CONVERGENCE OF PROBABILITY MEASURES

FRANÇOIS CHAPON

Université de Toulouse

2025–2026

Contents

1. Introduction 1
2. Probability measures on metric spaces 3
2.1. A bit of topology 3
2.2. Regularity of probability measures on metric spaces 5
3. Convergence of probability measures and the Portmanteau theorem 9
3.1. Weak convergence 9
3.2. The Portmanteau theorem 10
4. Topology on $\mathcal{M}_1(E)$ 12
5. Tightness and Prokhorov theorem 18
6. Convergence of stochastic processes 23
6.1. Borel σ-algebra and cylindrical σ-algebra 24
6.2. Tightness in $\mathscr{C}$ 26
6.3. Donsker theorem 31
7. Appendix. 33
7.1. Compactness 33
7.2. Dynkin π–λ theorem 35
References 35

1. Introduction

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 $n \geq 1$,

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

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

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

where N is normal random variable with distribution $\mathcal{N}(0,1)$, and where $\xrightarrow[n\to\infty]{(d)}$ 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 $t \in [0,1]$, we can consider the interpolated rescaled random walk defined by

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

where $\lfloor \cdot \rfloor$ denotes the floor function, i.e. $\lfloor u \rfloor$ is the greater integer such that

$$\lfloor u \rfloor \leq u < \lfloor u \rfloor + 1.$$

Thus, for any $t \in [\frac{k}{n}, \frac{k+1}{n})$,

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

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

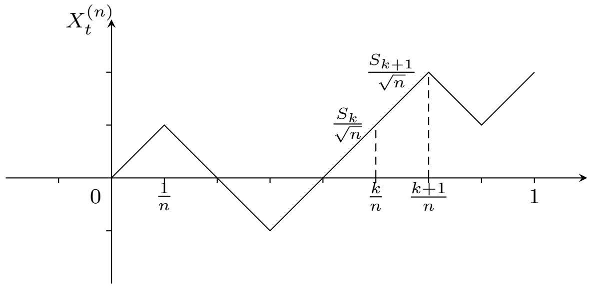


Figure 1. The interpolated random walk.

Now we can consider the random walk at different times. Let $t_0 = 0 < t_1 < \cdots < t_k \le 1$. The increments of the random walk being independent, we have

$$\left(X_{\lfloor t_1 \rfloor}^{(n)}, X_{\lfloor t_2 \rfloor}^{(n)} - X_{\lfloor t_1 \rfloor}^{(n)}, \ldots, X_{\lfloor t_k \rfloor}^{(n)} - X_{\lfloor t_{k-1} \rfloor}^{(n)} \right) \xrightarrow[n\to\infty]{(d)} (N_1, N_2, \ldots, N_k),$$

where the random variables $N_1, N_2, \ldots, N_k$ are independent and

$$N_j \sim \mathcal{N}(0, t_j - t_{j-1}),$$

for all j. By the continuous mapping theorem, we get that

$$\left(X_{\lfloor t_1 \rfloor}^{(n)}, X_{\lfloor t_2 \rfloor}^{(n)}, \ldots, X_{\lfloor t_k \rfloor}^{(n)} \right) \xrightarrow[n\to\infty]{(d)} (N_1, N_1 + N_2, \ldots, N_1 + N_2 + \cdots + N_k),$$

Note that for all j,

$$N_1 + \cdots + N_j \sim \mathcal{N}(0, t_j).$$

Now recall Slutsky lemma: if $(X_n)_n$ converges in distribution to some random variable X and $(Y_n)_n$ converges in probability to some *constant* y, then

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

in distribution. Since, for all j

$$X_{t_j}^{(n)} - X_{\lfloor t_j \rfloor}^{(n)} = \frac{nt_j - \lfloor nt_j \rfloor}{\sqrt{n}} \xi_{\lfloor nt_j \rfloor + 1} \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

from the previous convergence, Slutsky lemma, and the continuous mapping theorem, we also have

$$\left(X_{t_1}^{(n)}, X_{t_2}^{(n)}, \ldots, X_{t_k}^{(n)} \right) \xrightarrow[n\to\infty]{(d)} (N_1, N_1 + N_2, \ldots, N_1 + N_2 + \cdots + N_k).$$

Hence, for any fixed number of times, we have convergence of the *finite-dimensional distributions* of the sequence of random function $(X_t^{(n)}, t \in [0,1])$. But we would like to have the convergence in distribution of the whole random function $(X_t^{(n)}, t \in [0,1])$, which does not follow from the finite-dimensional distribution convergence. This is the content of Donsker theorem or Donsker invariance principle, that states that the rescaled random walk converges in distribution to the *Brownian motion*, see Fig 2:

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

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

- $B_0 = 0$,

- for all $k \geq 1$, for all $t_0 = 0 < t_1 < \cdots < t_k \leq 1$, the random variables
$$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}$$
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.

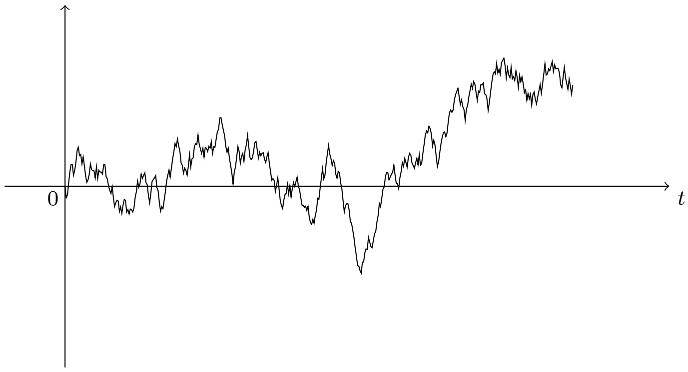


FIGURE 2. A Brownian trajectory.

The stochastic process $(B_t; t \in [0,1])$ is thus a random variable with values in $\mathscr{C}$, the space of continuous functions from $[0,1]$ to $\mathbb{R}$, whose distribution has some prescribed properties: for any k, and all $t_1 < \cdots < t_k$, the random vector $(B_{t_1}, \ldots, B_{t_k})$ 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 Donsker theorem is not obvious neither. The aim of the first part of this course is to prove Donsker theorem. To that end, we will need criteria to handle convergence in distribution for random variables with values in $\mathscr{C}$, i.e. random continuous functions. Since convergence in distribution of random variables depends only on the distribution of the random variables, we thus need to understand convergence of probability measures on $\mathscr{C}$, and more generally probability measures on a metric space. So we start by studying probability measures on metric spaces through their topological and measurable properties.

2. Probability measures on metric spaces

2.1. A bit of topology.

Definition 2.1. *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. Hence, any open set of E is the union of open balls of the form:*
$$B(x,r) = \{y \in E \mid d(x,y) < r\},$$
for $x \in E$ and $r > 0$.
- *We say that E is* **separable** *if it contains a countable dense subset.*
- *A metric space (E,d) is* **complete** *if any Cauchy sequence in E is converging in E.*

Definition 2.2. *A topological space E is said to be a* **Polish space** *if it is separable and metrizable and such that the metric d makes (E,d) complete.*

Remark 2.1. It is useful to known 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, the distance
$$d'(x,y) = |\arctan x - \arctan y|$$

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

Examples:

- $\mathbb{R}$ and more generally $\mathbb{R}^d$, equipped with the usual distance, are Polish spaces.
- A compact metric space K is a Polish space: for any $n \geq 1$,
$$K \subset \bigcup_{x\in K} B\left(x, \frac{1}{n}\right),$$
so by compactness, one can extract a finite subcover: there exists k_n and $x_1, \dots, x_{k_n}$ in K such that
$$K \subset \bigcup_{j=1}^{k_n} B\left(x_j, \frac{1}{n}\right).$$
The subset $\{x_j; 1 \leq j \leq k_n, n \geq 1\}$ is then countable and dense in K. Hence, K is separable. Moreover, K is complete: for any Cauchy sequence $(x_n)_n$, by compactness, there exists a subsequence $(x_{\varphi(n)})_n$ that converges in K. But a Cauchy sequence that admits an accumulation point converges, hence K is complete.
- The space
$$C([0,1]) = \{f : [0,1] \to \mathbb{R} \mid f \text{ continuous}\}$$
is a separable complete metric space. The metric on $C([0,1])$ is given by the uniform norm
$$d(f,g) = ||f-g||_\infty = \sup_{t\in[0,1]} |f(t) - g(t)|,$$
so the sequence $(f_n)_n$ converges to f if $(f_n)_n$ converges *uniformly* to f. By the Stone-Weierstrass theorem, the set of polynomials on $[0,1]$ is dense in $C([0,1])$, and thus the set of polynomials with rational coefficients, which is countable, is dense in $C([0,1])$, so $C([0,1])$ is separable. Moreover, let (f_n) be a Cauchy sequence in $C([0,1])$, 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 $(f_n(t))_n$ is a Cauchy sequence in $\mathbb{R}$ which is complete, hence, there exists $f(t) \in \mathbb{R}$ such that $f_n(t) \to f(t)$. Define f by $f : t \mapsto f(t)$. Letting m goes to infinity in the above majoration, one gets that for all $n \geq N$, for all $t \in [0,1]$,
$$|f_n(t) - f(t)| < \varepsilon,$$
i.e. $(f_n)_n$ converges uniformly to f. It is well known that the uniform limit of a sequence of continuous functions is continuous: by continuity of f_N, there exists $\delta > 0$ such that for all $|t-s| < \delta$, $|f_N(t) - f_N(s)| < \varepsilon$. Hence,
$$\begin{aligned}|f(t) - f(s)| &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| \\ &\leq 3\varepsilon.\end{aligned}$$
Thus f is continuous and $C([0,1])$ is complete.
- Note that the above is still true when $[0,1]$ is replace by a compact metric space K, i.e. $C(K)$, 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:

Theorem 2.1 (Stone-Weierstrass theorem)**.** *Let K be a compact metric space. Let $\mathcal{A}$ be a subalgebra of $C(K)$ such that $\mathcal{A}$ contains a non-zero constant function. Then $\mathcal{A}$ is dense in $C(K)$ if and only if it separates points: for any $x,y \in X$, there exists a function $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.*

(see [2]).

Let $\{x_n\}_n$ 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 $C(K)$ 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 $C(K)$.

- Note that the space $C(\mathbb{R}_+)$, 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 $C(\mathbb{R}_+)$ 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 $C(\mathbb{R}_+)$ is a Polish space.
- Any separable Hilbert space is a Polish space, so for instance $L^2(\mathbb{R})$ is a Polish space.
- Any separable Banach space is a Polish space, so for instance $L^p(\mathbb{R})$, for $p \in [1, +\infty)$ is a Polish space.
- A finite product of Polish spaces is Polish (exercise).
- As a counterexample, one has that $L^\infty(\mathbb{R})$, the space of essentially bounded functions on $\mathbb{R}$, is not separable: 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. $||f||_\infty = \inf\{a > 0 \mid \lambda(|f| > a) = 0\}$. Thus, the open balls $B(f_t, \frac{1}{2})$, for $t \in [0,1]$, are pairwise distincts. Now suppose that $L^\infty(\mathbb{R})$ is separable, and let $A \subset L^\infty(\mathbb{R})$ 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
$$C_b(\mathbb{R}) = \{f\colon \mathbb{R} \to \mathbb{R} \mid f \text{ continuous and bounded}\}$$
equipped with the uniform norm $||\cdot||_\infty$ is a complete normed space, but is not separable.

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

2.2. **Regularity of probability measures on metric spaces.** Let E be a topological space. We will always, unless specified, equipped 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 $(E, \mathcal{B}(E))$.

Exercise 2. Let (E,d) be a metric space. Let $A \subset E$, and define for all $x \in E$,
$$d(x,A) = \inf\{d(x,y) \mid y \in A\}.$$
Show that $x \mapsto d(x,A)$ is 1-Lipschitz on E. Show also, that we have the equivalence
$$x \in \overline{A} \;\Leftrightarrow\; d(x,A) = 0.$$
We will need the following useful lemma. We use the standard notations:
$$a \wedge b = \min\{a,b\} \;;\quad a \vee b = \max\{a,b\}.$$

Lemma 2.1. Let (E,d) be a metric space.

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

(2) Let $G \subset E$ be an open set. Define, for all n,

$$g_n(x) = nd(x, G^c) \wedge 1, \quad x \in E.$$

Then, for all n, 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. We have that f_n is bounded by 1. If $x \in F$, then $d(x, F) = 0$ since F is closed, hence $f_n(x) = 1$ for all n. If $x \notin F$, then $d(x, F) > 0$, 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 $f_n(x) = 0$. Hence, $f_n \searrow \mathbb{1}_F$, as $n \to \infty$ ($(f_n)_n$ is obviously nonincreasing). Moreover, since $x \mapsto \max(x, 0)$ is 1-Lispschitz and $x \mapsto nd(x, F)$ is n-Lispschitz, one obtains that f_n is n-Lipschitz for all n. The proof for $(g_n)_n$ is similar and left as an exercise. □

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 2.1. *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,*

$$\int_E f d\mu = \int_E f d\nu,$$

then $\mu = \nu$.

Proof. Let G be an open set. By lemma 2.1, there exists a nondecreasing sequence $(g_n)_n$ of bounded continuous functions such that $g_n \nearrow \mathbb{1}_G$ as $n \to \infty$. By assumption, we have

$$\int_E g_n d\mu = \int_E g_n d\nu,$$

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

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

Since probability measures on the Borel σ-algebra $\mathcal{B}(E)$ are characterized by their values on open sets (by Dynkin π–λ theorem 7.3, since the class of open sets is a π-system), we get $\mu = \nu$. □

Proposition 2.2. *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 2.2. 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 $\varepsilon > 0$ and let A be an open set. We have 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 $U = A$. For F, define

$$F_n = \left\{ x \in A \mid d(x, A^c) \geq \frac{1}{n} \right\}.$$

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

$$A = \bigcup_n F_n,$$

and thus,

$$\lim_n \mu(F_n) = \mu\left(\cup_n F_n\right) = \mu(A).$$

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 $\varepsilon > 0$. Let, for each n, $A_n \in \mathcal{A}$. 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

$$\cup_n F_n \subset \cup_n A_n \subset \cup_n U_n$$

and

$$(\cup_n U_n) \setminus (\cup_n F_n) \subset \cup_n (U_n \setminus F_n)$$

(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,

$$\mu\left((\cup_n U_n) \setminus (\cup_n F_n)\right) \leq \mu\left(\cup_n (U_n \setminus F_n)\right) \leq \sum_n \mu(U_n \setminus F_n) \leq \frac{\varepsilon}{2},$$

where we used sub-σ-additivity.

Set $U' = \bigcup_n U_n$, 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_ε such that

$$\mu\left(\cup_n F_n\right) \leq \mu\left(\cup_{k \leq n_\varepsilon} F_k\right) + \frac{\varepsilon}{2}.$$

Now set $F' = \cup_{k \leq n_\varepsilon} F_k$. Then $F' \subset A \subset U'$, and

$$\mu(U' \setminus F') = \mu(U') - \mu(F') = \mu(U') - \mu(\cup_n F_n) + \mu(\cup_n F_n) - \mu(F') \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

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

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 2.3. *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*

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

It says that on a Polish space, almost all the mass of a probability measure is concentrated on a compact set. Before proving this fact, a reminder about compactness may be useful, see Appendix 7.1.

Proof of Proposition 2.3. 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,

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

Since,

$$\bigcup_N \bigcup_{n=1}^N B\left(x_n, \frac{1}{k}\right) = \bigcup_n B\left(x_n, \frac{1}{k}\right),$$

we have

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

so there exists N_k 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

$$K = \bigcap_k \bigcup_{n=1}^{N_k} B\left(x_n, \frac{1}{k}\right).$$

By construction K is totally bounded, hence $\overline{K}$ is a compact set since E is complete. Then,

$$\begin{aligned}
\mu\left(\overline{K}^c\right) \leq \mu(K^c) &= \mu\left(\bigcup_k \bigcap_{n=1}^{N_k} B_f\left(x_n, \frac{1}{k}\right)^c\right) \\
&\leq \sum_k \mu\left(\bigcap_{n=1}^{N_k} B_f\left(x_n, \frac{1}{k}\right)^c\right) \\
&= \sum_k \mu\left(\left[\bigcup_{n=1}^{N_k} B_f\left(x_n, \frac{1}{k}\right)\right]^c\right) \\
&\leq \sum_k \frac{\varepsilon}{2^k} = \varepsilon,
\end{aligned}$$

which concludes the proof. □

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

Proposition 2.4. *Let E be a Polish space. Then every $\mu \in \mathcal{M}_1(E)$ is inner regular, i.e. for all $A \in \mathcal{B}(E)$,*

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

Proof. By exterior regularity, we have that

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

hence for all $A \in \mathcal{B}(E)$ and 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

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

□

3. Convergence of probability measures and the Portmanteau theorem

3.1. Weak convergence.

Definition 3.1. *Let E be a metric space. Let $(\mu_n)_n \subset \mathcal{M}_1(E)$ and $\mu \in \mathcal{M}_1(E)$. We say that $(\mu_n)_n$ converges weakly to μ, if for all bounded and continuous function $f \in C_b(E)$, we have*

$$\int_E f d\mu_n \underset{n\to\infty}{\longrightarrow} \int_E f d\mu.$$

We denote this convergence by $\mu_n \overset{weakly}{\underset{n\to\infty}{\longrightarrow}} \mu$, or $\mu \underset{n\to\infty}{\rightharpoonup} \mu$ or $\mu \underset{n\to\infty}{\Rightarrow} \mu$

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:

$$X_n \overset{\text{(d)}}{\underset{n\to\infty}{\longrightarrow}} X \quad \Leftrightarrow \quad \mu_n \overset{\text{weakly}}{\underset{n\to\infty}{\longrightarrow}} \mu.$$

Note that since integrals $\int_E f d\mu$ for all bounded continuous function f characterize the measure μ by Proposition 2.1, the limit is unique.
For instance, the sequence of Dirac masses on $\mathbb{R}$ $(\delta_{\frac{1}{n}})_n$ converges weakly to δ_0. But note that the sequence $(\delta_n)_n$ does not converge weakly to a probability measure. Indeed, if for instance f is a continuous function which goes to 0 at infinity, one has

$$\int f d\delta_n = f(n) \to 0.$$

The problem here is that the mass of the sequence $(\delta_n)_n$ "escapes" at infinity. An important part of this course will be to give a criterion in order to avoid this phenomenon. This is the notion of "tightness" for sequences that will be introduced later.
Note that one says that $(\mu)_n$ converges *vaguely* to a bounded measure (but not necessarily a probability measure) if for all continuous function vanishing at infinity f, one has

$$\int f d\mu_n \underset{n\to\infty}{\longrightarrow} \int f d\mu.$$

In the above example, one has that $(\delta_n)_n$ converges vaguely to the zero measure.

Example 3.1. Let (E, d) be a metric space. Let $(x_n)_n \subset E$ and $x \in E$. Then $(\delta_{x_n})_n$ converges weakly to δ_x if and only if $(x_n)_n$ converges to x.
Indeed, assume that $x_n \to x$. Then, for any continuous function f on E, we have

$$\int_E f d\delta_{x_n} = f(x_n) \underset{n\to\infty}{\longrightarrow} f(x) = \int_E f d\delta_x,$$

by continuity of f. Hence, $(\delta_{x_n})_n$ 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, $f(x_n)$ does not converge to $f(x)$.

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

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

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

$$\int_F f dg_\#\mu = \int_E f \circ g d\mu,$$

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 3.1 (Continuous mapping theorem)**.** *Let E and F be metric spaces. Let $(\mu_n)_n \subset \mathcal{M}_1(E)$ converging weakly to $\mu \in \mathcal{M}_1(E)$. Let $g\colon E \to F$ be a continuous function. Then the sequence of pushforward measures $(g_\#\mu_n)_n$ converges weakly to $g_\#\mu$.*

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

$$\int_F f dg_\#\mu_n = \int_E f \circ g d\mu_n \underset{n\to\infty}{\longrightarrow} \int_E f \circ g d\mu = \int_F f dg_\#\mu,$$

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

3.2. **The Portmanteau theorem.** We now give a criterion for weak convergence. This is known as the Portmanteau theorem (even if there is no mathematician named Portmanteau...).

Theorem 3.2 (Portmanteau theorem)**.** *Let (E, d) be a metric space and $(\mu_n)_n$ and μ be probability measures on E. The following assertions are equivalent:*

(1) The sequence $(\mu_n)_n$ converges weakly to μ.
(2) For all bounded and uniformly continuous function f, $\int f d\mu_n \underset{n\to\infty}{\longrightarrow} \int f d\mu$.
(3) For all bounded and Lipschitz function f, $\int f d\mu_n \underset{n\to\infty}{\longrightarrow} \int f d\mu$.
(4) For all open set $G \subset E$, $\mu(G) \leq \liminf_n \mu_n(G)$.
(5) For all closed set $F \subset E$, $\mu(F) \geq \limsup_n \mu_n(F)$.
(6) For all Borel set $A \in \mathcal{B}(E)$ of μ-continuity, i.e. such that $\mu(\partial A) = 0$, $\mu_n(A) \underset{n\to\infty}{\longrightarrow} \mu(A)$.
(7) For all bounded measurable function f which is continuous μ-almost everywhere, $\int f d\mu_n \underset{n\to\infty}{\longrightarrow} \int f d\mu$.

Proof. Since

$$\{\text{Lipschitz functions}\} \subset \{\text{uniformly continuous functions}\} \subset \{\text{continuous functions}\},$$

we have (1) $\Rightarrow$ (2) $\Rightarrow$ (3).
Obviously, (4) $\Leftrightarrow$ (5) using complementation.
Let us prove that (4) + (5) $\Rightarrow$ (6). Let $A \in \mathcal{B}(E)$ such that $\mu(\partial A) = 0$. Since $\mathring{A} \subset A \subset \overline{A}$, we have

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

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

$$\limsup_n \mu_n(A) \leq \limsup_n \mu_n(\overline{A}) \overset{(5)}{\leq} \mu(\overline{A}) = \mu(A) = \mu(\mathring{A}) \overset{(4)}{\leq} \liminf_n \mu_n(\mathring{A}) \leq \liminf_n \mu_n(A),$$

hence we get $\liminf_n \mu_n(A) = \limsup_n \mu_n(A) = \mu(A)$, i.e. $\mu_n(A) \underset{n\to\infty}{\longrightarrow} \mu(A)$.
Now we prove that (6) $\Rightarrow$ (7). 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 $f \geq 0$ (it not, just write $f = f_+ - f_-$). Then for any $\nu \in \mathcal{M}_1(E)$, using Fubini's theorem, we have

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

Hence, for all n,

$$\int_E f d\mu_n = \int_0^1 \mu_n(f \geq t)dt.$$

Denote by D the set of discontinuities of f. Now remark that for all $t \in \mathbb{R}$,

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

Indeed, let $x \in \partial\{f \geq t\} = \overline{\{f \geq t\}} \setminus \mathrm{Int}\{f \geq t\}$. 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 neighbourhood of x, hence $x \in \mathrm{Int}\{f > t\} \subset \mathrm{Int}\{f \geq t\}$, which contradicts the fact that $x \in \partial\{f \geq t\}$. Hence, $x \in \{f = t\} \cup D$.
Now remark that the set $\{t \in \mathbb{R} \,|\, \mu(f = t) > 0\}$ has Lebesgue measure equal to zero. Indeed, $\{t \in \mathbb{R} \,|\, \mu(f = t) > 0\}$ is the set of the atoms of the probability measure μ_f equal to the image measure of μ by f, and as such is at most countable (exercise). Hence, $\mu(\{f = t\}) = 0$ holds for Lebesgue-almost all $t \in \mathbb{R}$. Hence,

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

which is equal to 0 for Lebesgue almost all t. So, we get that:

- $\mu_n(f \geq t) \underset{n\to\infty}{\longrightarrow} \mu(f \geq t)$ Lebesgue almost everywhere, by assumption (6),
- $\mu_n(f \geq t) \leq 1$, which is integrable on $[0, 1]$ with respect to Lebesgue measure,

so by dominated convergence theorem,

$$\int_E f d\mu_n = \int_0^1 \mu_n(f \geq t)dt \underset{n\to\infty}{\longrightarrow} \int_0^1 \mu(f \geq t)dt = \int_E f d\mu.$$

This proves (7).
Moreover, (7) $\Rightarrow$ (1) is trivial.
It remains to prove (3) $\Rightarrow$ (4). Let G be an open set. From lemma 2.1, there exists a nondecreasing sequence $(g_k)_k$ such that g_k is bounded and k-Lispschitz for all k, such that

$$g_k \nearrow \mathbb{1}_G, \quad \text{as } k \to \infty.$$

Hence,

$$\mu_n(G) = \int \mathbb{1}_G d\mu_n \geq \int g_k d\mu_n \underset{n\to\infty}{\longrightarrow} \int g_k d\mu.$$

Using monotone convergence theorem, one obtains,

$$\liminf_n \mu_n(G) \geq \int g_k d\mu \underset{k\to\infty}{\longrightarrow} \int \mathbb{1}_G d\mu = \mu(G),$$

which concludes the proof. □

Example 3.2. Recall that if X is a real random variable, its cumulative distribution function is the function

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

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 $t \in \mathbb{R}$. It is well known that convergence in distribution is characterized by the following: for all $t \in \mathbb{R}$ such that F_X is continuous at t, we have

$$F_{X_n}(t) \underset{n\to\infty}{\longrightarrow} F_X(t).$$

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, the Portmanteau theorem implies that if $X_n \to X$ in distribution, i.e. $\mu_n \to \mu$ weakly, then, $F_{X_n}(t) \underset{n\to\infty}{\longrightarrow} F_X(t)$, for all t continuity point of F_X.

Conversely, let $\varepsilon > 0$ and consider first the open interval (a, b). Let D be the set of discontinuity points of F_X. Since D is countable (exercise: prove it), 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,

$$\begin{aligned}\liminf_n \mathbb{P}\left(X_n \in (a,b)\right) &\geq \liminf_n \mathbb{P}\left(X_n \in (a_k, b_k]\right) \\ &\geq \liminf_n \left(F_{X_n}(b_k) - F_{X_n}(a_k)\right) \\ &= F_X(b_k) - F_X(a_k) \\ &= \mathbb{P}\left(X \in (a_k, b_k]\right)\end{aligned}$$

by assumption. Letting $k \to \infty$, we get

$$\liminf_n \mathbb{P}\left(X_n \in (a,b)\right) \geq \mathbb{P}\left(X \in (a,b)\right).$$

Now let U be an open set. Since any open set is the countable union of disjoint open intervals, we write

$$U = \bigcup_i (a_i, b_i),$$

for some a_i and b_i. Now, choose 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

$$\liminf_n \mathbb{P}(X_n \in U) \geq \liminf_n \mathbb{P}(X_n \in V_N) = \mathbb{P}(X \in V_N) > \mathbb{P}(U) - \varepsilon.$$

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

Exercise 3. Show that the continuous mapping theorem is still true if one only assume that g is continuous μ-almost everywhere. *Hint: Let D be the set of discontinuities of g. Show that if A is a closed set in F, one has $\overline{g^{-1}(A)} \cap D^c \subset g^{-1}(A)$ and use the Portmanteau theorem, item (5).*

Exercise 4. Let (E, d) be a separable metric space. We say that the sequence of random variables $(X_n)_n$ converges in probability to the random variable X if for all $\varepsilon > 0$,

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

We denote this convergence by $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$. (The assumption that E is separable is to ensure that $d(X, Y)$ is measurable for two random variables X and Y).
Show that if $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ then $X_n \xrightarrow[n\to\infty]{(d)} X$.
Show that if $X_n \xrightarrow[n\to\infty]{(d)} x$, where $x \in E$ is deterministic, then $X_n \xrightarrow[n\to\infty]{\mathbb{P}} x$.

4. Topology on $\mathcal{M}_1(E)$

Let Y be a real normed space. Denote by Y^* the topological dual of Y, i.e. the space of *continuous* (i.e. bounded) real linear forms on Y. The topology on Y^* induced by the norm

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

is often too finer to be useful, for instance (in infinite dimension) there are very few sets that are compact for the norm topology. Another topology, which is coarser (i.e. smaller), is given by the *weak* topology*: it is the weakest topology on Y^* making all evaluation maps $\mathrm{ev}_x\colon Y^* \to \mathbb{R}$,

$\phi \mapsto \phi(x)$, continuous, as x ranges over Y. A basis of neighbourhood of $\phi \in Y^*$ for the weak* topology is given by the sets

$$\mathcal{V}(\phi, x_1, \ldots, x_k, \varepsilon) = \bigcap_{i=1}^{k} \left\{\psi \in Y^* \mid \left|\phi(x_i) - \psi(x_i)\right| < \varepsilon\right\},$$

indexed by $x_1, \ldots, x_k \in Y$, and $\varepsilon > 0$.
In particular, it can be proved that a sequence $(\phi_n)_n$ in Y^* converges to $\phi \in Y^*$ for the weak* topology, if and only if

$$\phi_n(x) \underset{n\to\infty}{\longrightarrow} \phi(x),$$

for all $x \in Y$, i.e. if ϕ_n converges pointwise to ϕ.

In our case, we consider the normed space $Y = C_b(E)$, where E is a metric space. By Proposition 2.1, the map from $\mathcal{M}_1(E)$ to the topological dual of $C_b(E)$ defined by

$$\begin{aligned} \mathcal{M}_1(E) &\to C_b(E)^* \\ \mu &\mapsto \left(f \mapsto \int_E f d\mu\right) \end{aligned}$$

is injective, so $\mathcal{M}_1(E)$ can be identified with a subset of $C_b(E)^*$. The weak topology on $\mathcal{M}_1(E)$ is defined as the restriction of the weak* topology of $C_b(E)^*$ to $\mathcal{M}_1(E)$ and the notion of weak convergence for sequences of probability measures corresponds to the notion of sequencial convergence of this topology.

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**-neighbourhood** of a subset $A \subset E$ by

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

It is useful to note that if A is a closed set, then

$$A = \bigcap_{\varepsilon>0} A^\varepsilon.$$

Definition 4.1. *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*

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

Note that since $B^\varepsilon = \overline{B}^\varepsilon$ (since $d(x, B) = d(x, \overline{B})$), one can replace in the definition of the Prokhorov distance $B \in \mathcal{B}(E)$ by $B \subset E$ closed. Note also that obviously, for all $B \in \mathcal{B}(E)$,

$$\mu(B) \leq \nu(B^1) + 1,$$

hence $\rho(\mu, \nu) \in [0, 1]$.
To manipulate this distance, it is useful to note that:

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

Before proving that ρ is indeed a distance, let us compute an example.

Example 4.1. Let us compute the Lévy-Prokhorov distance between two Dirac masses δ_x and δ_y, with $x \neq y$. By definition,

$$\rho(\delta_x, \delta_y) = \inf \{\varepsilon > 0 \mid \forall B \in \mathcal{B}(E), \delta_x(B) \leq \delta_y(B^\varepsilon) + \varepsilon\}.$$

Let $\varepsilon > d(x, y)$. Let B be a Borel set. If $x \notin B$, then obviously

$$0 = \delta_x(B) \leq \delta_y(B) + \varepsilon.$$

On the other hand, if $x \in B$, and since $y \in B^\varepsilon \Leftrightarrow d(x, y) < \varepsilon$, we also have

$$1 = \delta_x(B) \leq \delta_y(B^\varepsilon) + \varepsilon = 1 + \varepsilon.$$

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

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

hence, we obtain $\rho(\delta_x, \delta_y) \geq d(x, y) \wedge 1$. Finally,

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

Proposition 4.1. *The Prokhorov distance ρ is a metric on $\mathcal{M}_1(E)$.*

Proof. Let $\varepsilon > 0$. Then for all $B \in \mathcal{B}(E)$, since $B \subset B^\varepsilon$,

$$\mu(B) \leq \mu(B^\varepsilon) + \varepsilon.$$

Hence, for all $\varepsilon > 0$, $\rho(\mu, \mu) \leq \varepsilon$, and so $\rho(\mu, \mu) = 0$.
Let us show that ρ is symmetric (which is not obvious from the definition!). If $\rho(\mu, \nu) > \varepsilon$, there exists a Borel set B such that

$$\mu(B) > \nu(B^\varepsilon) + \varepsilon.$$

Now one remarks that

$$\left((B^\varepsilon)^c\right)^\varepsilon \subset B^c.$$

(exercise: prove it).
Thus,

$$\mu\left(\left((B^\varepsilon)^c\right)^\varepsilon\right) + \varepsilon \leq \mu\left(B^c\right) + \varepsilon < \nu\left((B^\varepsilon)^c\right),$$

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 $\rho(\mu, \nu) = 0$ implies that $\mu = \nu$. Suppose that for all $\varepsilon > 0$,

$$\rho(\mu, \nu) < \varepsilon.$$

Then for all closed set B, $\mu(B) \leq \nu(B^\varepsilon) + \varepsilon$. Take $\varepsilon = \frac{1}{k}$. One has that

$$B = \bigcap_k B^{\frac{1}{k}},$$

since $d(x, B) = 0 \Leftrightarrow x \in B$, since B is closed. Since $(B^{\frac{1}{k}})_k$ is decreasing in k, one gets that

$$\mu(B) \leq \lim_k \nu\left(B^{\frac{1}{k}}\right) = \nu(B).$$

By symmetry, we get $\mu(B) = \nu(B)$. Since probability measures on a metric space are characterized by their values on closed sets, we get $\mu = \nu$.
It remains to prove the triangle inequality. Let μ, ν, γ in $\mathcal{M}_1(E)$. Let $\varepsilon > 0$ and $\delta > 0$ such that

$$\rho(\mu, \nu) < \varepsilon \quad \text{and} \quad \rho(\nu, \gamma) < \delta.$$

Then for all Borel set B,

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

Now remark that $(B^\varepsilon)^\delta \subset B^{\varepsilon+\delta}$. Hence,

$$\nu(B^\varepsilon) \leq \gamma\left((B^\varepsilon)^\delta\right) + \delta \leq \gamma(B^{\varepsilon+\delta}) + \delta,$$

and thus

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

Hence, $\rho(\mu,\gamma) \leq \varepsilon+\delta$, and letting $\varepsilon \downarrow \rho(\mu,\nu)$ and $\delta \downarrow \rho(\nu,\gamma)$ gives the triangle inequality. $\square$

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

Theorem 4.1. *Let (E,d) be a separable metric space. Let $(\mu_n)_n$ and μ be probability measures on E. Then $\mu_n \overset{weakly}{\underset{n\to\infty}{\longrightarrow}} \mu$ if and only if $\rho(\mu_n,\mu) \underset{n\to\infty}{\longrightarrow} 0$.*

Proof. Suppose that $\rho(\mu_n,\mu) \underset{n\to\infty}{\longrightarrow} 0$. Let $\varepsilon>0$. Hence, for n large enough, for all closed set F,

$$\mu_n(F) \leq \mu(F^\varepsilon) + \varepsilon.$$

Now taking $\varepsilon = \frac{1}{k}$,

$$\limsup_n \mu_n(F) \leq \mu\left(F^{\frac{1}{k}}\right) + \frac{1}{k},$$

and noticing that $F = \cap_k F^{\frac{1}{k}}$ and that the intersection is decreasing, we get, letting $k\to\infty$, that

$$\limsup_n \mu_n(F) \leq \mu(F),$$

which proves that μ converges weakly to μ by the Portmanteau theorem.
Conversely, assume that $\mu_n \overset{\text{weakly}}{\underset{n\to\infty}{\longrightarrow}} \mu$. Let $\varepsilon>0$. Since E is separable, let $(x_n)_n$ be a dense sequence in E such that

$$E = \bigcup_n B(x_n,\varepsilon).$$

Define $B_1 = B(x_1,\varepsilon)$, and for $n\geq 1$, $B_{n+1} = B(x_{n+1},\varepsilon)\setminus(B_1\cup\cdots\cup B_n)$. The B_n's are pairwise disjoints (if there are empty sets among the B_n's, just reindex them), and

$$E = \bigcup_n B_n.$$

Now let N be large enough, such that

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

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

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

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

$$\liminf_n \mu_n(G) \geq \mu(G),$$

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, $I_A = \{k\in\{1,\dots,N\} \mid B_k\cap A\neq\varnothing\}$, and let

$$G(A) = \bigcup_{k\in I_A} B_k.$$

Then $G(A)^\varepsilon$ belongs to $\mathcal{G}$, and moreover $G(A)^\varepsilon \subset A^{3\varepsilon}$. Hence, for $n \geq n_0$,

$$\begin{aligned}\mu(A) &= \mu\left(A \cap G(A)\right) + \mu\left(A \cap G(A)^c\right)\\ &\leq \mu\left(G(A)^\varepsilon\right) + \mu\left(\bigcup_{k>N+1} B_k\right)\\ &\leq \mu_n\left(G(A)^\varepsilon\right) + 2\varepsilon\\ &\leq \mu_n\left(A^{3\varepsilon}\right) + 3\varepsilon.\end{aligned}$$

Hence, one obtains that for $n \geq n_0$, $\rho(\mu_n, \mu) \leq 3\varepsilon$, which concludes the proof. □

Proposition 4.2. *Let E be a separable metric space. Then $\mathcal{M}_1(E)$ is separable.*

Proof. Let $\mu \in \mathcal{M}_1(E)$. We use the same notations than in the previous proof. Let $(x_n)_n$ be a dense sequence in E such that

$$E = \bigcup_n B(x_n, \varepsilon).$$

and define $B_1 = B(x_1, \varepsilon)$, and for $n \geq 1$, $B_{n+1} = B(x_{n+1}, \varepsilon) \setminus (B_1 \cup \dots \cup B_n)$ so that for N large enough, we have

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

Let A be a Borel set and denote as before $G(A) = \bigcup_{k\in I_A} B_k$, where $I_A = \{k \in \{1, \dots, N\} \mid B_k \cap A \neq \varnothing\}$. Then,

$$\begin{aligned}\mu(A) &\leq \mu\left(G(A)\right) + \mu\left(A \setminus G(A)\right)\\ &\leq \sum_{k\in I_A} \mu(B_k) + \varepsilon.\end{aligned}$$

By density of $\mathbb{Q}$, choose $q_1, \dots, q_N \in \mathbb{Q} \cap [0, 1]$ such that

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

(First choose $r_1, \dots, r_N \in \mathbb{Q} \cap [0, 1]$ such that $\sum_{k=1}^N |r_k - \mu(B_k)| < \varepsilon$ and put $q_k = \frac{r_k}{\sum_{k=1}^N r_k}$ and use the fact that $\sum_{k=1}^N \mu(B_k) > 1 - \varepsilon$). Hence, denoting $\pi = \sum_{k=1}^N q_k\delta_{x_k}$, one obtains that

$$\mu(A) \leq \sum_{k\in I_A} q_k + 2\varepsilon = \pi\left(G(A)\right) + 2\varepsilon \leq \pi\left(A^{3\varepsilon}\right) + 3\varepsilon.$$

Hence, we get $\rho(\mu, \pi) \leq 3\varepsilon$, and finally the countable set

$$\left\{\sum_{k=1}^n q_k\delta_{x_k} \mid n \geq 1, q_1, \dots, q_n \in \mathbb{Q} \cap [0, 1], \sum_{k=1}^n q_k = 1\right\}$$

is dense in $\mathcal{M}_1(E)$. □

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

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

and

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

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

Definition 4.2. *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*

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

It is easy to verify that β is a metric on $\mathcal{M}_1(E)$, so the details are left as an exercise.

Both the Lévy-Prokhorov distance and the Kantorovich-Rubinstein distance induce the topology of weak convergence as a consequence of the following lemma:

Lemma 4.1. For all $\mu, \nu \in \mathcal{M}_1(E)$,

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

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

$$f_\varepsilon(x) = \max\left\{0, 1 - \frac{1}{\varepsilon}d(x,F)\right\},$$

as in lemma 2.1. Then

$$\mathbb{1}_B \leq f_\varepsilon \leq \mathbb{1}_{B^\varepsilon},$$

and $||f_\varepsilon||_{\text{BL}} \leq 1 + \frac{1}{\varepsilon}$, thus,

$$\begin{aligned}\mu(B) \leq \int f_\varepsilon d\mu &\leq \int f_\varepsilon d\nu + \left(1 + \frac{1}{\varepsilon}\right)\beta(\mu,\nu)\\ &\leq \nu(B^\varepsilon) + \left(1 + \frac{1}{\varepsilon}\right)\beta(\mu,\nu).\end{aligned}$$

Thus,

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

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

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

for $\beta(\mu,\nu) \leq \frac{1}{4}$ by a simple analysis of functions. If $\beta(\mu,\nu) > \frac{1}{4}$, the inequality is trivial since $\rho(\mu,\nu)$ is bounded by 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 such that $||f||_{\text{BL}} \leq 1$. Hence, $||f||_\infty \leq 1$ and $||f||_{\text{Lip}} \leq 1$. Then, using Fubini's theorem,

$$\begin{aligned}\left|\int f d\mu - \int f d\nu\right| &= \left|\int_{-1}^{1} \mu\left(\{f > t\}\right) dt - \int_{-1}^{1} \nu\left(\{f > t\}\right) dt\right|\\ &\leq \int_{-1}^{1} \left|\mu\left(\{f > t\}\right) - \nu\left(\{f > t\}\right)\right| dt.\end{aligned}$$

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

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

Hence,

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

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

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

Using Fubini's theorem again, one obtains

$$\int_{-1}^{1} \mu\left(\{t - \varepsilon < f \leq t\}\right) dt = \int_E \int_{-1}^{1} \mathbb{1}_{\{f \leq t < f + \varepsilon\}} dt d\mu \leq \varepsilon.$$

The same holds for the measure ν, so eventually, we get

$$\begin{aligned}\left|\int_E f d\mu - \int_E f d\nu\right| &\le \int_{-1}^{1} \mu\left(\{t-\varepsilon < f \le t\}\right) dt + \int_{-1}^{1} \nu\left(\{t-\varepsilon < f \le t\}\right) dt + 2\varepsilon \\ &\le \int_E \int_{-1}^{1} \mathbb{1}_{\{f\le t< f+\varepsilon\}} dt d\mu + \int_E \int_{-1}^{1} \mathbb{1}_{\{f\le t< f+\varepsilon\}} dt d\nu + \int_{-1}^{1} \varepsilon dt \\ &\le 4\varepsilon.\end{aligned}$$

Letting $\varepsilon \downarrow \rho(\mu,\nu)$ gives $\beta(\mu,\nu) \le 4\rho(\mu,\nu)$. □

Since we have seen that the Lévy-Prokhorov distance is a metric for weak convergence, we then have:

Proposition 4.3. *Let E be a separable metric space and let $(\mu_n)_n$ and μ be probability measures on E. Then, the following assertions are equivalent:*

(i) the sequence $(\mu_n)_n$ converges weakly to μ.
(ii) $\beta(\mu_n,\mu) \underset{n\to\infty}{\longrightarrow} 0$.
(iii) $\rho(\mu_n,\mu) \underset{n\to\infty}{\longrightarrow} 0$.

5. Tightness and Prokhorov theorem

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

Theorem 5.1 (Riesz representation theorem)**.** *Let (X,d) be a compact metric space. Let $\Lambda\colon C(X) \to \mathbb{R}$ be a linear form such that:*

(i) Λ is positive: for all $f \ge 0$, $\Lambda(f) \ge 0$,
(ii) $\Lambda(1) = 1$.

Then, there exists on X a unique Borel probability measure μ such that for all $f \in C(X)$,

$$\Lambda(f) = \int_X f d\mu.$$

Hence, the Riesz representation theorem asserts that the topological dual of $C(X)$ 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 locally compact (X is Hausdorff and any point of X admits a compact neighbourhood) and separable and with $C(X)$ replaced by $C_c(X)$, the space of continuous functions with compact support, or $C_0(X)$, the space of continuous functions vanishing at infinity. Of course, if X is compact, then $C_c(X) = C_0(X) = C(X)$. 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 $C_b(X)$ consists of objects more general that measures (in fact bounded finitely additive measures).

Theorem 5.2. *Let K be a compact space. Then, $\mathcal{M}_1(K)$ is compact.*

Proof. Recall that the space of continuous functions $C(K)$ is separable. Let $\mathcal{A} = \{f_k\}_k$ be a dense subset of $C(K)$. Since K is compact, every f_k is bounded, hence for all n,

$$\left|\int_K f_k d\mu_n\right| \le ||f_k||_\infty < \infty.$$

Thus, since $[-||f_k||_\infty, ||f_k||_\infty]$ is compact, there exists an extraction φ_k such that $\int f_k d\mu_{\varphi_k(n)}$ converges as $n\to\infty$. Using a diagonal extraction argument, one can construct an extraction φ such that for all k,

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

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

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

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

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

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

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

As such, Λ is a bounded linear form on $C(K)$, i.e. Λ belongs to the topological dual of $C(K)$. Moreover, one has $\Lambda(1) = 1$ and Λ is clearly a positive linear form, 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

$$\Lambda(f) = \int_K f d\mu,$$

for all $f \in C(K)$. Hence, for all $f \in C(K)$,

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

i.e. $(\mu_{\varphi(n)})_n$ converges weakly to μ, so $\mathcal{M}_1(K)$ is (sequentially) compact. □

Remark 5.1. In fact, we have proved the Banach-Alaoglu theorem in our particular case of $C(K)$, which states that the closed (for the norm) unit ball of the dual of a normed vector space Y is compact for the weak* topology. Recall that unless a normed vector space is finite dimensional, its closed unit ball is not compact for the norm topology!
But note that we cannot expect the above result to be true when the space K is not compact anymore (recall the counterexample of $(\delta_n)_n \subset \mathcal{M}_1(\mathbb{R})$). Indeed, the space $\mathcal{M}_1(E)$ is in general not closed for the weak* topology on $C_b(K)^*$.
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.

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

Definition 5.1. *Let (E, d) 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*

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

We have already seen that a single measure is tight on a Polish space. As an exercise, prove that a *finite* family $\mathcal{A} = \{\mu_1, \ldots, \mu_N\}$ of probability measures on a Polish space is tight.
One of the most important result in this course is then the following theorem.

Theorem 5.3 (Prokhorov theorem)**.** *Let (E, d) be a metric space and let $\mathcal{A} \subset \mathcal{M}_1(E)$. If $\mathcal{A}$ is tight then $\mathcal{A}$ is relatively compact. Moreover, if (E, d) is Polish, then if $\mathcal{A}$ is relatively compact, then $\mathcal{A}$ is tight.*

Proof. ($\Leftarrow$) We first prove that if $\mathcal{A}$ is relatively compact then it is tight when (E,d) is a Polish space. Let $(\mu_n)_n \subset \mathcal{A}$ converging weakly to some μ. Let $(G_n)_n$ be an increasing sequence of open sets such that $E = \bigcup_n G_n$. Then, for all $\varepsilon > 0$, there exist n such that for all $\mu \in \mathcal{A}$,

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

Indeed, if this is not the case, there exists $\varepsilon > 0$, such that for all n, there exists $\mu_n \in \mathcal{A}$, such that

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

Since $\mathcal{A}$ is relatively compact, there exists a subsequence $(\mu_{\varphi(n)})_n$ weakly converging to some probability measure μ. Hence, by the Portmanteau theorem, for all n

$$\mu(G_n) \leq \liminf_N \mu_{\varphi(N)}(G_n) \leq \liminf_N \mu_{\varphi(N)}(G_{\varphi(N)}) \leq 1 - \varepsilon,$$

which contradicts the fact that $E = \bigcup_n G_n$.
Now let $(x_n)_n$ be a dense sequence in E. Then, for all $k \geq 1$,

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

and let $G_n = \bigcup_{j=1}^n B(x_j, \frac{1}{k})$, so E is the increasing limit of the G_n's. By the previous claim, there exists n_k such that for all $\mu \in \mathcal{A}$,

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

Set

$$K = \bigcap_{k \geq 1} \bigcup_{j=1}^{n_k} B_f(x_j, \frac{1}{k}).$$

Then, we have, for all $\mu \in \mathcal{A}$,

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

and by construction K is totally bounded, hence since E is complete, $\overline{K}$ is compact, and $\mathcal{A}$ is tight.
($\Rightarrow$) Now we prove that if $\mathcal{A}$ is tight then it is relatively compact. So let $(\mu_n)_n \subset \mathcal{A}$. We have to prove that there exists a subsequence of $(\mu_n)_n$ that converges weakly. By tightness, for all $j \geq 1$, there exists a compact $K_j \subset E$ such that, for all n,

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

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 increasing. Now for all n and all j, define ν_n^j by

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

so ν_n^j is a probability measure on the compact set K_j. Hence, we have for all j, $(\nu_n^j)_n \subset \mathcal{M}_1(K_j)$ which is compact since K_j is compact, hence there exists a subsequence that converges weakly to some probability measure $\nu^j \in \mathcal{M}_1(K_j)$. By diagonal extraction, there exists an increasing φ such that, for all j,

$$\nu_{\varphi(n)}^j \underset{n \to \infty}{\rightharpoonup} \nu^j.$$

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

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

where $a_j \geq 1 - \frac{1}{j}$. Set $\phi = \varphi \circ \psi$. Hence, we have that for all j,

$$\begin{cases} \nu^j_{\phi(n)} \underset{n\to\infty}{\rightharpoonup} \nu^j \\ \mu_{\phi(n)}(K_j) \underset{n\to\infty}{\longrightarrow} a_j \in \left[1 - \frac{1}{j}, 1\right]. \end{cases}$$

Let now $B \in \mathcal{B}(E)$. Then, for all j

$$\begin{aligned} a_{j+1}\nu^{j+1}(B^{2\varepsilon}) &\geq a_{j+1}\nu^{j+1}(\overline{B^\varepsilon}) \\ &\text{(Portmanteau)} \\ &\geq \limsup_n \left(\mu_{\phi(n)}(K_{j+1})\nu^{j+1}_{\phi(n)}(\overline{B^\varepsilon})\right) \\ &= \limsup_n \mu_{\phi(n)}(\overline{B^\varepsilon} \cap K_{j+1}) \\ &\geq \limsup_n \mu_{\phi(n)}(\overline{B^\varepsilon} \cap K_j) \\ &= \limsup_n \left(\mu_{\phi(n)}(K_j)\nu^j_{\phi(n)}(\overline{B^\varepsilon})\right) \\ &\geq a_j \liminf_n \nu^j_{\phi(n)}(B^\varepsilon) \\ &\text{(Portmanteau)} \\ &\geq a_j\nu^j(B^\varepsilon). \end{aligned}$$

(We have used that if a_n and b_n are nonnegative and $a_n \to a$, then $\limsup_n(a_nb_n) = a\limsup_n(b_n)$). Now using that $\overline{B} = \bigcap_{\varepsilon>0} B^\varepsilon$, we get that for all closed set F,

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

and by exterior regularity of probability measures on a metric space, the fact that

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

for all Borel set B. We thus have a nondecreasing sequence of measures $(a_j\nu^j)_j$, 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). 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 and we have $\mu = \lim_{j\to\infty} \nu^j$. It remains to prove that $(\mu_{\phi(n)})_n$ converges weakly to μ. So let f be a continuous and bounded function on E. We have,

$$\begin{aligned} \left|\int_E f d\mu_{\phi(n)} - \int_E f d\mu\right| &\leq \left|\int_{K_j^c} f d\mu_{\phi(n)}\right| + \left|\int_{K_j} f d\mu_{\phi(n)} - \int_{K_j} f a_j d\nu^j\right| + \left|\int_{K_j} f a_j d\nu^j - \int_E f d\mu\right| \\ &\leq \frac{||f||_\infty}{j} + \left|\int_{K_j} f d\mu_{\phi(n)} - \int_{K_j} f a_j d\nu^j\right| + \left|\int_{K_j} f a_j d\nu^j - \int_E f d\mu\right| \\ &\leq \frac{||f||_\infty}{j} + \left|\int_{K_j} f \mu_{\phi(n)}(K_j) d\nu^j_{\phi(n)} - \int_{K_j} f a_j d\nu^j_{\phi(n)}\right| + \left|\int_{K_j} f a_j d\nu^j_{\phi(n)} - \int_{K_j} f a_j d\nu^j\right| \\ &\qquad + \left|\int_{K_j} f a_j d\nu^j - \int_E f d\mu\right| \\ &\leq \frac{||f||_\infty}{j} + \left|\mu_{\phi(n)}(K_j) - a_j\right| + \left|\int_{K_j} f a_j d\nu^j_{\phi(n)} - \int_{K_j} f a_j d\nu^j\right| \\ &\qquad + \left|\int_{K_j} f a_j d\nu^j - \int_E f d\mu\right|. \end{aligned}$$

Since $\mu_{\phi(n)}(K_j)$ converges to a_j as $n \to \infty$ and since $(\nu^j_{\phi(n)})_n$ converges weakly to ν^j, we get that

$$\limsup_n \left| \int_E f d\mu_{\phi(n)} - \int_E f d\mu \right| \leq \frac{||f||_\infty}{j} + \left| \int_{K_j} f a_j d\nu^j - \int_E f d\mu \right|.$$

But

$$\begin{aligned} \left| \int_{K_j} f a_j d\nu^j - \int_E f d\mu \right| &\leq \left| \int_{K_j} f a_j d\nu^j - \int_E f d\nu^j \right| + \left| \int_E f d\nu^j - \int_E f d\mu \right| \\ &\leq ||f||_\infty |a_j - 1| + + \left| \int_E f d\nu^j - \int_E f d\mu \right|, \end{aligned}$$

and since μ is the limit of ν^j, letting $j \to \infty$, we finally get

$$\limsup_n \left| \int_E f d\mu_{\phi(n)} - \int_E f d\mu \right| = 0.$$

Thus, $(\mu_{\phi(n)})_n$ converges weakly to μ, and $\mathcal{A}$ is relatively compact. □

Remark 5.2. The most interesting fact for us is that tightness implies relative compactness. Note that this holds under the assumption that (E, d) is only a metric space.

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 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:

Lemma 5.1. 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 5.1. 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:

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

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 5.4 (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

$$\mu_n \underset{n\to\infty}{\rightharpoonup} \mu.$$

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

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

Indeed, using Fubini theorem, we have

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

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

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

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

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

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

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

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

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

and thus

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

As such, the sequence $(\mu_n)_n$ is tight, so by Prokhorov 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. □

6. Convergence of stochastic processes

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

Definition 6.1. *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.

6.1. **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 $Cyl(\mathscr{C})$ is the smallest σ-algebra such that all the evaluation maps $\pi_t, t \in [0,1]$, are measurable, i.e.

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

Then we have the following.

Proposition 6.1. *We have,*

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

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 $Cyl(\mathscr{C})$, one has

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

Now we prove that any open ball of $\mathscr{C}$ lies in $Cyl(\mathscr{C})$. 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 $Cyl(\mathscr{C})$. But,

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

As such, $B_f(x,r) \in Cyl(\mathscr{C})$. 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 $Cyl(\mathscr{C})$, and finally $\mathcal{B}(\mathscr{C}) \subset Cyl(\mathscr{C})$. □

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 $Cyl(\mathscr{C})$. 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 $t_1 < \cdots < t_k$,

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

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 $k \geq 1$, and all $0 \leq t_1 < \cdots < t_k \leq 1$,

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

(as random vectors in $\mathbb{R}^k$!).
In the sequel, we will denote for all $k \geq 1$ and all $t_1, \ldots, t_k$ by

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

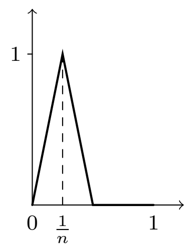
the finite-dimensional marginals of $\mu \in \mathcal{M}_1(\mathscr{C})$.
Note that by the continuous mapping theorem, if the sequence $(\mu_n)_n$ of probability measures on $\mathscr{C}$ converges weakly to some $\mu \in \mathcal{M}_1(\mathscr{C})$, then all the finite-dimensional marginals of μ_n converges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

$$\mu_n^{t_1,\ldots,t_k} \underset{n\to\infty}{\rightharpoonup} \mu^{t_1,\ldots,t_k}.$$

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

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

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



butions 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 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 $t_1, \dots, t_k$,

$$\mu_{n_i}^{t_1,\dots,t_k} \rightharpoonup \nu^{t_1,\dots,t_k}.$$

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

$$\mu_n^{t_1,\dots,t_k} \underset{n\to\infty}{\rightharpoonup} \mu^{t_1,\dots,t_k},$$

then,

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

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.

6.2. **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 6.2. *Let K be a compact metric space. A subset $\mathcal{A} \subset C(K)$ 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 $\mathcal{A} \subset C(K)$ 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 6.1 (Arzelà-Ascoli theorem)**.** *Let K be a compact metric space. Then $\mathcal{A} \subset C(K)$ 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 $\pi_x \colon f \mapsto f(x)$ 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 $g_1, \dots, g_r$ 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 $k \in \{1, \dots, r\}$ such that $d(f, g_k) < \varepsilon$. Since K is compact, g_k is uniformly continuous for all k, and since the family $g_1, \dots, g_r$ is finite, there exists $\delta > 0$, such that for all $k \in \{1, \dots, r\}$, $\omega_\delta(g_k) < \varepsilon$. 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 $j \in \{1, \ldots, J\}$ such that $d(x, q_j) < \delta$. Then, we have

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

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 $n \geq n_0$,

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

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

Remark 6.1. 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 $k = 1, \ldots, n$, $x_k = \frac{k}{n}x$, 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 5. 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 6.2. *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 $\mu \in \mathcal{A}$,

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

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

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

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 $\mu \in \mathcal{A}$,

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

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 $\mathscr{C}$. □

In more probabilistic terms, the previous proposition translates as:

Proposition 6.3. *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

$$\limsup_n \mathbb{P}\left(\omega_\delta(X^{(n)}) > \eta\right) \leq \varepsilon.$$

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 $f \in C([0,1])$, 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 6.4 (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 $s,t \in [0,1]$,*

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

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 $k \geq 0$,

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

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 $a_0, \ldots, a_l$ 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 $[0,1]$.

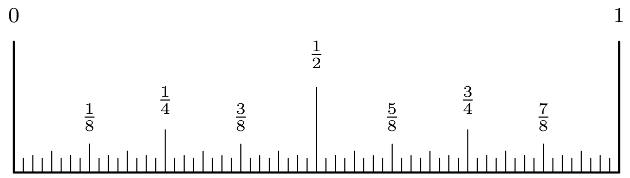


FIGURE 3. The dyadic rationals in $[0,1]$.

Lemma 6.1. 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

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

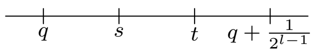
If moreover, $x \in \mathscr{C}$, then $N_\alpha(x) \leq \frac{2K}{1-2^{-\alpha}}$.

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:

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

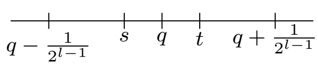
where a_k and b_k are in $\{0, 1\}$ and are equal to zero for k large enough. Let for $n \geq l$,

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

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 $N_\alpha(x) \leq \frac{2K}{1-2^{-\alpha}}$. □

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,

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

using Markov inequality. By assumption, we get

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

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

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

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

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

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

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

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

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

which implies, together with the first assumption of the proposition, the tightness of $(X^{(n)})_n$ by the "probabilistic" Ascoli criterion of Proposition 6.3. □

6.3. **Donsker theorem.**

Theorem 6.2 (Donsker theorem)**.** *Let $(\xi_k)_{k\geq 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}(\xi_1)=0$ and $\mathbb{V}\mathrm{ar}(\xi_1)=1$. 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*

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

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

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

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

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

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 $s, t \in [0, 1]$

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

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,

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

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

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

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

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

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 $(a+b+c)^4 \leq 3^{3/4}(a^4+b^4+c^4)$,

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

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. □

7. Appendix.

7.1. **Compactness.** To be clear, we recall here the (French!) definition of a compact space. Usually, in the Anglo-Saxon literature, the Hausdorff condition (i.e. distinct points have disjoint neighbourhoods) is omitted.

Definition 7.1. *A topological space X is called compact if it is Hausdorff and if every open cover of X has a finite subcover.*

We now introduce the notion of *total boundedness*, which is a generalization of compactness.

Definition 7.2. *A metric space (E, d) is* **totally bounded** *if and only if for all $\varepsilon > 0$, there exists a cover of E by balls of radius (at most) ε whose centers lie in E.*

For instance, 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 a space X equipped with the discrete metric $d(x, y) = 1$ if $x \neq y$ and 0 otherwise. Then it is bounded as $\sup_{x,y \in X} d(x, y) = 1$, 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 X.

Exercise 6. Let (E, d) be a metric space and let $X \subset E$ a totally bounded subset. Show that any subset Y of X is totally bounded.

Proposition 7.1. *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. □

Theorem 7.1. *Let (E, d) be a metric space. The following assertions are equivalent:*

- *(i) E is compact.*
- *(ii) E is sequentially compact: any sequence in E has a convergent subsequence converging to some point in E.*
- *(iii) E is complete and totally bounded.*

Proof. The equivalence between (i) and (ii) is the well known Bolzano-Weierstrass theorem.
(ii)$\Rightarrow$(iii): Let $(x_n)_n$ be a Cauchy sequence. By (ii), $(x_n)_n$ admits a converging subsequence, hence the whole sequence converges. Thus E is complete. Suppose by contradiction that E does not admit a finite subcover by balls of radius ε. Let $x_1 \in E$. Then $B(x_1, \varepsilon)$ does not cover E, hence there exists $x_2 \in E$ such that $d(x_2, x_1) \geq \varepsilon$. Suppose that we have constructed $x_1, \ldots, x_{n-1}$. Since $\bigcup_{i=1}^{n-1} B(x_i, \varepsilon)$ does not cover E there exists x_n such that $d(x_n, x_k) \geq \varepsilon$, for

all $k \leq n-1$. Hence, we obtain a sequence $(x_n)_{n\geq 1}$ such that any $k \neq l$, $d(x_k, x_l) \geq \varepsilon$, and thus cannot admit a converging subsequence. This contradicts (ii) and E is totally bounded.
(iii)⇒(ii): Let $(x_n)_n$ in E. By assumption, for all $k \geq 1$, there exists a finite cover of E by balls of radius $\frac{1}{k}$. Hence, for any k, there exists an infinite number of elements of $(x_n)_n$ that lie in the same ball B_k of radius $\frac{1}{k}$. Hence, for any k, there exists a subsequence $(x_{\varphi_k(n)})_n$ such that for all n, $x_{\varphi_k(n)} \in B_k$. By a diagonal extraction argument, there exists a subsequence $(x_{\varphi(n)})_n$ such that $d(x_{\varphi(m)}, x_{\varphi(n)}) \leq \frac{2}{m}$, for all $n \geq m$. We recall here how this works: for any k, there exists a subsequence $(x_{\varphi_k(n)})_n$ such that for all n, $x_{\varphi_k(n)} \in B_k$, hence for all m, n,

$$d(x_{\varphi_k(m)}, x_{\varphi_k(n)}) \leq \frac{2}{k}.$$

For $k = 1$, we get that there exists $(x_{\varphi_1(n)})_n$ such that for all m, n,

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

For $k = 2$, there exists an infinite number of elements of the sequence $(x_{\varphi_1(n)})_n$ that lie in the ball B_2. Hence, there exists $(x_{\varphi_1\circ\varphi_2(n)})_n$ such that for all m, n,

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

Repeating the argument, we obtain that there exists $(x_{\varphi_1\circ\cdots\circ\varphi_k(n)})_n$ such that for all m, n,

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

Now define φ by

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

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

$$d(x_{\varphi(m)}, x_{\varphi(n)}) \leq \frac{2}{m},$$

since $\varphi_1 \circ \cdots \circ \varphi_n$ is extracted from $\varphi_1 \circ \cdots \circ \varphi_m$ as $n \geq m$. Hence, $(x_{\varphi(n)})_n$ is a Cauchy sequence, and since E is complete it converges. This proves (ii). □

As an exercise, prove Tychonoff theorem using diagonal extraction:

Theorem 7.2 (Tychonoff theorem)**.**

$$[0, 1]^{\mathbb{N}}$$

is sequentially compact (for the product topology).

Exercise 7. Let (E, d) be a metric space. Show that

(i) If E is compact, then any continuous function is bounded.
(ii) If E is totally bounded, then any uniformly continuous function is bounded.
(iii) If E is bounded, then any Lipschitz function is bounded.

Definition 7.3. *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 $(x_n)_n$ in A, there exists a convergent subsequence $(x_{\varphi(n)})_n$, but the limit may not lie in A.

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

Proposition 7.2. *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 it is totally bounded (no completude needed here). Now, if A is totally bounded, then $\overline{A}$ is also totally bounded, and since it is closed in the complete metric space E, it is also complete. Hence, $\overline{A}$ is complete and totally bounded, hence compact. □

7.2. **Dynkin π–λ theorem.** Let E be a set. Recall that a λ-system Λ is a class of subsets of E that contains E, that is closed under countable increasing unions and under taking complement of subsets in supersets:

(i) $E \in \Lambda$,
(ii) if $(A_n)_n$ is a increasing countable collection of sets of Λ, i.e. $A_n \subset A_{n+1}$ for all n, then $\bigcup_n A_n \in \Lambda$,
(iii) if $A \subset B$ are two sets of Λ, $B \setminus A \in \Lambda$.

Then Dynkin π–λ theorem asserts that:

Theorem 7.3 (Dynkin π–λ theorem)**.** *If $\mathcal{A}$ is a π-system, i.e. is closed under finite intersections, then the smallest λ-system containing $\mathcal{A}$ is equal to the σ-algebra generated by $\mathcal{A}$.*

It has the following important consequence: if $\mathcal{A}$ is a π-system and if two probability measures on $\sigma(\mathcal{A})$ are equal on $\mathcal{A}$, they are equal on $\sigma(\mathcal{A})$. Indeed,

$$\Lambda = \{A \in \sigma(\mathcal{A}) \mid \mu(A) = \nu(A)\}$$

is a λ-system containing $\mathcal{A}$, hence by Dynkin π–λ theorem, it contains $\sigma(\mathcal{A})$.

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.