7. Topology in the spaces of measures

Let us fix a separable metrizable space X. Recall that for $\mu \in P(X)$ and $g \in C_b(X)$ we write $\mu(g) = \int_X g \, d\mu$ for simplicity. We shall endow $P(X)$ with a topology corresponding to convergence of sequences of measures. Write

$$N_\mu(g, \varepsilon) = \{\nu \in P(X) : |\nu(g) - \mu(g)| < \varepsilon\}.$$

We declare $N_\mu(g, \varepsilon)$ to be open in $P(X)$ (whenever $g \in C_b(X)$). We introduce the topology by taking all finite intersection of such sets as a base. Hence, basic open sets in $P(X)$ containing $\mu \in P(X)$ are of the form

$$N_\mu(g_1, \ldots, g_n, \varepsilon) = \bigcap_{k=1}^{n} N_\mu(g_i, \varepsilon), \text{ for } g_k \in C_b(X) \text{ and } \varepsilon > 0\}.$$

It should be clear that $\mu_n \longrightarrow \mu$ (in the sense of the previous section) if and only if the sequence of μ_n converges to μ in the topological spaces (every neighbourhood of μ contains almost all μ_n's). This topology (of weak convergence) is the only one we consider on $P(X)$. *The topology of weak convergence* is a traditional name in probability; as we mentioned, the terminology of functional analysis is different.

Our first aim is to show that $P(X)$ is metrizable; then we shall check that it is also separable. We start with a certain technical but useful fact on converging sequences.

Lemma 7.1. *Let $\mathcal{A} \subseteq Bor(X)$ be a family closed under taking finite intersections and such that every open $U \subseteq X$ is a countable union of sets from $\mathcal{A}$. For $\mu_n, \mu \in P(X)$,*

$$\textit{if } \lim_n \mu_n(A) = \mu(A) \textit{ for every } A \in \mathcal{A} \textit{ then } \mu_n \longrightarrow \mu.$$

Proof. Note that for $A, B \in \mathcal{A}$, we have $A \cap B \in \mathcal{A}$, so (by the assumption on convergence)

$$\mu_n(A \cup B) = \mu_n(A) + \mu_n(B) - \mu_n(A \cap B) \longrightarrow \mu(A) + \mu(B) - \mu(A \cap B) = \mu(A \cup B).$$

It follows that μ_n converge to μ on any finite union of sets from $\mathcal{A}$.

Take open $U \subseteq X$. We know that $U = \bigcup_k A_k$ for some $A_k \in \mathcal{A}$. Given $\varepsilon > 0$, there is p such that writing $B = \bigcup_{k=1}^p A_k$ we have $\mu(B) > \mu(U) - \varepsilon$. Hence

$$\mu(U) - \varepsilon < \mu(B) = \lim_n \mu_n(B) \le \liminf_n \mu_n(U),$$

so we can apply Theorem 6.5(iii). □

Corollary 7.2. *If X is zerodimensional (i.e. it has a base consisting of clopen sets) then $\mu_n \longrightarrow \mu$ in $P(X)$ if and only if $\lim_n \mu_n(C) = \mu(C)$ for every $C \in \mathrm{clop(X)}$.*

Proof. The condition is necessary because clopen sets have empty boundary (see Theorem 6.5). For the sufficiency, we apply Lemma 7.1 for $\mathcal{A} = \mathrm{clop(X)}$. □

Lemma 7.3. *For any separable metrizable X there is a countable family $\Phi \subseteq C_b(X)$ testing the convergence of sequences of measures in $P(X)$ in the following sense:*

$$\text{if } \lim_n \mu_n(\varphi) = \mu(\varphi) \text{ for every } \varphi \in \Phi \text{ then } \mu_n \longrightarrow \mu.$$

Proof. We first check the following.

Claim. There is a countable family $\Phi \subseteq C_b(X)$ of functions $X \to [0,1]$ such that for every open $U \subseteq X$ there is a nondecreasing sequence of $\varphi_k \in \Phi$ converging to χ_U pointwise.

Take a compatible metric ρ on X. Given $x \in X$ and $p < q$ define $\varphi(x,p,q)$ by

$$\varphi(x,p,q)(y) = \frac{\rho(y, X \setminus B(x,q))}{\rho(y, B(x,p)) + \rho(y, X \setminus B(x,q))}.$$

This is a formula for a function that is constant 1 on $B(x,p)$ and 0 on the complement of the bigger ball $B(x,q)$. Now we take the family of all such functions $\varphi(x,p,q)$ where x runs some countable dense set $D \subseteq X$ and $p,q \in \mathbb{Q}$. This is a countable family and we close it under taking maxima to get the family Φ.

Then Φ is as required: if $U \subseteq X$ is open and $z \in X$ then there is $x \in D$ and $p < q$ such that $z \in B(x,p)$ while $B(x,q) \subseteq U$. This means that χ_U is the supremum of some sequence from Φ; we closed Φ under taking maxima to get the required nondecreasing sequence.

Now, if $\lim_n \mu_n(\varphi) = \mu(\varphi)$ for every $\varphi \in \Phi$ then taking open $U \subseteq X$ and $\varepsilon > 0$, we have $\varphi_1 \le \varphi_2 \le \ldots \longrightarrow \chi_U$ for some $\varphi_k \in \Phi$ so

$$\mu(U) - \varepsilon < \mu(\varphi_k) < \mu_n(\varphi_k) + \varepsilon < \mu_n(U) + \varepsilon,$$

for large k (by the dominated convergence theorem) and n (by convergence on Φ). The line above gives $\mu(U) \le \liminf_n \mu_n(U)$, so $\mu_n \longrightarrow \mu$ by Theorem 6.5. □

Corollary 7.4. *The space $P(X)$ is metrizable for every metrizable and separable X.*

Proof. This follows from 7.3: We fix an enumeration $\Phi = \{\varphi_k : k \in \mathbb{N}\}$ and define a metric σ on $P(X)$ as

$$\sigma(\mu,\nu) = \sum_k 2^{-k} |\mu(\varphi_k) - \nu(\varphi_k)|.$$

The condition $\sigma(\mu,\mu_n) \to 0$ is equivalent to $\mu_n(\varphi) \to \mu(\varphi)$ for every $\varphi \in \Phi$ so $\mu_n \longrightarrow \mu$.

The fact that σ is a metric on $P(X)$ compatible with the topology of $P(X)$. follows from two observations. Note first that the ball $\{\nu \in P(X) : \sigma(\mu.\nu) < r\}$ is open because the condition is defined by a continuous function (on the right hand side of the definition of σ). On the other hand, the convergence with respect to the metric σ is equivalent to the convergence in the topological sense. □

Recall that $\delta_x \in P(X)$ is the Diract measure at $x \in X$. Write $\Delta_X = \{\delta_x : x \in X\}$. Note that conv$\Delta_X$, the convex hull of Δ_X, is the family of all measures supported by finite sets.

Theorem 7.5. *If X is separable and metrizable then so is $P(X)$.*

Proof. We only need to check separability of $P(X)$. We first prove that $\mathrm{conv}\Delta_X$ is dense in $P(X)$, it meets every basic open set

$$N_\mu(g_1,\ldots,g_n,\varepsilon)=\bigcap_{k=1}^{n}N_\mu(g_i,\varepsilon).$$

Note that if $g:X\to\mathbb{R}$ is a bounded continuous function then for $\varepsilon>0$ there is a finite partition of X into Borel sets such that the oscillation of g is $<\varepsilon$ on every piece. By simple induction, for a given finite family $g_1,\ldots,g_k\in C_B(X)$ there is a partition of X into Borel sets $B_j, j\le p$, such that $\mathrm{osc}_{B_j}(g_i)<\varepsilon$ for every $i\le k$ and $j\le p$. Pick $x_j\in B_j$ and consider

$$\nu=\sum_{j=1}^{p}\mu(B_j)\delta_{x_j}\in\mathrm{conv}\Delta_X.$$

For any $i\le k$, we have

$$|\nu(g_i)-\mu(g_i)|\le\sum_{j=1}^{p}\left|\int_{B_j}g_i\,\mathrm{d}\nu-\int_{B_j}g_i\,\mathrm{d}\mu\right|\le$$

$$\le\sum_{j=1}^{p}|\mu(B_j)g_i(x_j)-\int_{B_j}g_i\,\mathrm{d}\mu|\le\sum_{j=1}^{p}\int_{B_j}|g(x_j)-g(t)|\,\mathrm{d}\mu(t)<\varepsilon,$$

so, indeed, $\nu\in\mathrm{conv}\Delta_X\cap N_\mu(g_1,\ldots,g_k,\varepsilon)$.

Now if we take a countable dense set $D\subseteq X$; the obvious candidate for a dense subset of $P(X)$ is $\mathrm{conv}^{\mathbb{Q}}\Delta_D$, of all convex rational combinations of measures $\delta_x, x\in D$. Note that the set is indeed countable; it is dense in $P(X)$ since it is clearly dense in $\mathrm{conv}\Delta_X$. □

We shall check that many properties of X are carried over to $P(X)$ and vice versa.

Lemma 7.6. *The mapping $\delta:X\to P(X)$, $x\longrightarrow\delta_x$ is a homeomorphic embedding onto a closed set $\Delta_X\subseteq P(X)$.*

Proof. We already know that $\lim_n x_n=x$ is equivalent to $\delta_{x_n}\longrightarrow\delta_x$, so we check that Δ_X is closed. Note that, as $P(X)$ is metrizable, it is enough to check that if $\delta_{x_n}\longrightarrow\mu$ then $\mu\in\Delta_X$. If x_n has a converging subsequence $(x_{n_k})_k$ to $x\in X$ then, automatically, $\mu=\delta_x\in\Delta_X$.

Suppose that no subsequence of x_n' converges. Then for every x there is open $U_x\subseteq X$ containing x and such that $x_n\in U_x$ only for finitely many n's. Then

$$\mu(U_x)\le\liminf_n\delta_{x_n}(U_x)=0.$$

Hence X is covered by open sets of μ-measure zero. It follows (?!) that, $\mu(X)=0$, which is really a contradiction. □

Theorem 7.7. *The space $P(X)$ is compact if and only if X is compact.*

Proof. The forward implication follows immediately from Lemma 7.6.

Suppose that X is compact; since we know that $P(X)$ is a metrizable space, it is enough to check the sequential characterization of compactness, that every sequence has a converging subsequence.

Fix $\mu_n \in P(X)$. We know that $C(X)$ is separable so we can fix a sequence of g_k, of norm-one continuous functons which is uniformly dense in $C(X)$.

We use the diagonal principle to find an infinite $N \subseteq \mathbb{N}$ such that the limit $\lim_{n\in N} \mu_n(g_k)$ exists for every k (any doubts?— see the problem list). Then the limit $\lim_{n\in N} \mu_n(g)$ exists for every $g \in C(X)$; this is simple: For $\varepsilon > 0$ there is k such that $\|g - g_k\|_\infty < \varepsilon$ and so

$$|\mu_n(g) - \mu_m(g)|| \le |\mu_n(g_k) - \mu_m(g_k)| + |\mu_n(g - g_k)| + |\mu_m(g - g_k)| \le 3\varepsilon,$$

for large $n, m \in N$.

Finally, we define a functional φ on $C(X)$ by the formula $\varphi(g) = \lim_{n\in N} \mu_n)(g)$. This is clearly an additive positive norm-one functional so by the Riesz representation theorem it is represented by some measure $\mu \in P(X)$. We have $\mu_n \longrightarrow \mu$ as $n \in N$ which means that our subsequence has a limit in $P(X)$, and we are done. □

Remark 7.8. If Y is a subspace of X then we may treat $P(Y)$ as a subspace of $P(X)$. Simply, every measure $\mu \in P(Y)$ defines $\widetilde{\mu} \in P(X)$ by the formula $\widetilde{\mu}(B) = \mu(B \cap X)$, for $B \in Bor(X)$. To be sure that $P(Y)$ indeed becomes a subspace of $P(X)$ we need to check that the topology of $P(Y)$ agrees with the topology inherited from $P(X)$ — see the problem list.

Theorem 7.9. *The space $P(X)$ is Polish if and only if X is Polish.*

Proof. The forward implication follows again from Lemma 7.6 since the closed subspace of a Polish space is Polish.

Assume that X is Polish. Then X embeds into $[0,1]^{\mathbb{N}}$ as a G_δ subspace (by the Alexandrov theorem). In the sequel, we simply assume that X **is** such a subspace of the Hilbert cube. Write K for the closure of X in $[0,1]^{\mathbb{N}}$. Then $P(K)$ is compact (hence Polish), and it is enough to check that $P(X)$ is a G_δ-subspace of $P(K)$, because we conclude that $P(X)$ is Polish using Alexandrov again.

Write $X = \bigcap_k G_k$, where $G_k \subseteq [0,1]^{\mathbb{N}}$ are open. Then, using Remark 7.8,

$$P(X) = \bigcap_k \{\widetilde{\mu} \in P(K) : \widetilde{\mu}(K \setminus G_k) = 0\} = \bigcap_k \bigcap_r \{\widetilde{\mu} \in P(K) : \widetilde{\mu}(K \setminus G_k) < 1/r\}.$$

Now it remains to check that a set of the form $\{\nu \in P(K) : \nu(F) < a\}$ is open whenever $F \subseteq K$ is closed. In turn, its complement $\{\nu \in P(K) : \nu(F) \ge a\}$ is closed by Theorem 6.5(ii). □

It is useful to recognize compact subsets of $P(X)$ in case X is not compact; this is a subject of Prokhorov's[1] theorem below. A subset $M \subseteq P(X)$ is called *relatively compact* if its closure in $P(X)$ is compact; this is equivalent to saying that every sequence of $\mu_n \in M$ has a subsequence converging to some $\mu \in P(X)$.

Recall that every measure $\mu \in P(X)$ is tight whenever X is Polish (so for any $\varepsilon > 0$ there is a compact subset $K \subseteq X$ such that $\mu(K) > 1 - \varepsilon$). Tightness has its uniform version.

[1] or Prohorov; Prochorow in Polish tradition

Definition 7.10. If X is any separable metrizable space then the set $M \subseteq P(X)$ is said to be **uniformly tight** if for any $\varepsilon > 0$ there is a compact subset $K \subseteq X$ such that $\mu(K) > 1-\varepsilon$ for every $\mu \in M$.

Theorem 7.11 (Prokhorov). *Let X be any separable metrizable space and let $M \subseteq P(X)$.*
(a) If M is uniformly tight then M is relatively compact.
(b) If X is Polish and M is relatively compact then M is uniformly tight.

Proof. For (a) we again embed X into the Hilbert cube and write K for the closure of X in $[0,1]^{\mathbb{N}}$. Take $\mu_n \in P(X)$; since $P(K)$ is compact, there is a converging subsequence to some $\mu \in P(K)$. We only need to check that, in fact, $\mu \in P(X)$ and for this we need uniform tightness. For $\varepsilon > 0$ there is compact $L \subseteq X$ such that $\mu_n(L) > 1-\varepsilon$ for every n. The crucial point: Since L is compact, it is closed in K, so $\mu(L) \geq \limsup_n \mu_n(L) \geq 1-\varepsilon$ by Theorem 6.5(ii). In particular, $\mu(X) \geq 1-\varepsilon$ for every $\varepsilon > 0$. We get $\mu(X) = 1$, as required.

Recall again that every measure on a Polish space is tight, see Theorem 4.4; we shall augment its proof to check (b). Below I copy the previous argument and mark the additional think that we need.

Take a compatible complete metric ρ on X; we write $B(x,r)$ for the open ball $\{y \in X : \rho(x,y) < r\}$. Let $D \subseteq X$ be a countable dense set

Fix $\varepsilon > 0$; for every n we have $X = \bigcup_{d \in D} B(d, 1/n)$ (as D is dense). Hence there is a finite $I_n \subseteq D$ such that

$$\mu(\bigcup_{d \in I_n} B(d,1/n)) > 1 - \varepsilon/2^n.$$

for every $\mu \in M$. Put

$$F_n = \overline{\bigcup_{d \in I_n} B(d,1/n)},$$

and let $K = \bigcap_n F_n$. We have $\mu(X \setminus F_n) \leq \varepsilon/2^n$, so

$$\mu(X \setminus K) \leq \sum_n \varepsilon/2^n = \varepsilon.$$

It remains to check that K is compact. This follows from the fact that K is a closed set in a complete metric space (X,ρ), and K is totally bounded (for every $\delta > 0$, K has a finite δ-net). This is clear in view of the very definition of K.

Hence we need to verify that for any open cover $X = \bigcup_k V_k$ there is $p \in \mathbb{N}$ such that writing $U_p = \bigcup_{k=1}^p V_k$ we have $\mu(U_p) > 1-\varepsilon$ for every $\mu \in M$.

Suppose otherwise, that there is $\varepsilon > 0$ such that for every p there is $\mu_p \in M$ with $\mu_p(U_p) \leq 1-\varepsilon$. Since M is relatively compact, passing to a subsequence, we can assume that $\mu_p \longrightarrow \mu \in P(X)$. We get $\mu(U_p) \leq 1-\varepsilon$ (Theorem 6.5 again) for every p. But this means that $\mu(X) = \lim_p \mu(U_p) \leq 1-\varepsilon$, a contradiction. □

Part (b) of Prokhorov's theorem fails badly without Polishness, even for countable domains. By a clever result due to David Preiss, there is a relatively compact set $M \subseteq P(\mathbb{Q})$ with is not uniformly tight.[2]

[2]see Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 27 (1973), 109–116