# Lecture 16. Prokhorov's theorem

Paul Delatte
delatte@usc.edu
University of Southern California

Last updated: 31 December, 2022

To prove Prokhorov's theorem, we use the Riesz(–Markov–Kakutani) representation theorem which was proved in Lecture 13.

**Theorem 1 (Prokhorov's Theorem).** *Let $(X, d)$ be a metric space, $\mathcal{M}_1(X)$ the space of all Borel probability measures endowed with the topology of weak convergence, and $\mathcal{N}$ a subset of $\mathcal{M}_1(X)$.*

1. *If $\mathcal{N}$ is tight, then $\mathcal{N}$ is relatively sequentially compact.*
2. *If $(X, d)$ is Polish and $\mathcal{N}$ is relatively sequentially compact, then $\mathcal{N}$ is tight.*

*Proof.* 2. Suppose that $\overline{\mathcal{N}}$ is sequentially compact and $(X, d)$ is Polish. We first claim that if $U_1, U_2, \ldots$ are opens sets in $X$ that cover $X$ and if $\varepsilon > 0$, then there exists a $k \geq 1$ such that for all $\mu \in \mathcal{N}$,

$$\mu(\bigcup_{i=1}^{k} U_i) > 1 - \varepsilon.$$

By contradiction, suppose that for every $k \geq 1$ there is a $\mu_k \in \mathcal{N}$ such that $\mu_k(\bigcup_{i=1}^{k} U_i) \leq 1 - \varepsilon$. Since $\overline{\mathcal{N}}$ is sequentially compact, there is a $\mu \in \overline{\mathcal{N}}$ and a subsequence such that $\mu_{k_l} \xrightarrow{w} \mu$. For any $n \geq 1$, $\bigcup_{i=1}^{n} U_i$ is open, so

$$\mu(\bigcup_{i=1}^{n} U_i) \leq \liminf_{l\to\infty} \mu_{k_l}(\bigcup_{i=1}^{n} U_i) \leq \liminf_{l\to\infty} \mu_{k_l}(\bigcup_{i=1}^{k_l} U_i) \leq 1 - \varepsilon.$$

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

Now let $\varepsilon > 0$. Take $D = \{x_1, x_2, \ldots\}$ dense in $X$. For every $m \geq 1$, the open balls $B_{1/m}(x_i)$, $i = 1, 2, \ldots$, cover $X$, so by the claim there is a $k_m \geq 1$ such that for all $\mu \in \mathcal{N}$,

$$\mu(\bigcup_{i=1}^{k_m} B_{1/m}(x_i)) > 1 - \varepsilon 2^{-m}.$$

Take

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

Then $K$ is closed and for each $\delta > 0$, we can take $m > 1/\delta$ and obtain $K \subseteq \bigcup_{i=1}^{k_m} \overline{B}_{\delta}(x_i)$

so that $K$ is totally bounded. Since $X$ is complete, $K$ is compact. Now for each $\mu \in \mathcal{N}$,

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

This proves that $\mathcal{N}$ is tight.

(1.) Suppose that $\mathcal{N}$ is tight. Thus, for every $n \in \mathbb{N}$, there is a compact set $K_n \subseteq X$ such that $\mu(K_n) \ge 1 - n^{-1}$ for all $\mu \in \mathcal{N}$. Define $X_0 := \bigcup_{n\ge 1} K_n$. Then $\mu(X_0) = 1$ for all $\mu \in \mathcal{N}$ (and so there is no of loss of generality in assuming that $\mathcal{N} \subseteq \mathcal{M}_1(X_0)$). Since each $K_n$ is compact, (we can take a countable dense for each and so) $X_0$ is separable. By Lemma 8, $X_0$ is homeomorphic to a measurable subset of a compact metric space, i.e., it can be viewed as a subset of a compact metric space $\tilde{X}$.

We now consider $\mathcal{N} \subseteq \mathcal{M}_1(X_0)$ as a subset of the space $\mathcal{M}_1(\tilde{X})$, which is sequentially compact with respect to the topology of weak convergence by Lemma 7. Every sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{N}$ has thus a weakly convergent subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ with limit $\tilde{\mu} \in \mathcal{M}_1(\tilde{X})$. To finish the proof, we need to show that $\tilde{\mu}(X_0) = 1$, since then there is a $\mu \in \mathcal{M}_1(X)$ agreeing with $\tilde{\mu}$ on $X_0$ such that $\mu(X \setminus X_0) = 0$ and so $\mu_{n_k} \xrightarrow{w} \mu$ in $\mathcal{M}_1(X)$. Indeed, by portmanteau, we have

$$\tilde{\mu}(X_0) \ge \tilde{\mu}(K_N) \ge \limsup_{k\to\infty} \mu_{n_k}(K_N) \ge 1 - N^{-1}$$

and the claim follows by letting $N \to \infty$. This concludes the proof. □

*Reference.* T.5.2. in van Gaans' notes p.14 or T.6.26. in Cerny's notes p.49 or T.1.34. in Méliot's notes p.20.

*Remark.* There are several ways to prove the direct implication of Prokhorov's theorem which is the difficult part of the theorem: (a) one measure-theoretic way by construction of the weak limit using a diagonal argument and then following similar steps as in Carathéodory's theorem (see T.5.1. in CPM Billingsley p.59); (b) one functional-analytic way where it is first proved that compactness of the initial space implies compactness of the set of finite Borel measures on the space (see T.5.2. in van Gaans' notes p.14 or T.6.26. in Cerny's notes p.49). Another proof consists in extending the result from $\mathbb{R}^d$ (proved as in Helly's theorem or as in (a) using a diagonal argument) to $\mathbb{R}^{\mathbb{N}}$ and then embedding $X$ into $\mathbb{R}^{\mathbb{N}}$ (see T.23.2. in Kallenberg FMP p.507).

*Remark.* The direct implication (1.) is often stated for $X$ separable. There is no difference in assuming separability since "once you have a tight family, then all those measures live on a separable subset of $X$ anyway, so the rest of the space is irrelevant and might as well not be there" (Nathaniel Eldredge). See the proof for the details of the argument.

**Lemma 2** (**Portmanteau Theorem**). *Let $X$ be a metric space and $\mu, \mu_1, \mu_2, \cdots \in \mathcal{M}_1(X)$ be Borel probability measures on $X$. Then the following propositions are equivalent:*

*1. $\mu_n \xrightarrow{w} \mu$;*

*2. for all bounded and uniformly continuous function $f$ on $X$, $\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu$;*

*3. for all closed sets $F \subseteq X$, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$;*

*4. for all open sets $G \subseteq X$, $\liminf_{n\to\infty} \mu_n(F) \geq \mu(F)$;*

*5. for all Borel sets $A \in \mathcal{B}(X)$ with $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.*

*Proof.* The implication (1) $\implies$ (2) is clear and the equivalence (3) $\iff$ (4) as well (by taking complement).

(2) $\implies$ (3): Fix $F \subseteq X$ nonempty closed. Let $U_m := \{x \in X : d(x, F) < 1/m\}$ for any $m \geq 1$ (where $d(x, A) := \inf_{a\in A} d(x, a)$ if $A \neq \emptyset$). Then $U_m^c$ is closed and $\inf_{x\in F, y\in U_m^c} d(x, y) \geq 1/m$. Define $f_m$ as $f_m(x) := d(x, U_m^c)/(d(x, U_m^c) + d(x, F))$. Then $f_m$ is bounded and uniformly continuous on $X$ with $0 \leq f_m \leq 1$, $f_m = 1$ on $F$ and $f_m = 0$ on $U_m^c$. Hence, $\mathbb{1}_F \leq f_m \leq \mathbb{1}_{U_m}$. Then by (2),

$$\limsup_{n\tilde{\infty}} \mu_n(F) \leq \limsup_{n\to\infty} \int f_m\, d\mu_n = \int f_m\, d\mu \leq \mu(U_m).$$

Since $F$ is closed, $\bigcap_{m=1}^{\infty} U_m = F$. Since the sets $U_m$ are decreasing, then by continuity from below of measures

$$\mu(F) = \lim_{m\to\infty} \mu(U_m) \leq \limsup_{n\to\infty}(F).$$

(3) and (4) $\implies$ (5): Let $A \in \mathcal{B}(X)$ with $\mu(\partial A) = 0$. Then $\mu(A^o) = \mu(A) = \mu(\overline{A})$. Moreover, $A^o \subseteq A \subseteq \overline{A}$, $A^o$ is open, and $\overline{A}$ is closed. Hence,

$$\begin{aligned} \mu(A) = \mu(A^o) &\leq \liminf_{n\to\infty} \mu_n(A^o) \leq \liminf_{n\to\infty} \mu_n(A) \\ &\leq \limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty}(\overline{A}) \leq \mu(\overline{A}) = \mu(A). \end{aligned}$$

(5) $\implies$ (1): Fix $f \in C_b(X)$. Decompose the range of $f$ so that

$$\left[\inf_{x\in X} f(x), \sup_{x\in X} f(x)\right] \subseteq \bigcup_{i=1}^{N} [c_i, c_{i+1}]$$

with $\mu(\{f = c_i\})$ and $0 \leq c_{i+1} - c_i \leq \varepsilon$ for all $0 \leq i \leq N$. This is always possible since $f$ is bounded and the distribution function of $f$, $t \mapsto \mu(\{f \leq t\})$, has at most countably many discontinuities. The sets $A_i = \{c_i \leq f < c_{i+1}\}$ are measurable with $\mu(\delta A_i) = 0$. Define $g = \sum_{i=1}^{N} c_i \mathbb{1}_{A_i}$. Then (5) implies $\lim_{n\to\infty} \int g\, d\mu_n = \int g\, d\mu$. By construction, $\sup_{x\in X} |g(x) - f(x)| \leq \varepsilon$, and thus

$$\limsup_{n\to\infty} \left| \int f\, d\mu_n - \int f\, d\mu \right| \leq 2\varepsilon.$$

Since $\varepsilon$ and $f$ are arbitrary, this proves that $\mu_n \xrightarrow{w} \mu$. □

*Reference.* T.3.2. in van Gaans's notes and T.6.12. in Cerny's notes.

**Lemma 3** (**Tychonoff's theorem**). *Let $I$ be any set and, for each $i \in I$, let $K_i$ be a compact topological space. Then the product*

$$K := \prod_{i \in I} K_i = \{x = (x_i)_{i \in I} : x_i \in K_i \text{ for all } i \in I\}$$

*is compact with respect to the product topology (that is, the weakest topology on $K$ such that the obvious projection $\pi_i \colon K \to K_i$ is continuous for every $i \in I$).*

*Proof.* A family $\mathcal{G}$ of subsets of $K$ has the finite intersection property if the intersection of every finite subfamily of $\mathcal{G}$ has nonempty intersection. Let $\mathcal{F}$ be a family of closed subsets of $K$ that has the finite intersection property. The collection of all families of closed subsets of $K$ that have the finite intersection property and contain $\mathcal{F}$ is partially ordered by inclusion and every chain has an upper bound (the union of all sets in the chain). By Zorn's lemma, there is a maximal family $\mathcal{F}_M$ in this collection. Denote $\pi_i \colon K \to K_i$ the canonical projection onto the $i$ component. Then the family of closed sets $\{\overline{\pi_i(F)}\}_{F \in \mathcal{F}_M}$ has also the finite intersection property. Since $K_i$ is compact, there is some $x_i \in \bigcap_{F \in \mathcal{F}_M} \overline{\pi_i(F)}$. If $F_i$ is a closed neighborhood of $x_i$, then $\pi_i^{-1}(F_i) \in \mathcal{F}_M$ (since otherwise there would be some $F \in \mathcal{F}_M$ such that $F \cap \pi_i^{-1}(F_i) = \emptyset$ contradicting $\pi_i(F) \cap F_i \neq \emptyset$). For every finite subset $I_0 \subseteq I$, we have $\bigcap_{i \in I_0} \pi_i^{-1}(F_i) \in \mathcal{F}_M$. Thus for every $F \in \mathcal{F}_M$, we have $\bigcap_{i \in I_0} \pi_i^{-1}(F_i) \cap F \neq \emptyset$, and so every neighborhood of $x = (x_i)_{i \in I}$ intersects $F$. Since $F$ is closed, we have that $x \in F$. Since $x \in F$ for all $F \in \mathcal{F}_M \supseteq \mathcal{F}$, we have that $x \in F$ for all $F \in \mathcal{F}$, and so $K$ is compact. □

*Reference.* T.B.18. in Teschl TRFA p.535 or T.A.2.1. in Buhler&Salamon ETHZurich FA's notes p.450.

**Lemma 4** (**Banach–Alaoglu Theorem**). *Let $X$ be a (real) normed vector space and $X^*$ its dual topological space (that is, the space of all linear functionals on $V$) endowed with the norm*

$$\|x^*\| = \sup_{x \in X : \|x\| \leq 1} |x^*(x)|.$$

*Then the closed unit ball*

$$B^* := \{x^* \in X^* : \|x^*\| \leq 1\}$$

*in $X^*$ is compact with respect to the weak* topology.*

*Proof.* We apply Tychonoff's theorem with parameter space $I = X$. To each $x \in X$, associate the compact interval

$$K_x := [-\|x\|, \|x\|] \subseteq \mathbb{R}.$$

The product of these intervals is the space

$$K := \prod_{x \in X} K_x = \{f \colon X \to \mathbb{R} : |f(x)| \leq \|x\| \text{ for all } x \in X\} \subseteq \mathbb{R}^X.$$

Define

$$L := \{f \colon X \to \mathbb{R} : f \text{ is linear}\} \subseteq \mathbb{R}^X,$$

then

$$B^* = K \cap L.$$

By definition, the weak* topology on $B^*$ is induced by the product topology on $\mathbb{R}^X$. We now prove that the set $L$ is closed in $\mathbb{R}^X$ with respect to the product topology. Fix $x, y \in X$ and $\lambda \in \mathbb{R}$, and define the maps $\phi_{x,y} \colon \mathbb{R}^X \to \mathbb{R}$ and $\psi_{x,\lambda} \colon \mathbb{R}^X \to \mathbb{R}$ by

$$\phi_{x,y}(f) = f(x+y) - f(x) - f(y) \quad \text{and} \quad \psi_{x,\lambda}(f) = f(\lambda x) - \lambda f(x).$$

By definition of the product topology, these maps are continuous, and thus the set

$$L = \bigcap_{x,y\in X} \phi_{x,y}^{-1}(0) \cap \bigcap_{x\in X,\lambda\in\mathbb{R}} \psi_{x,\lambda}^{-1}(0)$$

is closed with respect to the product topology. By Tychonoff's theorem, $K$ is a compact subset of $\mathbb{R}^X$. Moreover, the space $\mathbb{R}^X$ is Hausdorff. It follows that $B^* = K \cap L$ is compact. □

*Reference.* T.3.2.4. in Buhler&Salamon ETHZurich FA's notes p.126 or T.5.10. in Teschl TRFA p.143.

**Lemma 5** (Riesz(–Markov–Kakutani) representation theorem)**.** *Let $X$ be a compact metric space. The spaces of continuous functions $C(X)$ and the space of finite signed Borel measures $\mathcal{M}_f^{\pm}(X)$ are isometrically isomorphic.*

*Proof.* See Lecture 13. □

**Lemma 6** ($X$ metrizable and separable implies $\mathcal{M}_1(X)$ metrizable and separable)**.** *Let $X$ be be a metrizable and separable topological space. Then the space $\mathcal{M}_1(X)$ of all Borel probability measures on $X$ endowed with the topology of weak convergence is metrizable and separable.*

*Proof.* Let $d$ be a metric on $X$. Define the Lévy–Prokhorov metric on $\mathcal{M}_1(X)$ by

$$\rho(\mu,\nu) := \inf\{\varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ and } \nu(A) \leq \mu(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(X)\}$$

where $A^\varepsilon = \{x \in X : d(x,y) < \varepsilon \text{ for some } y \in A\}$.

We first show that $\rho$ is a metric on $\mathcal{M}_1(X)$. Symmetry follows directly from the definition. Since $(A^\varepsilon)^\eta \subseteq A^{\varepsilon+\eta}$, the triangular inequality follows immediately. We now show that $\rho(\mu,\nu) = 0$ implies $\mu = \nu$. Let $A$ be a closed subset of $X$. Then $A = \bigcap_{\varepsilon>0} A^\varepsilon$. Since $\mu(A) \leq \nu(A^\varepsilon) + \varepsilon$ and $\nu(A) \leq \mu(A^\varepsilon) + \varepsilon$ for all $\varepsilon > 0$, then by letting $\varepsilon \to 0$ we have $\mu(A) = \nu(A)$. Since this holds for any closed subset $A$ of $X$, we have $\mu = \nu$. This proves that $\rho$ is a metric on $\mathcal{M}_1(X)$. Note that we can rewrite $\rho$ as

$$\rho(\mu,\nu) = \inf\{\varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(X)\}.$$

Indeed, let $\varepsilon > 0$ such that $\mu(A) \leq \nu(A^\varepsilon) + \varepsilon$. Note that $((A^\varepsilon)^c)^\varepsilon \subseteq A^c$. Then

$$\nu(A) = 1 - \nu(A^c) \leq 1 - \nu(((A^\varepsilon)^c)^\varepsilon) \leq \mu((A^\varepsilon)^c) + \varepsilon = \mu(A^\varepsilon) + \varepsilon.$$

We then prove that $\rho$ metrizes weak convergence. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}_1(X)$ such that $\rho(\mu_n,\mu) \to 0$ for some $\mu \in \mathcal{M}_1(X)$. Let $A$ be a closed set in $X$ and $\varepsilon > 0$. There exists an integer $n_0$ such that for all $n_0 \geq n$, $\mu_n(A) \leq \mu(A^\varepsilon) + \varepsilon$. Hence $\limsup_n \mu_n(A) \leq \mu(A^\varepsilon) + \varepsilon$. By letting $\varepsilon \to 0$, we get that $\limsup_n \mu_n(A) \leq \mu(A)$

since $A = \bigcap_{\varepsilon > 0} A^{\varepsilon}$ as a closed subset. By portmanteau, we have $\mu_n \xrightarrow{w} \mu$. Conversely, suppose $\mu_n \xrightarrow{w} \mu$ and let $(x_m)_{m \in \mathbb{N}}$ be a dense sequence in $X$. For all $\varepsilon > 0$, we have $X = \bigcup_{m \in \mathbb{N}} B_{\varepsilon}(x_m)$. Fix $\varepsilon > 0$ and a finite subset $M \subseteq \mathbb{N}$ such that $\mu(\bigcup_{m \in M} B_{\varepsilon}(x_m)) \geq 1 - \varepsilon$. The set $\mathcal{U}$ of all open sets of the form $U = B_{\varepsilon}(x_{m_1}) \cup \cdots \cup B_{\varepsilon}(x_{m_r})$ for $\{m_1, \ldots, m_r\} \subseteq M$ is finite, and so, by portmanteau, for $n$ sufficiently large, $\mu(U) \geq \mu(U) - \varepsilon$ for all $U \in \mathcal{U}$. Fix $A \in \mathcal{B}(X)$ and approximate $A$ by $U(A) := \bigcup_{m \in M : B_{\varepsilon}(x_m) \cap A \neq \emptyset} B_{\varepsilon}(x_m)$. If $x \in A \setminus U(A)$, then $x$ is not in any ball $B_{\varepsilon}(x_m)$, $m \in M$, and so $x$ is in a subset of $\mu$-measure less than $\varepsilon$. Therefore,

$$\mu(A) \leq \mu(U(A)) + \mu(A \setminus U(A)) \leq \mu_n(U(A)) + 2\varepsilon.$$

Since $U(A) \subseteq A^{2\varepsilon}$, then

$$\mu(A) \leq \mu_n(A^{2\varepsilon}) + 2\varepsilon.$$

Therefore, by asymmetric redefinition of the Lévy–Prokhorov metric, we have $\rho(\mu_n, \mu) \leq 2\varepsilon$. Since $\varepsilon > 0$ is arbitrary, this proves that $d(\mu_n, \mu) \to 0$.

We finally show that $\mathcal{M}_1(X)$ is separable. Let $(x_m)_{m \in \mathbb{N}}$ be a dense sequence in $X$. Consider the countable set of probability measures of the form $\nu = \sum_{m \in M} r_m \delta_{x_m}$ where $M \subseteq \mathbb{N}$ is finite and the $r_m$ are rational numbers such that $\sum_{m \in M} r_m = 1$. We want to approximate up to $\varepsilon > 0$ with respect to $\rho$ a measure $\mu \in \mathcal{M}_1(X)$ by a measure $\nu$ of this form. Fix $M$ such that $\mu(\bigcup_{m \in M} B_{\varepsilon}(x_m)) \geq 1 - \varepsilon$. By intersecting appropriately the balls $B_{\varepsilon}(x_m)$, we can find disjoint measurable subsets $B_m$ such that $x_m \in B_m \subseteq B_{\varepsilon}(x_m)$ for all $m \in M$ and such that $\bigcup_{m \in M} B_{\varepsilon}(x_m) = \bigsqcup_{m \in M} B_m$. Take $\nu = \sum_{m \in M} r_m \delta_{x_m}$ where the $r_m$ are chosen so that $\sum_{m \in M} |r_m - \mu(B_m)| \leq \varepsilon$. Let $A \in \mathcal{B}(X)$ and define $V(A) = \bigsqcup_{m \in M : B_m \cap A \neq \emptyset} B_m$. Then

$$\mu(A) \leq \mu(V(A)) + \mu(A \setminus V(A)) \leq \mu(V(A)) + \varepsilon.$$

Since $V(A) \subseteq A^{\varepsilon}$ and $|\mu(V(A)) - \nu(\nu(A))| \leq \sum_{m \in M : B_m \cap A \neq \emptyset} |\mu(B_m) - r_m| \leq \varepsilon$, we have

$$\mu(A) \leq \nu(V(A)) + 2\varepsilon \leq \nu(A^{2\varepsilon}) + 2\varepsilon.$$

By definition, this implies that $\rho(\mu, \nu) \leq 2\varepsilon$, and so we found a countable dense subset of $\mathcal{M}_1(X)$. This concludes the proof. □

*Reference.* P.1.25. in Méliot.

**Lemma 7** ($X$ compact implies $\mathcal{M}_1(X)$ compact)**.** *Let $X$ be a metrizable compact topological space. Then the space $\mathcal{M}_1(X)$ of all Borel probability measures on $X$ endowed with the topology of weak convergence is metrizable compact.*

*Proof.* We apply the Banach–Alaoglu theorem to the space $C(X)$ of continuous functionals on $X$. By Riesz's representation theorem, we can identify $\mathcal{M}_f^{\pm}(X) = C(X)^*$. Then the Banach–Alaoglu theorem guarantees that the ball $B^*$ in $\mathcal{M}_f^{\pm}(X)$ is compact. Moreover, $\mathcal{M}_1(X)$ is a subset of $B^*$ since $\|f\|_{\infty} \leq 1$ and $\mu \in \mathcal{M}_1(X)$ imply $|\mu(f)| \leq 1$. To show that $\mathcal{M}_1(X)$ is compact, it thus suffices to show that $\mathcal{M}_1(X)$ is closed. We have that $\mathcal{M}_1(X)$ is the intersection of

$$\{\mu \in \mathcal{M}(X) : \mu(1) = 1\}$$

and

$$\{\mu \in \mathcal{M}(X) : \mu(f) \geq 0 \text{ for all } f \in C(X) \text{ with } f \geq 0\},$$

which are closed with respect to the weak* topology as inverse images of closed sets in $\mathbb{R}$ under weak* continuous maps. It follows that $\mathcal{M}_1(X)$ is closed and hence compact. Since $X$ is metrizable compact, it is separable. By Lemma 6, it follows that $\mathcal{M}_1(X)$ is also metrizable (and separable), and hence metrizable compact. □

*Reference.* P.1.26. in Méliot.

**Lemma 8** (Separable Metric space embedded in Hilbert cube (part of Urysohn metrization theorem))**.** *Let $X$ be a separable metric space. Then there exists a compact metric space $Y$ and a map $T\colon X \to Y$ such that $T$ is a homeomorphism from $X$ onto $T(Y)$.*

*Proof.* Let $Y := [0,1]^{\mathbb{N}} = \{(\xi_i)_{i\in\mathbb{N}} : \xi_i \in [0,1] \text{ for all } i \in \mathbb{N}\}$. Since $[0,1]$ is compact, $Y$ is compact for the product topology by Tychonoff's theorem, and it is metrized by $d(\xi,\eta) = \sum_{i=1}^{\infty} 2^{-i}|\xi_i - \eta_i|$ for all $\xi,\eta \in Y$. Let $(x_i)_{i\in\mathbb{N}}$ be a dense sequence in $X$ and define $\alpha_i(x) := \min\{d(x,x_i),1\}$ for all $x \in X$ and all $i \in \mathbb{N}$. Then define the map $T\colon X \to Y$ by

$$T(x) := (\alpha_i)_{i\in\mathbb{N}} = (\min\{d(x,x_i),1\})_{i\in\mathbb{N}}.$$

The map $T$ is 1-Lipschitz since

$$\begin{aligned} d(T(x),T(y)) &= \sum_{i=1}^{\infty} 2^{-i}|\min\{d(x,x_i),1\} - \min\{d(y,x_i),1\}| \\ &\leq \sum_{i=1}^{\infty} 2^{-i}|d(x,x_i) - d(y,x_i)| \leq \sum_{i=1}^{\infty} 2^{-i} d(x,y) = d(x,y). \end{aligned}$$

In particular, the map $T$ is continuous. We now show that for any $C \subseteq X$ closed and $x \notin C$, there is $\varepsilon > 0$ and $i$ such that $\alpha_i(x) \leq \varepsilon/3$ and $\alpha_i(y) \geq 2\varepsilon/3$ for all $y \in C$. Take $\varepsilon := \min\{d(x,C),1\} \in (0,1]$ and $i$ such that $d(x,x_i) \leq \varepsilon/3$. Then $\alpha_i(x) \leq \varepsilon/3$ and for any $y \in C$,

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

In particular, if $x \neq y$, then by regularity of $X$ and the previous claim, there exists $i$ such that $\alpha_i(x) \neq \alpha_i(y)$. This proves that $T$ is injective, and so bijective on its image $T(X)$. We finally show that for any $(y_n)_{n\in\mathbb{N}}$ and $y$ in $X$, we have $y_n \to y$ if and only if $T(y_n) \to T(y)$. If $y_n \to y$, then $\alpha_i(y_n) \to \alpha_i(y)$ for all $i \in \mathbb{N}$ by continuity, and so $d(T(y_n),T(y)) \to 0$ as $n \to \infty$ (by the Moore–Osgood theorem for iterated limits). Conversely, suppose that $y_n \not\to y$. Then there is a subsequence such that $y \notin \overline{\{y_{n_1}, y_{n_2}, \dots\}}$. Then by the claim there is $i$ such that $\alpha_i(y) \leq \varepsilon/3$ and $\alpha_i(y_{n_k}) \geq 2\varepsilon/3$ for all $k$, so that $\alpha_i(y_{n_k}) \not\to \alpha_i(y)$ as $k \to \infty$, and hence $T(y_{n_k}) \not\to T(y)$. This concludes the proof. □

*Reference.* L.5.4. in van Gaans's notes p.16.

**Addendum. Definitions and use of Prokhorov's theorem.**

**Definition 9** (Tight (Family of) Measures)**.** A collection $M \subseteq \mathcal{M}_f(E)$ of finite measures on $(E,\mathcal{B}(E))$ is said to be **tight** if for all $\varepsilon > 0$, there exists a compact set $K \subseteq E$ such that

$$\sup\{\mu(E \setminus K) : \mu \in M\} < \varepsilon.$$

*Remark.* Prokhorov's theorem is of fundamental importance in probability theory as it can be used to: first, (a) determine weak convergence of a tight sequence since it suffices then to prove that any convergent subsequence has the same limit (see L.1.20. in Méliot's notes where 1. is verified due to sequential compactness); and then (b) to construct a measure with given properties as limit of a tight sequence of measures whose convergent subsequences have same limit.

**Lemma 10** (L.1.20. in Méliot)**.** *Let $X$ be a metric space and $(x_n)_{n\in\mathbb{N}}$ a sequence in $X$. Suppose $(x_n)_{n\in\mathbb{N}}$ satisfies the two properties:*

1. *every sub-sequence of $(x_n)_{n\in\mathbb{N}}$ has a convergent sub-subsequence;*
2. *if $(x_{n_k})_{k\in\mathbb{N}}$ is a convergent sub-sequence of $(x_n)_{n\in\mathbb{N}}$, then its limit is $x$.*

*Then $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence and its limit is $x$.*

*Proof.* Suppose $x_n \not\to x$. Then there exist $\varepsilon > 0$ and a sub-sequence $(x_{n_k})_{k\in\mathbb{N}}$ such that for all $k \in \mathbb{N}$, $d(x_{n_k}, x) \geq \varepsilon$. By assumption, there exists a convergent sub-sub-sequence $(x_{n_{k_l}})_{l\in BN}$ of $(x_{n_k})_{k\in\mathbb{N}}$ whose limit we denote $y$. Then $d(y, x) \geq \varepsilon$. This contradicts the fact that every convergent sub-sequence of $(x_n)_{n\in\mathbb{N}}$ has limit $x$. □

## References

Billingsley, P. (2013): *Convergence of probability measures*. John Wiley & Sons.

Bühler, T., and D. A. Salamon (2010): "Functional analysis," *Lecture notes*.

Cerny, J. (2016): "Advanced probability theory," *Lecture notes*.

Kallenberg, O. (2021): *Foundations of modern probability*. Springer.

Meliot, P.-L. (2016): "Convergence de mesures, processus de Poisson et processus de Lévy," *Lecture notes*.

Teschl, G. (2019): "Topics in real and functional analysis," *Unpublished lecture notes*.

van Gaans, O. (2003): "Probability measures on metric spaces," *Lecture notes*.