Probability measures on metric spaces

Onno van Gaans

These are some loose notes supporting the first sessions of the seminar *Stochastic Evolution Equations* organized by Dr. Jan van Neerven at the Delft University of Technology during Winter 2002/2003. They contain less information than the common textbooks on the topic of the title. Their purpose is to present a brief selection of the theory that provides a basis for later study of stochastic evolution equations in Banach spaces. The notes aim at an audience that feels more at ease in analysis than in probability theory. The main focus is on Prokhorov's theorem, which serves both as an important tool for future use and as an illustration of techniques that play a role in the theory.

The field of measures on topological spaces has the luxury of several excellent textbooks. The main source that has been used to prepare these notes is the book by Parthasarathy [6]. A clear exposition is also available in one of Bourbaki's volumes [2] and in [9, Section 3.2]. The theory on the Prokhorov metric is taken from Billingsley [1]. The additional references for standard facts on general measure theory and general topology have been Halmos [4] and Kelley [5].

Contents

The distribution of a random variable in a Banach space X will be a probability measure on X. When we study limit properties of stochastic processes we will be faced with convergence of probability measures on X. For certain aspects of the theory the linear structure of X is irrelevant and the theory of probability measures on metric spaces supplies some powerful tools. In view of the Banach space setting that we have in mind, it is not too restrictive to assume separability and completeness but we should avoid assuming compactness of the metric space.

1   Borel sets

Let (X, d) be a metric space. The *Borel σ-algebra* (σ-field) $\mathcal{B} = \mathcal{B}(X)$ is the smallest σ-algebra in X that contains all open subsets of X. The elements of $\mathcal{B}$ are called the *Borel sets* of X.

The metric space (X, d) is called *separable* if it has a countable dense subset, that is, there are $x_1, x_2, \ldots$ in X such that $\overline{\{x_1, x_2, \ldots\}} = X$. ($\overline{A}$ denotes the closure of $A \subset X$.)

Lemma 1.1. *If X is a separable metric space, then $\mathcal{B}(X)$ equals the σ-algebra generated by the open (or closed) balls of X.*

Proof. Denote

$$\mathcal{A} := \sigma\text{-algebra generated by the open (or closed) balls of } X.$$

Clearly, $\mathcal{A} \subset \mathcal{B}$.

Let D be a countable dense set in X. Let $U \subset X$ be open. For $x \in U$ take $r > 0$, $r \in \mathbb{Q}$ such that $B(x, r) \subset U$ ($B(x, r)$ open or closed ball with center x and radius r) and take $y_x \in D \cap B(x, r/3)$. Then $x \in B(y_x, r/2) \subset B(x, r)$. Set $r_x := r/2$. Then

$$U = \bigcup \{B(y_x, r_x) : x \in U\},$$

which is a countable union. Therefore $U \in \mathcal{A}$. Hence $\mathcal{B} \subset \mathcal{A}$. $\square$

Lemma 1.2. *Let (X, d) be a separable metric space. Let $\mathcal{C} \subset \mathcal{B}$ be countable. If $\mathcal{C}$ separates closed balls from points in the sense that for every closed ball B and every $x \in X \setminus B$ there exists $C \in \mathcal{C}$ such that $B \subset C$ and $x \notin C$, then the σ-algebra generated by $\mathcal{C}$ is the Borel σ-algebra.*

Proof. Clearly $\sigma(\mathcal{C}) \subset \mathcal{B}$, where $\sigma(\mathcal{C})$ denotes the σ-algebra generated by $\mathcal{C}$. Let B be a closed ball in X. Then $B = \bigcap \{C \in \mathcal{C} : B \subset C\}$, which is a countable intersection and hence a member of $\sigma(\mathcal{C})$. By the previous lemma we obtain $\mathcal{B} \subset \sigma(\mathcal{C})$. $\square$

If $f : S \to T$ and $\mathcal{A}_S$ and $\mathcal{A}_T$ are σ-algebras in S and T, respectively, then f is called *measurable* (w.r.t. $\mathcal{A}_S$ and $\mathcal{A}_T$) if

$$f^{-1}(A) = \{x \in S : f(x) \in A\} \in \mathcal{A}_S \text{ for all } A \in \mathcal{A}_T.$$

Proposition 1.3. *Let (X, d) be a metric space. $\mathcal{B}(X)$ is the smallest σ-algebra with respect to which all (real valued) continuous functions on X are measurable (w.r.t. $\mathcal{B}(X)$ and $\mathcal{B}(\mathbb{R})$).*
(See [6, Theorem I.1.7, p. 4].)

2 Borel probability measures

Let (X, d) be a metric space. A *finite Borel measure* on X is a map $\mu : \mathcal{B}(X) \to [0, \infty)$ such that

$$\mu(\emptyset) = 0, \text{ and}$$
$$A_1, A_2, \ldots \in \mathcal{B} \text{ mutually disjoint} \implies \mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i).$$

μ is called a *Borel probabiliy measure* if in addition $\mu(X) = 1$.

The following well known continuity properties will be used many times.

Lemma 2.1. *Let X be a metric space and μ a finite Borel measure on X. Let $A_1, A_2, \ldots$ be Borel sets.*

(1) If $A_1 \subset A_2 \subset \cdots$ and $A = \bigcup_{i=1}^{\infty} A_i$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

(2) If $A_1 \supset A_2 \supset \cdots$ and $A = \bigcap_{i=1}^{\infty}$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

The next observation is important in the proof of Theorems 3.2 and 4.2.

Lemma 2.2. *If μ is a finite Borel measure on X and $\mathcal{A}$ is a collection of mutually disjoint Borel sets of X, then at most countably many elements of $\mathcal{A}$ have nonzero μ-measure.*

Proof. For $m \geq 1$, let $\mathcal{A}_m := \{A \in \mathcal{A} : \mu(A) > 1/m\}$. For any distinct $A_1, \ldots, A_k \in \mathcal{A}_m$ we have

$$\mu(X) \geq \mu(\bigcup_{i=1}^{k} A_i) = \mu(A_1) + \cdots + \mu(A_k) > k/m,$$

hence $\mathcal{A}_m$ has at most $m\mu(X)$ elements. Thus

$$\{A \in \mathcal{A} : \mu(A) > 0\} = \bigcup_{m=1}^{\infty} \mathcal{A}_m$$

is countable. $\qquad\square$

Example. If μ is a finite Borel measure on $\mathbb{R}$, then $\mu(\{t\}) = 0$ for all except at most countably many $t \in \mathbb{R}$.

Proposition 2.3. *Any finite Borel measure on X is* regular, *that is, for every $B \in \mathcal{B}$*

$$\begin{aligned} \mu(B) &= \sup\{\mu(C) : C \subset B, \ C \ closed\} \quad \textit{(inner regular)} \\ &= \inf\{\mu(U) : U \supset B, \ U \ open\} \quad \textit{(outer regular)}. \end{aligned}$$

Proof. Define the collection $\mathcal{R}$ by

$$A \in \mathcal{R} \quad \Longleftrightarrow \quad \begin{array}{l} \mu(A) = \sup\{\mu(C) : C \subset A,\ C \text{ closed}\} \text{ and} \\ \mu(A) = \inf\{\mu(C) : U \supset A,\ U \text{ open}\}. \end{array}$$

We have to show that $\mathcal{R}$ contains the Borel sets. *step 1: $\mathcal{R}$ is a σ-algebra:*
$\emptyset \in \mathcal{R}$. Let $A \in \mathcal{R}$, let $\varepsilon > 0$. Take C closed and U open with $C \subset A \subset U$ and
$\mu(A) < \mu(C) + \varepsilon$, $\mu(A) > \mu(U) - \varepsilon$. Then $U^c \subset A^c \subset C^c$, U^c is closed, C^c is
open, and

$$\begin{array}{l} \mu(A^c) = \mu(X) - \mu(A) > \mu(X) - \mu(C) - \varepsilon = \mu(C^c) - \varepsilon, \\ \mu(A^c) = \mu(X) - \mu(A) < \mu(X) - \mu(U) + \varepsilon = \mu(U^c) + \varepsilon. \end{array}$$

Hence $A^c \in \mathcal{R}$.

Let $A_1, A_2, \ldots \in \mathcal{R}$ and let $\varepsilon > 0$. Take for each i

$$\begin{array}{l} U_i \text{ open }, C_i \text{ closed with} \\ C_i \subset A \subset U_i, \\ \mu(U_i) - \mu(A_i) < 2^{-i}\varepsilon,\ \mu(A_i) - \mu(C_i) < 2^{-i}\varepsilon/2. \end{array}$$

Then $\bigcup_i C_i \subset \bigcup_i A_i \subset \bigcup_i U_i$ and $\bigcup_i U_i$ is open, and

$$\begin{aligned} \mu(\bigcup_i U_i) - \mu(\bigcup_i A_i) \quad &\leq \quad \mu\Big(\bigcup_{i=1}^{\infty} U_i \setminus \bigcup_{i=1}^{\infty} A_i\Big) \\ &\leq \quad \mu\Big(\bigcup_{i=1}^{\infty}(U_i \setminus A_i)\Big) \leq \sum_{i=1}^{\infty} \mu(U_i \setminus A_i) \\ &= \quad \sum_{i=1}^{\infty}(\mu(U_i) - \mu(A_i)) < \sum_{i=1}^{\infty} 2^{-i}\varepsilon = \varepsilon. \end{aligned}$$

Further, $\mu(\bigcup_{i=1}^{\infty} C_i) = \lim_{k \to \infty} \mu(\bigcup_{i=1}^{k} C_i)$, hence for some large k, $\mu(\bigcup_{i=1}^{\infty} C_i) - \mu(\bigcup_{i=1}^{k} C_i) < \varepsilon/2$. Then $C := \bigcup_{i=1}^{k} C_i \subset \bigcup_{i=1}^{\infty} A_i$, C is closed, and

$$\begin{aligned} \mu(\bigcup_{i=1}^{\infty} A_i) - \mu(C) \quad &< \quad \mu(\bigcup_{i=1}^{\infty} A_i) - \mu(\bigcup_{i=1}^{\infty} C_i) + \varepsilon/2 \\ &\leq \quad \mu\Big(\bigcup_{i=1}^{\infty} A_i \setminus \bigcup_{i=1}^{\infty} C_i\Big) + \varepsilon/2 \\ &\leq \quad \mu\Big(\bigcup_{i=1}^{\infty}(A_i \setminus C_i)\Big) + \varepsilon/2 \\ &\leq \quad \sum_{i=1}^{\infty} \mu(A_i \setminus C_i) + \varepsilon/2 \\ &= \quad \sum_{i=1}^{\infty}\Big(\mu(A_i) - \mu(C_i)\Big) + \varepsilon/2 \leq \varepsilon/2 + \varepsilon/2. \end{aligned}$$

4

Hence $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$. Thus $\mathcal{R}$ is a σ-algebra.

step2: $\mathcal{R}$ contains all open sets: We prove: $\mathcal{R}$ contains all closed sets. Let $A \subset X$ be closed. Let $U_n := \{x \in X : d(x, A) < 1/n\} = \{x \in X : \exists\, a \in A$ with $d(a, x) < 1/n\}$, $n = 1, 2, \ldots$. Then U_n is open, $U_1 \supset U_2 \supset \cdots$, and $\bigcap_{i=1}^{\infty} U_i = A$, as A is closed. Hence $\mu(A) = \lim_{n \to \infty} \mu(U_n) = \inf_n \mu(U_n)$. So

$$\mu(A) \leq \inf\{\mu(U) : U \supset A, \ U \text{ open}\} \leq \inf_n \mu(U_n) = \mu(A).$$

Hence $A \in \mathcal{R}$.

Conclusion: $\mathcal{R}$ is a σ-algebra that contains all open sets, so $\mathcal{R} \supset \mathcal{B}$. $\qquad\square$

Corollary 2.4. *If μ and ν are finite Borel measures on the metric space X and $\mu(A) = \nu(A)$ for all closed A (or all open A), then $\mu = \nu$.*

A finite Borel measure μ on X is called *tight* if for every $\varepsilon > 0$ there exists a compact set $K \subset X$ such that $\mu(X \setminus K) < \varepsilon$, or, equivalently, $\mu(K) \geq \mu(X) - \varepsilon$. A tight finite Borel measure is also called a *Radon measure*.

Corollary 2.5. *If μ is a tight finite Borel measure on the metric space X, then*

$$\mu(A) = \sup\{\mu(K) : K \subset A, \ K \text{ compact}\}$$

for every Borel set A in X.

Proof. Take for every $\varepsilon > 0$ a compact set K_ε such that $\mu(X \setminus K_\varepsilon) < \varepsilon$. Then

$$\mu(A \cap K_\varepsilon) = \mu(A \setminus K_\varepsilon^c) \geq \mu(A) - \mu(K_\varepsilon^c) > \mu(A) - \varepsilon$$

and

$$
\begin{aligned}
\mu(A \cap K_\varepsilon) \ &= \ \sup\{\mu(C) : C \subset K_\varepsilon \cap A, \ C \text{ closed}\} \\
&\leq \ \sup\{\mu(K) : K \subset A, \ K \text{ compact}\},
\end{aligned}
$$

because each closed subset contained in a compact set is compact. Combination completes the proof. $\qquad\square$

Of course, if (X, d) is a compact metric space, then every finite Borel measure on X is tight. There is another interesting case. A complete separable metric space is sometimes called a *Polish space*.

Theorem 2.6. *If (X, d) is a complete separable metric space, then every finite Borel measure on X is tight.*

We need a lemma from topology.

Lemma 2.7. *If (X, d) is a complete metric space, then a closed set K in X is compact if and only if it is totally bounded, that is, for every $\varepsilon > 0$ the set K is covered by finitely many balls (open or closed) of radius less than or equal to ε.*

Proof. $\Rightarrow$) Clear: the covering with all ε-balls with centers in K has a finite subcovering.

$\Leftarrow$) Let $(x_n)_n$ be a sequence in K. For each $m \geq 1$ there are finitely many $1/m$-balls that cover K, at least one of which contains x_n for infinitely many n. For $m = 1$ take a ball B_1 with radius ≤ 1 such that $N_1 := \{n : x_n \in B_1\}$ is infinite, and take $n_1 \in N_1$. Take a ball B_2 with radius $\leq 1/2$ such that $N_2 := \{n > n_1 : x_n \in B_2 \cap B_1\}$ is infinite, and take $n_2 \in N_2$. Take B_3, radius $\leq 1/3$, $N_3 := \{n > n_2 : x_n \in B_3 \cap B_2 \cap B_1\}$ infinite, $n_3 \in N_3$. And so on.

Thus $(x_{n_k})_k$ is a subsequence of $(x_n)_n$ and since $x_{n_{ell}} \in B_k$ for all $\ell \geq k$, $(x_{n_k})_k$ is a Cauchy sequence. As X is complete, $(x_n)_n$ converges in X and as K is closed, the limit is in K. So $(x_n)_n$ has a convergent subsequence and K is compact. $\qquad\square$

Proof of Theorem 2.6. We have to prove that for every $\varepsilon > 0$ there exists a compact set K such that $\mu(X \setminus K) < \varepsilon$. Let $D = \{a_1, a_2, \ldots\}$ be a countable dense subset of X. Then for each $\delta > 0$, $\bigcup_{k=1}^{\infty} B(a_k, \delta) = X$. Hence $\mu(X) = \lim_{n \to \infty} \mu(\bigcup_{k=1}^{n} B(a_k, \delta))$ for all $\delta > 0$. Let $\varepsilon > 0$. Then there is for each $m \geq 1$ an n_m such that

$$\mu\Big(\bigcup_{k=1}^{n_m} B(a_k, 1/m) \Big) > \mu(X) - 2^{-m}\varepsilon.$$

Let

$$K := \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m).$$

Then K is closed and for each $\delta > 0$,

$$K \subset \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m) \subset \bigcup_{k=1}^{n_m} B(a_k, \delta)$$

if we choose $m > 1/\delta$. So K is compact, by the lemma. Further,

$$\begin{aligned} \mu(X \setminus K) &= \mu\Big(\bigcup_{m=1}^{\infty} (X \setminus \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m)) \Big) \leq \sum_{m=1}^{\infty} \mu\Big(X \setminus \bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m) \Big) \\ &= \sum_{m=1}^{\infty} \Big(\mu(X) - \mu(\bigcup_{k=1}^{n_m} \overline{B}(a_k, 1/m)) \Big) < \sum_{m=1}^{\infty} 2^{-m}\varepsilon = \varepsilon. \end{aligned}$$

$\qquad\square$

3 Weak convergence of measures

Let (X, d) be a metric space and denote

$$C_b(X) := \{f : X \to \mathbb{R} : f \text{ is continuous and bounded}\}.$$

Each $f \in C_b(X)$ is integrable with respect to any finite Borel measure on X.

Definition 3.1. Let $\mu, \mu_1, \mu_2, \ldots$ be finite Borel measures on X. We say that $(\mu_i)_i$ *converges weakly* to μ if

$$\int f d\mu_i \to \int f d\mu \text{ as } i \to \infty \text{ for all } f \in C_b(X).$$

Notation: $\mu_i \Rightarrow \mu$. (There is at most one such a limit μ, as follows from the metrization by the Prokhorov metric, which is discussed in the next section.)

Theorem 3.2. *Let (X, d) be a metric space, let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on X. The following statements are equivalent:*

(a) $\mu_i \Rightarrow \mu$

(b) $\int g d\mu_i \to \int g d\mu$ *for all* $g \in UC_b(X) := \{f : X \to \mathbb{R} : f \text{ is uniformly continuous and bounded}\}$

(c) $\limsup_{i \to \infty} \mu_i(C) \le \mu(C)$ *for all closed* $C \subset X$

(d) $\liminf_{i \to \infty} \mu_i(U) \ge \mu(U)$ *for all open* $U \subset X$

(e) $\mu_i(A) \to \mu(A)$ *for every Borel set A in X with $\mu(\partial A) = 0$. ($\partial A = \overline{A} \setminus A^\circ$).*

Proof. (a)$\Rightarrow$(b) is clear

(b)$\Rightarrow$(c): Let C be a closed set, nonempty. Let $U_m := \{x : d(x, C) < 1/m\}$, $m \ge 1$. (Here $d(x, A) := \inf_{a \in A} d(x, a)$ if $A \ne \emptyset$, and "$d(x, \emptyset) := \infty$".) Then U_m^c is closed and

$$\inf_{x \in C, y \in U_m^c} d(x, y) \ge 1/m.$$

Hence there is an $f_m \in UC_b(X)$ with $0 \le f \le 1$ on X, $f_m = 1$ on C, and $f_m = 0$ on U_m^c. (Indeed, $f_m(x) := \frac{d(x, U_m^c)}{d(x, U_m^c) + d(x, C)}$ does the job.) Since

$$\mu_i(C) = \int \mathbb{1}_C d\mu_i \le \int f_m d\mu_i,$$

we get by assumption (b)

$$\limsup_{i \to \infty} \mu_i(C) \le \limsup_{i \to \infty} \int f_m d\mu_i = \int f_m d\mu \le \int \mathbb{1}_{U_m} d\mu = \mu(U_m).$$

Because $\bigcap_{m=1}^{\infty} U_m = C$ (since C is closed) we find

$$\mu(C) = \lim_{m \to \infty} \mu(U_m) \ge \limsup_{i \to \infty} \mu_i(C).$$

(c)$\Rightarrow$(d): By complements,

$$\begin{aligned} \liminf_{i \to \infty} \mu_i(U) &= \liminf_{i \to \infty} \left(\mu_i(X) - \mu_i(U^c) \right) = 1 - \limsup_{i \to \infty} \mu_i(U^c) \\ &\ge 1 - \mu(U^c) = \mu(X) - \mu(U^c) = \mu(U). \end{aligned}$$

(d)⇒(c): Similarly.

(c)+(d)⇒(e): $A^\circ \subset A \subset \overline{A}$, A° is open and $\overline{A}$ is closed, so by (c) and (d)

$$
\begin{aligned}
\limsup \mu_i(A) &\leq \limsup \mu_i(\overline{A}) \leq \mu(\overline{A}) = \mu(A \cup \partial A) \\
&\leq \mu(A) + \mu(\partial A) = \mu(A), \\
\liminf \mu_i(A) &\geq \liminf \mu_i(A^\circ) \geq \mu(A^\circ) = \mu(A \setminus \partial A) \\
&\geq \mu(A) - \mu(\partial A) = \mu(A),
\end{aligned}
$$

hence $\mu_i(A) \to \mu(A)$.

(e)⇒(a): Let $g \in C_b(X)$. Idea: we have $\int f d\mu_i \to \int f d\mu$ for suitable simple functions; we want to approximate g to get $\int g d\mu_i \to \int g d\mu$.

Define

$$
\nu(E) := \mu(\{x : g(x) \in E\}) = \mu(g^{-1}(E)), \quad E \text{ Borel set in } \mathbb{R}.
$$

Then ν is a finite Borel measure (probability measure) on $\mathbb{R}$ and if we take $a < -\|g\|_\infty$, $b > \|g\|_\infty$, then $\nu(\mathbb{R} \setminus (a,b)) = 0$. As ν is finite, there are at most countably many α with $\nu(\{\alpha\}) > 0$. Hence for $\varepsilon > 0$ there are $t_0, \ldots, t_m \in \mathbb{R}$ such that

(i) $a = t_0 < t_1 < \cdots < t_m = b$,

(ii) $t_j - t_{j-1} < \varepsilon$, $j = 1, \ldots, m$,

(iii) $\nu(\{t_j\}) = 0$, i.e., $\mu(\{x : g(x) = t_j\}) = 0$, $j = 0, \ldots, m$.

Take

$$
A_j := \{x \in X : t_{j-1} \leq g(x) < t_j\} = g^{-1}([t_{j-1}, t_j)), \ j = 1, \ldots, m.
$$

Then $A_j \in \mathcal{B}(X)$ for all j and $X = \bigcup_{j=1}^m A_j$. Further,

$$
\begin{aligned}
\overline{A}_j &\subset \{x : t_{j-1} \leq g(x) \leq t_j\} \text{ (since this set is closed and } \supset A_j), \\
A_j^\circ &\supset \{x : t_{j-1} < g(x) < t_j\} \text{ (since this set is open and } \subset A_j),
\end{aligned}
$$

so

$$
\begin{aligned}
\mu(\partial A_j) &= \mu(\overline{A}_j \setminus A_j^\circ) \leq \mu(\{x : g(x) = t_{j-1} \text{ or } g(x) = t_j\}) \\
&= \mu(\{x : g(x) = t_{j-1}\}) + \mu(\{x : g(x) = t_j\}) = 0 + 0.
\end{aligned}
$$

Hence by (e), $\mu_i(A_j) \to \mu(A_j)$ as $i \to \infty$ for $j = 1, \ldots, m$. Put

$$
h := \sum_{j=1}^m t_{j-1} \mathbb{1}_{A_j},
$$

then $h(x) \leq g(x) \leq h(x) + \varepsilon$ for all $x \in X$. Hence

$$
\begin{aligned}
|\int g d\mu_i - \int g d\mu| &= |\int (g - h) d\mu_i + \int h d\mu_i - \int (g - h) d\mu - \int h d\mu| \\
&\leq \int |g - h| d\mu_i + |\int h d\mu_i - \int h d\mu| + \int |g - h| d\mu \\
&\leq \varepsilon \mu_i(X) + |\sum_{j=1}^m t_{j-1} \Big(\mu_i(A_j) - \mu(A_j)\Big)| + \varepsilon \mu(X).
\end{aligned}
$$

8

It follows that $\limsup_{i \to \infty} |\int g\,d\mu_i - \int g\,d\mu| \leq 2\varepsilon$. Thus $\int g\,d\mu_i \to \int g\,d\mu$ as $i \to \infty$. □

Remark. The condition that the measures $\mu, \mu_1, \mu_2, \ldots$ in the above theorem are probability measures can be weakened to finite Borel measures such that $\mu_i(X) \to \mu(X)$ as $i \to \infty$. The same proof can be used with only minor modifications in the proof of the equivalence (c)$\Leftrightarrow$(d).

4 The Prokhorov metric

Let (X, d) be a metric space. Denote

$$\mathcal{P} = \mathcal{P}(X) := \text{all Borel probability measures on } X.$$

We have defined the notion of weak convergence in $\mathcal{P}$. Define for $\mu, \nu \in \mathcal{P}$

$$d_P(\mu, \nu) := \inf\{\alpha > 0 : \mu(A) \leq \nu(A_\alpha) + \alpha \text{ and } \nu(A) \leq \mu(A_\alpha) + \alpha \ \forall A \in \mathcal{B}(X)\},$$

where

$$A_\alpha := \{x : d(x, A) < \alpha\} \text{ if } A \neq \emptyset, \quad \emptyset_\alpha := \emptyset \text{ for all } \alpha > 0.$$

(Here $d(x, A) = \inf\{d(x, a) : a \in A\}$.) The function d_P is called the *Prokhorov metric on $\mathcal{P}$* (induced by d), which makes sense because of the next theorem. If X is separable, then convergence in the metric d_P is the same as weak convergence in $\mathcal{P}$.

Theorem 4.1. *Let (X, d) be a metric space.*

(1) d_P is a metric on $\mathcal{P} = \mathcal{P}(X)$.

(2) Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{P}$. Then $d_P(\mu_i, \mu) \to 0$ implies $\mu_i \Rightarrow \mu$.

Proof. (1): Any $\alpha \geq 1$ is in the set of the defining formula of d_P, so the infimum is well defined. Clearly $d_P(\mu, \nu) \geq 0$ and $d_P(\mu, \nu) = d_P(\nu, \mu)$ for all $\mu, \nu \in \mathcal{P}$.

$d_P(\mu, \mu) = 0$: Let $\mu \in \mathcal{P}$. For every Borel set A and $\alpha > 0$, $A_\alpha \supset A$, so $\mu(A) \leq \mu(A_\alpha) + \alpha$, hence $d_P(\mu, \nu) \leq \alpha$, whence $d_P(\mu, \mu) = 0$.

$d_P(\mu, \nu) = 0 \Rightarrow \mu = \nu$: If $d_P(\mu, \nu) = 0$, then there is a sequence $\alpha_n \downarrow 0$ such that $\mu(A) \leq \nu(A_{\alpha_n}) + \alpha_n$ and $\nu(A) \leq \mu(A_{\alpha_n}) + \alpha_n$ for all n. As $\overline{A} = \bigcap_n A_{\alpha_n}$, it follows that $\mu(A) \leq \nu(\overline{A})$ and $\nu(A) \leq \mu(\overline{A})$. In particular, $\mu(A) = \nu(A)$ for all closed sets A and therefore $\mu = \nu$ (by inner regularity).

Triangle inequality: Let $\mu, \nu, \eta \in \mathcal{P}$. Let $\alpha > 0$ be such that

$$\mu(A) \leq \eta(A_\alpha) + \alpha, \quad \eta(A) \leq \mu(A_\alpha) + \alpha \quad \text{for all } A \in \mathcal{B}$$

and $\beta > 0$ such that

$$\nu(A) \leq \eta(A_\beta) + \beta, \quad \eta(A) \leq \nu(A_\beta) + \beta \quad \text{for all } A \in \mathcal{B}.$$

Then for $A \in \mathcal{B}$:

$$\mu(A) \le \eta(A_\alpha) + \alpha \le \nu((A_\alpha)_\beta) + \alpha + \beta,$$
$$\nu(A) \le \eta(A_\beta) + \beta \le \mu((A_\beta)_\alpha) + \beta + \alpha.$$

Now notice that $(A_\alpha)_\beta \subset A_{\alpha+\beta}$. (Indeed, $x \in (A_\alpha)_\beta \Rightarrow d(x, A_\alpha) < \beta \Rightarrow \exists \ y \in A_\alpha : d(x,y) < \beta$, and $y \in A_\alpha \Rightarrow \exists \ a \in A : d(y,a) < \alpha$, so that $d(x,a) \le d(x,y) + d(y,a) < \alpha + \beta$, and $x \in A_{\alpha+\beta}$.) Of course also $(A_\beta)_\alpha \subset A_{\alpha+\beta}$. Hence for all $A \in \mathcal{B}$,

$$\mu(A) \le \nu(A_{\alpha+\beta}) + \alpha + \beta,$$
$$\nu(A) \le \mu(A_{\alpha+\beta}) + \alpha + \beta.$$

Thus, by the definition, $d_P(\mu, \nu) \le \alpha + \beta$. The infimum over the α under consideration is $d_P(\mu, \eta)$ and the infimum over the β is $d_P(\eta, \nu)$. Thus taking infimum over α and β yields

$$d_P(\mu, \nu) \le d_P(\mu, \eta) + d_P(\eta, \nu).$$

The proof of (1) is complete.

(2): Assume that $d_P(\mu_i, \mu) \to 0$ as $i \to \infty$. Then there are $\alpha_i \downarrow 0$ with $\mu_i(A) \le \mu(A_{\alpha_i}) + \alpha_i$ and $\mu(A) \le \mu_i(A_{\alpha_i}) + \alpha_i$ for all $A \in \mathcal{B}$. Hence for $A \in \mathcal{B}$,

$$\begin{aligned} \limsup_{i \to \infty} \mu_i(A) & \le \limsup_{i \to \infty} \left(\mu(A_{\alpha_i}) + \alpha_i \right) \\ & = \lim_{i \to \infty} \mu(A_{\alpha_i}) = \mu(\overline{A}). \end{aligned}$$

In particular, for any closed $C \subset X$, $\limsup_{i \to \infty} \mu_i(C) \le \mu(C)$, and therefore $\mu_i \Rightarrow \mu$. $\qquad\blacksquare$

Theorem 4.2. *If (X, d) is a separable metric space, then for any $\mu, \mu_1, \mu_2, \ldots \in \mathcal{P}(X)$ one has*

$$\mu_i \Rightarrow \mu \qquad \text{if and only if} \qquad d_P(\mu_i, \mu) \to 0.$$

For the proof we need a lemma on existence of special coverings with small balls.

Lemma 4.3. *Let X be a separable metric space and μ be a finite Borel measure on X. Then for each $\delta > 0$ there are countably many open (or closed) balls $B_1, B_2, \ldots$ such that*

$\bigcup_{i=1}^{\infty} B_i = X$,
radius of B_i is $< \delta$ for all i,
$\mu(\partial B_i) = 0$ *for all i.*

Proof. Let D be a countable dense set in X. Let $x \in D$. Let $S(x, r) := \{y \in X : d(y, x) = r\}$. Observe that the boundary of the open or closed ball centered at x and with radius r is contained in $S(x, r)$. Given $\delta > 0$, the collection

10

$\mathcal{S} := \{S(x,r) : \delta/2 < r < \delta\}$ is disjoint and therefore at most countably many of its members have μ-measure > 0. As $\mathcal{S}$ is uncountable, there exists an $r \in (\delta/2, \delta)$ such that $\mu(S(x,r)) = 0$. In this way we find for each $x \in D$ an open (or closed) ball $B(x,r)$ centered at x with radius $r \in (\delta/2, \delta)$ and $\mu(\partial B(x,r)) = 0$. As D is dense these balls cover X, and as D is countable we have countably many, say $B_1, B_2, \ldots$. $\qquad \square$

Proof of Theorem 4.2. $\Leftarrow$) already done.

$\Rightarrow$) Let $\varepsilon > 0$. We want to show that there exists an N such that for every $i \geq N$ we have $d_P(\mu_i, \mu) \leq \varepsilon$, which means that $\mu_i(B) \leq \mu(B_\varepsilon) + \varepsilon$ and $\mu(B) \leq \mu_i(B_\varepsilon) + \varepsilon$ for all $B \in \mathcal{B}$.

Take $\delta > 0$ with $\delta < \varepsilon/3$ and take with aid of the previous lemma open balls $B_1, B_2, \ldots$ with radius $< \delta/2$ such that $\bigcup_{j=1}^\infty B_j = X$ and $\mu(\partial B_j) = 0$ for all j. Fix k such that

$$\mu\Big(\bigcup_{j=1}^k B_j \Big) \geq 1 - \delta.$$

Consider the collection of sets that can be built by combining the balls $B_1, \ldots, B_k$:

$$\mathcal{A} := \{\bigcup_{j \in J} B_j : J \subset \{1, \ldots, k\}\},$$

which is a finite collection. We are going to use this collection to approximate arbitrary Borel sets. For each $A \in \mathcal{A}$, $\partial A \subset \partial B_1 \cup \cdots \cup \partial B_k$, so $\mu(\partial A) \leq \mu(\partial B_1) + \cdots + \mu(\partial B_k) = 0$. Since $\mu_i \Rightarrow \mu$, we have $\mu_i(A) \to \mu(A)$ for all $A \in \mathcal{A}$. Fix N such that

$$|\mu_i(A) - \mu(A)| < \delta \quad \text{for all } i \geq N \text{ and for all } A \in \mathcal{A}.$$

Then in particular $\mu_i(\bigcup_{j=1}^k B_j) \geq \mu(\bigcup_{j=1}^k B_j) - \delta \geq 1 - 2\delta$ for all $i \geq N$.

Let now $B \in \mathcal{B}$ be given. Take as approximation of B the set

$$A := \bigcup\{B_j : j \in \{1, \ldots, k\} \text{ such that } B_j \cap B \neq \emptyset\} \in \mathcal{A}.$$

We find

- $A \subset B_\delta = \{x : d(x, B) < \delta\}$ because the diameter of each B_j is $< \delta$,

- $B = [B \cap \bigcup_{j=1}^k B_j] \cup [B \cap (\bigcup_{j=1}^k B_j)^c] \subset A \cup (\bigcup_{j=1}^k B_j)^c$, because $B \cap \bigcup_{j=1}^k B_j = \bigcup_{j=1}^k (B \cap B_j) \subset A$,

- $|\mu_i(A) - \mu(A)| < \delta$ for all $i \geq N$,

- $\mu\Big((\bigcup_{j=1}^k B_j)^c\Big) \leq \delta$, $\mu_i\Big((\bigcup_{j=1}^k B_j)^c\Big) \leq 2\delta$ for all $i \geq N$.

11

Hence for every $i \geq N$:

$$
\begin{aligned}
\mu(B) &\leq \mu(A) + \mu\Big(\big(\bigcup_{j=1}^{k} B_j\big)^c\Big) \leq \mu(A) + \delta \leq \mu_i(A) + 2\delta \\
&\leq \mu_i(B_\delta) + 2\delta \leq \mu_i(B_\varepsilon) + \varepsilon, \\
\mu_i(B) &\leq \mu_i(A) + \mu_i\Big(\big(\bigcup_{j=1}^{k} B_j\big)^c\Big) \leq \mu_i(A) + 2\delta \leq \mu(A) + 3\delta \\
&\leq \mu(B_\delta) + 3\delta \leq \mu(B_\varepsilon) + \varepsilon.
\end{aligned}
$$

This is true for every $B \in \mathcal{B}$, so $d_P(\mu_i, \mu) \leq \varepsilon$ for all $i \geq N$. $\qquad\square$

Proposition 4.4. *Let (X, d) be a separable metric space. Then $\mathcal{P} = \mathcal{P}(X)$ with the Prokhorov metric d_P is separable.*

Proof. Let $D := \{a_1, a_2, \ldots\}$ be a countable set in X. Let

$$
\mathcal{M} := \{\alpha_1 \delta_{a_1} + \cdots + \alpha_k \delta_{a_k} : \alpha_1, \ldots, \alpha_k \in \mathbb{Q} \cap [0, 1], \ \sum_{j=1}^{k} \alpha_j = 1, \ k = 1, 2, \ldots\}.
$$

(Here δ_a denotes the Dirac measure at $a \in X$: $\delta_a(A) = 1$ if $a \in A$, 0 otherwise.) Clearly, $\mathcal{M} \subset \mathcal{P}$ and $\mathcal{M}$ is countable.

Claim: $\mathcal{M}$ is dense in $\mathcal{P}$. Indeed, let $\mu \in \mathcal{P}$. For each $m \geq 1$, $\bigcup_{j=1}^{\infty} B(a_j, 1/m) = X$. Take k_m such that

$$
\mu(\bigcup_{j=1}^{k_m} B(a_j, 1/m)) \geq 1 - 1/m.
$$

Modify the balls $B(a_j, 1/m)$ into disjoint sets by taking $A_1^m := B(a_1, 1/m)$, $A_j^m := B(a_j, 1/m) \setminus \big[\bigcup_{i=1}^{j-1} B(a_i, 1/m)\big]$, $j = 2, \ldots, k_m$. Then $A_1^m, \ldots, A_{k_m}^m$ are disjoint and $\bigcup_{i=1}^{j} A_i^m = \bigcup_{i=1}^{j} B(a_i, 1/m)$ for all j. In particular, $\mu(\bigcup_{j=1}^{k_m} A_j^m) \geq 1 - 1/m$, so

$$
\sum_{j=1}^{k_m} \mu(A_j^m) \in [1 - 1/m, 1].
$$

We approximate

$$
\mu(A_1^m)\delta_{a_1} + \cdots + \mu(A_{k_m}^m)\delta_{a_{k_m}}
$$

by

$$
\mu_m := \alpha_1^m \delta_{a_1} + \cdots + \alpha_{k_m}^m \delta_{a_{k_m}},
$$

where we choose $\alpha_j^m \in [0, 1] \cap \mathbb{Q}$ such that $\sum_{j=1}^{k_m} \alpha_j^m = 1$ and

$$
\sum_{j=1}^{k_m} |\mu(A_j^m) - \alpha_j^m| < 2/m.
$$

(First take $\beta_j \in [0,1] \cap \mathbb{Q}$ with $\sum_{j=1}^{k_m} |\beta_j - \mu(A_j^m)| < 1/2m$, then $\sum_j \beta_j \in [1 - 3/2m, 1 + 1/2m]$. Take $\alpha_j := \beta_j / \sum_i \beta_i \in [0,1] \cap \mathbb{Q}$, then $\sum_j \alpha_j = 1$ and $\sum_{j=1}^{k_m} |\beta_j - \alpha_j| = |1 - 1/\sum_i \beta_i| \sum_{j=1}^{k_m} \beta_j = |\sum_i \beta_i - 1| \le 3/2m$, so $\sum_{j=1}^{k_m} |\alpha_j - \mu(A_j^m)| < 1/2m + 3/2m = 2/m$.)

Then for each m, $\mu_m \in \mathcal{M}$. To show: $\mu_m \Rightarrow \mu$. Let $g \in UC_b(X)$. Then

$$\left| \int g \, d\mu_m - \int g \, d\mu \right| = \left| \sum_{j=1}^{k_m} \alpha_j^m g(a_j) - \int g \, d\mu \right|$$

$$\le \left| \sum_{j=1}^{k_m} \mu(A_j^m) g(a_j) - \int g \, d\mu \right| + (2/m) \sup_j |g(a_j)|$$

$$\le \left| \int \sum_{j=1}^{k_m} g(a_j) \mathbb{1}_{A_j^m} \, d\mu - \int g \, d\mu \right| + (2/m) \|g\|_\infty$$

$$\le \left| \sum_{j=1}^{k_m} \int \Big(g(a_j) \mathbb{1}_{A_j^m} - g \mathbb{1}_{A_j^m} \Big) d\mu - \int g \mathbb{1}_{(\bigcup_{j=1}^{k_m})^c} \, d\mu \right| + (2/m) \|g\|_\infty$$

$$\le \sum_{j=1}^{k_m} \sup_{x \in A_j^m} |g(a_j) - g(x)| \mu(A_j^m) + \|g\|_\infty \mu \Big((\bigcup_{j=1}^{k_m} A_j^m)^c \Big) + (2/m) \|g\|_\infty.$$

Each A_j^m is contained in a ball with radius $1/m$ around a_j. Since g is uniformly continuous, for every $\varepsilon > 0$ there is a $\delta > 0$ such that $|g(y) - g(x)| < \varepsilon$ whenever $|x - y| < \delta$, so $|g(x) - g(a_j)| < \varepsilon$ for all $x \in A_j^m$ for all j. Then for $m > 1/\delta$ it follows from the above computation that

$$\left| \int g \, d\mu_m - \int g \, d\mu \right| \le \varepsilon + \|g\|_\infty (1/m) + (2/m) \|g\|_\infty.$$

Hence $\int g \, d\mu_m \to \int g \, d\mu$ as $m \to \infty$. Thus, $\mu_m \Rightarrow \mu$. $\qquad \square$

Conclusion. If (X, d) is a separable metric space, then so is $\mathcal{P}(X)$ with the induced Prokhorov metric. Moreover, a sequence in $\mathcal{P}(X)$ converges in metric if and only if it converges weakly and to the same limit.

5 Prokhorov's theorem

Let (X, d) be a metric space and let $\mathcal{P}(X)$ be the set of Borel probability measures on X. Endow $\mathcal{P}(X)$ with the Prokhorov metric induced by d.

In the study of limit behavior of stochastic processes one often needs to know when a sequence of random variables is convergent in distribution or, at least, has a subsequence that converges in distribution. This comes down to finding a good description of the sequences in $\mathcal{P}(X)$ that have a convergent subsequence or rather of the relatively compact sets of $\mathcal{P}(X)$. Recall that a subset S of a metric space is called *relatively compact* if its closure $\overline{S}$ is compact.

The following theorem by Yu.V. Prokhorov [7] gives a useful description of the relatively compact sets of $\mathcal{P}(X)$ in case X is separable and complete. Let us first attach a name to the equivalent condition.

Definition 5.1. A set Γ of Borel probability measures on X is called *tight* if for every $\varepsilon > 0$ there exists a compact subset K of X such that

$$\mu(K) \geq 1 - \varepsilon \quad \text{for all } \mu \in \Gamma.$$

(Also other names and phrases are in use instead of 'Γ is tight': 'Γ is uniformly tight', 'Γ satisfies Prokhorov's condition', 'Γ is uniformly Radon', and maybe more).

Remark. We have shown already: if (X, d) is a complete separable metric space, then $\{\mu\}$ is tight for each $\mu \in \mathcal{P}(X)$ (see Theorem 2.6).

Theorem 5.2 (Prokhorov). *Let (X, d) be a complete separable metric space and let Γ be a subset of $\mathcal{P}(X)$. Then the following two statements are equivalent:*

(a) *$\overline{\Gamma}$ is compact in $\mathcal{P}(X)$.*

(b) *Γ is tight.*

Let us first remark here that completeness of X is not needed for the implication (b)$\Rightarrow$(a). The proof of the theorem is quite involved. We start with the more straightforward implication (a)$\Rightarrow$(b).

Proof of (a)$\Rightarrow$(b). Claim: If $U_1, U_2, \ldots$ are open sets in X that cover X and if $\varepsilon > 0$, then there exists a $k \geq 1$ such that

$$\mu(\bigcup_{i=1}^{k} U_i) > 1 - \varepsilon \quad \text{for all } \mu \in \Gamma.$$

To prove the claim by contradiction, suppose that for every $k \geq 1$ there is a $\mu_k \in \Gamma$ with $\mu_k(\bigcup_{i=1}^{k} U_i) \leq 1 - \varepsilon$. As $\overline{\Gamma}$ is compact, there is a $\mu \in \overline{\Gamma}$ and a subsequence with $\mu_{k_j} \Rightarrow \mu$. For any $n \geq 1$, $\bigcup_{i=1}^{n} U_i$ is open, so

$$
\begin{aligned}
\mu(\bigcup_{i=1}^{n} U_i) &\leq \liminf_{j \to \infty} \mu_{k_j}(\bigcup_{i=1}^{n} U_i) \\
&\leq \liminf_{j \to \infty} \mu_{k_j}(\bigcup_{i=1}^{k_j} U_i) \leq 1 - \varepsilon.
\end{aligned}
$$

But $\bigcup_{i=1}^{\infty} U_i = X$, so $\mu(\bigcup_{i=1}^{n} U_i) \to \mu(X) = 1$ as $n \to \infty$, which is a contradiction. Thus the claim is proved.

Now let $\varepsilon > 0$ be given. Take $D = \{a_1, a_2, \ldots\}$ dense in X. For every $m \geq 1$ the open balls $B(a_i, 1/m)$, $i = 1, 2, \ldots$, cover X, so by the claim there is a k_m such that

$$\mu\Big(\bigcup_{i=1}^{k_m} B(a_i, 1/m)\Big) > 1 - \varepsilon 2^{-m} \quad \text{for all } \mu \in \Gamma.$$

Take

$$K := \bigcap_{m=1}^{\infty} \bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m).$$

Then K is closed and for each $\delta > 0$ we can take $m > 1/\delta$ and obtain $K \subset \bigcup_{i=1}^{k_m} B(a_i, \delta)$, so that K is totally bounded. Hence K is compact, since X is complete. Moreover, for each $\mu \in \Gamma$

$$
\begin{aligned}
\mu(X \setminus K) &= \mu\Big(\bigcup_{m=1}^{\infty} \Big[\bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m) \Big]^c \Big) \\
&\leq \sum_{m=1}^{\infty} \mu\Big(\Big[\bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m) \Big]^c \Big) \\
&= \sum_{m=1}^{\infty} \Big(1 - \mu\Big(\bigcup_{i=1}^{k_m} \overline{B}(a_i, 1/m) \Big) \Big) \\
&< \sum_{m=1}^{\infty} \varepsilon 2^{-m} = \varepsilon.
\end{aligned}
$$

Hence Γ is tight. $\qquad\square$

The proof that condition (b) implies (a) is more difficult. We will follow the proof from [6], which is based on compactifications and the Riesz representation theorem. The latter will be discussed in a later section and, accordingly, we invoke it with almost no explanation here.

Observe that if X is a compact metric space, every set of Borel probability measures on X is tight, so that in particular $\mathcal{P}(X)$ itself is tight. Thus, the implication (b)$\Rightarrow$(a) entails the assertion that $\mathcal{P}(X)$ is compact whenever X is compact. We choose the latter as an important intermediate step in the proof of (b)$\Rightarrow$(a).

Proposition 5.3. *If (X, d) is a compact metric space, then $(\mathcal{P}(X), d_P)$ is a compact metric space.* (Note that any compact metric space is separable.)

Proof. (Revisited in Corollary 6.8.) As X is compact, $C_b(X) = C(X) = \{f : X \to \mathbb{R} : f \text{ is continuous}\}$, which is a Banach space under the supremum norm defined by

$$\|f\|_\infty = \sup_{x \in X} |f(x)|.$$

Denote by $C(X)'$ the Banach dual space of $C(X)$ and consider

$$\Phi := \{\varphi \in C(X) : \|\varphi\| \leq 1, \ \varphi(\mathbb{1}) = 1, \ \varphi(f) \geq 0 \ \forall f \in C(X) \text{ with } f \geq 0\}.$$

For $\mu \in \mathcal{P}(X)$ define $\varphi_\mu(f) := \int f d\mu$, $f \in C(X)$. According to the Riesz representation theorem, the map $T : \mu \to \varphi_\mu$ is a bijection from $\mathcal{P}(X)$ onto Φ. Moreover, T is a sequential homeomorphism relative to the weak* topology on

Φ. By Alaoglu's theorem, $B' := \{\varphi \in C(X)' : \|\varphi\| \le 1\}$ is weak* compact and therefore Φ is weak* compact, since Φ is weak* closed in B'. Hence Φ is weak* sequentially compact and hence $\mathcal{P}(X)$ is compact. $\qquad\square$

Remark. Also the converse is true: if $\mathcal{P}(X)$ is compact then so is X. This comes from the fact that $x \mapsto \delta_x$ is a homeomorphism from X onto $\{\delta_x : x \in X\} \subset \mathcal{P}(X)$, and $\{\delta_x : x \in X\}$ is closed in $\mathcal{P}(X)$. (See Proposition 9.3.)

In the cases that we want to consider, X is typically not compact. We can make use of the previous proposition by considering a compactification of X.

Lemma 5.4. *If (X, d) is a separable metric space, then there exist a compact metric space (Y, δ) and a map $T : X \to Y$ such that T is a homeomorphism from X onto $T(X)$.*

(T is in general not an isometry. If it were, then X complete $\Rightarrow T(X)$ complete $\Rightarrow T(X) \subset Y$ closed $\Rightarrow T(X)$ compact, which is not true for, e.g., $X = \mathbb{R}$.)

Proof. Let $Y := [0, 1]^{\mathbb{N}} = \{(\xi_i)_{i=1}^\infty : \xi_i \in [0, 1] \ \forall i\}$ and

$$\delta(\xi, \eta) := \sum_{i=1}^\infty 2^{-i} |\xi_i - \eta_i|, \qquad \xi, \eta \in Y.$$

Then δ is a metric on Y, its topology is the topology of coordinatewise convergence, and (Y, δ) is compact.

Let $D = \{a_1, a_2, \dots\}$ be dense in X and define

$$\alpha_i(x) := \min\{d(x, a_i), 1\}, \qquad x \in X, \ i = 1, 2, \dots.$$

Then for each k, $\alpha_k : X \to [0, 1]$ is continuous. For $x \in X$ define

$$T(x) := (\alpha_i(x))_{i=1}^\infty \in Y.$$

Claim: for any $C \subset X$ closed and $x \notin C$ there exist $\varepsilon > 0$ and i such that

$$\alpha_i(x) \le \varepsilon/3, \quad \alpha_i(y) \ge 2\varepsilon/3 \quad \text{for all } y \in C.$$

To prove the claim, take $\varepsilon := \min\{d(x, C), 1\} \in (0, 1]$. Take i such that $d(a_i, x) < \varepsilon/3$. Then $\alpha_i(x) \le \varepsilon/3$ and for $y \in C$ we have

$$
\begin{aligned}
\alpha_i(y) &= \min\{d(y, a_i), 1\} \ge \min\{(d(y, x) - d(x, a_i)), 1\} \\
&\ge \min\{(d(x, C) - \varepsilon/3), 1\} \\
&\ge \min\{2\varepsilon/3, 1\} = 2\varepsilon/3.
\end{aligned}
$$

In particular, if $x \ne y$ then there exists an i such that $\alpha_i(x) \ne \alpha_i(y)$, so T is injective. Hence $T : X \to T(X)$ is a bijection. It remains to show that for $(x_n)_n$ and x in X:

$$x_n \to x \iff T(x_n) \to T(x).$$

If $x_n \to x$, then $\alpha_i(x_n) \to \alpha_i(x)$ for all i, so $\delta(T(x_n), T(x)) \to 0$ as $n \to \infty$.

16

Conversely, suppose that $x_n \nrightarrow x$. Then there is a subsequence such that $x \notin \overline{\{x_{n_1}, x_{n_2}, \dots\}}$. Then by the claim there is an i such that $\alpha_i(x) \le \varepsilon/3$ and $\alpha_i(x_{n_k}) \ge 2\varepsilon/3$ for all k, so that $\alpha_i(x_{n_k}) \nrightarrow \alpha_i(x)$ as $k \to \infty$ and hence $T(x_{n_k}) \nrightarrow T(x)$. $\qquad\square$

We can now complete the proof of Prokhorov's theorem.

Proof of (b)⇒(a). We will show more: *If (X, d) is a separable metric space and $\Gamma \subset \mathcal{P}(X)$ is tight, then $\overline{\Gamma}$ is compact.* Let $\Gamma \subset \mathcal{P}(X)$ be tight. First observe that $\overline{\Gamma}$ is tight as well. Indeed, let $\varepsilon > 0$ and let K be a compact subset of X such that $\mu(K) \ge 1 - \varepsilon$ for all $\mu \in \Gamma$. Then for every $\mu \in \overline{\Gamma}$ there is a sequence $(\mu_n)_n$ in Γ that converges to μ and then we have $\mu(K) \ge \limsup_{n\to\infty} \mu_n(K) \ge 1 - \varepsilon$.

Let $(\mu_n)_n$ be a sequence in $\overline{\Gamma}$. We have to show that it has a convergent subsequence. Let (Y, δ) be a compact metric space and $T : X \to Y$ be such that T is a homeomorphism from X onto $T(X)$. For $B \in \mathcal{B}(Y)$, $T^{-1}(B)$ is Borel in X. Define
$$\nu_n(B) := \mu_n(T^{-1}(B)), \quad B \in \mathcal{B}(Y), \ n = 1, 2, \dots.$$

Then $\nu \in \mathcal{P}(Y)$ for all n. As Y is a compact metric space, $\mathcal{P}(X)$ is a compact metric space, hence there is a $\nu \in \mathcal{P}(Y)$ and a subsequence such that $\nu_{n_k} \Rightarrow \nu$ in $\mathcal{P}(Y)$. We want to translate ν back to a measure on X. Set $Y_0 := T(X)$.

Claim: ν is concentrated on Y_0 in the sense that there exists a set $E \in \mathcal{B}(Y)$ with $E \subset Y_0$ and $\nu(E) = 1$.

If we assume the claim, define
$$\nu_0(A) := \nu(A \cap E), \quad A \in \mathcal{B}(Y_0).$$

(Note: $A \in \mathcal{B}(Y_0) \Rightarrow A \cap E$ Borel in $E \Rightarrow A \cap E$ Borel in Y, since E is a Borel subset of Y.) The measure ν_0 is a finite Borel measure on Y_0 and $\nu_0(E) = \nu(E) = 1$. Now we can translate ν_0 back to
$$\mu(A) := \nu_0(T(A)) = \nu_0((T^{-1})^{-1}(A)), \quad A \in \mathcal{B}(X).$$

Then $\mu \in \mathcal{P}(X)$. We want to show that $\mu_{n_k} \Rightarrow \mu$ in $\mathcal{P}(X)$. Let C be closed in X. Then $T(C)$ is closed in $T(X) = Y_0$. ($T(C)$ need not be closed in Y.) Therefore there exists $Z \subset Y$ closed with $Z \cap Y_0 = T(C)$. Then $C = \{x \in X : T(x) \in T(C)\} = \{x \in X : T(x) \in Z\} = T^{-1}(Z)$, because there are no points in $T(C)$ outside Y_0, and $Z \cap E = T(C) \cap E$. Hence

$$
\begin{aligned}
\limsup_{k\to\infty} \mu_{n_k}(C) \ &= \ \limsup_{k\to\infty} \nu_{n_k}(Z) \\
&\le \ \nu(Z) \\
&= \ \nu(Z \cap E) + \nu(Z \cap E^c) = \nu(T(C) \cap E) + 0 \\
&= \ \nu_0(T(C)) = \mu(C).
\end{aligned}
$$

So $\mu_{n_k} \Rightarrow \mu$.

Finally, to prove the claim we use tightness of $\overline{\Gamma}$. For each $m \geq 1$ take K_m compact in X such that $\mu(K_m) \geq 1 - 1/m$ for all $\mu \in \Gamma$. Then $T(K_m)$ is a compact subset of Y hence closed in Y, so

$$\begin{aligned}
\nu(T(K_m)) &\geq \limsup_{k \to \infty} \nu_{n_k}(T(K_m)) \\
&\geq \limsup_{k \to \infty} \mu_{n_k}(K_m) \geq 1 - 1/m.
\end{aligned}$$

Take $E := \bigcup_{m=1}^{\infty} K_m$. Then $E \in \mathcal{B}(Y)$ and $\nu(E) \geq \nu(K_m)$ for all m, so $\nu(E) = 1$. $\square$

Example. Let $X = \mathbb{R}$, $\mu_n(A) := n^{-1}\lambda(A \cap [0, n])$, $A \in \mathcal{B}(\mathbb{R})$. Here λ denotes Lebesgue measure on $\mathbb{R}$. Then $\mu_n \in \mathcal{P}(\mathbb{R})$ for all n. The sequence $(\mu_n)_n$ has no convergent subsequence. Indeed, suppose $\mu_{n_k} \Rightarrow \mu$, then

$$\begin{aligned}
\mu((-N, N)) &\leq \liminf_{n \to \infty} \mu_n((-N, N)) \\
&= \liminf_{n \to \infty} n^{-1}\lambda([0, N]) = \liminf_{n \to \infty} N/n = 0,
\end{aligned}$$

so $\mu(\mathbb{R}) = \sup_{N \geq 1} \mu((-N, N)) = 0$. There is leaking mass to infinity; the set $\{\mu_n : n = 1, 2, \ldots\}$ is not tight.

6 Riesz representation theorem

In the proof of Prokhorov's theorem we have used the Riesz representation theorem. It yields a correspondence between functionals on a space of continuous functions and measures on the underlying set. The standard theorem deals with compact spaces and will be discussed in this section. The next section derives via compactification an extension for non-compact spaces.

Let (X, d) be a metric space. For each finite Borel measure μ on X, the map φ_μ defined by

$$\varphi_\mu(f) := \int f \, d\mu, \quad f \in C_b(X),$$

is linear from $C_b(X)$ to $\mathbb{R}$ and

$$|\varphi_\mu(f)| \leq \int |f| \, d\mu \leq \|f\|_\infty \mu(X).$$

Hence $\varphi_\mu \in C_b(X)'$, where $C_b(X)'$ denotes the Banach dual space of the Banach space $(C_b(X), \|\cdot\|_\infty)$. (Here $\|f\|_\infty = \sup_{x \in X} |f(x)|$.) Further, $\|\varphi_\mu\| \leq \mu(X)$ and since $\varphi_\mu(\mathbb{1}) = \mu(X) = \|\mathbb{1}\|_\infty \mu(X)$ we have

$$\|\varphi_\mu\| = \mu(X).$$

Moreover,

$$f \geq 0 \implies \varphi_\mu(f) \geq 0.$$

Definition 6.1. A linear map $\varphi : C_b(X) \to \mathbb{R}$ is called *positive* if

$$\varphi(f) \geq 0 \text{ for all } f \in C_b(X) \text{ with } f \geq 0.$$

(Then $f \leq g \Rightarrow \varphi(f) \leq \varphi(g)$.)

Lemma 6.2. *For every positive $\varphi \in C_b(X)'$ one has*

$$\|\varphi\| = \varphi(\mathbb{1}).$$

Proof. Clearly, $\varphi(\mathbb{1}) \leq \|\varphi\| \, \|\mathbb{1}\|_\infty = \|\varphi\|$. For $f \in C_b(X)$,

$$-\|f\|_\infty \mathbb{1} \leq f \leq \|f\|_\infty \mathbb{1},$$

so

$$-\|f\|_\infty \varphi(\mathbb{1}) \leq \varphi(f) \leq \|f\|_\infty \varphi(\mathbb{1}),$$

so

$$|\varphi(f)| \leq \varphi(\mathbb{1}) \|f\|_\infty,$$

hence $\|\varphi\| \leq \varphi(\mathbb{1})$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

If X is compact, then $C_b(X) = C(X) = \{f : X \to \mathbb{R} : f \text{ is continuous}\}$ and every positive bounded linear functional on $C(X)$ is represented by a finite Borel measure on X. The truth of this statement does not depend on X being a metric space. In the extension to the non-compact case that we will discuss in the next section we need the generality of non-metrizable compact Hausdorff spaces. Formally we have not defined Borel sets, Borel measures, $C_b(X)$, etc. for topological spaces that are not metrizable. The appropriate definitions are literally the same and omitted.

Theorem 6.3 (Riesz representation theorem). *If (X, d) is a compact Hausdorff space and $\varphi \in C(X)'$ is positive and $\|\varphi\| = 1$, then there exists a unique Borel probability measure μ on X such that*

$$\varphi(f) = \int f d\mu \quad \text{for all } f \in C(X).$$

(See [8, Theorem 2.14, p. 40].)

By obvious scaling, the Riesz representation theorem can be extended to a correspondence between not necessarily normalized positive bounded functionals on $C(X)$ and finite Borel measures on X. More than that, there is also a correspondence of topologies.

Consider the *weak* topology* on $C_b(X)'$, which is the coarsest topology such that the functional $\varphi \mapsto \varphi(f)$ on $C_b(X)'$ is continuous for every $f \in C_b(X)$. A sequence $\varphi_1, \varphi_2, \ldots$ in $C_b(X)'$ converges in the *weak* topology* to $\varphi \in C_b(X)'$ if and only if $\varphi_n(f) \to \varphi(f)$ for all $f \in C_b(X)$. The following observation is immediate.

Proposition 6.4. *Let (X, d) be a compact metric space and let $\mu, \mu_1, \mu_2, \ldots$ be finite Borel measures on X. Then the following two statemts are equivalent:*

(a) *$\mu_n \Rightarrow \mu$, that is, $\int f d\mu_n \to \int f d\mu$ for all $f \in C_b(X)$.*

(b) *$\varphi_{\mu_n} \to \varphi_\mu$ in the weak* topology, that is, $\varphi_{\mu_n}(f) \to \varphi_\mu(f)$ for all $f \in C_b(X)$.*

With a suitable notion of nonpositive measure, the representation by a measure can be extended to every member of $C_b(X)'$. We include the statements without proofs.

Definition 6.5. A *signed Borel measure* on a metric space (X, d) is a map $\mu : \mathcal{B}(X) \to \mathbb{R}$ of the form

$$\mu = \mu_1 - \mu_2$$

where μ_1 and μ_2 are finite Borel measures on X. This is equivalent to

$\mu(\emptyset) = 0$,
μ is σ-additive,
    i.e., $A_1, A_2, \ldots \in \mathcal{B}(X)$disjoint $\implies \mu\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mu(A_i)$,
$\sup_{A \in \mathcal{B}(X)} |\mu(A)| < \infty$.

Theorem 6.6. *Let (X, d) be a compact metric space. For a finite Borel measure μ on X let $\varphi_\mu(f) := \int f d\mu$, $f \in C(X)$, and let T be the map $\mu \mapsto \varphi_\mu$. Then*

(1) *$T(\mu + \nu) = T(\mu) + T(\nu)$ and $T(c\mu) = cT(\mu)$ for all finite Borel measures μ and ν on X and all $c \geq 0$,*

(2) *T is a sequential homeomorphism from $\{\mu : \mu \text{ finite Borel measure on } X\}$ onto $\{\varphi \in C(X)' : \varphi \text{ positive}\}$ with the weak* topology, and $T(\mathcal{P}(X)) = \{\varphi \in C_b(X)' : \|\varphi\| = 1, \ \varphi \text{ positive}\}$,*

(3) *T extends uniquely to a linear sequential homeomorphism from the signed Borel measures on X onto $C(X)'$ with the weak* topology.*

Remark. One can show that $C(X)$ is separable if X is compact and metrizable ([5, 7.S(d), p. 245]) and one can derive from the separability of $C(X)$ that $\{\varphi \in C(X)' : \|\varphi\| \leq 1\}$ is metrizable ([3, Theorem V.5.1, p. 426]). Therefore T in the above theorem is a homeomorphism and not only a sequential homeomorphism.

We are now in a position to have a closer look at the proof of Proposition 5.3.

Theorem 6.7. *Let (X, d) be a metric space. Then*

(1) *$B' := \{\varphi \in C_b(X)' : \|\varphi\| \leq 1\}$ is weak* compact (Alaoglu's theorem, see [3, Theorem V.4.2, p. 424]),*

(2) *$\{\varphi \in C_b(X)' : \|\varphi\| = 1, \ \varphi \text{ is positive}\}$ is weak* closed in B'.*

Proof of (2). For positive $\varphi \in B'$, we have $\|\varphi\| = 1 \Leftrightarrow \varphi(\mathbb{1}) = 1$. Hence

$$\{\varphi \in C_b(X)' : \|\varphi\| = 1, \ \varphi \text{ is positive}\}$$
$$= \{\varphi \in C_b(X)' : \|\varphi\| \leq 1, \ \varphi(\mathbb{1}) = 1, \ \varphi(f) \geq 0, \ \forall f \in C_b(X) \text{ s.t. } f \geq 0\}$$
$$= \{\varphi \in B' : \varphi(\mathbb{1}) = 1\} \cap \bigcap_{f \in C_b(X), \ f \geq 0} \{\varphi \in B' : \varphi(f) \geq 0\}.$$

Since $\varphi \mapsto \varphi(f)$ is weak* continuous for all $f \in C_b(X)$, this set is weak* closed in B'. $\qquad\square$

Corollary 6.8. *If (X, d) is a compact metric space, then $(\mathcal{P}(X), d_P)$ is a compact metric space.*

Proof. The map $T : \mathcal{P}(X) \to \{\varphi \in C(X)' : \|\varphi\| = 1, \ \varphi \text{ positive}\} =: \Phi$ is a sequential homeomorphism with respect to the weak* topology on Φ. By the previous theorem, Φ is weak* compact, hence sequentially weak* compact. So $\mathcal{P}(X)$ is sequentially compact. As $\mathcal{P}(X)$ is a metric space, $\mathcal{P}(X)$ is compact. $\quad\square$

7 Riesz representation for non-compact spaces

As we are mainly interested in metric spaces that are not compact, it is natural for us to study an extension of the Riesz representation theorem to non-compact spaces. Such an extension can be obtained by means of a compactification of the space.

The compactification of Lemma 5.4 has the advantage of being metrizable, but it is not suitable for the present purposes. We have to step outside metric topology for a moment. We want a connection between the continuous functions on the compactification and the bounded continuous functions on the original space. Such a compactification is the famous Stone-Čech compactification.

Theorem 7.1. *Let (X, d) be a metric space. There exists a compact Hausdorff space Y and a map $T : X \to Y$ such that*

(i) *T is a homeomorphism from X onto $T(X)$,*

(ii) *$T(X)$ is dense in Y,*

(iii) *for every $f \in C_b(X)$ there exists one and only one $g \in C(Y)$ 'that extends f', that is, $g \circ T = f$.*

The pair (Y, T) of the above theorem is essentially unique and called the *Stone-Čech compactification* of X (see [5, 5.24, p. 152–3; 5.P, p. 166]). We will not be unnecessarily cautious, and view X as a subspace of Y. Then the above theorem says that every metric space X is a dense subspace of a compact Hausdorff space Y such that $C_b(X) \simeq C(Y)$ under the natural isomorphism of extension and restriction. From the Riesz representation theorem for compact Hausdorff spaces we thus have the next conclusion.

Corollary 7.2. *Let (X, d) be a metric space. If $\varphi : C_b(X) \to \mathbb{R}$ is bounded linear and positive, then there exists a unique finite Borel measure μ on the Stone-Čech compactification Y of X such that*

$$\varphi(f) = \int \overline{f} d\mu \quad \text{for all } f \in C_b(X),$$

where $\overline{f} \in C(Y)$ denotes the extension of f.

Thus the positive bounded linear functionals on $C_b(X)$ correspond to the finite Borel measures on the Stone-Čech compactification of X. It is interesting to know when such a measure is concentrated on X itself. It turns out to be connected with a stronger continuity property of the functional than mere norm continuity. The precise statement is in the next theorem, which is an extension of the Riesz representation theorem for compact spaces (cf. [2, 5.2 Proposition 5, p. 58, and 5.6 Proposition 12, p. 65]. For theory on convergence of nets (generalized sequences), see [3, I.7, p. 26–31].

Theorem 7.3. *Let (X, d) be a metric space and let $\varphi \in C_b(X)'$ be positive. The following statements are equivalent:*

(a) *There exists a tight finite Borel measure μ on X such that*

$$\varphi(f) = \int f d\mu \quad \text{for all } f \in C_b(X).$$

(b) *For every $\varepsilon > 0$ there exists a compact $K \subset X$ such that $|\varphi(f)| \leq \varepsilon$ for all $f \in C_b(X)$ with $\|f\|_\infty \leq 1$ and $f = 0$ on K.*

(c) *The restriction of φ to the unit ball $B = \{f \in C_b(X) : \|f\|_\infty \leq 1\}$ is continuous with respect to the topology of uniform convergence on compact sets.*

If (a) holds, then the measure μ is unique.

Proof. The proof of the uniqueness is routine. It also follows from the denseness theorem in Section 8.

   (a)$\Rightarrow$(c): Let $(f_i)_{i \in I}$ be a net in B and let $f \in B$ be such that $f_i \to f$ uniformly on compact sets. Let $\varepsilon > 0$. We want to show that there is an $i_0 \in I$ such that $|\varphi(f_i) - \varphi(f)| < \varepsilon$ for all $i \in I$ with $i \geq i_0$. Since μ is tight, there is a compact $K \subset X$ with $\mu(X \setminus K) < \varepsilon/3$. Then $f_i \to f$ uniformly on K, so there is an $i_0 \in I$ such that

$$|f_i - f| < \varepsilon/(3\mu(K) + 1) \text{ on } K \text{ for all } i \geq i_0.$$

Then for $i \geq i_0$,

$$
\begin{aligned}
|\varphi(f_i) - \varphi(f)| &\leq \int_K |f_i - f| d\mu + \int_{X \setminus K} |f_i - f| d\mu \\
&\leq \tfrac{\varepsilon}{3\mu(K)+1} \mu(K) + \|f_i - f\|_\infty \mu(X \setminus K) \\
&< \varepsilon/3 + 2\varepsilon/3 = \varepsilon.
\end{aligned}
$$

22

Hence $\varphi(f_i) \to \varphi(f)$, and φ is continuous on B.

(c)$\Rightarrow$(b): Suppose that (b) is not true. Then there exists an $\varepsilon > 0$ such that for every compact $K \subset X$ there is an $f_K \in C_b(X)$ with $\|f_K\|_\infty \le 1$ and $f_K = 0$ on K and such that $|\varphi(f_K)| > \varepsilon$. Then $(f_K)_{K \in \mathcal{K}}$, where $\mathcal{K} = \{K \subset X : K$ compact$\}$ with inclusion as ordering, is a net in B that converges to zero in the topology of uniform convergence on compact sets. Indeed, for each compact $K_0 \subset X$, $f_K = 0$ on K_0 for all $K \supset K_0$. Since $|\varphi(f_K)| > \varepsilon$ for all $K \in \mathcal{K}$, it follows that φ is not continuous on B.

(b)$\Rightarrow$(a): Take for each $m \ge 1$ a compact set $K_m \subset X$ such that $|\varphi(f)| \le 1/m$ for all $f \in C_b(X)$ with $\|f\|_\infty \le 1$ and $f = 0$ on K_m. Let Y be the Stone-Čech compactification of X. For every $g \in C(Y)$ its restriction to X is an element of $C_b(X)$ and we can define

$$\psi(g) := \varphi(g|_X), \quad g \in C(Y).$$

Then $\psi : C(Y) \to \mathbb{R}$ is a bounded linear and positive functional, so by the Riesz representation theorem there exists a finite Borel measure ν on Y such that

$$\psi(g) = \int g d\nu \quad \text{for all } g \in C(Y).$$

We want to restrict ν to a measure μ on X that represents φ. Therefore we need that ν has no mass outside X.

Let $E := \bigcup_m K_m \subset X$. Since every K_m is compact, E is a Borel set in Y. To show that ν has no mass outside E we exploit the assumption (b) by means of an approximation of $\mathbb{1}_{K_m^c}$ by continuous functions. Let

$$h_m(x) := \min\{d(x, K_m), 1\}, \quad x \in Y, \ m \ge 1.$$

Then $h_m \in C(Y)$, $0 \le h_m \le \mathbb{1}_{K_m^c}$ and $\sqrt[n]{h_m} \uparrow \mathbb{1}_{K_m^c}$ as $n \to \infty$, since $h_m(x) > 0$ for every $x \in K_m^c$. Hence by the monotone convergence theorem,

$$\begin{aligned}
\nu(Y \setminus K_m) &= \int \mathbb{1}_{K_m^c} d\nu = \lim_{n\to\infty} \int \sqrt[n]{h_m} d\nu \\
&= \lim_{n\to\infty} \psi(\sqrt[n]{h_m}) = \lim_{n\to\infty} \varphi(\sqrt[n]{h_m}|_X) \le 1/m,
\end{aligned}$$

by assumption (b). Therefore

$$\nu(Y \setminus E) = \nu\left(\bigcap_{m=1}^\infty K_m^c\right) = 0.$$

Define

$$\mu(A) := \nu(A \cap E), \quad A \in \mathcal{B}(X).$$

(Notice that $A \in \mathcal{B}(X) \Rightarrow A \cap E$ Borel in E hence Borel in Y.) Then μ is a finite Borel measure on X. To show that μ represents φ, let $f \in C_b(X)$ and let $\overline{f} \in C(Y)$ be its extension. Since

$$\nu(Y \setminus E) = 0 \text{ and } \mu(X \setminus E) = 0,$$

it follows that

$$\int f d\mu = \int f \mathbb{1}_E d\mu = \int \overline{f} \mathbb{1}_E d\nu = \int \overline{f} d\nu = \psi(\overline{f}) = \varphi(f).$$

Finally, notice that $\mu(X \setminus K_m) = \nu(E \setminus K_m) = \nu(Y \setminus K_m) \leq 1/m$ for all m, so that μ is tight. $\qquad\square$

Remark. (1) If X is compact, then every $\varphi \in C_b(X)'$ satisfies condition (c). Thus we retrieve the Riesz representation theorem for compact metric spaces.

(2) We have shown earlier that if (X, d) is a complete separable metric space, then each finite Borel measure on X is tight. Hence for such a space condition (c) is necessary to have representation by any finite Borel measure.

Example. Let $X = \mathbb{N}$, $d(x, y) = |x - y|$, $x, y \in X$. We will show that there exists a $\varphi \in C_b(X)'$ that is not represented by a finite Borel measure. Observe that $C_b(X) = \ell^\infty(\mathbb{N})$ and define

$$\varphi_0(x) := \lim_{k \to \infty} x(k)$$

for all $x \in c := \{y \in \ell^\infty(\mathbb{N}) : \lim_{k \to \infty} y(k) \text{ exists}\}$. The set c is a closed subspace of $\ell^\infty(\mathbb{N})$ and φ_0 is a positive bounded linear functional on c. Let

$$p(x) := \max\{\limsup_{k \to \infty} x(k), 0\}, \quad x \in \ell^\infty(\mathbb{N}).$$

Then $p(x + y) \leq p(x) + p(y)$ and $p(\lambda x) = \lambda p(x)$ for all $x, y \in \ell^\infty(\mathbb{N})$, $\lambda \geq 0$. Further, $\varphi_0(x) \leq p(x)$ for all $x \in c$. Hence by the Hahn-Banach theorem (see [3, II.3.10, p. 62]) there exists a linear functional $\varphi : \ell^\infty(\mathbb{N}) \to \mathbb{R}$ that extends φ_0 and such that $\varphi(x) \leq p(x)$ for all $x \in \ell^\infty(\mathbb{N})$. Then $|\varphi(x)| \leq |p(x)| \leq \|x\|_\infty$ for all $x \in \ell^\infty(\mathbb{N})$ so φ is bounded, and for $x \in \ell^\infty(\mathbb{N})$ with $x \geq 0$ we have $\varphi(x) = -\varphi(-x) \geq -p(-x) = 0$, so φ is positive.

Let now

$$x_n := \mathbb{1}_{\{n,n+1,\dots\}} \in c, \quad n = 1, 2, \dots.$$

Then $\varphi(x_n) = \varphi_0(x_n) = 1$ for all n, but for any finite Borel measure μ on $\mathbb{N}$ we have $\int x_n d\mu \to 0$ as $n \to \infty$, since $x_n \to 0$ pointwise and $0 \leq x_n \leq \mathbb{1}$ for all n. Hence φ cannot be represented by a finite Borel measure.

8 Integrable functions on metric spaces

Let (X, d) be a metric space and let μ be a finite Borel measure on X. Is $C_b(X)$ dense in $\mathcal{L}^1(\mu)$? The answer is positive and we can show more. Let

$$\text{Lip}_b(X) \quad := \quad \{f : X \to \mathbb{R} : f \text{ is bounded and Lipschitz continuous}$$
$$\text{with respect to } d\}.$$

Lemma 8.1. *Let (X, d) be a metric space and let μ be a finite Borel measure on X.*

24

(1) For every $U \subset X$ open and every $\varepsilon > 0$ there exists an $f \in \mathrm{Lip}_b(X)$ with $0 \le f \le \mathbb{1}_U$ and $\int (\mathbb{1}_U - f) d\mu < \varepsilon$.

(2) For every $A \in \mathcal{B}(X)$ and every $\varepsilon > 0$ there exists an $f \in \mathrm{Lip}_b(X)$ with $\int |f - \mathbb{1}_A| d\mu < \varepsilon$.

Proof. (1): If $U = X$, take $f = \mathbb{1}$. If $U \subset X$ is open and $U \ne X$, let

$$h(x) := \min\{d(x, U^c), 1\}, \quad x \in X.$$

Then $h \in \mathrm{Lip}_b(X)$. Indeed, observe that $\min\{a, c\} - \min\{b, c\} \le a - b$ if $a \ge b$, so that

$$|h(x) - h(y)| \le |d(x, U^c) - d(y, U^c)| \le d(x, y),$$

for all $x, y \in X$. Further, $0 \le h \le \mathbb{1}_U$ on X and $h(x) > 0$ for all $x \in U$.

Take a strictly concave Lipschitz continuous function $\rho : [0, 1] \to [0, 1]$ with $\rho(0) = 0$ and $\rho(1) = 1$. For instance, $\rho(x) = x(2 - x)$. Denote the iterates of ρ by $\rho^1 := \rho$, $\rho^n := \rho \circ \rho^{n-1}$, $n = 2, 3, \ldots$. For $0 < \alpha < 1$,

$$\rho(\alpha) = \rho((1 - \alpha)0 + \alpha 1) > (1 - \alpha)\rho(0) + \alpha\rho(1) = \alpha,$$

so $\rho^n(\alpha)$ is increasing in n and its limit must be 1. Thus, $\rho^n(0) = 0$, $\rho^n(1) = 1$ for all n, and $\rho^n(\alpha) \uparrow 1$ for every $0 < \alpha < 1$.

For each n, $\rho^n \circ h \in \mathrm{Lip}_b(X)$, $\rho^n \circ h \ge 0$, and $\rho^n(h(x)) \uparrow \mathbb{1}_U(x)$ for all $x \in X$. By the monotone convergence theorem we therefore find

$$\int (\rho^n \circ h) d\mu \to \int \mathbb{1}_U d\mu \quad \text{as } n \to \infty.$$

So for large n the function $f := \rho^n \circ h$ has the desired properties.

(2): With aid of the outer regularity of μ, take $U \subset X$ open with $A \subset U$ and $\mu(U \setminus A) < \varepsilon/2$. Take, by (1), $f \in \mathrm{Lip}_b(X)$ with $\int |\mathbb{1}_U - f| d\mu < \varepsilon/2$. Then $\int |f - \mathbb{1}_A| d\mu < \varepsilon$. $\square$

Theorem 8.2. *If (X, d) is a metric space and μ is a finite Borel measure on X, then $\mathrm{Lip}_b(X)$ is dense in $\mathcal{L}^1(\mu)$. Consequently, $C_b(X)$ is dense in $\mathcal{L}^1(\mu)$.*

Proof. 1. Let $A_1, \ldots, A_n \in \mathcal{B}(X)$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R} \setminus \{0\}$, and let $\varepsilon > 0$. Take for each $k \in \{1, \ldots, n\}$ an $h_k \in \mathrm{Lip}_b(X)$ with

$$\int |h_k - \mathbb{1}_{A_k}| d\mu < \frac{\varepsilon}{n|\alpha_k|}.$$

Then $\sum_{k=1}^n \alpha_k h_k \in \mathrm{Lip}_b(X)$ and

$$\int |\sum_{k=1}^n h_k - \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}| d\mu \le \sum_{k=1}^n |\alpha_k| \int |h_k - \mathbb{1}_{A_k}| d\mu < \varepsilon.$$

2. Stepfunctions as in 1. are dense in $\mathcal{L}^1(\mu)$.
(See also [2, 5.2 Proposition 3, p. 57].) $\square$

Corollary 8.3. *Let (X, d) be a metric space and let μ and ν be finite Borel measures on X. If*

$$\int f d\mu = \int f d\nu \quad \text{for all } f \in \text{Lip}_b(X),$$

then $\mu = \nu$.

9 More properties of the space of probability measures

Let (X, d) be a metric space, let $\mathcal{P}(X)$ be the set of Borel probability measures on X, and let d_P be the Prokhorov metric on $\mathcal{P}(X)$ as defined in Section 4. With aid of Prokhorov's theorem we can show that $(\mathcal{P}(X), d_P)$ is complete if (X, d) is complete (cf. [7, Lemma 1.4, p. 169]).

Lemma 9.1. *Let (X, d) be a complete metric space and let $\Gamma \subset \mathcal{P}(X)$. In order that Γ is tight, it suffices that for every $\varepsilon, \delta > 0$ there are $a_1, \ldots, a_n \in X$ such that*

$$\mu\Big(\bigcup_{i=1}^{n} B(a_i, \delta) \Big) \geq 1 - \varepsilon \quad \text{for all } \mu \in \Gamma.$$

Proof. Let $\varepsilon > 0$. Assume that for each $m \geq 1$ the points $a_1^m, \ldots, a_{n_m}^m \in X$ are such that

$$\mu\Big(\bigcup_{i=1}^{n_m} B(a_i^m, 1/m) \Big) \geq 1 - 2^{-m}\varepsilon \quad \text{for all } \mu \in \Gamma.$$

Take

$$K := \bigcap_{m=1}^{\infty} \bigcup_{i=1}^{n_m} \overline{B}(a_i^m, 1/m).$$

Then K is closed and for a given $\delta > 0$ we can take $m > 1/\delta$ and obtain

$$K \subset \bigcup_{i=1}^{n_m} \overline{B}(a_i^m, 1/m) \subset \bigcup_{i=1}^{n_m} B(a_i^m, \delta).$$

Hence K is totally bounded and thus compact. Further, for $\mu \in \Gamma$,

$$
\begin{aligned}
\mu(K) &= \lim_{M \to \infty} \mu\Big(\bigcap_{m=1}^{M} \bigcup_{i=1}^{n_m} \overline{B}(a_i^m, 1/m) \Big) \\
&= 1 - \lim_{M \to \infty} \mu\Big(\bigcup_{i=1}^{M} \Big[\bigcup_{m=1}^{n_m} \overline{B}(a_i^m, 1/m) \Big]^c \Big) \\
&\geq 1 - \lim_{M \to \infty} \sum_{m=1}^{M} \mu\Big(\Big[\bigcup_{i=1}^{n_m} \overline{B}(a_i^m, 1/m) \Big]^c \Big) \\
&\geq 1 - \sum_{m=1}^{\infty} 2^{-m}\varepsilon = 1 - \varepsilon.
\end{aligned}
$$

26

So Γ is tight. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Theorem 9.2. *Let (X,d) be a separable metric space. If (X,d) is complete, then $(\mathcal{P}(X), d_P)$ is complete.*

Proof. Let $(\mu_k)_k$ be a Cauchy sequence in $(\mathcal{P}(X), d_P)$. We will show that $\{\mu_k : k = 1, 2, \ldots\}$ is tight. Take $D = \{a_1, a_2, \ldots\}$ dense in X. Let $\varepsilon, \delta > 0$. Set

$$\gamma := \min\{\varepsilon, \delta\}/2$$

and fix N such that

$$d_P(\mu_k, \mu_\ell) < \gamma \quad \text{for all } k, \ell \geq N.$$

Then for $k, \ell \geq N$ we have

$$\mu_k(A) \leq \mu_\ell(A_\gamma) + \gamma \text{ and } \mu_\ell(A) \leq \mu_k(A_\gamma) + \gamma \text{ for all } A \in \mathcal{B}(X).$$

Take now $n \geq 1$ such that for $k \in \{1, \ldots, N\}$

$$\mu_k\Big(\bigcup_{i=1}^{n} B(a_i, \delta/2)\Big) \geq 1 - \gamma.$$

(Such an n exists because $\bigcup_{i=1}^{\infty} B(a_i, \delta/2) = X$ so that $\lim_{m \to \infty} \mu_k(\bigcup_{i=1}^{m} B(a_i, \delta/2)) \to 1$ for each of the finitely many $k \in \{1, \ldots, N\}$.) Observe that

$$\Big(\bigcup_{i=1}^{n} B(a_i, \delta/2)\Big)_\gamma \subset \bigcup_{i=1}^{n} B(a_i, \delta/2 + \gamma) \subset \bigcup_{i=1}^{n} B(a_i, \delta).$$

Therefore,

$$
\begin{aligned}
\mu_N\Big(\bigcup_{i=1}^{n} B(a_i, \delta/2)\Big) &\leq \mu_k\Big(\Big(\bigcup_{i=1}^{n} B(a_i, \delta/2)\Big)_\gamma\Big) + \gamma \\
&\leq \mu_k\Big(\bigcup_{i=1}^{n} B(a_i, \delta)\Big) + \gamma
\end{aligned}
$$

for all $k \geq N$. Then

$$\mu_k\Big(\bigcup_{i=1}^{n} B(a_i, \delta)\Big) \geq 1 - 2\gamma \geq 1 - \varepsilon \text{ for } k \geq N$$

and

$$
\begin{aligned}
\mu_k\Big(\bigcup_{i=1}^{n} B(a_i, \delta)\Big) &\geq \mu_k\Big(\bigcup_{i=1}^{n} B(a_i, \delta/2)\Big) \\
&\geq 1 - \gamma \geq 1 - \varepsilon
\end{aligned}
$$

for $k = 1, \ldots, N$. By the previous lemma it follows that the set $\{\mu_k : k = 1, 2, \ldots\}$ is tight and therefore relatively compact in $\mathcal{P}(X)$ by Prokhorov's theorem. Hence there is a subsequence $(\mu_{k_i})_i$ that converges to some $\mu \in \mathcal{P}(X)$. As $(\mu_k)_k$ is Cauchy it follows that $\mu_k \Rightarrow \mu$. Thus, $(\mathcal{P}(X), d_P)$ is complete. $\quad\square$

Finally, we want to see that completeness of (X, d) is necessary for completeness of $(\mathcal{P}(X), d_P)$. We can derive this by embedding X in $\mathcal{P}(X)$. More specifically, we show that X and the set of Dirac measures $\Delta := \{\delta_x : x \in X\}$ are in a suitable sense isomorphic. (δ_x denotes the Dirac measure at x.)

Proposition 9.3. *Let (X, d) be a separable metric space. Then:*

(1) $d_P(\delta_x, \delta_y) = \min\{d(x, y), 1\}$ *for every $x, y \in X$,*

(2) $x \mapsto \delta_x$ *is a homeomorphism from X onto $\Delta := \{\delta_x : x \in X\} \subset \mathcal{P}(X)$,*

(3) *a sequence $(x_n)_n$ is Cauchy in (X, d) if and only if $(\delta_{x_n})_n$ is Cauchy in $(\mathcal{P}(X), d_P)$,*

(4) *Δ is closed in $\mathcal{P}(X)$.*

Proof. (1): From the very definition of d_P, $d_P(\mu, \nu) \leq 1$ for all $\mu, \nu \in \mathcal{P}(X)$. Let $\alpha > d(x, y)$. Then for each $A \in \mathcal{B}(X)$,

$$x \in A \Rightarrow y \in A_\alpha \quad \text{and} \quad y \in A \Rightarrow x \in A_\alpha,$$

so

$$\delta_x(A) \leq \delta_y(A_\alpha) + \alpha, \qquad \delta_y(A) \leq \delta_x(A_\alpha) + \alpha,$$

and hence $d_P(\delta_x, \delta_y) \leq \alpha$. Thus, $d_P(\delta_x, \delta_y) \leq d(x, y)$.

Assume $d_P(\delta_x, \delta_y) < 1$ and let $d_P(\delta_x, \delta_y) < \alpha < 1$. Then

$$\delta_x(A) \leq \delta_y(A_\alpha) + \alpha \text{ and } \delta_y(A) \leq \delta_x(A_\alpha) + \alpha \text{ for all } A \in \mathcal{B}(X).$$

Hence for $A = \{x\}$ we find

$$1 \leq \delta_y(B(x, \alpha)) + \alpha.$$

As $\alpha < 1$ it follows that $y \in B(x, \alpha)$, so $d(x, y) < \alpha$. Thus $d(x, y) \leq d_P(\delta_x, \delta_y)$.

(2) and (3) are clear from (1).

(4): Let $(x_n)_n$ be a sequence in X such that $\delta_{x_n} \Rightarrow \mu$ for some $\mu \in \mathcal{P}(X)$. We have to show that $\mu \in \Delta$. Suppose $(x_n)_n$ has no convergent subsequence. Then $S := \{x_1, x_2, \ldots\}$ is closed and so is every subset of S. Hence for every nonempty subset C of S we have

$$\mu(C) \geq \limsup_{n \to \infty} \delta_{x_n}(C) \geq 1.$$

This is only possible if S consists of one point, but that yields a contradiction. Hence there is a subsequence and an $x \in X$ such that $x_{n_k} \to x$. By (2), $\delta_{x_n} \Rightarrow \delta_x$, so $\mu = \delta_x \in \Delta$. $\qquad\square$

With aid of the above proposition we can add the 'only if' counterparts to Proposition 5.3 and Theorem 9.2.

Theorem 9.4. *Let (X, d) be a separable metric space.*

(1) (X, d) *is compact* $\Longleftrightarrow$ $(\mathcal{P}(X), d_P)$ *is compact.*

(2) (X, d) *is complete* $\Longleftrightarrow$ $(\mathcal{P}(X), d_P)$ *is complete.*

Remark. There are other metrics on $\mathcal{P}(X)$ in use than the Prokhorov metric. For instance the *bounded Lipschitz metric*, which is defined by

$$d_{\mathrm{BL}}(\mu, \nu) := \sup\{|\int f d\mu - \int f d\nu| : f \in \mathrm{Lip}_b(X),\ \|f\|_{\mathrm{Lip}} \leq 1\},\ \mu, \nu \in \mathcal{P}(X),$$

where

$$\|f\|_{\mathrm{Lip}} = \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}, \quad f \in \mathrm{Lip}_b(X).$$

If (X, d) is separable, then a sequence in $\mathcal{P}(X)$ converges weakly if and only if it converges in the metric d_{BL}. Further, $(\mathcal{P}(X), d_{\mathrm{BL}})$ is separable and complete if (X, d) is separable and complete. (See [10, 1.12, p. 73–74].)

References

[1] Billingsley, Patrick, *Convergence of Probability Measures*, Wiley & Sons, New York - London, 1968.

[2] Bourbaki, N., *Éléments de mathématique. Livre 6. Intégration. Chapitre 9*, Hermann, Paris, 1969.

[3] Dunford, Nelson, and Schwartz, Jacob T., *Linear Operators. Part I: General Theory*, Wiley-Interscience, New York, 1957.

[4] Halmos, Paul R., *Measure Theory* [GTM 18], Springer, New York - Berlin - Heidelberg, 1974.

[5] Kelley, John L., *General Topology* [GTM 27], Springer, New York - Heidelberg - Berlin, 1975.

[6] Parthasarathy, K.R., *Probability Measures on Metric Spaces*, Academic Press, New York - London, 1967.

[7] Prokhorov, Yu.V., Convergence of random processes and limit theorems in probability theory, *Theor. Prob. Appl.* **1** (1956), 157–214.

[8] Rudin, Walter, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1987.

[9] Stroock, Daniel W., *Probability Theory, an Analytic View*, Cambridge University Press, Cambridge, 1993.

[10] van der Vaart, Aad W. and Wellner, Jon A., *Weak convergence and empirical processes, with applications to statistics*, Springer, New York, 1996.