PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 144, Number 7, July 2016, Pages 3123–3133
http://dx.doi.org/10.1090/proc/12932
Article electronically published on November 20, 2015

SKOROHOD'S REPRESENTATION THEOREM FOR SETS OF PROBABILITIES

MARTIN DUMAV AND MAXWELL B. STINCHCOMBE

(Communicated by David Asher Levin)

Abstract. We characterize sets of probabilities, $\mathbf{\Pi}$, on a measure space $(\Omega, \mathcal{F})$, with the following representation property: for every measurable set of Borel probabilities, A, on a complete separable metric space, (M, d), there exists a measurable $X : \Omega \to M$ with $A = \{X(P) : P \in \mathbf{\Pi}\}$. If $\mathbf{\Pi}$ has this representation property, then: if $K_n \to K_0$ is a sequence of compact sets of probabilities on M, there exists a sequence of measurable functions, $X_n : \Omega \to M$ such that $X_n(\mathbf{\Pi}) \equiv K_n$ and for all $P \in \mathbf{\Pi}$, $P(\{\omega : X_n(\omega) \to X_0(\omega)\}) = 1$; if the K_n are convex as well as compact, there exists a jointly measurable $(K, \omega) \mapsto H(K, \omega)$ such that $H(K_n, \mathbf{\Pi}) \equiv K_n$ and for all $P \in \mathbf{\Pi}$, $P(\{\omega : H(K_n, \omega) \to H(K_0, \omega)\}) = 1$.

1. Introduction

Throughout, (M, d) is a complete separable metric (Polish) space, $\mathcal{M}$ is its Borel σ-field, $\Delta(M)$ is the set of countably additive probabilities on $\mathcal{M}$, and $C_b(M)$ the continuous and bounded $\mathbb{R}$-valued functions on M. Let $\rho(\cdot, \cdot)$ be a metric on $\Delta(M)$ making $(\Delta(M), \rho)$ Polish and inducing the weak* topology, that is, $\rho(\mu_n, \mu_0) \to 0$ iff $\int u\, d\mu_n \to \int u\, d\mu_0$ for every $u \in C_b(M)$. Let $\mathcal{D}_{\mathcal{M}}$ denote the Borel σ-field on the set of probabilities $\Delta(M)$; equivalently $\mathcal{D}_{\mathcal{M}}$ is the minimal σ-field containing the sets $\{\mu : \mu(E) \le r\}$, $E \in \mathcal{M}$, $r \in [0, 1]$ (which follows from $\mu \mapsto \int f\, d\mu$ being $\mathcal{D}_{\mathcal{M}}$-measurable if f is measurable [4, Theorem III.62]).

Throughout, all probabilities are assumed countably additive. The set of probabilities on a measure space $(\Omega, \mathcal{F})$ is always given the minimal σ-field containing the sets $\{P : P(E) \le r\}$, $E \in \mathcal{F}$, and $r \in [0, 1]$. For any measurable mapping, f, between measure spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ and any probability Q on $(\Omega, \mathcal{F})$, the image law of Q under f is denoted $f(Q)$ and defined as the probability on $(\Omega', \mathcal{F}')$ giving mass $Q(f^{-1}(E'))$ to every $E' \in \mathcal{F}'$. If $\mathbf{\Pi}$ is a set of probabilities, then $f(\mathbf{\Pi})$ denotes $\{f(Q) : Q \in \mathbf{\Pi}\}$. This paper will give a set of necessary and sufficient conditions for a set of probabilities to have the following property.

Definition 1. A set of probabilities, $\mathbf{\Pi}$, on a measure space $(\Omega, \mathcal{F})$ is **descriptively complete** if for every Polish space (M, d) and every non-empty measurable $A \subset \Delta(M)$, there exists a measurable $X : \Omega \to M$ such that $X(\mathbf{\Pi}) = A$.

Received by the editors May 18, 2012 and, in revised form, July 16, 2014, January 27, 2015 and August 19, 2015.

2010 *Mathematics Subject Classification.* Primary 60B10, 60F99, 91B06.

Key words and phrases. Skorohod's representation theorem, strongly zero-one sets of probabilities, multiple prior models of choice.

Descriptive completeness is a set-valued version of the first part of Skorohod's representation theorem [10, Thm. 3.1.1]: if λ is Lebesgue measure (aka the uniform distribution) on $[0,1]$ with its usual Borel σ-field and $\rho(\mu_n, \mu_0) \to 0$ in $\Delta(M)$, then there exist measurable functions, $X_n, X_0 : [0,1] \to M$ such that

Sko(a) $X_n(\lambda) = \mu_n$, $X_0(\lambda) = \mu_0$, and
Sko(b) $\lambda(\{s \in [0,1] : X_n(s) \to X_0(s)\}) = 1$.

Blackwell and Dubins [2] give the following simultaneous Skorohod theorem:[1] There exists a jointly measurable $h : \Delta(M) \times [0,1] \to M$ such that

Bl-Du(a) for all $\mu \in \Delta(M)$, $h(\mu, \lambda) = \mu$, and
Bl-Du(b) if $\rho(\mu_n, \mu_0) \to 0$, then $\lambda(\{s \in [0,1] : h(\mu_n, s) \to h(\mu_0, s)\}) = 1$.

Taking $X_n(s) = h(\mu_n, s)$ in Bl-Du(a) and Bl-Du(b) recovers Sko(a) and Sko(b).

Since the composition of measurable functions is measurable, we can replace the probability λ on $[0,1]$ with any non-atomic P on any measure space $(\Omega, \mathcal{F})$: because P is non-atomic, there exists a measurable $f : \Omega \to [0,1]$ such that $f(P) = \lambda$; if $X : [0,1] \to M$ is measurable and $X(\lambda) = \mu$, then the composition $X \circ f$ is measurable and $X(f(P)) = \mu$; if $(\mu, s) \mapsto h(\mu, s)$ is jointly measurable and $h(\mu, \lambda) = \mu$, then $(\mu, \omega) \mapsto h(\mu, f(\omega))$ is jointly measurable and $h(\mu, f(P)) = \mu$.

A useful example of a descriptively complete set of probabilities has $\Omega^\circ = [0,1] \times [0,1]$, $\mathcal{F}^\circ$ its usual Borel σ-field, and $\mathbf{\Pi}^\circ = \{\lambda_r : r \in [0,1]\}$ where for each $r \in [0,1]$, λ_r denotes the uniform distribution on $\{r\} \times [0,1]$.

Lemma 1. $\mathbf{\Pi}^\circ$ *is descriptively complete.*

Throughout, we use the Borel isomorphism theorem: if B is a measurable subset of a Polish space (M, d) and B' a measurable subset of a Polish space (M', d'), then there is a measurable bijection with measurable inverse between B and B' iff they have the same cardinality (e.g. [4, Theorem III.20] or [5, Theorem 13.1.1]).

Proof. For any non-empty measurable $A \subset \Delta(M)$, the Borel isomorphism theorem guarantees the existence of a measurable onto function, $\psi_A : [0,1] \to A$. Let $h(\cdot, \cdot)$ be the Blackwell-Dubins function. The mapping $X(r, s) := h(\psi_A(r), s)$ is measurable, and for each $r \in [0,1]$, $X(\lambda_r) = \psi_A(r)$ so that $X(\mathbf{\Pi}^\circ) = A$. □

Lemma 1 contains a characterization of descriptive completeness. Because Ω° is a Polish space, a necessary condition for $\mathbf{\Pi}$ to be descriptively complete is the existence of a measurable $X : \Omega \to \Omega^\circ$ such that $X(\mathbf{\Pi}) = \mathbf{\Pi}^\circ$. Since the composition of measurable functions is measurable, by Lemma 1, the existence of such an X is also sufficient for $\mathbf{\Pi}$ to be descriptively complete. We will show that there exists a measurable X with $X(\mathbf{\Pi}) = \mathbf{\Pi}^\circ$ iff $\mathbf{\Pi}$ satisfies the following.

Definition 2. A set of probabilities, $\mathbf{\Pi}$, on a measure space $(\Omega, \mathcal{F})$ is **measurably mutually orthogonal and simultaneously Skorohod (mmosS)** if

(a) there exists a measurable, onto $d : \mathbf{\Pi} \to [0,1]$ and a measurable, onto $\varphi : \Omega \to [0,1]$ such that for all $r \in [0,1]$, for all $P \in d^{-1}(r)$, $P(\varphi^{-1}(r)) = 1$, and
(b) for every Polish space (M, d) and every $\mu \in \Delta(M)$, there exists a measurable $f : \Omega \to M$ such that for all $P \in \mathbf{\Pi}$, $f(P) = \mu$.

[1] After replacing filter (*filtre*) with sequence (*suite*), a slightly stronger statement and complete proof is in Théorème 1.3.2. in [6]

The condition in Definition 2(a) is a measurable version of the requirement that $\mathbf{\Pi}$ can be partitioned into sets of mutually orthogonal probabilities: if $P_r \in d^{-1}(r)$ and $P_s \in d^{-1}(s)$, $r \neq s$, then P_r and P_s are mutually orthogonal (have disjoint carriers) because $\varphi^{-1}(r)$ and $\varphi^{-1}(s)$ are disjoint. The condition in Definition 2(b) requires that each $P \in \mathbf{\Pi}$ be non-atomic, which guarantees that for every P and $\mu \in \Delta(M)$ there exists a measurable $f_P : \Omega \to M$ with $f_P(P) = \mu$. The simultaneity condition is that a single measurable f serve for all of the $P \in \mathbf{\Pi}$.

The weak* Hausdorff metric is given

$$d_H(A, B) = \inf\{\epsilon \geq 0 : (\forall \mu \in A)(\exists \nu \in B)[\rho(\mu, \nu) < \epsilon], \\ (\forall \nu \in B)(\exists \mu \in A)[\rho(\mu, \nu) < \epsilon]\}, \tag{1}$$

for $A, B \in \mathbf{K}(\Delta(M))$, the class of non-empty, compact subsets of $\Delta(M)$. Let $\mathbf{KCon}(\Delta(M))$ denote the d_H-closed subclass of non-empty, compact, and convex subsets of $\Delta(M)$. Restricted to the compact subsets, the Hausdorff metric is equivalent for equivalent metrics on $\Delta(M)$. Give $\mathbf{K}(\Delta(M))$ the Borel σ-field generated by the weak* Hausdorff metric, and give every product of measurable spaces the product σ-field.

Theorem 1. *A set of probabilities, $\mathbf{\Pi}$, on $(\Omega, \mathcal{F})$ is descriptively complete iff it is mmosS, and if $\mathbf{\Pi}$ is descriptively complete, then*

(a) *for any sequence K_n, K_0 of compact subsets of $\Delta(M)$ with $d_H(K_n, K_0) \to 0$, there exists a sequence of measurable functions, $X_n, X_0 : \Omega \to M$ such that $X_n(\mathbf{\Pi}) = K_n$, $X_0(\mathbf{\Pi}) = K_0$, and for all $P \in \mathbf{\Pi}$, $P(\{\omega : X_n(\omega) \to X_0(\omega)\}) = 1$, and*

(b) *there exists a jointly measurable $H : \mathbf{KCon}(\Delta(M)) \times \Omega \to M$ such that for all $K \in \mathbf{KCon}(\Delta(M))$, $H(K, \mathbf{\Pi}) = K$, and if $d_H(K_n, K_0) \to 0$, then for all $P \in \mathbf{\Pi}$, $P(\{\omega : H(K_n, \omega) \to H(K_0, \omega)\}) = 1$.*

Theorem 1(a) is the extension of Skorohod's representation theorem to compact sets of probabilities while Theorem 1(b) is the extension of the Blackwell-Dubins simultaneous Skorohod theorem to compact convex sets of probabilities. The next section gives the proof, the following discusses related work and extensions.

2. Proof

Theorem 1 concerns a representation result for measurable sets and two continuity results: the extension of the Skorohod continuity result to arbitrary compact sets of probabilities; and the extension of the simultaneous Skorohod result to compact and convex sets of probabilities. We cover these in turn.

Throughout, in a metric space (X, d), for $\epsilon > 0$ and $x \in X$, $B_\epsilon(x) := \{y \in X : d(x, y) < \epsilon\}$, is the open ball of radius ϵ around x. The following alternative notion of the convergence of sets will appear frequently below.

Definition 3. For a sequence F_n of closed subsets of a metric space (X, d), x is a **limit point of** F_n if for all $\epsilon > 0$, $\{n : B_\epsilon(x) \cap F_n \neq \emptyset\}$ has finite complement and x is an **accumulation point of** F_n if for all $\epsilon > 0$, $\{n : B_\epsilon(x) \cap F_n \neq \emptyset\}$ is infinite. The set of limit points is the **lower closed limit**, denoted $\mathrm{Li}(F_n)$, the set of accumulation points is the **upper closed limit**, denoted $\mathrm{Ls}(F_n)$. The sequence F_n converges to F_0 in the **Kuratowski-Painlevé** sense, written $F_0 = KP\text{-}\lim_n F_n$, if $\mathrm{Li}(F_n) = \mathrm{Ls}(F_n) = F_0$.

2.1. **Representation.** Since the composition of measurable functions is measurable, by Lemma 1, it is sufficient to show that $\mathbf{\Pi}$ is mmosS iff there exists a measurable $X : \Omega \to \Omega^\circ$ such that $X(\mathbf{\Pi}) = \mathbf{\Pi}^\circ$. Suppose first that $\mathbf{\Pi}$ is mmosS, let $f : \Omega \to [0,1]$ have the property that for all $P \in \mathbf{\Pi}$, $f(P) = \lambda$, and let φ and d be the functions given in Definition 2. Define $X(\omega) \in \Omega^\circ$ by $X(\omega) = (\varphi(\omega), f(\omega))$. Defining $\mathbf{\Pi}_r = d^{-1}(r)$, for $r \in [0,1]$ and all $P \in \mathbf{\Pi}_r$, $X(P) = \lambda_r$ so that $X(\mathbf{\Pi}) = \mathbf{\Pi}^\circ$.

Now suppose that there exists a measurable $X : \Omega \to \Omega^\circ$ such that $X(\mathbf{\Pi}) = \mathbf{\Pi}^\circ$. For each ω, $X(\omega)$ is a vector in $[0,1]\times[0,1]$, denoted $X(\omega) = (X_r(\omega), X_s(\omega))$. Define $\varphi(\omega) = X_r(\omega)$. We first verify the measurable mutual orthogonality. The mapping $Q \mapsto X(Q)$ from probabilities on $(\Omega, \mathcal{F})$ to probabilities on Ω° is measurable, and the mapping from $\mathbf{\Pi}^\circ$ to $[0,1]$ defined by $\gamma(\lambda_r) = r$ is measurable. For $P \in \mathbf{\Pi}$, define $d(P) = \gamma(X(P))$. If $d(P) = r$, then $P(X^{-1}(\{r\} \times [0,1])) = P(\varphi^{-1}(r)) = 1$. We now verify the simultaneous Skorohod property. Skorohod's representation theorem implies that for any $\mu \in \Delta(M)$, there exists a measurable $g : [0,1] \to M$ such that $g(\lambda) = \mu$. Define $f(\omega) = g(X_s(\omega))$. Since $X_s(P) = \lambda$ for all $P \in \mathbf{\Pi}$, $f(P) = \mu$ for all $P \in \mathbf{\Pi}$. $\square$

2.2. **Compact sets of probabilities.** We now turn to the proof of Theorem 1(a). Let K_n, K_0 be a sequence of compact subsets of $\Delta(M)$ with $d_H(K_n, K_0) \to 0$. Since the composition of measurable functions is measurable, it is sufficient to prove the result for $\mathbf{\Pi}^\circ$, that is, it is sufficient to show that there exists a sequence of measurable functions, $X_n, X_0 : \Omega^\circ \to M$ such that $X_n(\mathbf{\Pi}^\circ) = K_n$, $X_0(\mathbf{\Pi}^\circ) = K_0$, and for all $\lambda_r \in \mathbf{\Pi}^\circ$, $\lambda_r(\{(r,s) : X_n(r,s) \to X_0(r,s)\}) = 1$.

If $d_H(K_n, K_0) \to 0$, then $K_0 = \mathrm{Li}(K_n) = \mathrm{Ls}(K_n)$, i.e. $K_0 = KP\text{-}\lim_n K_n$.[2] Because $(\Delta(M), \rho)$ is complete, a sequence in $\Delta(M)$ converges iff it is Cauchy. Let $\mathcal{X} = \times_{n=1}^\infty K_n$ and let $\mathcal{X}^C$ denote the set of $x = (\mu_n)_{n=1}^\infty \in \mathcal{X}$ such that $n \mapsto \mu_n$ is a Cauchy sequence in $\Delta(M)$. Because $x = (\mu_n)_{n=1}^\infty$ being a Cauchy sequence puts no restriction on the value of any particular μ_n, $\mathrm{proj}_n\, \mathcal{X}^C = K_n$ for each $n \in \{1, 2, \ldots\}$. Therefore, since $K_0 = KP\text{-}\lim_n K_n$, K_0 is the set of limits of sequences in $\mathcal{X}^C$.

Because each K_n is compact, $\mathcal{X}$ is compact, hence Polish, when given the product topology. The set $\mathcal{X}^C$ is measurable,[3] hence by the Borel isomorphism theorem, there exists a measurable onto $\psi : [0,1] \to \mathcal{X}^C$. For $n \geq 1$, let $X_n(r,s) = h(\psi_n(r), s)$ where $h(\cdot,\cdot)$ is the Blackwell-Dubins function and $\psi_n(r) = \mathrm{proj}_n(\psi(r))$. For each n, $X_n(\lambda_r) = \psi_n(r)$ so that $X_n(\mathbf{\Pi}^\circ) = K_n$. For each r, $\lim_n \psi_n(r)$ exists because $\psi(r)$ is a Cauchy sequence. Let $\psi_0(r) = \lim_n \psi_n(r)$ and note that $\psi_0([0,1]) = K_0$ so that $X_0(r,s) := h(\psi_0(r), s)$ has the property that $X_0(\mathbf{\Pi}^\circ) = K_0$. By the continuity of the Blackwell-Dubins functions, Bl-Du(b), for each $\lambda_r \in \mathbf{\Pi}^\circ$, $\lambda_r(\{(r,s) : X_n(r,s) \to X_0(r,s)\}) = 1$. $\square$

2.3. **Compact convex sets of probabilities.** We now turn to Theorem 1(b). Let K_n belong to $\mathbf{KCon}(\Delta(M))$, the class of non-empty, compact, and convex subsets of $\Delta(M)$ and suppose that $d_H(K_n, K_0) \to 0$. As above, it is sufficient to prove the result for $\mathbf{\Pi}^\circ$. The following simultaneous retract result is useful.

Lemma 2. *There exists a jointly continuous $(K, \mu) \mapsto f_K(\mu)$ from $\mathbf{KCon}(\Delta(M)) \times \Delta(M)$ to $\Delta(M)$ such that for all (K, μ), $f_K(\mu) \in K$, and if $\mu \in K$, then $f_K(\mu) = \mu$.*

---

[2]If $x \in K_0$, then for all $\epsilon > 0$, $B_\epsilon(x) \cap K_n \neq \emptyset$ for all large n. If $x \notin K_0$, pick $\epsilon > 0$ so that $B_{2\cdot\epsilon}(x) \cap K_0 = \emptyset$. For n large enough that $d_H(K_n, K_0) < \epsilon$, $B_\epsilon(x) \cap K_n = \emptyset$.

[3]$\mathcal{X}^C = \cap_m \cup_N \cap_{j,k \geq N} G(m,j,k)$ where $G(m,j,k) = \{(\mu_n)_{n=1}^\infty \in \mathcal{X} : \rho(\mu_j, \mu_k) < 1/m\}$.

Proof. The following ℓ_2-based metric on $\Delta(M)$ induces weak* convergence and has strictly convex closed balls: let $\{u_i : i \in \mathbb{N}\}$ be a set of continuous functions with $|u_i(x)| \leq 1$ for all $x \in M$ such that μ_n weak* converges to μ_0 iff $\int u_i\, d\mu_n \to \int u_i\, d\mu_0$ for all $i \in \mathbb{N}$; define $\rho_2(\mu, \nu) = \left(\sum_i \frac{1}{2^i}(\int u_i\, d\mu - \int u_i\, d\nu)^2\right)^{\frac{1}{2}}$. For each $\mu \in \Delta(M)$ and each non-empty, compact, convex $K \subset \Delta(M)$, there exists a minimizer for the problem $\min_{\nu \in K} \rho_2(\mu, \nu)$ because K is compact and the distance function is continuous, and the minimizer is unique because K is convex and ρ_2 has strictly convex closed balls. Let $f_K(\mu)$ be that minimizer. □

Returning to the proof of Theorem 1(b), because $\Delta(M)$ is an uncountable Polish space, there exists a measurable bijection with measurable inverse, $\psi : [0,1] \leftrightarrow \Delta(M)$. Define $H(K, (r,s)) = h(f_K(\psi(r)), s)$ where $h(\cdot, \cdot)$ is the Blackwell-Dubins function. The joint measurability is clear, and for any $\lambda_r \in \mathbf{\Pi}^\circ$ and any $K \in \mathbf{KCon}(\Delta(M))$, $H(K, \lambda_r) = f_K(\psi(r))$ so $H(K, \mathbf{\Pi}^\circ) = K$. Finally, $d_H(K_n, K_0) \to 0$ implies that $f_{K_n}(\mu) \to f_{K_0}(\mu)$ for every μ, so $\lambda_r(\{(r,s) : H(K_n, (r,s)) \to H(K_0, (r,s))\}) = 1$. □

3. Related work and extensions

We first discuss how interest in both the representation and the continuity parts of Theorem 1 arose from models of choice in the presence of ambiguity. We then discuss the relation between descriptive completeness and a necessary and sufficient condition for the existence of consistent estimators in statistics. Finally, we turn to the possibilities of generalizing the continuity parts of Theorem 1 from compact sets to closed sets. This is related to the continuity of the cores of random closed sets and leads to a counter-example to [11, Theorem 1].

3.1. **On the relation with decision theory.** Decision theory in the face of uncertainty has two main models, related by change of variables, one due to von Neumann and Morgenstern [12], the other to Savage [9]. Both models use a space of consequences, usually a Polish space in applications, and one of them also has a measure space of states, $(\Omega, \mathcal{F})$.

Seventy years ago, von Neumann and Morgenstern [12, Ch. 3.6] gave a short axiomatic foundation for preferences over distributions on M. A preference, $\succsim$, on $\Delta(M)$ is a complete, transitive, binary relation on $\Delta(M)$. Preferences satisfying their axioms can be represented by $\mu \succsim \nu$ iff

$$\int_M u(x)\, d\mu(x) \geq \int_M u(x)\, d\nu(x), \tag{2}$$

where $u \in C_b(M)$ is unique up to positive affine transformations. A decade later, Savage's [9] work provided an axiomatic foundation for preference over measurable functions (random variables) from a state space, $(\Omega, \mathcal{F})$, to M. Preferences over measurable functions satisfying his axioms can be represented by $X \succsim Y$ iff

$$\int_\Omega u(X(\omega))\, dP(\omega) \geq \int_\Omega u(Y(\omega))\, dP(\omega), \tag{3}$$

where P is a unique non-atomic probability, often interpreted as a Bayesian prior distribution, and $u \in C_b(M)$ is unique up to positive affine transformations.

The approaches are directly related by change of variables, taking $\mu = X(P)$ and $\nu = Y(P)$, the integrals on both sides of the inequalities in (2) and (3) are the same. If M is a Polish space and P is non-atomic, then Sko(a) implies *domain equivalence*, i.e. the set of choice situations that can be modeled by the two approaches is the same, and the continuity result, Sko(b), implies that problems involving convergence in distribution $\mu_n \to \mu_0$ can be analyzed using the stronger convergence condition $X_n \to X_0$ a.e.

Over the last several decades, the systematic inability of either approach to explain behavior in the face of ambiguity, understood as partially known probabilities, has led many to replace Savage's single prior, P, with a set of priors, $\mathbf{\Pi}$ (see e.g. the monograph [7]). A well-studied class of preferences over functions is given by $X \succsim Y$ iff

$$\alpha \inf_{P \in \mathbf{\Pi}} \int u(X)\, dP(\omega) + (1-\alpha) \sup_{Q \in \mathbf{\Pi}} \int u(X)\, dQ(\omega) \geq \alpha \inf_{P \in \mathbf{\Pi}} \int u(Y)\, dP(\omega) + (1-\alpha) \sup_{Q \in \mathbf{\Pi}} \int u(Y)\, dQ(\omega). \tag{4}$$

The parallel von Neumann and Morgenstern approach works with preferences over the sets of probabilities, $X(\mathbf{\Pi})$ and $Y(\mathbf{\Pi})$. The motivating question for the present paper in the context of decision theoretic models is "What conditions on $\mathbf{\Pi}$ yield both domain equivalence and continuity?"

3.2. **On the relation with consistent estimators.** Breiman *et al.* [3] show that the following condition is necessary and sufficient for the existence of consistent estimators.

Definition 4. A set of probabilities $\mathbf{\Pi}$ on a measure space $(\Omega, \mathcal{F})$ is **strongly zero-one** if there exists a measurable $\Omega' \subset \Omega$ and an onto $\Phi : \Omega' \to \mathbf{\Pi}$ such that for all $P \in \mathbf{\Pi}$, $P(\varphi^{-1}(P)) = 1$.

If $\mathbf{\Pi}$ is mmosS and the function $d : \mathbf{\Pi} \to [0,1]$ in Definition 2(a) is one-to-one as well as onto, then $\mathbf{\Pi}$ is strongly zero-one. For any measurable E, the mapping $\omega \mapsto \Phi(\omega)(E)$ is measurable and $\mathcal{C} = \{E \in \mathcal{F} : \omega \mapsto \Phi(\omega)(E) \text{ is constant}\}$ is a sub-σ-field of $\mathcal{F}$. Since $\emptyset$ and Ω belong to $\mathcal{C}$, the set of constant values of the functions $\omega \mapsto \Phi(\omega)(E)$ always contains both 0 and 1. The simultaneous Skorohod condition of Definition 2(b) is equivalent to the set of constant values being all of $[0,1]$.

A canonical example illustrates the connection between strongly zero-one sets of probabilities and consistent estimators.

Example 1. Let $\Omega = \{0,1\}^{\mathbb{N}}$, let $\mathcal{F}$ be the product σ-field, and for each $n \in \mathbb{N}$, let $\mathcal{F}_n$ be the minimal sub-σ-field making the mapping $\omega \mapsto (\omega_1, \ldots, \omega_n)$ measurable. Define $\phi(\omega) = \liminf_n \frac{1}{n}\#\{k \leq n : \omega_k = 1\} \in [0,1]$. For each $r \in (0,1)$, let P_r be the distribution on Ω of an independent and identically distributed sequence of Bernoulli(r) random variables, set $\mathbf{\Pi} = \{P_r : r \in (0,1)\}$, and define $\Phi(\omega) = P_{\phi(\omega)}$. By the strong law of large numbers, for each r, $P_r(\phi^{-1}(r)) = 1$ so that $\mathbf{\Pi}$ is strongly zero-one. If ω is distributed according to one of the $P_r \in \mathbf{\Pi}$, interest centers on finding consistent estimators of r, that is, a sequence of $\mathcal{F}_n$-measurable functions $\widehat{r}_n : \Omega \to [0,1]$ such that for each $P_r \in \mathbf{\Pi}$, $P_r(\{\omega : \widehat{r}_n(\omega) \to r\}) = 1$. An obvious choice is $\widehat{r}_n = \frac{1}{n}\#\{k \leq n : \omega_k = 1\}$.

A measure space is **standard** if it is measurably isomorphic to an uncountable Borel measurable subset of a Polish space. It can be shown that every uncountable, measurable, strongly zero-one set of non-atomic probabilities on a standard space is descriptively complete. Thus, the set $\{P_r : r \in (0,1)\}$ in Example 1 is descriptively complete.

3.3. **Relaxing compactness to closedness.** The strategy of proof for Theorem 1(a) can be adapted to sequences of closed sets. There are, however, some subtle issues of the choice of topology on the class of closed sets that must be considered. It seems difficult to adapt the strategy of proof for Theorem 1(b) to closed convex sets. For a metric space (X,d), let $\mathbf{Cl}(X)$ denote the set of non-empty closed subsets of X. Of interest are the relations between $\mathbf{Cl}(M)$ and $\mathbf{Cl}(\Delta(M))$.

3.3.1. *Sequences of Closed Sets.* The following is a direct parallel to Theorem 1(a).

Corollary 1(a). *If $\mathbf{\Pi}$ is descriptively complete, then for any sequence F_n, F_0 of closed subsets of $\Delta(M)$ with $F_0 = KP\text{-}\lim_n F_n$, there exists a sequence of measurable functions, $X_n, X_0 : \Omega \to M$ such that $X_n(\mathbf{\Pi}) = F_n$, $X_0(\mathbf{\Pi}) = F_0$, and for all $P \in \mathbf{\Pi}$, $P(\{\omega : X_n(\omega) \to X_0(\omega)\}) = 1$.*

Proof. As before, it is sufficient to prove this for $\mathbf{\Pi}^\circ$. Let $\mathcal{X} = \times_{n=1}^\infty F_n$ and let $\mathcal{X}^C$ denote the set of $x = (\mu_n)_{n=1}^\infty \in \mathcal{X}$ such that $n \mapsto \mu_n$ is a Cauchy sequence. Being the product of Polish spaces, $\mathcal{X}$ is Polish when given the product topology, $\mathcal{X}^C$ is a measurable subset, hence there exists a measurable onto $\psi : [0,1] \to \mathcal{X}^C$. Define $X_n(r,s) = h(\psi_n(r), s)$ and $X_0(r,s) = h(\psi_0(r), s)$ where $\psi_0(r) := \lim_n \psi_n(r)$ and $h(\cdot,\cdot)$ is the Blackwell-Dubins function. $\square$

In the compact case, $K_0 = KP\text{-}\lim_n K_n$ if $d_H(K_n, K_0) \to 0$, so Theorem 1(a) is useful for problems in which compactness of the sets of probabilities and continuity with respect to the Hausdorff topology can be used. For Corollary 1(a) to have a similar utility, we need a topology on $\mathbf{Cl}(\Delta(M))$ yielding Kuratowski-Painlevé convergence of closed sets. However, Kuratowski-Painlevé convergence of closed subsets of (X,d) is not topological unless X is locally compact (e.g. [8]), and $\Delta(M)$ is locally compact in the weak* topology iff $\Delta(M)$ is compact,[4] in which case closed sets are compact and there is no extra value to Corollary 1(a). Alternative choices of topology include the Wijsman, the Fell, and the Vietoris topology, all three of which agree when $\Delta(M)$ is (locally) compact.

The Wijsman topology is Polish and can be metrized by the (generally non-complete) metric $\rho_W(F,F') = \sum_i \frac{1}{2^i} \min\{1, |\rho(\mu_i, F) - \rho(\mu_i, F')|\}$ where $\{\mu_i : i \in \mathbb{N}\}$ is a dense subset of $\Delta(M)$. Unlike the Hausdorff topology for compact sets, the Wijsman topology depends on the metric: metrics on $\Delta(M)$ with the same uniformity can lead to different Wijsman topologies; metrics leading to the same Wijsman topology can have different uniformities. However, if $\rho_W(F_n, F_0) \to 0$, then $F_0 = KP\text{-}\lim_n F_n$ (see [1, Sections 2.1, 2.2, and 2.5] for these). Thus, Corollary 1(a) is useful for problems in which closedness of the sets of probabilities and continuity with respect to one of the Wijsman topologies can be used.

The Vietoris and the Fell topologies are unsatisfactory for different reasons. A sequence F_n converges to F_0 in $\mathbf{Cl}(\Delta(M))$ in the Vietoris topology iff it converges in each Wijsman topology generated by a metric on $\Delta(M)$ inducing the weak* topology. This kind of convergence is extremely demanding, sequences that "should" converge in applications often do not when M, hence $\Delta(M)$, fails to be compact.

[4]If $\Delta(M)$ is locally compact, then for any $\mu \in M$, there exists an open neighborhood G_μ of μ with compact closure. For small enough $r > 0$, $A_r := ((1-r)p + r\Delta(M)) \subset G_\mu$. For the closure of A_r to be compact, $\Delta(M)$ must be compact.

While the Vietoris topology is larger (has more open sets) than the Wijsman topology, the Fell topology is smaller. The Fell topology is not metrizable: the sequential closure of a class of sets does not imply closure of the class; and Fell-continuity of functions on $\mathbf{Cl}(\Delta(M))$ is not determined by sequences of closed sets. This means that Fell continuity is not a good match with Corollary 1(a) even though, for sequences, not nets, if F_n converges in the Fell topology to F_0, then $F_0 = KP\text{-}\lim_n F_n$ as well (see [1, Section 5.1-2] for these).

3.3.2. *Sequences of closed convex sets.* The essential problem in extending the proof strategy used for the simultaneous Skorohod result, Theorem 1(b), to closed sets is finding a substitute for the simultaneous retract result given in Lemma 2. The continuous linear imbedding of $\Delta(M)$ in a Hilbert space works for compact convex sets because compactness survives the imbedding, and this guarantees that a distance minimizer exists. For closed convex sets of probabilities that imbed not only as *relatively* closed subsets of the range but as closed subsets of the entire Hilbert space, one can guarantee the existence of minimizers, but not all closed convex sets belong to this class. Our attempts at alternative proof strategies for Lemma 2 have foundered on this point, but we do not have a counterexample.

3.3.3. *The closed sets* $\Delta(B)$. For B a non-empty, measurable subset of M, the set of probabilities with $\mu(B) = 1$ is denoted $\Delta(B)$, and $\Delta(B)$ is a closed subset of $\Delta(M)$ iff B is a closed subset of M. In choice problems, knowing only that the probability belongs to $\Delta(B)$ corresponds to knowing that the true distribution concentrates on B but having no knowledge of the relative likelihoods of different subsets of B. When the B_n are compact, $B_0 = KP\text{-}\lim_n B_n$ iff $\Delta(B_n) = KP\text{-}\lim_n \Delta(B_n)$. The relation is more delicate when the B_n are closed but not necessarily compact.

If we replace the metric $d(x, y)$ by the equivalent metric $d'(x, y) = d(x, y)/(1 + d(x, y))$, the metric space (M, d') is also Polish and the Wijsman topology on $\mathbf{Cl}(M)$ is unchanged (see [1, Theorem 2.1.10]). Associated with the metric d' on M is the Prohorov metric ρ' on $\Delta(M)$, given by

$$\rho'(\mu, \nu) = \inf\{\epsilon > 0 : (\forall E \in \mathcal{X})[\mu(E) \leq \nu(E^\epsilon) + \epsilon]\}, \tag{5}$$

where $E^\epsilon := \cup_{x \in E}\{y \in X : d'(x, y) < \epsilon\}$ is the ϵ-enlargement of the set E using the metric $d'(\cdot, \cdot)$.

The proof of Lemma 3 will use the following observations: $\rho'(\mu_n, \mu) \to 0$ iff $\int u\, d\mu_n \to \int u\, d\mu$ for all bounded continuous $u : M \to \mathbb{R}$, that is, iff μ_n converges to μ in the weak* topology ([5, Theorem 11.3.3]); working with complements, one can show that $\mu(E) \leq \nu(E^\epsilon) + \epsilon$ for all measurable E iff $\nu(E) \leq \mu(E^\epsilon) + \epsilon$ for all measurable E ([5, Theorem 11.3.1]); $E^\epsilon = (\mathrm{cl}(E))^\epsilon$ so one can replace measurable sets with closed ones in (5); $(\Delta(M), \rho')$ is Polish because (M, d') is Polish ([5, Corollary 11.5.5]); for $x \in X$, letting $\delta_x(E) = 1_E(x) \in \Delta(X)$ denote point mass on x, $\rho'(\delta_x, \delta_y) = d'(x, y)$, and consequently, for $F \in \mathbf{Cl}(M)$, $\rho'(\delta_x, \Delta(F)) = d'(x, F)$; and finally, if $g : M \to M$ is a measurable function satisfying $d'(x, g(x)) < r$ for all $x \in M$, then for any $\mu \in \Delta(M)$, if $\nu = g(\mu)$, then $\rho'(\mu, \nu) \leq r$ because for all measurable E, $g^{-1}(E) \subset E^r$.

Let "$\to_{\tau_W(d')}$" denote convergence in the Wijsman topology on $\mathbf{Cl}(M)$ generated by the metric d', and "$\to_{\tau_W(\rho')}$" denote convergence in the Wijsman topology on

$\mathbf{Cl}(\Delta(M))$ generated by the metric ρ'. We have the following relation between these two modes of convergence.[5]

Lemma 3. *$\Delta(B_n) \to_{\tau_W(\rho')} \Delta(B_0)$ iff $B_n \to_{\tau_W(d')} B_0$.*

Proof. Suppose first that $\Delta(B_n) \to_{\tau_W(\rho')} \Delta(B_0)$. Pick an arbitrary $x \in X$ and let δ_x denote point mass on x. By assumption, $\rho'(\delta_x, \Delta(B_n)) \to \rho'(\delta_x, \Delta(B_0))$. This implies that $d'(x, B_n) \to d'(x, B_0)$, that is, $B_n \to_{\tau_W(d')} B_0$.

Now suppose that $B_n \to_{\tau_W(d')} B_0$ and fix an arbitrary $\mu \in \Delta(M)$. We show that

$$\limsup_n \rho'(\mu, \Delta(B_n)) \leq_{(1)} \rho'(\mu, \Delta(B_0)) \leq_{(2)} \liminf_n \rho'(\mu, \Delta(B_n)). \tag{6}$$

Let $r = \rho'(\mu, \Delta(B_0))$.

(1) Picking arbitrary $\epsilon > 0$, it is sufficient to show that $\limsup_n \rho'(\mu, \Delta(B_n)) < r + \epsilon$. The denseness of the compactly supported probabilities in $\Delta(B_0)$ implies that there is a compact $K \subset B_0$ and a $\gamma \in \Delta(B_0)$ with $\gamma(K) = 1$ such that $|r - \rho'(\mu, \gamma)| < \epsilon/2$. By the assumption that $B_n \to_{\tau_W(d)} B_0$, in particular, for every $x \in K$, $d'(x, B_n) \to d'(x, B_0)$. Because $K \subset B_0$, $d'(x, B_0) = 0$. The functions $x \mapsto d'(x, B_n)$ on K have Lipschitz constant 1, hence converge uniformly to 0 on K. Pick N such that for all $n \geq N$, $|d'(x, B_n)| < \epsilon/4$. By measurable selection, for each $n \geq N$, there exists a measurable function $g_n : K \to B_n$ such that $d'(x, g_n(x)) < \epsilon/2$ for all $x \in K$, and this can be extended to M by setting $g_n(x) = x$ for $x \notin K$. Letting $\gamma_n = g_n(\gamma)$, we have $\rho'(\gamma, \gamma_n) \leq \epsilon/2$. Since $\gamma_n \in \Delta(B_n)$, we have $\rho'(\mu, \Delta(B_0)) \leq \rho'(\mu, \gamma_n)$ and $|\rho'(\mu, \gamma_n) - r| < \epsilon/2 + \epsilon/2$, we have $\limsup \rho'(\mu, \Delta(B_n)) < r + \epsilon$.

(2) Picking arbitrary $\epsilon > 0$, it is sufficient to show that $\liminf_n \rho'(\mu, \Delta(B_n)) > r - \epsilon$. The denseness of the probabilities with finite support implies that there is no loss in assuming that μ is of the form $\sum_{i=1}^I \alpha_i \delta_{x_i}$. We first show that (a) there is a finitely supported $\gamma = \sum_{i=1}^I \alpha_i \delta_{y_i}$, $y_i \in B_0$, such that $|\rho'(\mu, \gamma) - r| < \epsilon/2$, and then show that (b) for large n, there exists $\gamma_n \in \Delta(B_n)$ such that $\rho'(\gamma, \gamma_n) < \epsilon/2$. These two imply that for large n, $\rho'(\mu, \Delta(B_n)) > r - (\epsilon/2 + \epsilon/2)$, completing the proof.

(a) The Ky Fan metric for measurable functions g, h mapping a probability space $(\Omega, \mathcal{F}, P)$ to (M, d) is $\alpha_{d'}(g, h) = \inf\{\delta > 0 : P(\{d'(f, g) > \delta\}) < \delta\}$. Strassen's Theorem [5, Corollary 11.6.4] connects the metric $\alpha_{d'}(\cdot, \cdot)$ to the metric $\rho'(\cdot, \cdot)$ as follows: if $(\Omega, \mathcal{F}, P)$ is non-atomic and (M, d) is Polish, then for any two probabilities $\mu, \nu \in \Delta(M)$, there exist measurable functions $g, h : \Omega \to M$ such that $g(P) = \mu$, $h(P) = \nu$, and $\alpha_{d'}(g, h) = \rho(\mu, \nu)$. If $g(P)$ is equal to μ, that is, $g(P) = \sum_{i=1}^I \alpha_i \delta_{x_i}$, then (ignoring sets of measure 0) there exists a measurable partition, $\{E_i : i = 1, \ldots, I\}$, of Ω such that $g(\omega) = \sum_{i=1}^I x_i 1_{E_i}(\omega)$ with $P(E_i) = \alpha_i$. Any measurable $h : \Omega \to M$ gives rise to the function $f_h : \Omega \to \mathbb{R}_+$ defined by $f_h(\omega) = \sum_i d'(x_i, h(\omega)) 1_{E_i}(\omega)$, and $\alpha_{d'}(g, h) = \inf\{\delta > 0 : P(\{f_h(\omega) > \delta)\}) < \delta\}$. If $h(\omega) \in B_0$ for all ω, then $0 \leq \sum_i d'(x_i, B_0) 1_{E_i}(\omega) \leq f_h(\omega)$. Picking $y_i \in B_0$ such that $|d'(x_i, y_i) - d'(x_i, B_0)| < \epsilon/2$, setting $h(\omega) = \sum_i y_i 1_{E_i}(\omega)$ and $\gamma = h(P)$ yields $\gamma \in \Delta(B_0)$ such that $|\rho'(\mu, \gamma) - r| < \epsilon/2$.

(b) By the assumption that $B_n \to_{\tau_W(d)} B_0$, there exists N such that for each $n \geq N$ and each y_i, $i = 1, \ldots, I$, $|d'(y_i, B_n) - d'(y_i, B_0)| < \epsilon/2$. Since $y_i \in B_0$, this implies that for all $n \geq N$, there exists $y_{n,i} \in B_n$ such that $d'(y_i, y_{n,i}) < \epsilon/2$. Define $\gamma_n \in \Delta(B_n)$ by $\gamma_n = \sum_i \alpha_i \delta_{y_{n,i}}$ such that $\rho'(\gamma, \gamma_n) < \epsilon/2$. □

[5]We thank Professor Pedro Terán for noticing that our earlier version of this lemma was not, in fact, "immediate," and that our result and [11, Theorem 1] could not simultaneously be true.

3.3.4. *On non-convergence in the Vietoris topology.* If $\omega \mapsto B(\omega)$ is a measurable mapping from a non-atomic probability space $(\Omega, \mathcal{F}, P)$ to $\mathbf{Cl}(M)$, its **closed core** can be defined as the closure of the set of $\xi(P)$ where $\xi : \Omega \to M$ is measurable and $P(\{\omega : \xi(\omega) \in B(\omega)\}) = 1$. Theorem 1 in [11] claims that if $P(\{\omega : B_n(\omega) \to_{\tau_W(d_0)} B_0(\omega)\}) = 1$, then the closed cores of the B_n converge, in the Vietoris topology for weak*-closed sets of probability measures in $\Delta(M)$, to the closed core of B_0. A sequence converges in the Vietoris topology on closed subsets of $\Delta(M)$ iff it converges in every Wijsman topology generated by every metric equivalent to the Prohorov metric.

In the special case that $B(\omega) = B$ for all $\omega \in \Omega$, the core of B is $\Delta(B)$. If [11, Theorem 1] were true, we could reach the much stronger conclusion in Lemma 3 that $B_n \to_{\tau_W(d')} B_0$ implies that $\Delta(B_n)$ converges to $\Delta(B_0)$ in the Vietoris topology for the weak*-closed subsets of $\Delta(M)$. The following variant of [1, Example 2.1.3] leads to a counter-example to that result.

Example 2. The metric space $M = \mathbb{N} = \{1, 2, \ldots\}$ is complete under the equivalent metrics $d_0(\cdot, \cdot)$ and $d_1(\cdot, \cdot)$ defined by

$$d_0(m,n) = \begin{cases} 1 & \text{if } m \neq n \\ 0 & \text{if } m = n \end{cases} \text{ and } d_1(m,n) = \begin{cases} 1 & \text{if } m = 1, n > 1 \\ 0.5 & \text{if } 1 < m, n\ m \neq m \\ 0 & \text{if } m = n \end{cases}. \tag{7}$$

d_0 and d_1 are equivalent metrics, which implies that $d_0' := d_0/(1 + d_0)$ and $d_1' := d_0/(1 + d_1)$ are also equivalent, induce the same Wijsman topologies on $\mathbf{Cl}(M)$, and their associated Prohorov metrics, ρ_0' and ρ_1', induce the same weak* topology on $\Delta(M)$. Let $B_0 = \{1\}$ and let $B_n = \{1, n, n + 1, n + 2, \ldots\}$. For every $m > 1$, $d_0'(m, B_n) = 0.5 = d_0(m, B_0)$ for every $n > m$, and $d_0'(1, B_n) = d_0(1, B_0) = 0$ so $B_n \to_{\tau_W(d_0')} B_0$. However, $d_1'(2, B_n) = 0.5/1.5$ for all $n > 2$ but $d_1'(2, B_0) = 0.5$. By Lemma 3, $\Delta(B_n) \to_{\tau_W(\rho_0')} \Delta(B_0)$ but $\Delta(B_n) \not\to_{\tau_W(\rho_1')} \Delta(B_0)$.

References

[1] Gerald Beer, *Topologies on closed and closed convex sets*, Mathematics and its Applications, vol. 268, Kluwer Academic Publishers Group, Dordrecht, 1993. MR1269778 (95k:49001)

[2] David Blackwell and Lester E. Dubins, *An extension of Skorohod's almost sure representation theorem*, Proc. Amer. Math. Soc. **89** (1983), no. 4, 691–692, DOI 10.2307/2044607. MR718998 (86b:60005)

[3] Leo Breiman, Lucien Le Cam, and Lorraine Schwartz, *Consistent estimates and zero-one sets*, Ann. Math. Statist. **35** (1964), 157–161. MR0161413 (28 #4620)

[4] Claude Dellacherie and Paul-André Meyer, *Probabilities and potential*, North-Holland Mathematics Studies, vol. 29, North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, 1978. MR521810 (80b:60004)

[5] R. M. Dudley, *Real analysis and probability*, Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original. MR1932358 (2003h:60001)

[6] Xavier Fernique, *Un modèle presque sûr pour la convergence en loi* (French, with English summary), C. R. Acad. Sci. Paris Sér. I Math. **306** (1988), no. 7, 335–338. MR934613 (89a:60010)

[7] Itzhak Gilboa, *Theory of decision under uncertainty*, Econometric Society Monographs, vol. 45, Cambridge University Press, Cambridge, 2009. MR2493167 (2010i:91002)

[8] S. Mrówka, *On the convergence of nets of sets*, Fund. Math. **45** (1958), 237–246. MR0098359 (20 #4820)

[9] Leonard J. Savage, *The foundations of statistics*, Second revised edition, Dover Publications, Inc., New York, 1972. MR0348870 (50 #1364)

[10] A. V. Skorohod, *Limit theorems for stochastic processes* (Russian, with English summary), Teor. Veroyatnost. i Primenen. **1** (1956), 289–319. MR0084897 (18,943c)

[11] Pedro Terán, *A continuity theorem for cores of random closed sets*, Proc. Amer. Math. Soc. **136** (2008), no. 12, 4417–4425, DOI 10.1090/S0002-9939-08-09412-4. MR2431058 (2009h:60029)

[12] John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, New Jersey, 1944. MR0011937 (6,235k)

Department of Economics, Universidad Carlos III de Madrid, Av. de la Universidad, 30, 28911 Leganés, Madrid, Spain
E-mail address: mdumav@gmail.com

Department of Economics, University of Texas, Austin, Texas 78712-0301
E-mail address: max.stinchcombe@gmail.com