Appendix A

Space of Probability Measures

The space of all probability measures on a given sample space can be equipped with multiple topologies and metrics and corresponding measurable structures. In this appendix we study foremost the weak topology and its Borel σ-field, and next turn attention to a number of different metrics.

Throughout $\mathfrak{M}$ stands for the set of all probability measures on a "sample space" $(\mathfrak{X}, \mathscr{X})$.

A.1 Borel Sigma Field

The *Borel σ-field $\mathscr{X}$* on a topological space $\mathfrak{X}$ is defined as the smallest σ-field making all open subsets of $\mathfrak{X}$ measurable. To avoid complications, we restrict to separable metric spaces $(\mathfrak{X}, d)$, unless otherwise specified, and often to a *Polish space*, a space with a topology that is generated by a metric that renders it complete and separable. The following result gives alternative characterizations of $\mathscr{X}$; its proof can be found in standard texts.

Theorem A.1 (Borel sigma field) *The Borel σ-field $\mathscr{X}$ on a separable metric space $\mathfrak{X}$ is also*

(i) *the smallest σ-field containing a generator $\mathscr{G}$ for the topology on $\mathfrak{X}$;*
(ii) *the smallest σ-field containing containing all balls of rational radius with center located in a countable dense subset of $\mathfrak{X}$;*
(iii) *the smallest σ-field making all continuous, real-valued functions on $\mathfrak{X}$ measurable.*

A *Borel isomorphism* between two measurable spaces is a one-to-one, onto, bimeasurable map; that is, the map and its inverse are both measurable. A measurable space is called a standard Borel space if it is Borel isomorphic to a Polish space. The cardinality of such a set is at most c, the cardinality of $\mathbb{R}$. A remarkable theorem in descriptive set theory says that any uncountable standard Borel space is Borel isomorphic to $\mathbb{R}$ (or equivalently the unit interval); see Theorem 3.3.13 of Srivastava (1998).

A.2 Weak Topology

The *weak topology* on $\mathfrak{M}$ is defined by the convergence of all nets $\int \psi\, P_\alpha \to \int \psi\, dP$, for ψ ranging over $\mathfrak{C}_b(\mathfrak{X})$, the space of all bounded continuous real-valued functions on $\mathfrak{X}$. Equivalently, it is the weakest topology that makes all maps $P \mapsto \int \psi\, dP$ continuous.

We denote the weak topology on $\mathfrak{M}$ by $\mathscr{W}$ and the weak convergence of a net P_α on $\mathfrak{M}$ to a $P \in \mathfrak{M}$ by $P_\alpha \rightsquigarrow P$.

The *portmanteau theorem* characterizes weak convergence in several different ways.

Theorem A.2 (Portmanteau) *The following statements are equivalent for any net P_α in $\mathfrak{M}$ and $P \in \mathfrak{M}$:*

(i) *P_α converges weakly to P;*
(ii) *$\int \psi\, dP_\alpha \to \int \psi\, dP$ for all bounded uniformly continuous $\psi: \mathfrak{X} \to \mathbb{R}$;*
(iii) *$\int \psi\, dP_\alpha \to \int \psi\, dP$ for all bounded continuous $\psi: \mathfrak{X} \to \mathbb{R}$ with compact support;*
(iv) *$\int \psi\, dP_\alpha \to \int \psi\, dP$ for all bounded Lipschitz continuous $\psi: \mathfrak{X} \to \mathbb{R}$;*
(v) *$\limsup_\alpha P_\alpha(F) \le P(F)$ for every closed subset F;*
(vi) *$\liminf_\alpha P_\alpha(G) \ge P(G)$ for every open subset G;*
(vii) *$\lim_\alpha P_\alpha(A) = P(A)$ for every Borel subset A with $P(\partial A) = 0$[1] where $\partial A = \bar{A} \cap \bar{A^c}$ stands for the topological boundary of A.*

If $\mathfrak{X}$ is a Euclidean space, then these statements are also equivalent to, for F_α and F the distribution functions of P_α and P,

(viii) *$F_\alpha(x) \to F(x)$ at all x such that $P(\{y: y_i \le x_i$ for all $i,\ y_i = x_i$ for some $i\}) = 0$.*

A base for the neighborhood system of the weak topology at a given $P_0 \in \mathfrak{M}$ is given by open sets of the form $\{P: |\int \psi_i\, dP - \int \psi_i\, dP_0| < \epsilon, i = 1, \ldots, k\}$, where $0 \le \psi_i \le 1$, $i = 1, \ldots, k$, are continuous functions, $k \in \mathbb{N}$ and $\epsilon > 0$. It follows that a subbase for the neighborhood system is given by $\{P: \int \psi\, dP < \int \psi\, dP_0 + \epsilon\}$, where $0 \le \psi \le 1$ is a continuous function. These neighborhoods are convex, and hence the weak topology is locally convex, that is, the weak topology admits a convex base for the neighborhood system.

The following theorem shows that the weak topology has simple and desirable topological properties.

Theorem A.3 (Weak topology) *The weak topology $\mathscr{W}$ on the set $\mathfrak{M}$ of Borel measures on a separable metric space $\mathfrak{X}$ is metrizable and separable. Furthermore,*

(i) *$\mathfrak{M}$ is complete if and only if $\mathfrak{X}$ is complete.*
(ii) *$\mathfrak{M}$ is Polish if and only if $\mathfrak{X}$ is Polish.*
(iii) *$\mathfrak{M}$ is compact if and only if $\mathfrak{X}$ is compact.*

The set of measures $\sum_{i=1}^k w_i \delta_{x_i}$ for $(w_i) \in (\mathbb{Q}^+)^k$, with $\sum_{i=1}^k w_i = 1$ and $k \in \mathbb{N}$, and (x_i) ranging over a countable dense subset of $\mathfrak{X}$ is dense in $\mathfrak{M}$.

Two metrics inducing $\mathscr{W}$ are the *Lévy-Prohorov distance* and *bounded Lipschitz distance*, given by

$$d_L(P, Q) = \inf\Big\{\epsilon > 0: P(A) < Q(A^\epsilon) + \epsilon,\ Q(A) < P(A^\epsilon) + \epsilon\Big\}, \tag{A.1}$$

$$d_{BL}(P, Q) = \sup_{\|f\|_{\mathfrak{C}^1} \le 1} \Big|\int f\, dP - \int f\, dQ\Big|, \tag{A.2}$$

[1] Such a set is known as a P-continuity set.

where $A^\epsilon = \{y: d(x, y) < \epsilon \text{ for some } x \in A\}$ and the supremum is over all functions $f: \mathfrak{X} \to [-1, 1]$ with $|f(x) - f(y)| \le d(x, y)$ for all $x, y \in \mathfrak{X}$. The bounded Lipschitz distance can also be defined on the set of all signed Borel measures, and then is induced by the *bounded Lipschitz norm* $\|P\| = \sup\{|\int f\, dP|: \|f\|_{\mathfrak{C}^1} \le 1\}$. Under this norm the signed Borel measures form a Banach space, with the set of all bounded Lipschitz functions as its dual space. This justifies the name *weak topology*.[2]

Prohorov's theorem characterizes the weakly compact subsets of $\mathfrak{M}$. A subset $\Gamma \subset \mathfrak{M}$ is called *tight* if, given any $\epsilon > 0$, there exists a compact subset K_ϵ of $\mathfrak{X}$ such that $P(K_\epsilon) \ge 1 - \epsilon$ for every $P \in \Gamma$.

Theorem A.4 (Prohorov) *If $\mathfrak{X}$ is Polish, then $\Gamma \subset \mathfrak{M}$ is pre-compact if and only if Γ is tight.*

Let $\mathscr{M}$ be the Borel σ-field on $\mathfrak{M}$ generated by the weak topology $\mathscr{W}$. If $\mathfrak{X}$ is Polish, then this is also the σ-field generated by the evaluation maps or "projections" $P \mapsto P(A)$.

Proposition A.5 (Borel sigma field weak topology) *If $\mathfrak{X}$ is Polish, and $\mathscr{X}_0$ is a generator of $\mathscr{X}$, then the Borel σ-field $\mathscr{M}$ on $\mathfrak{M}$ for the weak topology is also*

(i) *the smallest σ-field on $\mathfrak{M}$ making all maps $P \mapsto P(A)$ measurable, for $A \in \mathscr{X}$;*
(ii) *the smallest σ-field on $\mathfrak{M}$ making all maps $P \mapsto P(A)$ measurable, for $A \in \mathscr{X}_0$;*
(iii) *the smallest σ-field on $\mathfrak{M}$ making all maps $P \mapsto \int \psi\, dP$ measurable, for $\psi \in \mathfrak{C}_b(\mathfrak{X})$.*

Consequently, a finite measure on $(\mathfrak{M}, \mathscr{M})$ is completely determined by the set of distributions induced under the maps (a) or (b) or (c) given by

(a) $P \mapsto (P(A_1), \dots, P(A_k))$, *for $A_1, \dots, A_k \in \mathscr{X}_0$ and $k \in \mathbb{N}$;*
(b) $P \mapsto (P(A_1), \dots, P(A_k))$, *for every partition $A_1, \dots, A_k$ of $\mathfrak{X}$ in sets in $\mathscr{X}$ and $k \in \mathbb{N}$;*
(c) $P \mapsto \int \psi\, dP$, *for $\psi \in \mathfrak{C}_b(\mathfrak{X})$.*

Proof The maps $P \mapsto \int \psi\, dP$ in (iii) are $\mathscr{M}$-measurable, because weakly continuous, and hence generate a σ-field smaller than the weak Borel σ-field $\mathscr{M}$. Conversely, these maps generate the weak topology, so that finite intersections of sets of the form $\{P: \int \psi\, dP < c\}$ are a basis for the weakly open sets. Because the weak topology is Polish, every open set is a countable union of these basis sets, and hence is contained in the σ-field generated by the maps in (iii). Then so is $\mathscr{M}$, and the proof of (iii) is complete.

For (i) we first show that the map $P \mapsto P(C)$ is $\mathscr{M}$-measurable for every closed subset C of $\mathfrak{X}$. Because the set $C_n = \{y: d(x, y) < n^{-1} \text{ for some } x \in C\}$ is open, for $n \in \mathbb{N}$, by Urysohn's lemma there exists a continuous function $0 \le \psi_n \le 1$ such that $\psi_n(x) = 1$ for all $x \in C$ and $\psi_n(x) = 0$ for all $x \notin C_n$. Thus $\mathbb{1}\{x \in C\} \le \psi_n(x) \le \mathbb{1}\{x \in C_n\}$, and hence $P(C) \le \int \psi_n\, dP \le P(C_n) \to P(C)$, for any $P \in \mathfrak{M}$. Thus $P \mapsto P(C)$ is the limit of the sequence of $P \mapsto \int \psi_n\, dP$, and hence is $\mathscr{M}$-measurable by (iii).

[2] However, this name was used before this justification was found, and some authors prefer weak*-convergence over weak convergence, considering $\mathfrak{M}$ as embedded in the dual space of $\mathfrak{C}_b(\mathfrak{X})$.

Now the collection of sets $\{A \in \mathscr{X}: P \mapsto P(A)$ is measurable$\}$ can be checked to be a Λ-system.[3] As it contains the Π-system[4] of all closed sets, it must be equal to $\mathscr{X}$, by Dynkin's Π-Λ theorem. Thus *all* evaluation maps are $\mathscr{M}$-measurable.

Conversely, let $\bar{\mathscr{M}}$ be the smallest σ-field on $\mathfrak{M}$ making all maps $\{P \mapsto P(A),\ A \in \mathscr{X}\}$ measurable. For every continuous function $\psi: \mathfrak{X} \to \mathbb{R}$ there exists a sequence of simple functions with $0 \le \psi_n \uparrow \psi$. Clearly $P \mapsto \int \psi_n \, dP$ is $\bar{\mathscr{M}}$-measurable for every n, and so $P \mapsto \int \psi \, dP = \lim_{n\to\infty} \int \psi_n \, dP$ is also $\bar{\mathscr{M}}$-measurable. This implies that $\mathscr{M} \subset \bar{\mathscr{M}}$, in view of (iii).

The equivalence of (i) and (ii) follows from standard measure theoretic arguments using the good sets principle.

It remains to prove the assertions (a)–(c). The induced distribution of $(P(A_1), \ldots, P(A_k))$ determines the finite measure on the σ-algebra generated by the maps $P \mapsto P(A_i)$, for $i = 1, \ldots, k$. Thus the collection of all induced distributions of this type determines the measure on the union of all these σ-fields, which is a field, that generates $\mathscr{M}$ by (ii). Assertion (a) follows because fields are measure-determining. Next assertion (b) follows, because the coordinates of vectors in (a) can be written as finite sums over measures of partitions as in (b). The proof of (c) is similar to the proof of (a), after we first note that the set of *univariate* distributions of maps of the type $\int \sum_i a_i \psi_i \, dP$ for a given finite set $\psi_1, \ldots, \psi_k$ determines the joint distribution of the vector $(\int \psi_1 \, dP, \ldots, \int \psi_k \, dP)$. ☐

By Proposition A.5 a measurable function $P: (\Omega, \mathscr{A}) \to (\mathfrak{M}, \mathscr{M})$ from some measurable space $(\Omega, \mathscr{A})$ in $(\mathfrak{M}, \mathscr{M})$ is a Markov kernel from Ω to $\mathfrak{X}$: a map $P: \Omega \times \mathscr{X} \to [0, 1]$ such that, for every $A \in \mathscr{X}$, the map $\omega \mapsto P(\omega, A)$ is measurable and, for every $\omega \in \Omega$, the map $A \mapsto P(\omega, A)$ is a probability measure on $\mathfrak{X}$. If $(\Omega, \mathscr{A})$ is equipped with a probability measure, then the induced probability measure of the random measure P is called a *random probability distribution* on $(\mathfrak{M}, \mathscr{M})$.

The *expectation measure* $\mu(A) = \int P(A) \, d\Pi(P)$ of a given random probability distribution Π on $(\mathfrak{M}, \mathscr{M})$ is a well-defined probability measure on $(\mathfrak{X}, \mathscr{X})$. The tightness of a family of probability measures is equivalent to the tightness of the corresponding collection of expectation measures.

Theorem A.6 *A family of distributions $\{\Pi_\lambda: \lambda \in \Lambda\}$ on $(\mathfrak{M}, \mathscr{M})$ is tight if and only if the corresponding family $\{\mu_\lambda: \lambda \in \Lambda\}$ of expectation measures on $(\mathfrak{X}, \mathscr{X})$ is tight.*

Proof Fix $\epsilon > 0$. If $\{\Pi_\lambda: \lambda \in \Lambda\}$ is tight, there exists a compact set $\Gamma \subset \mathfrak{M}$ such that $\Pi_\lambda(\Gamma^c) < \epsilon/2$ for all $\lambda \in \Lambda$. By Prohorov's theorem, there exists a compact set $K \subset \mathfrak{X}$ such that $P(K^c) < \epsilon/2$ for all $P \in \Gamma$. Then, for all $\lambda \in \Lambda$,

$$\mu_\lambda(K^c) = \left(\int_{\Gamma^c} + \int_{\Gamma}\right) P(K^c) \, d\Pi_\lambda(P) \le \Pi_\lambda(\Gamma^c) + \sup_{P\in\Gamma} P(K^c) < \epsilon.$$

Conversely, given $\epsilon > 0$ and $m \in \mathbb{N}$, let K_m be a compact subset of $\mathfrak{X}$ such that $\mu_\lambda(K_m^c) < 6\epsilon\pi^{-2}m^{-3}$ for all $\lambda \in \Lambda$. Define closed sets $\Gamma_m = \{P \in \mathfrak{M}: P(K_m^c) \le m^{-1}\}$, and set

[3] A class of sets closed under countable disjoint union and proper differencing.
[4] A class of sets closed under finite intersection.

$\Gamma = \cap_{m=1}^{\infty} \Gamma_m$. If $\eta > 0$ is given and $m > \eta^{-1}$, then $P(K_m^c) < \eta$ for every $P \in \Gamma$. Therefore Γ is a tight family, and closed. By Prohorov's theorem it is compact. Now for all $\lambda \in \Lambda$,

$$\Pi_\lambda(\Gamma^c) \le \sum_{m=1}^{\infty} \Pi_\lambda(\Gamma_m^c) \le \sum_{m=1}^{\infty} m\mu_\lambda(K_m^c) \le 6\epsilon\pi^{-2} \sum_{m=1}^{\infty} m^{-2} = \epsilon.$$

Here the second inequality follows, because $\Pi_\lambda(P\colon P(K_m^c) > m^{-1}) \le m\mathrm{E}P(K_m^c)$. ☐

Many interesting subsets of $\mathfrak{M}$ or maps on this space are $\mathscr{M}$-measurable. The following results give some examples.

Proposition A.7 *The following sets are contained in $\mathscr{M}$:*

(i) *The set $\{P \in \mathfrak{M}\colon P\{x\colon P(\{x\}) > 0\} = 1\}$ of all discrete probability measures.*
(ii) *The set $\{P \in \mathfrak{M}\colon P\{x\colon P(\{x\}) > 0\} = 0\}$ of all atomless probability measures.*

Proof The set $\mathcal{E}$ as defined in Lemma A.8 (ii) is measurable in the product σ-field and hence the map $P \mapsto \int \mathbb{1}\{(P, x) \in \mathcal{E}\}\, dP(x) = P\{x\colon P(\{x\}) > 0\}$ is measurable by Lemma A.8 (i). The sets in (i) and (ii) are the sets where this map is 1 and 0, respectively. ☐

Lemma A.8 (i) *If $f\colon \mathfrak{M} \times \mathfrak{X} \to \mathbb{R}$ is $\mathscr{M} \otimes \mathscr{X}$-measurable and integrable, then the map $P \mapsto \int f(P, x)\, dP(x)$ is $\mathscr{M}$-measurable.*
(ii) *The set $\mathcal{E} = \{(P, x) \in \mathfrak{M} \times \mathfrak{X}\colon P(\{x\}) > 0\}$ is $\mathscr{M} \otimes \mathscr{X}$-measurable.*

Proof (i). For $f = \mathbb{1}_{\mathcal{U}}$ for a set $\mathcal{U}$ in $\mathscr{M} \otimes \mathscr{X}$, we have $\int f(P, x)\, dP(x) = P(\mathcal{U}_P)$, for $\mathcal{U}_P := \{x\colon (P, x) \in \mathcal{U}\}$ the P-cross section of $\mathcal{U}$. It can be checked that $\mathscr{G} = \{\mathcal{U} \in \mathscr{M} \otimes \mathscr{X}\colon P \mapsto P(\mathcal{U}_P)$ is $\mathscr{M}$-measurable$\}$ is a Λ-class. It contains the Π-class of all product type sets $\mathcal{U} = C \times A$, with $C \in \mathscr{M}$, $A \in \mathscr{X}$, since for a set of this type $P(\mathcal{U}_P) = P(A)\mathbb{1}\{P \in C\}$ is a measurable function of P, by Proposition A.5. By Dynkin's theorem $\mathscr{G}$ is equal to the product σ-field $\mathscr{M} \otimes \mathscr{X}$.

It follows that the lemma is true for every f of the form $f = \mathbb{1}_{\mathcal{U}}$, for $\mathcal{U}$ in $\mathscr{M} \otimes \mathscr{X}$. By routine arguments the result extends to simple functions, nonnegative measurable functions, and finally to all jointly measurable functions.

(ii). It suffices to show that $(P, x) \mapsto P(\{x\})$ is measurable. The class of sets $\{F \in \mathscr{X} \otimes \mathscr{X}\colon (P, x) \mapsto P(F_x)$ is measurable$\}$ can be checked to be a Λ-class of subsets of $\mathfrak{X} \times \mathfrak{X}$. It contains the Π-system of all measurable rectangles of $\mathfrak{X} \times \mathfrak{X}$, which generates the product σ-field. Hence by Dynkin's theorem this class is equal to $\mathscr{X} \otimes \mathscr{X}$. The result follows, because the diagonal $\{(x, x)\colon x \in \mathfrak{X}\}$, is a product measurable subset. ☐

Proposition A.9 *The map $\psi\colon \mathfrak{M}(\mathbb{R}) \times \mathbb{R} \to \mathfrak{M}(\mathbb{R})$ defined by $\psi(P, \theta) = P(\cdot - \theta)$ is continuous. Furthermore, the restriction of ψ to $\mathfrak{F} \times \mathbb{R}$ for $\mathfrak{F} \subset \mathfrak{M}(\mathbb{R})$ a set of probability measures with a unique qth quantile at a given number a is one-to-one, and has an inverse ψ^{-1} which is also continuous. In particular, the map $P(\cdot - \theta) \mapsto \theta$ is continuous.*

Proof For $\theta_n \to \theta$ and a bounded uniformly continuous function $\phi\colon \mathbb{R} \to \mathbb{R}$, the sequence $\phi(x + \theta_n)$ converges uniformly in x to $\phi(x + \theta)$. Hence if $P_n \rightsquigarrow P$, then

$$\int \phi(x)\,dP_n(x-\theta_n) = \int \phi(x+\theta_n)\,dP_n(x) \to \int \phi(x+\theta)\,dP(x) = \int \phi(x)\,dP(x-\theta).$$

This proves the continuity of ψ.

If $P_n(\cdot - \theta_n) \rightsquigarrow P(\cdot - \theta)$ and every P_n and P have the same unique qth quantile, then the unique qth quantile $a + \theta_n$ of $P_n(\cdot - \theta_n)$ converges to the unique qth quantile $a + \theta$ of $P(\cdot - \theta)$. Thus $\theta_n \to \theta$, and for any uniformly continuous function ϕ,

$$\int \phi\,dP_n = \int \phi(x-\theta_n)\,dP_n(x-\theta_n) \to \int \phi(x-\theta)\,dP(x-\theta) = \int \phi\,dP.$$

We conclude that also $P_n \rightsquigarrow P$, whence ψ^{-1} is continuous. $\square$

A.3 Other Topologies

Next to the weak topology some other topologies on $\mathfrak{M}$ are of interest. The *total variation distance* is defined as

$$d_{TV}(P, Q) = \|P - Q\|_{TV} = \sup_{A \in \mathscr{X}} |P(A) - Q(A)|. \tag{A.3}$$

Convergence $P_n \to P$ in the corresponding topology is equivalent to the uniform convergence of $P_n(A)$ to $P(A)$, for A ranging over all Borel sets in $\mathfrak{X}$. The supremum in the definition of d_{TV} does not change if it is restricted to a field that generates $\mathscr{X}$, as the probability of any set can be approximately closely by the probability of a set from the field. In particular, it can be restricted to a countable collection. It may also be checked that

$$d_{TV}(P, Q) = \sup\Big\{\int \psi\,dP - \int \psi\,dQ\colon\ \psi\colon \mathfrak{X} \to [0, 1], \text{measurable}\Big\}. \tag{A.4}$$

The space $\mathfrak{M}$ is complete under the total variation metric, but it is separable only if $\mathfrak{X}$ is countable, since $d_{TV}(\delta_x, \delta_y) = 1$ for all $x \neq y$.

This nonseparability causes that the Borel σ-field relative to d_{TV}, generated by all d_{TV}-open subsets, is not equal to the ball σ-field on $\mathfrak{M}$, generated by all open d_{TV}-balls. In fact, in general the d_{TV}-ball σ-field is strictly smaller than $\mathscr{M}$, which in turn is strictly smaller than the d_{TV}-Borel σ-field (Problems A.2 and A.3). These properties make the total variation topology on the whole of $\mathfrak{M}$ somewhat clumsy. However, when restricted to subsets of $\mathfrak{M}$ the total variation topology may be separable, and the Borel σ-field more manageable. In particular, this is true for the set of all probability measures that are absolutely continuous with respect to a given measure ν.

Proposition A.10 (Domination) *The subset $\mathfrak{M}_0 \subset \mathfrak{M}$ of all probability measures on a standard Borel space that are absolutely continuous relative to a given σ-finite measure is complete and separable relative to the total variation metric. The trace of the d_{TV}-Borel σ-field on $\mathfrak{M}_0$ coincides with the trace of $\mathscr{M}$.*

Proof The first assertion follows, because $\mathfrak{M}_0$ is isomorphic to $\mathbb{L}_1(\nu)$, for ν the dominating measure, which is well known to be complete and separable. See Appendix B for details.

Because the supremum in the definition of d_{TV} can be restricted to a countable set, and every set $\{P\colon |P(A) - Q(A)| < \epsilon\}$ is contained in $\mathfrak{M}$ by Proposition A.5, every open total variation ball is contained in $\mathscr{M}$. In a separable metric space any open set is a countable union of open balls. Hence every (relatively) d_{TV}-open set in $\mathfrak{M}_0$ is contained in $\mathscr{M} \cap \mathfrak{M}_0$. ☐

For a Euclidean space $\mathfrak{X} = \mathbb{R}^k$ another topology on $\mathfrak{M}$ is given by the uniform convergence of the corresponding cumulative distribution functions (c.d.f.) $F(x) = P((-\infty, x])$, metrized by the familiar *Kolmogorov-Smirnov distance*

$$d_{KS}(P, Q) = \sup_{x \in \mathbb{R}} |P((-\infty, x]) - Q((-\infty, x])|. \tag{A.5}$$

This topology is intermediate between the weak and total variation topologies. Like the total variation distance, d_{KS} makes $\mathfrak{M}$ complete, but not separable (as $d_{KS}(\delta_x, \delta_y) = 1$ for all $x \neq y$, as before). Its main appeal is its importance in the Glivenko-Cantelli theorem. On the subspace of continuous probability measures the Kolmogorov-Smirnov and weak topologies coincide. This is a consequence of *Pólya's theorem.*

Proposition A.11 (Pólya) *Any sequence of distribution functions that converge weakly to a continuous distribution function converges also in the Kolmogorov-Smirnov metric. Consequently any d_{KS}-open neighborhood of an atomless $P_0 \in \mathfrak{M}(\mathbb{R}^k)$ contains a weak neighborhood of P_0.*

A fourth topology, also intermediate between the weak and total variation topologies, is the *topology of setwise convergence*: the weakest topology such that the maps $A \mapsto P(A)$, for $A \in \mathscr{X}$ are continuous. In this topology a net P_α converges to a limit P if and only if $P_\alpha(A) \to P(A)$, for all $A \in \mathscr{X}$. This topology has the inconvenience of being not metrizable (see Problem A.4). It is inherited from the product topology on $[0, 1]^{\mathscr{X}}$ after embedding $\mathfrak{M}$ through the map $P \mapsto (P(A): A \in \mathscr{X})$.[5] A second inconvenience is that $\mathfrak{M}$ is not measurable as a subset of $[0, 1]^{\mathscr{X}}$ relative to the product σ-field, since every product measurable subset is determined by only countably many coordinates. These disadvantages disappear by restricting to a countable collection $\{A_i : i \in \mathbb{N}\}$ of sets. The pointwise convergence $P_\alpha(A_i) \to P(A_i)$ for every $i \in \mathbb{N}$ is equivalent to convergence for the semimetric

$$d(P, Q) = \sum_{i=1}^{\infty} 2^{-i} |P(A_i) - Q(A_i)|.$$

For a sufficiently rich collection of sets, this will be stronger than weak convergence. For instance, all intervals in $\mathfrak{X} = \mathbb{R}^k$ with rational endpoints, or all finite unions of balls around points in a countable dense set of radii from a countable dense set in a general space $\mathfrak{X}$.

A.4 Support

Intuitively, the *support* of a measure on a measurable space $(\mathfrak{X}, \mathscr{X})$ is the smallest subset outside which no set gets positive measure. Unfortunately, such a smallest subset may not exist if $\mathfrak{X}$ is uncountable. For a topological space $\mathfrak{X}$ with its Borel σ-field a more fruitful notion of support of a measure μ is

[5] According to the *Vitale-Hahn-Sacks theorem*, $\mathfrak{M}$ is sequentially closed as a subset of $[0, 1]^{\mathscr{X}}$.

$$\mathrm{supp}(\mu) = \Big(\bigcup_{\substack{U \text{ open},\\ \mu(U)=0}} U\Big)^c = \{x \in \mathfrak{X}: \mu(U) > 0 \text{ for every open } U \ni x\}. \tag{A.6}$$

Thus the support is the residual closed set that remains after removing all open sets of measure zero. For a probability measure P, the definition can be rewritten as

$$\mathrm{supp}(P) = \bigcap_{\substack{C \text{ closed},\\ P(C)=1}} C = \{x \in \mathfrak{X}: P(U) > 0 \text{ for every open } U \ni x\}. \tag{A.7}$$

Clearly $\mathrm{supp}(P)$ is closed and no proper closed subset of it can have full probability content. For separable metric spaces the support itself also has probability one, and hence can be characterized as the *smallest closed subset of probability one*, or equivalently, as the complement of the largest open subset of probability zero. (This is true more generally if there exists a separable closed subset with probability one. For general topological spaces the support may fail to have full probability, and may even be empty.) In particular, the support of a prior distribution Π on the space $(\mathfrak{M}, \mathscr{M})$ of Borel probability measures on a separable metric space with the weak topology is given by

$$\mathrm{supp}(\Pi) = \bigcap_{\substack{\mathcal{C} \text{ weakly closed},\\ \Pi(\mathcal{C})=1}} \mathcal{C} = \{P \in \mathfrak{M}: \Pi(\mathcal{U}) > 0 \text{ for every open } \mathcal{U} \ni P\}. \tag{A.8}$$

A.5 Historical Notes

The results in this appendix are all well known, albeit not easy to find in one place. General references include Parthasarathy (2005), Billingsley (1968), van der Vaart and Wellner (1996) and Dudley (2002). In particular for a proof of the Portmanteau theorem, see Parthasarathy (2005), pages 40–42, or van der Vaart (1998), Lemma 2.2; for Prohorov's theorem, see Parthasarathy (2005), Theorem 6.7, or van der Vaart and Wellner (1996), Theorem 1.3.8; for metrizability and separability, see van der Vaart and Wellner (1996), Chapter 1.12, or Parthasarathy (2005), pages 39–52.

Problems

A.1 Show that $x \mapsto \delta_x$ is a homeomorphism between $\mathfrak{X}$ and the space of all degenerate probability measures under the weak topology. In particular, the correspondence is a Borel isomorphism.

A.2 Show that the trace of the σ-field generated by all total variation open sets on the space of all degenerate probability measures is the power set, and hence has cardinality bigger than $\mathbb{R}$. As $\mathfrak{M}$ is a Polish space under the weak topology, its cardinality equals that of $\mathbb{R}$. Conclude that the Borel σ-field corresponding to the total variation topology is strictly bigger than $\mathscr{M}$.

A.3 Show that the σ-field generated by all total variation open balls is strictly smaller than $\mathscr{M}$.

A.4 Consider the topology of setwise convergence on $\mathfrak{M}$.

(a) Find an uncountable collection of disjoint open sets under the topology. Thus $\mathfrak{M}$ is not separable under the topology induced by the setwise convergence.

(b) Show that a continuous probability measure P_0 does not have a countable base of its neighborhood system. Thus $\mathfrak{M}$ is neither separable nor metrizable.

A.5 Kuratowski's theorem implies that the inverse of a Borel measurable one-to-one map from standard Borel space onto a metrizable space is also measurable (see Corollary 4.5.5 of Srivastava 1998). Use this to give another proof of the fact that the strong and the weak topologies on the space of absolutely continuous probability measures with respect to a σ-finite measure ν generate the same σ-field.