_n, \frac{1}{k}\Big),$$

By the previous claim applied with $G_n = \bigcup_{j=1}^n B(x_j, \frac{1}{k})$, for each $k \geq 1$, there exists n_k such that for all $\mu \in \mathcal{A}$,

$$\mu\left(\bigcup_{j=1}^{n_k} B\Big(x_j, \frac{1}{k}\Big)\right) > 1 - \frac{\varepsilon}{2^k}.$$

Set $K = \overline{A}$ with

$$A := \bigcap_{k\geq 1} \bigcup_{j=1}^{n_k} B\Big(x_j, \frac{1}{k}\Big).$$

Then, by construction A is totally bounded, hence since E is complete, K is compact. Moreover, for all $\mu \in \mathcal{A}$,

$$\mu(K^c) \leq \mu(A^c) \leq \sum_{k\geq 1} \mu\left(\left[\bigcup_{j=1}^{n_k} B(x_j, \frac{1}{k})\right]^c\right) \leq \sum_k \frac{\varepsilon}{2^k} = \varepsilon,$$

so $\mathcal{A}$ is tight.

Part (i). Assume $\mathcal{A}$ is tight. Let $(\mu_n)_{n\geq 0} \in \mathcal{A}^{\mathbb{N}}$. We have to prove that there exists a subsequence of $(\mu_n)_{n\geq 0}$ that converges weakly. By tightness, for all $j \geq 1$, there exists a compact $K_j \subset E$ such that, for all $n \geq 0$,

$$\mu_n(K_j) > 1 - \frac{1}{j}.$$

Up to replacing K_j by $\bigcup_{i=1}^j K_i$, one can assume that the sequence of compact sets $(K_j)_j$ is nondecreasing. Let $j \geq 1$ Now, for all $n \geq 0$, define ν_n^j by

$$\nu_n^j(A) = \frac{\mu_n(A \cap K_j)}{\mu_n(K_j)},$$

so that ν_n^j is a probability measure on the compact set K_j. Since $\mathcal{M}_1(K_j)$ is sequentially compact by Theorem 4.2, any subsequence of $(\nu_n^j)_{n\geq 0}$ has a sub-subsequence that converges weakly in $\mathcal{M}_1(K_j)$ By diagonal extraction, there exists an increasing φ such that, for all $j \geq 1$, there exists $\nu^j \in \mathcal{M}_1(K_j)$ such that

$$\nu^j_{\varphi(n)} \xrightarrow[n\to\infty]{\text{(w)}} \nu^j.$$

Now, for all $j \geq 1$, the sequence $(\mu_{\varphi(n)}(K_j))_{n\geq 0}$ is in the (sequentially) compact set $[0, 1]$. By diagonal extraction again (or the fact that $[0, 1]^{\mathbb{N}}$ is sequentially compact), there exists a subsequence $(\mu_{\varphi\circ\psi(n)}(K_j))_{n\geq 0}$ such that for all $j \geq 1$,

$$\mu_{\varphi\circ\psi(n)}(K_j) \underset{n\to\infty}{\longrightarrow} a_j \in [0, 1],$$

where $a_j \geq 1 - \frac{1}{j}$. Set $\phi = \varphi \circ \psi$. As a summary of the preceding discussion, we have that for all $j \geq 1$,

$$\begin{cases} \nu^j_{\phi(n)} \xrightarrow[n\to\infty]{\text{(w)}} \nu^j, \\ \mu_{\phi(n)}(K_j) \xrightarrow[n\to\infty]{} a_j \in \left[1 - \frac{1}{j}, 1\right]. \end{cases} \tag{4.1}$$

We still need to build the potential limit μ for our subsequence of $(\mu_n)_{n\geq 0}$. To do so, we show that $(a_j\nu^j)_{j\geq 1}$ is a nondecreasing sequence of measures and define μ as its limit. Then we prove that $\mu_{\phi(n)} \to \mu$ weakly as $n \to \infty$.

Let $j \geq 1$. For any $B \in \mathcal{B}(E)$ and $\varepsilon > 0$,

$$\begin{aligned}
a_{j+1}\nu^{j+1}(B^{2\varepsilon}) &\geq a_{j+1}\nu^{j+1}(\overline{B^\varepsilon}) && (\text{because } \overline{B^\varepsilon} \subset B^{2\varepsilon}) \\
&\geq \limsup_{n\to\infty}\Big(\mu_{\phi(n)}(K_{j+1})\nu^{j+1}_{\phi(n)}(\overline{B^\varepsilon})\Big) && (\text{by Portmanteau theorem}) \\
&= \limsup_{n\to\infty} \mu_{\phi(n)}(\overline{B^\varepsilon} \cap K_{j+1}) && (\text{by definition of } \nu^{j+1}_{\phi(n)}) \\
&\geq \limsup_{n\to\infty} \mu_{\phi(n)}(\overline{B^\varepsilon} \cap K_j) && (\text{because } K_j \subset K_{j+1}) \\
&= \limsup_{n\to\infty}\Big(\mu_{\phi(n)}(K_j)\nu^j_{\phi(n)}(\overline{B^\varepsilon})\Big) && (\text{by definition of } \nu^j_{\phi(n)}) \\
&= a_j \limsup_{n\to\infty} \nu^j_{\phi(n)}(\overline{B^\varepsilon}) && (\text{because } \mu_{\phi(n)}(K_j) \to a_j) \\
&\geq a_j \liminf_{n\to\infty} \nu^j_{\phi(n)}(B^\varepsilon) && (\text{because } B^\varepsilon \subset \overline{B^\varepsilon}) \\
&\geq a_j \nu^j(B^\varepsilon) && (\text{by Portmanteau theorem}).
\end{aligned}$$

(In the 3rd equality, we have used that if a_n and b_n are nonnegative and $a_n \to a$, then $\limsup_n(a_n b_n) = a\limsup_n b_n$). Applying this to a closed set F (so that $F = \bigcap_{\varepsilon>0} F^\varepsilon$) and letting $\varepsilon \downarrow 0$, we get

$$a_{j+1}\nu^{j+1}(F) \geq a_j\nu^j(F).$$

Then, by exterior regularity of probability measures on a metric space (Proposition 2.18), we deduce, for any $B \in \mathcal{B}(E)$,

$$a_{j+1}\nu^{j+1}(B) \geq a_j\nu^j(B).$$

We thus have a nondecreasing sequence of measures $(a_j\nu^j)_{j\geq 1}$, hence μ defined by

$$\mu(B) = \lim_{j\to\infty} a_j\nu^j(B), \quad B \in \mathcal{B}(E),$$

is a measure on E (since the pointwise limit of a nondecreasing sequence of measure is a measure by monotone convergence theorem). Moreover, since $1 - \frac{1}{j} \leq a_j \leq 1$, we have

$$\mu(E) = \lim_{j\to\infty} a_j = 1.$$

Hence, μ is a probability measure on E.

It remains to prove that $(\mu_{\phi(n)})_{n\geq 0}$ converges weakly to μ. For any closed set $F \subset E$ and $j \geq 1$,

$$\begin{aligned}
\limsup_{n\to\infty} \mu_{\phi(n)}(F) &\leq \limsup_{n\to\infty} \mu_{\phi(n)}(F \cap K_j) + \frac{1}{j} && (\text{because } \mu_{\phi(n)}(K_j^c) \leq 1/j) \\
&= \limsup_{n\to\infty} \mu_{\phi(n)}(K_j)\nu^j_{\phi(n)}(F) + \frac{1}{j} && (\text{by definition of } \nu^j_{\phi(n)}) \\
&= a_j\nu^j(F) + \frac{1}{j} && (\text{by (4.1)}).
\end{aligned}$$

By definition of μ, letting $j \to \infty$ yields $\limsup_n \mu_{\phi(n)}(F) \leq \mu(F)$. Thus, by Portmanteau theorem, $\mu_{\phi(n)} \to \mu$ weakly, and so $\mathcal{A}$ is sequentially relatively compact. □

4.3 Applications of Prokhorov's theorem

The typical use of this theorem will be the following: if one can prove that a sequence $(\mu_n)_n$ of probability measures on E is tight, then by Prokhorov's theorem, it is relatively compact, hence for any subsequence, one can extract a weakly convergent sub-subsequence. If moreover, one can prove that there is a unique accumulation point, then the whole sequence converges in view of the following easy but useful lemma:

Lemme 4.9. *Let (E, d) be a metric space and let $(x_n)_n \subset E$ and $x \in E$. Then $x_n \to x$ if and only if for all subsequence $(x_{\varphi(n)})_n$, there exists a further sub-subsequence $(x_{\varphi\circ\psi(n)})_n$ that converges to x.*

Proof. The direct half is trivial (and not very useful...). Now, if $x_n \not\to x$, there exists $\varepsilon > 0$, and some φ such that $d(x_{\varphi(n)}, x) > \varepsilon$, for all n. But then, no subsequence of $(x_{\varphi(n)})_n$ can converge to x. □

Example 4.10. Recall that for μ a probability measure on $\mathbb{R}$ (or more generally on $\mathbb{R}^d$), its characteristic function or Fourier transform is the function $\hat{\mu}$

$$\hat{\mu}(t) = \int_{\mathbb{R}} e^{itx} \mu(dx), \quad t \in \mathbb{R}.$$

By Fourier inversion formula, μ is characterized by $\hat{\mu}$. Moreover, let μ_n and μ be probability measures on $\mathbb{R}$. It is well known that weak convergence is characterized by pointwise convergence of characteristic functions:

$$\mu_n \xrightarrow[n\to\infty]{(\mathrm{w})} \mu \quad \Leftrightarrow \quad \hat{\mu}_n(t) \xrightarrow[n\to\infty]{} \hat{\mu}(t), \quad \forall t \in \mathbb{R}.$$

Lévy's theorem asserts that in fact, we don't need to know a priori that the limit is the characteristic function of a probability measure, we just need continuity at 0:

Theorem 4.11 (Lévy's theorem)**.** *Let μ_n and μ be probability measures on $\mathbb{R}$. Suppose that there exists a function ϕ, which is continuous at 0, such that,*

$$\forall t \in \mathbb{R}, \quad \hat{\mu}_n(t) \underset{n\to\infty}{\to} \phi(t).$$

Then ϕ is the characteristic function of a probability measure μ on $\mathbb{R}$, and

$$\mu_n \xrightarrow[n\to\infty]{(\mathrm{w})} \mu.$$

Proof. Let us show that for any $\mu \in \mathcal{M}_1(\mathbb{R})$, and all $\delta > 0$,

$$\mu(|x| > 2/\delta) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \hat{\mu}(t)) \, \mathrm{d}t.$$

Indeed, using Fubini theorem, we have

$$\frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \hat{\mu}(t)) \, \mathrm{d}t = \frac{1}{\delta} \int_{\mathbb{R}} \int_{-\delta}^{\delta} \left(1 - e^{itx}\right) \mathrm{d}t \mu(dx) = 2 \int_{\mathbb{R}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right) \mu(dx).$$

Now, since the function $f : u \mapsto 1 - \frac{\sin u}{u}$ is nonnegative and satisfies $f(u) > \frac{1}{2}$ for $|u| > 2$, we get

$$2 \int_{\mathbb{R}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right) \mu(dx) \geq 2 \int_{|\delta x| > 2} \left(1 - \frac{\sin(\delta x)}{\delta x}\right) \mu(dx) \geq \mu(|\delta x| > 2),$$

which proves the claim.

Let $\varepsilon > 0$. By assumption, $\hat{\mu}_n$ converges pointwise to ϕ. Since $\hat{\mu}_n(0) = 1$, we also have $\phi(0) = 1$. Moreover, since ϕ is continuous at 0, there exists $\delta > 0$ such that

$$\frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi(t)) \, \mathrm{d}t < \frac{\varepsilon}{2}.$$

By dominated convergence theorem, we also have that there exists n_0 such that for all $n \geq n_0$,

$$\left| \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \hat{\mu}_n(t)) \, \mathrm{d}t - \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi(t)) \, \mathrm{d}t \right| < \frac{\varepsilon}{2}.$$

Hence, for all $n \geq n_0$, one has

$$\frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \hat{\mu}_n(t))\, \mathrm{d}t < \varepsilon,$$

and thus

$$\sup_{n \geq n_0} \mu_n(|x| > 2/\delta) \leq \varepsilon.$$

As such, the sequence $(\mu_n)_n$ is tight, so by Prokhorov's theorem, there exists a subsequence $(\mu_{n_k})_k$ that converges weakly to some probability measure μ. By dominated convergence theorem, this implies pointwise convergence of the corresponding characteristic functions:

$$\hat{\mu}_{n_k}(t) \underset{k \to \infty}{\to} \hat{\mu}(t), \quad \forall t \in \mathbb{R}.$$

Hence, one obtains that $\phi(t) = \hat{\mu}(t)$, for all $t \in \mathbb{R}$, so ϕ is the characteristic function of μ. Since μ is characterized by its characteristic function, we get that any weak accumulation point of $(\mu_n)_n$ is equal to μ, so the whole sequence converges weakly to μ. $\square$

Exercise 17. Let E be a separable complete metric space. Show that $(\mathcal{M}_1(E), \rho)$ is also a separable complete metric space:

(i) Let $(\mu_n)_n$ be a Cauchy sequence in $(\mathcal{M}_1(E), \rho)$. Let $\varepsilon > 0$ and $\{x_k\}_k$ be a dense sequence in E. Show that there exists $M > 0$ such that for all n,

$$\mu_n\left(\bigcup_{i=1}^{M} B(x_i, 2\varepsilon)\right) \geq 1 - 2\varepsilon.$$

(ii) Deduce that $(\mu_n)_n$ is tight.

(iii) Conclude.

5 Convergence of stochastic processes

In the sequel, we denote by $\mathscr{C} = \mathcal{C}([0,1])$ the space of continuous functions from $[0,1]$ to $\mathbb{R}$.

Definition 5.1. *A continuous stochastic process $X = (X_t)_{t\in[0,1]}$ is a random variable with values in $\mathscr{C}$.*

It means that the map

$$\begin{aligned} X\colon \Omega &\to \mathscr{C} \\ \omega &\mapsto X(\omega) \end{aligned}$$

is a measurable map from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the measurable space $(\mathscr{C}, \mathcal{B}(\mathscr{C}))$. Hence, for almost all $\omega \in \Omega$,

$$\begin{aligned} X(\omega)\colon [0,1] &\to \mathbb{R} \\ t &\mapsto X_t(\omega) \end{aligned}$$

is a continuous function on $[0,1]$, i.e. the map $t \mapsto X_t$ is a.s. a continuous random function. We will use the notations $(X_t)_{t\in[0,1]}$ or $(X(t))_{t\in[0,1]}$ interchangeably.

5.1 Borel σ-algebra and cylindrical σ-algebra

Let, for all $t \in [0,1]$,

$$\begin{aligned} \pi_t\colon \mathscr{C} &\to \mathbb{R} \\ x &\mapsto x_t \end{aligned}$$

be the evaluation map at time t. The cylindrical σ-algebra $\mathrm{Cyl}(\mathscr{C})$ is the smallest σ-algebra such that all the evaluation maps $\pi_t, t \in [0,1]$, are measurable, i.e.

$$\begin{aligned} \mathrm{Cyl}(\mathscr{C}) &= \sigma(\pi_t,\ t \in [0,1]) \\ &= \sigma\left(\bigcup_{t\in[0,1]} \pi_t^{-1}(\mathcal{B}(\mathbb{R}))\right). \end{aligned}$$

Then we have the following.

Proposition 5.2. *We have,*

$$\mathcal{B}(\mathscr{C}) = \mathrm{Cyl}(\mathscr{C}).$$

Proof. For all $t \in [0,1]$, π_t is continuous (since if $X_n \to X$ in $\mathscr{C}$, i.e. $X_n \to X$ uniformly, then $X_n \to X$ pointwise, i.e. for all $t \in [0,1]$, $\pi_t(X_n) \to \pi_t(X)$). Thus, by definition of the cylindrical σ-algebra $\mathrm{Cyl}(\mathscr{C})$, one has

$$\mathrm{Cyl}(\mathscr{C}) \subset \mathcal{B}(\mathscr{C}).$$

Now we prove that any open ball of $\mathscr{C}$ lies in $\mathrm{Cyl}(\mathscr{C})$. Let $x \in \mathscr{C}$ and let $r > 0$. Since

$$B(x,r) = \bigcup_{\substack{s\in\mathbb{Q} \\ 0<s<r}} B_f(x,s),$$

it suffices to prove that any closed ball lies in $\mathrm{Cyl}(\mathscr{C})$. But,

$$\begin{aligned}
B_f(x,r) &= \{y \in \mathscr{C} \mid d(x,y) \le r\} \\
&= \left\{ y \in \mathscr{C} \mid \sup_{t\in[0,1]} |x(t)-y(t)| \le r \right\} \\
&= \bigcap_{t\in[0,1]} \{y \in \mathscr{C} \mid |x(t)-y(t)| \le r\} \\
&= \bigcap_{t\in[0,1]\cap\mathbb{Q}} \{y \in \mathscr{C} \mid |x(t)-y(t)| \le r\} && \text{(by continuity)} \\
&= \bigcap_{t\in[0,1]\cap\mathbb{Q}} \pi_t^{-1}(B_f(x(t),r)).
\end{aligned}$$

As such, $B_f(x,r) \in \mathrm{Cyl}(\mathscr{C})$. Since $\mathscr{C}$ is separable, any open set is a countable union of open balls or closed balls. Hence, the sets of open balls is included in $\mathrm{Cyl}(\mathscr{C})$, and finally $\mathcal{B}(\mathscr{C}) \subset \mathrm{Cyl}(\mathscr{C})$. □

This has the following important consequence. For all $k \ge 1$, and all $0 \le t_1 < \cdots < t_k \le 1$, define

$$\begin{aligned}
\pi_{t_1,\dots,t_k} : \mathscr{C} &\to \mathbb{R}^k \\
f &\mapsto (f(t_1),\dots,f(t_k)),
\end{aligned}$$

and consider the class of cylinders

$$\mathcal{C} = \left\{ \pi_{t_1,\dots,t_k}^{-1}(B_1 \times \cdots \times B_k) \mid k \ge 1, 0 \le t_1 < \cdots < t_k \le 1, B_1,\dots,B_k \in \mathcal{B}(\mathbb{R}) \right\}.$$

Then, $\mathcal{C}$ is a π-system that generates $\mathrm{Cyl}(\mathscr{C})$. Hence,

$$\mathcal{B}(\mathscr{C}) = \sigma(\mathcal{C}).$$

Since by Dynkin π–λ theorem, two probability measures are equal if and only if they are equal on a π-system that generates the Borel σ-algebra, we get the for all μ and ν in $\mathcal{M}_1(\mathscr{C})$, μ and ν are equal if and only if for all $t_1 < \cdots < t_k$,

$$\pi_{t_1,\dots,t_k\#}\mu = \pi_{t_1,\dots,t_k\#}\nu,$$

where $T_{\#}\mu$ denotes the image measure of μ by the map T. From a probabilistic point of view, let $X = (X_t)_{t\in[0,1]}$ and $Y = (Y_t)_{t\in[0,1]}$ be two continuous processes, i.e. two random elements of $\mathscr{C}$. Then,

$$X \overset{(d)}{=} Y$$

if and only if they have the same *finite-dimensional distributions*: for all $k \ge 1$, and all $0 \le t_1 < \cdots < t_k \le 1$,

$$(X_{t_1},\dots,X_{t_k}) \overset{(d)}{=} (Y_{t_1},\dots,Y_{t_k})$$

(as random vectors in $\mathbb{R}^k$!).

In the sequel, we will denote for all $k \ge 1$ and all $t_1,\dots,t_k$ by

$$\mu^{t_1,\dots,t_k} = \pi_{t_1,\dots,t_k\#}\mu$$

the finite-dimensional marginals of $\mu \in \mathcal{M}_1(\mathscr{C})$.

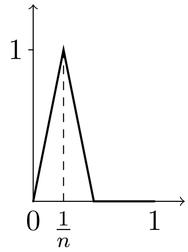
Note that by the continuous mapping theorem, if the sequence $(\mu_n)_n$ of probability measures on $\mathscr{C}$ converges weakly to some $\mu \in \mathcal{M}_1(\mathscr{C})$, then all the finite-dimensional marginals of μ_n converges to that of μ: for all $k \ge 1$, and all $0 \le t_1 < \cdots < t_k \le 1$,

$$\mu_n^{t_1,\dots,t_k} \xrightarrow[n\to\infty]{\text{(w)}} \mu^{t_1,\dots,t_k}.$$

The above convergence is thus weak convergence in $\mathcal{M}_1(\mathbb{R}^k)$ and is called *convergence of the finite-dimensional distributions*. But the converse does not hold! For instance, consider the sequence of continuous functions $(z_n)_n$, where

$$z_n(t) = \max\left\{1 - n|t - \frac{1}{n}|, 0\right\}, \quad t \in [0,1].$$

Then, the sequence of Dirac masses $(\delta_{z_n})_n$ converges in the sense of finite dimensional distributions



to the Dirac mass at the zero function δ_0. Indeed, recall that δ_{z_n} converges weakly to δ_0 if and only if $z_n \to 0$ (the zero function). But $\|z_n\|_\infty = 1$, hence z_n does not converge to the zero function. But clearly, for any fixed $t \in [0,1]$, $z_n(t) \to 0$ as $n \to \infty$, and so δ_{z_n} converges in finite-dimensional distributions to the zero function.

Strategy: The strategy for proving that a sequence of probability measures $(\mu_n)_n \subset \mathcal{M}_1(\mathscr{C})$ converges weakly to some probability measure μ is thus the following:

- Prove that $(\mu_n)_n$ is tight. By Prokhorov's theorem, it implies that $(\mu_n)_n$ is relatively compact: for all subsequence of $(\mu_n)_n$, there exists a further sub-subsequence that converges weakly.
- Prove that there is a unique accumulation point. To that end, prove that the finite-dimensional distributions converge to that of μ.

Indeed, if we known that $(\mu_n)_n$ is relatively compact, then for all subsequence of $(\mu_n)_n$, there exists a further sub-subsequence, say μ_{n_i}, that converges weakly to some ν (depending on the subsequence). By the continuous mapping theorem, the finite dimensional distributions will also converge: for all k, for all $t_1, \ldots, t_k$,

$$\mu_{n_i}^{t_1,\ldots,t_k} \xrightarrow[i\to\infty]{(\mathrm{w})} \nu^{t_1,\ldots,t_k}.$$

But if we also known that for all $t_1, \ldots, t_k$,

$$\mu_n^{t_1,\ldots,t_k} \xrightarrow[n\to\infty]{(\mathrm{w})} \mu^{t_1,\ldots,t_k},$$

then,

$$\nu^{t_1,\ldots,t_k} = \mu^{t_1,\ldots,t_k},$$

for all k, and thus $\nu = \mu$ since finite-dimensional distributions characterize probability measures on $\mathscr{C}$. Hence, for all subsequence, there exists a further sub-subsequence that converges weakly to the same μ, hence the whole sequence $(\mu_n)_n$ converges weakly to μ.

5.2 Tightness in $\mathscr{C}$

To understand tightness of families of probability measures on $\mathscr{C}$, we first need the understand compactness in $\mathscr{C}$. This is done using the Arzelà-Ascoli theorem, which will be of fundamental importance in the sequel. First recall the definition of equicontinuity.

Definition 5.3. *Let K be a compact metric space. A subset $\mathcal{A} \subset \mathcal{C}(K)$ is said to be (uniformly) equicontinuous if for all $\varepsilon > 0$, there exists $\delta > 0$, such that for all $f \in \mathcal{A}$, for all $x, y \in K$ such that $d(x, y) < \delta$, we have*

$$|f(x) - f(y)| < \varepsilon.$$

We can restate this definition using the *modulus of continuity.* Define, for f continuous,

$$\omega_\delta(f) = \sup_{d(x,y)<\delta} |f(x) - f(y)|,$$

so f is uniformly continuous if and only if $\omega_\delta(f) \to 0$ as $\delta \to 0$. Then $\mathcal{A} \subset \mathcal{C}(K)$ is equicontinuous if and only if

$$\sup_{f\in\mathcal{A}} \omega_\delta(f) \underset{\delta\to 0}{\longrightarrow} 0.$$

Roughly speaking, a sequence of functions is equicontinuous if and only if it is "uniformly uniformly continuous".

Theorem 5.4 (Arzelà-Ascoli theorem)**.** *Let K be a compact metric space. Then $\mathcal{A} \subset \mathcal{C}(K)$ is relatively compact if and only if*

(1) *$\mathcal{A}$ is "pointwise uniformly bounded":*

$$\forall x \in K, \quad \sup_{f\in\mathcal{A}} |f(x)| < \infty.$$

(2) *$\mathcal{A}$ is equicontinuous:*

$$\sup_{f\in\mathcal{A}} \omega_\delta(f) \underset{\delta\to 0}{\longrightarrow} 0.$$

Proof. Suppose that $\mathcal{A}$ is relatively compact. Since the evaluation map $\pi_x : f \mapsto f(x)$ is continuous, $\pi_x(\overline{\mathcal{A}})$ is a compact subset of $\mathbb{R}$. Since $\pi_x(\overline{\mathcal{A}})$ contains $\pi_x(\mathcal{A}) = \{f(x) \mid f \in \mathcal{A}\}$, we get that for all $f \in \mathcal{A}$, $|f(x)| < \infty$. This proves (1). Replacing $\mathcal{A}$ by its closure, one may suppose that $\mathcal{A}$ is compact (a subfamily of an equicontinuous family is equicontinuous). Let $\varepsilon > 0$. Let $g_1, \ldots, g_r$ a finite family of functions in $\mathcal{A}$ such that

$$\mathcal{A} \subset \bigcup_{k=1}^{r} B(g_k, \varepsilon).$$

Let $f \in \mathcal{A}$ and let $k \in \{1, \ldots, r\}$ such that $d(f, g_k) < \varepsilon$. Since K is compact, g_k is uniformly continuous for all k, and since the family $g_1, \ldots, g_r$ is finite, there exists $\delta > 0$, such that for all $k \in \{1, \ldots, r\}$, $\omega_\delta(g_k) < \varepsilon$. Hence, for all $x, y \in K$, such that $d(x, y) < \delta$,

$$|f(x) - f(y)| \leq |f(x) - g_k(x)| + |g_k(x) - g_k(y)| + |g_k(y) - f(y)| \leq 3\varepsilon.$$

This proves the equicontinuity of $\mathcal{A}$.

Conversely, assume (1) and (2). Let $(f_n)_n \subset \mathcal{A}$. Our aim is to prove that there exists a subsequence of $(f_n)_n$ that converges uniformly to some f. Since K is compact, K is separable, and denote by $Q = \{q_1, q_2, \ldots\}$ a dense countable subset of K. By assumption (1), for all $q \in Q$, the sequence $(f_n(q))_n$ is included in a compact subset of $\mathbb{R}$. Hence, there exists a subsequence $(f_{\phi_q(n)}(q))_n$ that converges to some $f(q) \in \mathbb{R}$. Using a diagonal extraction argument, one has that there exists an extraction ϕ such that for all $q \in Q$, $f_{\phi(n)}(q) \to f(q)$, as $n \to \infty$. We have that f is uniformly continuous on Q. Indeed, by (2), for all $\varepsilon > 0$, there exists $\delta > 0$, such that for all $x, y \in Q$ such that $d(x, y) < \delta$, for all n, we have

$$|f_{\phi(n)}(x) - f_{\phi(n)}(y)| < \varepsilon,$$

and letting $n \to \infty$ gives that $|f(x) - f(y)| < \varepsilon$. Hence, f is uniformly continuous on a dense subset of K, so there exists a unique extension, still denoted by f, to K, which is uniformly continuous (exercise). It remains to prove that $(f_{\phi(n)})_n$ converges uniformly to f (exercise). Let $\delta > 0$ such that

$$\sup_{g \in \mathcal{A}} \omega_\delta(g) < \varepsilon \quad \text{and} \quad \omega_\delta(f) < \varepsilon.$$

Since Q is dense, one has

$$K \subset \bigcup_{q \in Q} B(q, \delta).$$

By compactness of K, there exists a finite subcover, say:

$$K \subset \bigcup_{j=1}^{J} B(q_j, \delta).$$

Let $x \in K$, and let $j \in \{1, \dots, J\}$ such that $d(x, q_j) < \delta$. Then, we have

$$\begin{aligned}
|f_{\phi(n)}(x) - f(x)| &\leq |f_{\phi(n)}(x) - f_{\phi(n)}(q_j)| + |f_{\phi(n)}(q_j) - f(q_j)| + |f(q_j) - f(x)| \\
&\leq \sup_{g \in \mathcal{A}} \omega_\delta(g) + \max_{j=1,\dots,J} |f_{\phi(n)}(q_j) - f(q_j)| + \omega_\delta(f) \\
&\leq 2\varepsilon + \max_{j=1,\dots,J} |f_{\phi(n)}(q_j) - f(q_j)|.
\end{aligned}$$

Now, since $f_{\phi(n)}(q) \underset{n \to \infty}{\longrightarrow} f(q)$ for any $q \in Q$, there exists n_0 (depending only on ε), such that for all $n \geq n_0$,

$$\max_{j=1,\dots,J} |f_{\phi(n)}(q_j) - f(q_j)| \leq \varepsilon.$$

Finally, $\sup_{x \in K} |f_{\phi(n)}(x) - f(x)| \leq 3\varepsilon$, for all $n \geq n_0$. This concludes the proof. $\square$

Remark 5.5. When $K = [0, 1]$ (or more generally a compact convex subset of $\mathbb{R}^d$), one can replace assumption (1) in Arzelà-Ascoli theorem by:

(1') $\sup_{f \in \mathcal{A}} |f(0)| < \infty$.

Indeed, (1')+(2) implies easily (1). Let $x \in [0, 1]$ and write

$$|f(x)| \leq |f(0)| + \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|,$$

where for $k = 1, \dots, n$, $x_k = \frac{k}{n} x$, so $|x_k - x_{k-1}| = \frac{x}{n} \leq \frac{1}{n}$. Choose n large enough so that $\frac{1}{n} < \delta$. Then, for all $f \in \mathcal{A}$,

$$|f(x)| \leq \sup_{f \in \mathcal{A}} |f(0)| + n \sup_{f \in \mathcal{A}} \omega_\delta(f),$$

which is finite by (1) and (2).

Exercise 18. Let X a metric space and Y a *complete* metric space. Let $f \colon A \to Y$ be a uniformly continuous function on a dense subset $A \subset X$. Then, there exists a unique uniformly continuous extension of f to the whole X.

The next proposition is then more or less a direct translation of the Arzelà-Ascoli theorem.

Proposition 5.6. *Let $\mathcal{A} \subset \mathcal{M}_1(\mathscr{C})$. Then, $\mathcal{A}$ is tight if and only if one has:*

(i) *the family $\{\mu^0 \mid \mu \in \mathcal{A}\}$ is a tight family in $\mathcal{M}_1(\mathbb{R})$,*

(ii) *for all $\varepsilon > 0$ and all $\eta > 0$, there exists $\delta > 0$ such that for all $\mu \in \mathcal{A}$,*

$$\mu(\{x \in \mathscr{C} \mid \omega_\delta(x) > \eta\}) \leq \varepsilon.$$

Proof. Assume $\mathcal{A}$ is tight. Let $\varepsilon > 0$, and let $\mathcal{K}$ a compact set in $\mathscr{C}$ such that for all $\mu \in \mathcal{A}$,

$$\mu(\mathcal{K}) \geq 1 - \varepsilon.$$

By continuity $\mathcal{K}_0 := \pi_0(\mathcal{K})$ is compact in $\mathbb{R}$ and then

$$\mu^0(\mathcal{K}_0) = \mu\Big(\pi_0^{-1}(\mathcal{K}_0)\Big) = \mu(\{x \in \mathscr{C} \mid x_0 \in \mathcal{K}_0\}) \geq \mu(\mathcal{K}) \geq 1 - \varepsilon,$$

for all $\mu \in \mathcal{A}$. This proves the first item. Now let also $\eta > 0$. Since $\mathcal{K} \subset \mathscr{C}$ is compact, by Arzelà-Ascoli theorem $\mathcal{K}$ is equicontinuous. Hence, there exists $\delta > 0$ such that $\sup_{x \in \mathcal{K}} \omega_\delta(x) < \eta$. Hence, for all $\mu \in \mathcal{A}$,

$$\mu(\{x \in \mathscr{C} \mid \omega_\delta(x) \leq \eta\}) \geq \mu(\mathcal{K}) \geq 1 - \varepsilon,$$

which proves the second assertion.

Conversely, assume (i) and (ii). Let $\varepsilon > 0$ and for all n, let $\delta_n > 0$ such that for all $\mu \in \mathcal{A}$,

$$\mu\left(\left\{x \in \mathscr{C} \mid \omega_{\delta_n}(x) \leq \frac{1}{n}\right\}\right) \geq 1 - \frac{\varepsilon}{2^n}.$$

Let also $K_0 \subset \mathbb{R}$ a compact set such that, for all $\mu \in \mathcal{A}$,

$$\mu^0(K_0) \geq 1 - \varepsilon.$$

Now, define,

$$\mathcal{K} = \bigcap_n \left\{x \in \mathscr{C} \mid \omega_\delta(x) \leq \frac{1}{n}\right\} \cap \{x \in \mathscr{C} \mid x_0 \in K_0\}.$$

Then $\mathcal{K}$ is closed. Indeed, $\{x \in \mathscr{C} \mid x_0 \in K_0\} = \pi_0^{-1}(K_0)$ is the inverse image of the compact set K_0 by the continuous function π_0, hence is closed. Moreover, for all n, $\left\{x \in \mathscr{C} \mid \omega_\delta(x) \leq \frac{1}{n}\right\}$ is closed, since if $(x_k)_k \subset \mathscr{C}$ converges uniformly to x, then for all $s, t \in [0,1]$,

$$\begin{aligned}
|x(s) - x(t)| &\leq |x(s) - x_k(s)| + |x_k(s) - x_k(t)| + |x_k(t) - x(t)| \\
&\leq \|x - x_k\|_\infty + \omega_\delta(x_k) + \|x_k - x\|_\infty \\
&\leq 2\|x - x_k\|_\infty + \frac{1}{n} \\
&\leq \frac{1}{n} \quad \text{as } k \to \infty.
\end{aligned}$$

By Arzelà-Ascoli theorem, $\mathcal{K}$ is relatively compact, hence compact, and we have, for all $\mu \in \mathcal{A}$,

$$\begin{aligned}
\mu(\mathcal{K}^c) &\leq \sum_{n \geq 1} \mu\left(\left\{x \in \mathscr{C} \mid \omega_{\delta_n}(x) > \frac{1}{n}\right\}\right) + \mu(K_0^c) \\
&\leq \sum_{n \geq 1} \frac{\varepsilon}{2^n} + \varepsilon = 2\varepsilon.
\end{aligned}$$

Hence, $\mathcal{A}$ is tight in $\mathscr{C}$. $\square$

In more probabilistic terms, the previous proposition translates as:

Proposition 5.7. *Let $(X^{(n)})_n$ be a sequence of stochastic processes in $\mathscr{C}$. Then, $(X^{(n)})_n$ is tight if and only if one has*

(i) *the sequence $(X_0^{(n)})_n$ is tight (in $\mathbb{R}$),*

(ii) *for all $\varepsilon > 0$ and all $\eta > 0$, there exists $\delta > 0$ such that*

$$\limsup_{n \to \infty} \mathbb{P}\Big(\omega_\delta(X^{(n)}) > \eta\Big) \leq \varepsilon.$$

The previous proposition is not very convenient to use, since it involves the modulus of continuity, which is generally difficult to handle. Several criteria for tightness are known; the following one, due to Kolmogorov, is perhaps the easiest to apply.

For $f \in \mathcal{C}([0,1])$, define the α-Hölder norm (it's not a norm) by

$$N_\alpha(f) = \sup_{\substack{s,t\in[0,1] \\ s\neq t}} \frac{|f(s)-f(t)|}{|s-t|^\alpha}.$$

Proposition 5.8 (Kolmogorov criterion)**.** *Let $(X^{(n)})_{n\geq 1}$ be a sequence of random variables in $\mathscr{C}$. Suppose that*

- *the sequence $(X^{(n)}(0))_{n\geq 1}$ is tight in $\mathbb{R}$,*
- *there exists β, p, C positive such that for all $s,t \in [0,1]$,*

$$\sup_{n\geq 1} \mathbb{E}\Big|X^{(n)}(s) - X^{(n)}(t)\Big|^p \leq C|t-s|^{1+\beta}.$$

Then, $(X^{(n)})_{n\geq 1}$ is tight. More precisely, for all $\alpha \in]0, \frac{\beta}{p}[$ and all $\varepsilon > 0$, there exists $M > 0$ such that

$$\sup_{n\geq 1} \mathbb{P}(N_\alpha(X^{(n)}) > M) \leq \varepsilon.$$

To prove this criterion, we will need the following lemma. Denote, for all $k \geq 0$,

$$\mathbb{D}_k = \left\{ \frac{i}{2^k} \mid i = 0, \dots, 2^k \right\}$$

the dyadic rationals of order k, and let $\mathbb{D} = \bigcup_{k\geq 0} \mathbb{D}_k$ the dyadic rationals in $[0,1]$. Any dyadic rational $d \in \mathbb{D}_l$ can be written

$$d = \sum_{k=0}^{l} \frac{a_k}{2^k},$$

where $a_0, \dots, a_l$ are in $\{0,1\}$. Note that $\mathbb{D}_k \subset \mathbb{D}_{k+1}$ for all k, and recall also that $\mathbb{D}$ is dense in $[0,1]$.

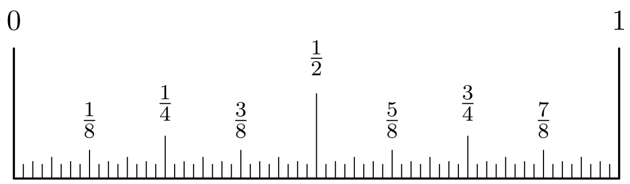


Figure 3: The dyadic rationals in $[0,1]$.

Lemme 5.9. *Let $x\colon [0,1] \to \mathbb{R}$. Suppose there exists $\alpha > 0$ and $K \geq 0$ such that for all $k \geq 0$ and for any two consecutive elements s and t of $\mathbb{D}_k$, one has*

$$|x_t - x_s| \leq K|t-s|^\alpha.$$

Then, for all $s,t \in \mathbb{D}$, we have

$$|x_t - x_s| \leq \frac{2K}{1-2^{-\alpha}}|t-s|^\alpha.$$

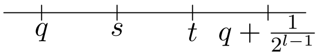
If moreover, $x \in \mathscr{C}$, then $N_\alpha(x) \leq \frac{2K}{1-2^{-\alpha}}$.

Proof. Let $s, t \in \mathbb{D}$ and assume without loss of generality that $s < t$. Note that if s and t are consecutive in some $\mathbb{D}_l$, then the inequality is trivial since $K \leq \frac{2K}{1-2^{-\alpha}}$. If not, the idea is to write the interval $[s, t]$ as a union of intervals $[\tau_k, \tau_{k+1}]$, where τ_k and τ_{k+1} are consecutive dyadic rationals in some $\mathbb{D}_l$.

So let $l \geq 1$ be the unique integer such that

$$\frac{1}{2^l} < t - s \leq \frac{1}{2^{l-1}}.$$

(Draw a picture).



Suppose first that there exists $q \in \mathbb{D}_{l-1}$ such that $q \leq s < t < q + \frac{1}{2^{l-1}}$, i.e. s and t lie between two consecutive dyadic rationals in $\mathbb{D}_{l-1}$. Thus, we can write s and t as:

$$s = q + \sum_{k \geq l} \frac{a_k}{2^k} \quad \text{and} \quad t = q + \sum_{k \geq l} \frac{b_k}{2^k},$$

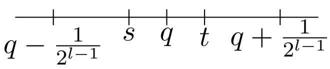
where a_k and b_k are in $\{0, 1\}$ and are equal to zero for k large enough. Let for $n \geq l$,

$$s_n = q + \sum_{k=l}^{n} \frac{a_k}{2^k} \quad \text{and} \quad t_n = q + \sum_{k=l}^{n} \frac{b_k}{2^k},$$

and let $L \geq 1$ large enough such that $s_L = s$ and $t_L = t$. Then, s_n and s_{n+1} are equal or are two consecutive dyadic of $\mathbb{D}_{n+1}$ and the same holds for t_n and t_{n+1}, hence, using the convention that $s_{l-1} = t_{l-1} = q$,

$$\begin{aligned}
d(x_s, x_t) &= d(x_{s_L}, x_{t_L}) \\
&\leq \sum_{k=l-1}^{L} d(x_{s_k}, x_{s_{k+1}}) + \sum_{k=l-1}^{L} d(x_{t_k}, x_{s_{t+1}}) \\
&\leq 2K \sum_{k \geq l-1} \frac{1}{2^{(k+1)\alpha}} \\
&= \frac{2K}{1-2^{-\alpha}} \frac{1}{2^{l\alpha}} \\
&\leq \frac{2K}{1-2^{-\alpha}} (t-s)^{\alpha}.
\end{aligned}$$

Now suppose that there exists $q \in \mathbb{D}_{l-1}$ such that $q - \frac{1}{2^{l-1}} < s < q \leq t < q + \frac{1}{2^{l-1}}$. This time, we



write

$$s = q - \sum_{k \geq l} \frac{a_k}{2^k} \quad \text{and} \quad t = q + \sum_{k \geq l} \frac{b_k}{2^k},$$

and the same reasoning as in the previous case gives the result.

If moreover $x \in \mathscr{C}$, by density of $\mathbb{D}$ in $[0, 1]$, one obtains $N_\alpha(x) \leq \frac{2K}{1-2^{-\alpha}}$. $\square$

Proof of Kolmogorov's criterion. For all $k \geq 0$ and all $n \geq 0$, define

$$Z_{n,k} = \max_{0 \leq i \leq 2^k - 1} \left| X^{(n)}\left(\frac{i}{2^k}\right) - X^{(n)}\left(\frac{i+1}{2^k}\right) \right|.$$

Then,

$$\begin{aligned}\mathbb{P}\Big(Z_{n,k} \geq K2^{-k\alpha}\Big) &= \mathbb{P}\left(\exists i \in \{0,\dots,2^k-1\}, \left|X^{(n)}\left(\frac{i}{2^k}\right) - X^{(n)}\left(\frac{i+1}{2^k}\right)\right| \geq K2^{-k\alpha}\right)\\ &\leq \sum_{i=0}^{2^k-1} \mathbb{P}\left(\left|X^{(n)}\left(\frac{i}{2^k}\right) - X^{(n)}\left(\frac{i+1}{2^k}\right)\right| \geq K2^{-k\alpha}\right)\\ &\leq 2^k \max_{0\leq i\leq 2^k-1} \mathbb{P}\left(\left|X^{(n)}\left(\frac{i}{2^k}\right) - X^{(n)}\left(\frac{i+1}{2^k}\right)\right| \geq K2^{-k\alpha}\right)\\ &\leq 2^k K^{-p} 2^{kp\alpha} \mathbb{E}\left(\left|X^{(n)}\left(\frac{i}{2^k}\right) - X^{(n)}\left(\frac{i+1}{2^k}\right)\right|^p\right),\end{aligned}$$

using Markov inequality. By assumption, we get

$$\mathbb{P}\Big(Z_{n,k} \geq K2^{-k\alpha}\Big) \leq CK^{-p}2^{-k(\beta-p\alpha)}.$$

Since $0 < \alpha < \frac{\beta}{p}$, by the previous lemma, we have

$$\mathbb{P}\left(N_\alpha(X^{(n)}) > \frac{2K}{1-2^{-\alpha}}\right) \leq \mathbb{P}\Big(\exists k \geq 0, Z_{n,k} > K2^{-k\alpha}\Big) \leq CK^{-p}\sum_{k\geq 0} 2^{-k(\beta-p\alpha)} = \frac{CK^{-p}}{1-2^{\beta-p\alpha}},$$

the right hand side being independent of n and goes to 0 as $K \to \infty$. Hence, we obtain that for all $\varepsilon > 0$, there exists $M > 0$ such that

$$\sup_{n\geq 0} \mathbb{P}\Big(N_\alpha(X^{(n)}) > M\Big) \leq \varepsilon.$$

Now from the inequality $\omega_\delta(X) \leq N_\alpha(X)\delta^\alpha$, we deduce that

$$\mathbb{P}\Big(\omega_\delta(X^{(n)}) > \eta\Big) \leq \mathbb{P}\Big(N_\alpha(X^{(n)}) > \eta\delta^{-\alpha}\Big),$$

hence for all $\varepsilon > 0$, there exists $\delta > 0$, such that

$$\sup_n \mathbb{P}\Big(\omega_\delta(X^{(n)}) > \eta\Big) \leq \varepsilon,$$

which implies, together with the first assumption of the proposition, the tightness of $(X^{(n)})_n$ by the "probabilistic" Ascoli criterion of Proposition 5.7. □

5.3 Donsker theorem

Theorem 5.10 (Donsker theorem)**.** *Let $(\xi_k)_{k\geq 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}(\xi_1) = 0$ and $\mathrm{Var}(\xi_1) = 1$. Define $(S_n)_{n\geq 0}$ by $S_0 = 0$, and $S_n = \xi_1 + \cdots + \xi_n$, for $n \geq 1$. Consider the continuous stochastic process $(X_t^{(n)})_{t\in[0,1]}$ defined by*

$$X_t^{(n)} = \frac{1}{\sqrt{n}}\Big(S_{\lfloor nt\rfloor} + (nt - \lfloor nt\rfloor)\xi_{\lfloor nt\rfloor+1}\Big),$$

for $t \in [0,1]$ and $n \geq 0$. Then, as $n \to \infty$, $X^{(n)}$ converges in distribution to a Brownian motion $(B_t)_{t\in[0,1]}$.

We will prove Donsker theorem under the additional assumption that $\mathbb{E}(\xi_1^4) < \infty$ in order to apply Kolmogorov criterion with $p = 4$ (it does not work for $p = 2$). We refer to Billingsley [1] for the general statement.

Proof of Donsker theorem. Recall that, for $t \in [0,1]$, the interpolated random walk is

$$X_t^{(n)} = \frac{1}{\sqrt{n}}\Big(S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\xi_{\lfloor nt \rfloor + 1}\Big),$$

that is if $nt \in [k, k+1)$, then

$$X_t^{(n)} = \frac{1}{\sqrt{n}}(S_k + (nt-k)\xi_{k+1}).$$

The proof that the finite-dimensional distributions converge has been seen in the introduction (for the symmetric random walk, but the proof is the same here). Hence, it remains to prove that $(X^{(n)})_n$ is tight. To that end, we are going to prove that $(X^{(n)})_n$ satisfies Kolmogorov criterion with $p = 4$ and $\beta = 1$, i.e. for all $s, t \in [0,1]$

$$\sup_{n \geq 1} \mathbb{E}\Big|X_t^{(n)} - X_s^{(n)}\Big|^4 \leq C|t-s|^2.$$

Note that the first condition in Kolmogorov criterion is trivially satisfied since $X_0^{(n)} = 0$ for all n. So assume without loss of generality that $s < t$. First assume that nt and ns are integers, say $nt = m$ and $nt = k$. Then,

$$\Big|X_t^{(n)} - X_s^{(n)}\Big| = \frac{1}{\sqrt{n}}|S_m - S_k| = \frac{1}{\sqrt{n}}\left|\sum_{j=k+1}^{m} \xi_j\right|.$$

But,

$$\mathbb{E}\left(\sum_{j=k+1}^{m} \xi_j\right)^4 = \sum_{a,b,c,d} \mathbb{E}(\xi_a \xi_b \xi_c \xi_d),$$

and due to the independence and the centering of the ξ_j's, the only contributing terms are those for which the four indices are equal or only two of them are different. Hence,

$$\begin{aligned}
\mathbb{E}\left(\sum_{j=k+1}^{m} \xi_j\right)^4 &= \sum_{a=k+1}^{m} \mathbb{E}(\xi_a^4) + \binom{4}{2} \sum_{\substack{a,b=k+1 \\ a \neq b}}^{m} \mathbb{E}(\xi_a^2)\mathbb{E}(\xi_b^2) \\
&= (m-k)\mathbb{E}(\xi_1^4) + 6(m-k)(m-k-1)\mathbb{E}(\xi_1^4),
\end{aligned}$$

since $(\mathbb{E}(x_1^2))^2 \leq \mathbb{E}(\xi_1^4)$ by Cauchy-Schwarz inequality. Hence, we obtain

$$\mathbb{E}\Big|X_t^{(n)} - X_s^{(n)}\Big|^4 \leq \frac{1}{n^2}(m-k)^2\mathbb{E}(\xi_1^4) = (t-s)^2\mathbb{E}(\xi_1^4).$$

Now if $k \leq ns < nt < k+1$, i.e. nt and ns are in the same interval, then

$$\begin{aligned}
\mathbb{E}\Big|X_t^{(n)} - X_s^{(n)}\Big|^4 &= \frac{1}{n^2}|nt - ns|^4 \mathbb{E}\xi_1^4 \\
&= n^2(t-s)^2(t-s)^2\mathbb{E}\xi_1^4 \\
&\leq C(t-s)^2,
\end{aligned}$$

for some constant $C > 0$, since in this case $(nt - ns) \leq 1$. If now ns and nt are in different interval, say $k \leq ns < k+1 \leq m \leq nt < m+1$, then, using $(a+b+c)^4 \leq 3^{3/4}(a^4+b^4+c^4)$,

$$\begin{aligned}
\mathbb{E}\Big|X_t^{(n)} - X_s^{(n)}\Big|^4 &\leq C\left(\mathbb{E}\Big|X_t^{(n)} - X_m^{(n)}\Big|^4 + \mathbb{E}\Big|X_m^{(n)} - X_{k+1}^{(n)}\Big|^4 + \mathbb{E}\Big|X_{k+1}^{(n)} - X_s^{(n)}\Big|^4\right) \\
&= C\left(\frac{1}{n^2}(nt-m)^4\mathbb{E}\xi_1^4 + \frac{1}{n^2}\mathbb{E}\left|\sum_{j=k+2}^{m} \xi_j\right|^4 + \frac{1}{n^2}(ns-(k+1))^4\mathbb{E}\xi_1^4\right) \\
&\leq C'(t-s)^2,
\end{aligned}$$

for some constant $C' > 0$, since $ns - (k+1) \leq 1$, $nt - m \leq 1$, and by the previous case (a sum over an empty set is 0 by convention). Finally, we have obtained that for all $s, t \in [0,1]$, for all n,

$$\mathbb{E}\left|X_t^{(n)} - X_s^{(n)}\right|^4 \leq C(t-s)^2.$$

Hence, by Kolmogorov criterion the sequence of random variables $(X^{(n)})_n$ is tight in $\mathscr{C}$. This concludes the proof of Donsker theorem. $\square$

6 Solutions to some exercises

6.1 Exercises of Section 2

Solution to Exercise 1. Let $\varepsilon > 0$. Since X is totally bounded, there exists $a_1, \dots, a_n \in X$ such that

$$X \subset \bigcup_{i=1}^{n} B(a_i, \varepsilon).$$

Let $I = \{i \mid B(a_i, \varepsilon) \cap Y \neq \varnothing\}$ (I is not empty since $Y \subset X$, except if Y is empty...). For all $i \in I$, let $y_i \in B(a_i, \varepsilon) \cap Y$. Then $B(a_i, \varepsilon) \subset B(y_i, 2\varepsilon)$ since if $x \in B(a_i, \varepsilon)$, then,

$$d(x, y_i) \leq d(x, a_i) + d(a_i, y_i) < 2\varepsilon.$$

Hence,

$$Y = X \cap Y \subset \bigcup_{i \in I} B(a_i, \varepsilon) \subset \bigcup_{i \in I} B(y_i, 2\varepsilon).$$

Solution to Exercise 3. Let $x = (x_n)_n$ a sequence in $[0,1]^{\mathbb{N}}$. For all n, $x_n \in [0,1]^{\mathbb{N}}$, i.e. x_n is a sequence of numbers in $[0,1]$:

$$x_n = (x_n^k)_k.$$

We want to prove that there exists a convergent subsequence: $\exists \phi \uparrow$ such that $(x_{\phi(n)})_n$ converges to some $x = (x^k)_k \in [0,1]^{\mathbb{N}}$ for the product topology, i.e. for all k,

$$x^k_{\phi(n)} \underset{n \to \infty}{\longrightarrow} x^k.$$

Since for all k, $(x_n^k)_n$ is a sequence in $[0,1]$ which is compact, there exists ϕ_k (which depends on k!) such that

$$x^k_{\phi_k(n)} \underset{n \to \infty}{\longrightarrow} x^k.$$

This is where we use a diagonal extraction argument. Start with $k = 1$. Since $(x_n^1)_n \subset [0,1]$, there exists ϕ_1 such that

$$x^1_{\phi_1(n)} \underset{n \to \infty}{\longrightarrow} x^1.$$

Now let $k = 2$. Now $(x^2_{\phi_1(n)})_n$ is a subsequence of $(x^2_n)_n \subset [0,1]$, hence we have that there exists ϕ_2 such that

$$x^2_{\phi_1 \circ \phi_2(n)} \underset{n \to \infty}{\longrightarrow} x^2.$$

But $(x^1_{\phi_1 \circ \phi_2(n)})_n$ is a subsequence of $(x^1_{\phi_1(n)})_n$, hence we also have that

$$x^1_{\phi_1 \circ \phi_2(n)} \underset{n \to \infty}{\longrightarrow} x^1.$$

Hence, we have that

$$\begin{cases} x^1_{\phi_1 \circ \phi_2(n)} \underset{n \to \infty}{\longrightarrow} x^1 \\ x^2_{\phi_1 \circ \phi_2(n)} \underset{n \to \infty}{\longrightarrow} x^2. \end{cases}$$

Now repeat the argument. There exists ϕ_k such that for all $j \leq k$,

$$x^j_{\phi_1 \circ \phi_2 \circ \cdots \circ \phi_k(n)} \underset{n \to \infty}{\longrightarrow} x^j.$$

Now define, for all n,

$$\phi(n) = \phi_1 \circ \phi_2 \circ \cdots \circ \phi_n(n).$$

Then ϕ is increasing, and as $n \to \infty$, for all j,

$$x^j_{\phi(n)} \longrightarrow x^j,$$

i.e. $x_{\phi(n)} = (x^j_{\phi(n)})_j$ converges as $n \to \infty$ to $x = (x^j)_j$, hence $[0,1]^{\mathbb{N}}$ is sequentially compact.

Solution to Exercise 4.

1. Let f be a continuous function. Let $(y_n)_n$ in $f(E)$. For all n, there exists $x_n \in E$ such that $y_n = f(x_n)$. Since E is compact, there exists a subsequence $(x_{\varphi(n)})_n$ of (x_n) does converge to some $x \in E$. By continuity of f, $y_{\varphi(n)} = f(x_{\varphi(n)})$ converges to $f(x)$. Hence, there exists a subsequence of $(y_n)_n$ that converges to some $y = f(x) \in f(E)$, that is $f(E)$ is compact and f is bounded.
2. Let f be a uniformly continuous function. Let $\varepsilon > 0$. Since f is uniformly continuous, there exists $\delta > 0$ such that for all $x, y \in E$ such that $d(x,y) < \delta$, $d(f(x), f(y)) < \varepsilon$. Moreover, since E is totally bounded,
$$E \subset \bigcup_{i=1}^{n} B(x_i, \delta),$$
for some $x_1, \ldots, x_n$ in E. Hence,
$$f(E) \subset \bigcup_{i=1}^{n} f(B(x_i, \delta)) \subset \bigcup_{i=1}^{n} B(f(x_i), \varepsilon),$$
since if $y \in f(B(x_i,\delta))$, there exists $x \in B(x_i, \delta)$ such that $y = f(x)$, hence $d(y, f(x_i)) = d(f(x), f(x_i)) < \varepsilon$ since $d(x, x_i) < \delta$. Hence, $f(E)$ is totally bounded, and in particular f is bounded.
3. Let f be a Lipschitz function, i.e. there exists $C > 0$ such that for all $x, y \in E$,
$$d(f(x), f(y)) \leq C d(x,y).$$
Since E is bounded, there exists $M > 0$, such that $\sup_{x,y \in E} d(x,y) \leq M$. Hence,
$$\sup_{x,y\in E} d(f(x), f(y)) \leq CM,$$
and f is bounded.

Solution to Exercise 6. Let A be the subset of $\mathcal{C}_b(\mathbb{R})$ consisting of functions that are 0 or 1 at the integers and affine in between. There is an uncountable number of such functions (indeed there are $2^{\mathbb{N}}$ such functions), and for any $f, g \in A$, if $f \neq g$, they will differ by 1 at some integer, hence

$$\|f - g\|_\infty \geq 1.$$

This implies that $\mathcal{C}_b(\mathbb{R})$ is not separable (in the same way as for the proof that $L^\infty([0,1])$ is not separable).

Solution to Exercise 7.

1. By the triangle inequality, for all $z \in A$, for all $x, y \in E$,
$$d(x, A) \leq d(x, z) \leq d(x, y) + d(y, z),$$
so, taking the infimum over $z \in A$,
$$d(x, A) \leq d(x, y) + d(y, A),$$
thus showing that $d(x,A) - d(y,A) \leq d(x,y)$. The roles of x and y being symmetric, we obtain
$$|d(x,A) - d(y,A)| \leq d(x,y),$$
hence $x \mapsto d(x, A)$ is a 1-Lipschitz function.

2. If $x \in \overline{A}$, then for all $\varepsilon > 0$, there exists $y \in A$ such that $d(x,y) < \varepsilon$, and therefore $d(x,A) \leq \varepsilon$ for all $\varepsilon > 0$, hence $d(x,A) = 0$.

   Conversely, if $d(x,A) = 0$, then, for all $\varepsilon > 0$, there exists $y \in A$ such that $d(x,y) < \varepsilon$, which means that $x \in \overline{A}$.

3. Since $A \subset \overline{A}$, we have $d(x,\overline{A}) \leq d(x,A)$, so it remains to prove the reversed inequality. Let $\varepsilon > 0$ and $y \in \overline{A}$. Then, there exists $z \in A$, such that $d(y,z) < \varepsilon$. Then, for any x,

$$d(x,A) \leq d(x,z) \leq d(x,y) + d(y,z) \leq d(x,y) + \varepsilon.$$

   Letting $\varepsilon \to 0$ and taking the infimum over $y \in \mathcal{A}$ gives that $d(x,A) \leq d(x,\overline{A})$.

6.2 Exercises of Section 3

Solution to Exercise 8. We first comment on the separability assumption. If X and Y are random variables in E, then (X,Y) is a r.v. in $(E^2, \mathcal{B}(E) \otimes \mathcal{B}(E))$. On the other hand, $d\colon E^2 \to \mathbb{R}$ is continuous, so it is measurable from $(E^2, \mathcal{B}(E^2))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Therefore, $d(X,Y)$ is a real r.v. as soon as $\mathcal{B}(E^2) \subset \mathcal{B}(E) \otimes \mathcal{B}(E)$. This is not always true for general metric spaces, but it is true if E is separable.

The proof of this inclusion makes use of the fact that an open set of a separable metric space can be written as a countable union of balls. The product topology on E^2 is generated by the metric $d_{E^2}((x_1,y_1),(x_2,y_2)) := d(x_1,x_2) \vee d(y_1,y_2)$, for which balls are products of balls in E so are in $\mathcal{B}(E) \otimes \mathcal{B}(E)$. If E is separable, then E^2 is separable, so any open set can be written as a countable union of balls and therefore is in $\mathcal{B}(E) \otimes \mathcal{B}(E)$. This implies $\mathcal{B}(E^2) \subset \mathcal{B}(E) \otimes \mathcal{B}(E)$.

Note that the reversed inclusion $\mathcal{B}(E) \otimes \mathcal{B}(E) \subset \mathcal{B}(E^2)$ is true for any topological space E.

1. Assume $X_n \xrightarrow[n\to\infty]{(\mathbb{P})} X$. Let $\varepsilon > 0$. Then, there exists n_0 such that for all $n \geq n_0$,

$$\mathbb{P}(d(X_n, X) \geq \varepsilon) \leq \varepsilon.$$

   We want to prove Criterion (v) of Portmanteau theorem. Let F be a closed set. Then,

$$\begin{aligned}\mathbb{P}(X_n \in F) &= \mathbb{P}(X_n \in F, d(X_n,X) < \varepsilon) + \mathbb{P}(X_n \in F, d(X_n,X) \geq \varepsilon)\\ &\leq \mathbb{P}(d(X,F) < \varepsilon) + \varepsilon,\end{aligned}$$

   for $n \geq n_0$. Taking the $\limsup$ as $n \to \infty$ and then letting $\varepsilon \downarrow 0$ gives that

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(d(X,F) = 0) = \mathbb{P}(X \in F),$$

   because F is closed so $d(x,F) = 0 \Leftrightarrow x \in F$. By Portmanteau theorem, $X_n \xrightarrow[n\to\infty]{(\mathrm{d})} X$.

2. Assume $X_n \xrightarrow[n\to\infty]{(\mathrm{d})} x$, where $x \in E$ is deterministic. Let $\varepsilon > 0$. By the Portmanteau theorem, using that the ball $B(x,\varepsilon)$ is open, we have

$$\liminf_{n\to\infty} \mathbb{P}(X_n \in B(x,\varepsilon)) \geq \delta_x(B(x,\varepsilon)) = 1,$$

   hence

$$\lim_{n\to\infty} \mathbb{P}(d(X_n,x) < \varepsilon) = 1,$$

   i.e. $X_n \xrightarrow[n\to\infty]{(\mathbb{P})} x$.

3. *Hint:* Follow the ideas of Question 1.

4. Assume $X_n \xrightarrow[n\to\infty]{(\mathrm{d})} X$ and $Y_n \xrightarrow[n\to\infty]{(\mathbb{P})} y$, where $y \in E$ is deterministic. Let $f \in \mathcal{C}_b(E^2)$. Then, setting $g\colon x \in E \mapsto f(x,y)$, one has $g \in \mathcal{C}_b(E)$, and so

$$\mathbb{E}[f(X_n,y)] = \mathbb{E}[g(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[g(X)] = \mathbb{E}[f(X,y)].$$

This proves $(X_n,y) \xrightarrow[n\to\infty]{(\mathrm{d})} (X,y)$. Now recall we equip E^2 with the metric $d_{E^2}((x_1,y_1),(x_2,y_2)) := d(x_1,x_2) \vee d(y_1,y_2)$. Then

$$d_{E^2}((X_n,Y_n),(X_n,y)) = d(Y_n,y) \xrightarrow[n\to\infty]{(\mathbb{P})} 0,$$

because $Y_n \xrightarrow[n\to\infty]{(\mathbb{P})} y$ by Question 2. Thus, applying Question 3 (but in E^2) we deduce that

$$(X_n,Y_n) \xrightarrow[n\to\infty]{(\mathrm{d})} (X,y).$$

Solution to Exercise 12.

1. By density of $\mathbb{Q}$ in $\mathbb{R}$, there exist $r_1,\dots,r_N \in \mathbb{Q}\cap(0,1]$ such that $|r_k - \mu(B_k)| < \varepsilon/N$ for each $k \in [\![1,N]\!]$. Then, we consider
$$q_k := \frac{r_k}{\sum_{k=1}^N r_k}.$$
Note that $\sum_{k=1}^N q_k = 1$. Then, we bound
$$\sum_{k=1}^N |q_k - \mu(B_k)| < \sum_{k=1}^N |q_k - r_k| + \sum_{k=1}^N |r_k - \mu(B_k)|.$$
The 2nd term is smaller than ε by choice of the r_k's. For the first term, note that all the $q_k - r_k$ have the same sign so that
$$\sum_{k=1}^N |q_k - r_k| = \left|\sum_{k=1}^N q_k - \sum_{k=1}^N r_k\right| = \left|1 - \sum_{k=1}^N r_k\right| \le \left|1 - \sum_{k=1}^N \mu(B_k)\right| + \sum_{k=1}^N |\mu(B_k) - r_k| < 2\varepsilon,$$
using that $\sum_{k=1}^N \mu(B_k) = \mu(\bigcup_{k=1}^N B_k) > 1-\varepsilon$. This proves that $\sum_{k=1}^N |q_k - \mu(B_k)| < 3\varepsilon$.
2. Note that π is indeed a probability measure because $q_k \ge 0$ and $\sum_{k=1}^N q_k = 1$. Let $A \in \mathcal{B}(E)$. As in the proof of Theorem 3.19, we have
$$\mu(A) = \mu(A\cap G(A)) + \mu(A \cap G(A)^c) \le \mu(A \cap G(A)) + \varepsilon.$$
Furthermore, note that
$$\mu(A\cap G(A)) \le \mu(G(A)) = \sum_{i\in I_A} \mu(B_i) < \sum_{i\in I_A} q_i + 3\varepsilon,$$
using that $\sum_{k=1}^N |q_k - \mu(B_k)| < 3\varepsilon$. Then, for each $i \in I_A$, there exists $x \in B_i \cap A$ and, since $B_i \subset B(x_i,\varepsilon)$, $d(x,x_i) < \varepsilon$, so that $x_i \in A^\varepsilon$. It follows that
$$\pi(A^\varepsilon) \ge \sum_{i\in I_A} q_i.$$
We thus obtain $\mu(A) \le \pi(A^\varepsilon) + 4\varepsilon$, and therefore $\rho(\mu,\pi) \le 4\varepsilon$.

3. We have just proved that the countable set

$$\left\{ \sum_{k=1}^{N} q_k \delta_{x_k} \,\middle|\, N \geq 1, q_1, \ldots, q_n \in \mathbb{Q} \cap [0,1], \sum_{k=1}^{N} q_k = 1 \right\}$$

is dense in $\mathcal{M}_1(E)$ for the Lévy–Prokhorov distance, so $(\mathcal{M}_1(E), \rho)$ is separable

Solution to Exercise 14. Let $\|f\|_{\mathrm{BL}} \leq 1$. Denote $d := d(x,y)$, $L := \|f\|_{\mathrm{Lip}}$ and $M := \|f\|_\infty$. Then,

$$\left| \int f \, \mathrm{d}\delta_x - \int f d\delta_y \right| = |f(x) - f(y)| \leq \min(Ld, 2M).$$

If $Ld \leq 2M$, then

$$(2+d)L \leq 2L + 2M \leq 2\|f\|_{\mathrm{BL}}.$$

If $2M \leq Ld$, then

$$(2+d)M \leq Ld + Md = d\|f\|_{\mathrm{BL}}.$$

Thus,

$$\min(Ld, 2M) \leq \frac{2d}{2+d}\|f\|_{\mathrm{BL}},$$

giving that

$$\beta(\delta_x, \delta_y) \leq \frac{2d(x,y)}{2+d(x,y)}.$$

Now let $f(z) = \frac{d(z,y)-d(z,x)}{2+d(x,y)}$. Since $x \mapsto d(\cdot, x)$ is 1-Lipschitz, we have

$$|f(z)| \leq \frac{d(x,y)}{2+d(x,y)}$$

and

$$|f(z_1) - f(z_2)| \leq \frac{1}{2+d(x,y)}(|d(z_1,y) - d(z_2,y)| + |d(z_1,x) - d(z_2,x)|) \leq \frac{1}{2+d(x,y)} 2d(z_1,z_2)$$

Thus, $\|f\|_\infty \leq \frac{d(x,y)}{2+d(x,y)}$ and $\|f\|_{\mathrm{Lip}} \leq \frac{2}{2+d(x,y)}$, so $\|f\|_{\mathrm{BL}} \leq 1$. And we have

$$\left| \int f \, \mathrm{d}\delta_x - \int f d\delta_y \right| = \left| \frac{d(x,y)}{2+d(x,y)} + \frac{d(x,y)}{2+d(x,y)} \right| = \frac{2d(x,y)}{2+d(x,y)}.$$

Hence, $\beta(\delta_x, \delta_y) = \frac{2d(x,y)}{2+d(x,y)}$.

6.3 Exercises of Section 4

Solution to Exercise 15. An easy way to see it is to use the Lévy–Prokhorov distance: for any $m \neq n$, $\rho(\delta_m, \delta_n) = |m-n| \wedge 1 = 1$, so the sequence $(\delta_n)_{n\geq 0}$ cannot have a convergent subsequence.