Skorohods representation theorem for sets of probabilities

Martin Dumav and Maxwell B. Stinchcombe
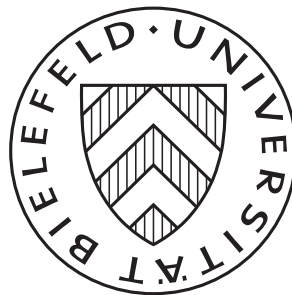
SKOROHOD'S REPRESENTATION THEOREM FOR SETS OF PROBABILITIES

MARTIN DUMAV, MAXWELL B. STINCHCOMBE

ABSTRACT. From Breiman *et al.* [3], a set of probabilities, $\mathbf{\Pi}$, on a measure space, $(\Omega, \mathcal{F})$, is *strongly zero-one* if there exists an $E \in \mathcal{F}$, a measurable, onto $\varphi : \Omega \to \mathbf{\Pi}$ such that for all $p \in \mathbf{\Pi}$, $p(\varphi^{-1}(p)) = 1$. Suppose that $\mathbf{\Pi}$ is an uncountable, measurable, strongly zero-one set of non-atomic probabilities on a standard measure space, that M is a complete, separable metric space, $\Delta_\mathcal{M}$ is the set of Borel probabilities on M and $Comp(\Delta_\mathcal{M})$ is the class of non-empty, compact subsets of $\Delta_\mathcal{M}$ with the Hausdorff metric. There exists a jointly measurable $H : Comp(\Delta_\mathcal{M}) \times \Omega \to M$ such that for all $K \in Comp(\Delta_\mathcal{M})$, $H(K, \mathbf{\Pi}) = K$, and if $d_H^\rho(K_n, K_0) \to 0$, then for all $p \in \mathbf{\Pi}$, $p(\{\omega : H(K_n, \omega) \to H(K_0, \omega)\}) = 1$. When each K_n and $\mathbf{\Pi}$ are singleton sets, this is the Blackwell and Dubins [2] version of Skorohod's representation theorem.

1. EXTENDING SKOROHOD'S REPRESENTATION THEOREM TO SETS

Let (M, d) be a complete separable metric (Polish) space, $\mathcal{M}$ the Borel σ-field on M, $\Delta_\mathcal{M}$ the set of countably additive probabilities on $\mathcal{M}$, and $C_b(M)$ the continuous, $\mathbb{R}$-valued functions on M. In $\Delta_\mathcal{M}$, let $\rho(\cdot, \cdot)$ be any metric inducing the weak* topology, that is, $\rho(\mu_n, \mu_0) \to 0$ iff $\int f \, d\mu_n \to \int f \, d\mu_0$ for every $f \in C_b(M)$. The Borel σ-field, $\mathcal{D}_\mathcal{M}$ on $\Delta_\mathcal{M}$ is the smallest σ-field containing all of the ρ-open sets, and it can alternatively be characterized as the smallest σ-field containing all sets of the form $\{\mu : \mu(E) \leq r\}$, $E \in \mathcal{M}$, $r \in [0, 1]$.

A measurable isomorphism between the measure spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ is a bijection that is measurable and has a measurable inverse. A measure space $(\Omega, \mathcal{F})$ is called *standard* if it is measurably isomorphic to a Borel measurable subset of a Polish space. Let $\Delta_\mathcal{F}$ denote the set of countably additive probabilities on $\mathcal{F}$, and $\mathcal{D}_\mathcal{F}$ the sigma-field generated by sets of the form $\{p \in \Delta(\mathcal{F}) : p(E) \leq r\}$, $E \in \mathcal{F}$, $r \in [0, 1]$. It is known that the measure space $(\Delta_\mathcal{F}, \mathcal{D}_\mathcal{F})$ is standard iff $(\Omega, \mathcal{F})$ is ([5, Theorem III.60]. In particular, $(\Delta_\mathcal{M}, \mathcal{D}_\mathcal{M})$ is standard.

Let $(\Omega, \mathcal{F})$ be a standard measure space and p a non-atomic, countably additive probability on $\mathcal{F}$. Skorohod [10] showed that if (M, d) is a complete separable metric space and $\rho(\mu_n, \mu_0) \to 0$, then there exist random variables, $X_n, X_0 : \Omega \to M$ such that

Sko(a) $X_n(p) = \mu_n$, $X_0(p) = \mu_0$, and
Sko(b) $p(\{\omega : X_n(\omega) \to X(\omega)\}) = 1$

where $X_n(p)$ is the image law of the distribution p under the random variable X_n, that is, $X_n(p)(B) = p(X_n^{-1}(B))$ for each $B \in \mathcal{M}$. Blackwell and Dubins

[2] extended Skorohod's result, proving the existence of a jointly measurable $h : \Delta_{\mathcal{M}} \times \Omega \to M$ such that

Bl-Du(a) for all $\mu \in \Delta_{\mathcal{M}}$, $h(\mu, p) = \mu$, and

Bl-Du(b) for all $\mu_n \to \mu_0$, $p(\{\omega : h(\mu_n, \omega) \to h(\mu_0, \omega)\}) = 1$.

Setting $X_n(\omega) = h(\mu_n, \omega)$ and $X_0(\omega) = h(\mu_0, \omega)$ recovers Skorohod's result. This paper gives a parallel to the Blackwell-Dubins result: the single non-atomic p is replaced with a strongly zero-one set of non-atomic probabilities, $\mathbf{\Pi}$; and the $\mu_n \to \mu_0$ in $\Delta_{\mathcal{M}}$ are replaced with compact subsets $K_n \to K_0$ in $\Delta_{\mathcal{M}}$.

Let $Comp(\Delta_{\mathcal{M}})$ denote the class of non-empty, compact subsets of $\Delta_{\mathcal{M}}$ with the Hausdorff metric, $d_H^\rho(A, B) = \inf\{\epsilon \geq 0 : (\forall \mu \in A)(\exists \nu \in B)[\rho(\mu, \nu) < \epsilon], \ (\forall \nu \in B)(\exists \mu \in A)[\rho(\mu, \nu) < \epsilon]\}$. Restricted to the compact subsets, the Hausdorff metric is equivalent for equivalent metrics on $\Delta_{\mathcal{M}}$, that is, if ρ and ρ' both induce the weak* topology on $\Delta_{\mathcal{M}}$, then $d_H^\rho(K_n, K_0) \to 0$ iff $d_H^{\rho'}(K_n, K_0) \to 0$. We use the following from Breiman *et al.* [3]

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

Theorem 1. *If (M, d) is a Polish space, $(\Omega, \mathcal{F})$ is a standard measure space, $\mathbf{\Pi} \subset \Delta_{\mathcal{F}}$ is a measurable, uncountable, and strongly zero-one set of non-atomic probabilities, then*

   (a) *for all measurable $A \subset \Delta_{\mathcal{M}}$, there exists a measurable $H_A : \Omega \to M$ such that $H_A(\mathbf{\Pi}) = A$, and*

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

Theorem 1(a) generalizes Skorohod's representation result in Sko(a) for a singleton probability distribution $\mu \in \Delta_{\mathcal{M}}$ to a measurable $A \subset \Delta_{\mathcal{M}}$; and Theorem 1(b) generalizes Bl-Du(a) and Bl-Du(b) from singleton probability distributions to compact sets of probability distributions. The next section gives the proof, the following discusses related work and extensions.

2. Proof

The argument for the representability of measurable sets of probabilities in Theorem 1(a) sets up a measurable isomorphism $\psi_A : \mathbf{\Pi} \leftrightarrow A$, and then uses strongly zero-one condition of $\mathbf{\Pi}$ and the $h(\mu, p) \equiv \mu$ property of Blackwell-Dubins function $h : \Delta_{\mathcal{M}} \times \Omega \to M$ to form the composition of measurable functions that is the requisite measurable $H_A : \Omega \to M$. The argument for the representation that is continuous for compact and convex in Theorem 1(b), begins with convergent sequences of compact convex subsets of $\Delta_{\mathcal{M}}$, and then extends to general compact subsets of $\Delta_{\mathcal{M}}$.

The arguments below make use of the following:

   (R1) The Borel isomorphism theorem: standard measure spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, are measurably isomorphic iff they have the same cardinality (e.g. [5, Theorem III.20] or [6, Theorem 13.1.1]).

   (R2) An implication of the Borel isomorphism theorem is that any uncountable standard probability space $(\Omega, \mathcal{F}, p)$ with non-atomic p will suffice for the

Skorohod and the Blackwell-Dubins results, and we take it to be the unit interval with the Borel σ-field and the uniform distribution, $([0,1], \mathcal{B}, \lambda)$.

(R3) If X is any random variable taking values in $[0,1]$ with $P(X = u) = 0$ for each $u \in [0,1]$, then its cumulative distribution function (cdf), F_P defined by $F_P(t) = P(X \le t)$, is an element of $C[0,1]$, the continuous functions on $[0,1]$. Further, the random variable $F_P(X)$ has the uniform distribution, i.e. $P(F_P(X) \le t) = t$ for all $t \in [0,1]$.

(R4) With the uniform metric, $d_\infty(f,g) = \max_{t \in [0,1]} |f(t) - g(t)|$, $C[0,1]$ is a Polish space, and its Borel σ-field is also generated by sets of the form $\{f \in C[0,1] : f(t) \le r\}$, $t \in [0,1]$, $r \in [0,1]$. Further, the mapping $(f,t) \to f(t)$ is jointly continuous, so that if $\omega \mapsto (f_\omega, u_\omega) \in C[0,1] \times [0,1]$ is measurable, then so is the function $\omega \mapsto f_\omega(u_\omega)$.

Proof of Theorem 1(a). Let A be a measurable subset of $\Delta_\mathcal{M}$. Because Π is an uncountable measurable subset of the standard measure space $(\Delta_\mathcal{F}, \mathcal{D}_\mathcal{F})$, (R1) yields the existence of a measurable isomorphism $\psi_A : \Pi \leftrightarrow A$ if A is uncountable, while ψ_A can be taken to be measurable and onto if A is countable.

Let $g : \Omega \leftrightarrow [0,1]$ be a measurable isomorphism and consider the mapping $p \mapsto F_{g(p)}(t) := p(g^{-1}([0,t])$ from Π to the cdf of the distribution $g(p)$. For any $t \in [0,1]$, $\{p \in \Pi : F_g(p)(t) \le r\} = \{p : p(g^{-1}[0,t]) \le r\}$. By (R4) this set and the mapping $p \mapsto F_{g(p)}$ is measurable.

The requisite $H_A : \Omega \to M$ is defined by

(1) $$H_A(\omega) = h(\psi_A(\varphi(\omega)), F_{g(\varphi(\omega))}(g(\omega)))$$

where $h : \Delta_\mathcal{M} \times [0,1] \to M$ is the jointly measurable Blackwell-Dubins function and φ is from Definition 1.

The measurability of $H_A(\cdot)$ being clear, to show that $H_A(\Pi) = A$, it is sufficient to show that for each $p \in \Pi$, $H_A(p) = \psi_A(p)$. For any $p \in \Pi$, $p(\varphi^{-1}(p)) = 1$, so that for a set of ω having p-probability 1, $\psi_A(\varphi(\omega)) = \psi_A(p)$. Further, by (R3) the mapping $\omega \mapsto F_{g(\varphi(\omega))}(g(\omega))$ induces the uniform distribution on $[0,1]$. Therefore, for each $p \in \Pi$, the distribution of $\omega \mapsto H_A(\omega)$ is equal to $h(\psi_A(p), \lambda)$, which is, by (R2), equal to $\psi_A(p)$. $\square$

The proof of Theorem 1(b) begins with compact convex sets of probabilities on M, $Conv(\Delta_\mathcal{M})$, and uses the following simultaneous retract result.

Lemma 1. *There exists a jointly continuous* $(A, \mu) \mapsto f_A(\mu)$ *from* $Conv(\Delta_\mathcal{M}) \times \Delta_\mathcal{M}$ *to* $\Delta_\mathcal{M}$ *such that for all* (A, μ), $f_A(\mu) \in A$, *and if* $\mu \in A$, *then* $f_A(\mu) = \mu$.

Proof. The correspondence (point-to-set map)

(2) $$\Gamma(A, \mu) = \begin{cases} \{(A, \mu)\} & \text{if } \mu \in A, \\ \{(A, \nu) : \nu \in A\} & \text{else.} \end{cases}$$

from $Conv(\Delta_\mathcal{M}) \times \Delta_\mathcal{M}$ to $Conv(\Delta_\mathcal{M}) \times \Delta_\mathcal{M}$ is lower hemicontinuous, defined on a metric, hence paracompact, space, and takes values in a topologically complete, locally convex, vector space. By Michael's selection theorem [8, Theorem 3.2 (p. 364 *et. seq.*)] Γ has a continuous selection γ. Define $f_A(\mu) = \text{proj}_{\Delta_\mathcal{M}}(\gamma(A, \mu))$. $\square$

Proof of Theorem 1(b) *for compact convex sets.* Suppose that $d_H(A_n, A_0) \to 0$ in $Conv(\Delta_\mathcal{M})$. Let $\psi : \Pi \leftrightarrow \Delta_\mathcal{M}$ be a measurable isomorphism, for any $A \in Conv(\Delta_\mathcal{M})$, define $H = h(f_A \circ \psi \circ \varphi, F_{g \circ \varphi} \circ g)$, i.e.

(3) $$H(A, \omega) = h(f_A(\psi(\varphi(\omega))), F_{g(\varphi(\omega))}(g(\omega)))$$

3

where h is the Blackwell-Dubins function, φ is from Definition 1, f_A is the jointly continuous function of Lemma 1, g is the measurable isomorphism between Ω and $[0,1]$ from the first half of the Theorem, and $F_{g(\varphi(\omega))}$ is the cdf of $g(\varphi(\omega))$. The joint measurability is clear, and the arguments for part (a) of the proof deliver $H(A, \mathbf{\Pi}) = A$ for all closed convex A. For continuity, fix an arbitrary $p \in \mathbf{\Pi}$ and suppose that $d_H^\rho(A_n, A_0) \to 0$. $\varphi(\cdot)$ being constant and equal to p on the set $\varphi^{-1}(p)$, $\psi(\varphi(\omega)) \equiv \psi(p)$ for all $\omega \in \varphi^{-1}(p)$, and the distribution of $F_{g(p)}(g(p)) \equiv \lambda$. Thus, $p(\{\omega : H(A_n, \omega) \to H(A_0, \omega)\}) = 1$ by the continuity of $(A, \mu) \mapsto f_A(\mu)$ and Bl-Du(b), the almost everywhere convergence property of the Blackwell-Dubins function. $\square$

To extend this proof to sequences K_n, K_0 of compact but not necessarily convex subsets of $\Delta_\mathcal{M}$, we: replace the compact metric space (M, d) with the compact metric space $(\Delta_\mathcal{M}, \rho)$; replace the K_n with $A_n := \Delta(K_n)$, the set of probability distributions on the set K_n of probability distributions; use Borel isomorphism theorem to map $\mathbf{\Pi} \subset \Delta_\mathcal{F}$ to $\mathbf{\Pi}'$, a tractable set of probabilities on $\mathcal{F}$; and use the tractability to change random variables whose image under $\mathbf{\Pi}$ is in $Conv(\Delta_\mathcal{M})$ to random variables whose image is $Comp(\Delta_\mathcal{M})$.

Proof of Theorem 1(b) *for compact sets.* Let $\mathbf{\Pi}^\circ$ be the set of probabilities on $\Omega^\circ := [0,1]^3$ defined by setting p_r to be the uniform distribution on $\{r\} \times [0,1] \times [0,1]$ and $\mathbf{\Pi}^\circ = \{p_r : r \in [0,1]\}$. The typical element of Ω° will be denoted (r, ω_1, ω_2), and setting $\varphi(r, \omega_1, \omega_2) = p_r$, shows that $\mathbf{\Pi}^\circ$ is strongly zero-one. Also, let Ω' be the projection of Ω° onto its first two axes, $(r, \omega_1, \omega_2) \mapsto (r, \omega_1)$, let q_r be the image of p_r under this projection, that is, q_r is the uniform distribution on $\{r\} \times [0,1]$, and let $\mathbf{\Pi}'' = \{q_r : r \in [0,1]\}$, another strongly zero-one set. Theorem 1(a) guarantees the existence of a measurable $H_{\mathbf{\Pi}^\circ} : \Omega \to \Omega^\circ$ such that $H_{\mathbf{\Pi}^\circ}(\mathbf{\Pi}) = \mathbf{\Pi}^\circ$.

Let $A_n = \Delta(K_n)$ be the convex and weak* compact set of probability distributions on K_n, $n = 0, 1, \ldots$. Let $e(\cdot)$ be any metric on $\Delta(\Delta_\mathcal{M})$ that induces weak* convergence in the set of Borel probabilities on the Polish space $\Delta_\mathcal{M}$, and note that $D_H^e(A_n, A_0) \to 0$ iff $d_H^\rho(K_n, K_0) \to 0$ where $D_H^e(\cdot)$ is the Hausdorff distance corresponding to the metric $e(\cdot)$.

In parallel with the previous arguments, we start with a measurable isomorphism $\psi : \mathbf{\Pi}'' \leftrightarrow \Delta(\Delta_\mathcal{M})$ and construct jointly measurable $(\Delta(K), (r, \omega_1)) \mapsto \mathcal{H}(\Delta(K), (r, \omega_1)) \in \Delta(\Delta_\mathcal{M})$ such that $\mathcal{H}(\Delta(K), \mathbf{\Pi}'') \equiv \Delta(K)$ and $q_r(\{(r, \omega_1) \in \Omega' : \mathcal{H}(A_n, (r, \omega_1)) \to \mathcal{H}(A, (r, \omega_1))\}) = 1$ for all $q_r \in \mathbf{\Pi}''$.

Let $\Delta^{pt}(\Delta_\mathcal{M})$ be the uncountable, closed set of point masses in $\Delta(\Delta_\mathcal{M})$, let $\mathbf{\Pi}'$ be the uncountable, measurable set $\{p' \in \mathbf{\Pi}'' : \psi(p') \in \Delta^{pt}(\Delta_\mathcal{M})\}$, and let $\psi' : \mathbf{\Pi}'' \leftrightarrow \mathbf{\Pi}'$ be a measurable isomorphism. The previous arguments using $\psi' \circ \psi$ in the role of ψ deliver a jointly measurable mapping $(A, (r, \omega_1) \mapsto \mathcal{H}'(A, (r, \omega_1)) \in \Delta(\Delta_\mathcal{M})$ such that $\mathcal{H}'(A, \mathbf{\Pi}'') = \{\delta_\mu : \mu \in K\}$. The map ι defined by $\iota(\delta_\mu) = \mu$ is clearly continuous. The composite function $\mathcal{H}''(\Delta(K), (r, \omega_1, \omega_2)) := h(\iota(\mathcal{H}'(\Delta(K), (r, \omega_1))), \omega_2)$, where h is, as usual, the Blackwell-Dubins function, is jointly measurable and has the property that $\mathcal{H}''(\Delta(K), \mathbf{\Pi}^\circ) \equiv K$ and $p_r(\{\mathcal{H}''(\Delta(K_n), \cdot) \to \mathcal{H}''(\Delta(K_0), \cdot)\}) = 1$ for all $p_r \in \mathbf{\Pi}^\circ$. The requisite function is $H(K_n, \omega) = \mathcal{H}''(\Delta(K_n), H_{\mathbf{\Pi}^\circ}(\omega))$. $\square$

3. Related Work and Extensions

We first illustrate the role of strongly zero-one condition plays in consistent estimation, then discuss how interest in Theorem 1 arose from models of choice in the presence of ambiguity. The next parts of this section discuss in turn the possibility

of dropping: the completeness requirement on M; the compactness requirement on the $K_n \to K_0$; and the standardness assumption on $(\Omega, \mathcal{F})$. The section ends with a discussion of the minimality of strongly zero-one sets $\mathbf{\Pi}$.

3.1. **Consistent Estimation.** Breiman *et al.* [3] show that the strong zero-one condition is necessary and sufficient for the existence of consistent estimators. A canonical example illustrates this result.

Example 1. *For $\Omega = \{0,1\}^{\mathbb{N}}$ and $r \in [0,1]$, let $E_r \subset \Omega$ be the set of $(\omega_1, \omega_2, ...)$ such that $\lim_n \frac{1}{n} \#\{k \leq n : \omega_k = 1\} = r\}$. For $\omega \in E := \cup_{r \in [0,1]} E_r$, define $f(\omega)$ as the unique r for which $\omega \in E_r$. Let p_r be the distribution on Ω of an independent and identically distributed sequence of Bernoulli(r) random variables, and set $\mathbf{\Pi} = \{p_r : r \in [0,1]\}$. By the strong law of large numbers, for each p_r, $p_r(E_r) = 1$ so that defining $\varphi(\omega) = p_{f(\omega)}$ gives $p_r(\varphi^{-1}(p_r)) = 1$. If ω is distributed according to one of the $p_r \in \mathbf{\Pi}$, interest centers on finding consistent estimators, that is, a sequence of functions $(\omega_1, \ldots, \omega_n) \mapsto \widehat{\varphi}_n(\omega_1, \ldots, \omega_n)$ such that for each p_r, $p_r(\{\omega : \widehat{\varphi}_n(\omega_1, \ldots, \omega_n) \to \varphi(\omega)\}) = 1$. An obvious choice is $\widehat{\varphi}_n = p_{\widehat{r}_n}$ where $\widehat{r}_n = \frac{1}{n} \#\{k \leq n : \omega_k = 1\}$.*

3.2. **Decision Theory.** Decision theory in the face of uncertainty has two main models, related by change of variables. One of them is due to von Neumann and Morgenstern [12] and the other due 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, $(\Omega, \mathcal{F})$.

von Neumann and Morgenstern [12] gave a short axiomatic foundation for preferences over distributions on M. A preference, $\succsim$, on $\Delta_{\mathcal{M}}$ is a complete, transitive, binary relation on $\Delta_{\mathcal{M}}$. In their approach, it can be represented by $\mu \succsim \nu$ iff

$$(4) \qquad \int_M u(x) \, d\mu(x) \geq \int_M u(x) \, d\nu(x)$$

where the $u \in C_b(M)$ is unique up to positive affine transformations. By contrast, Savage's [9] work provides an axiomatic foundation for preference over measurable functions from a state space, $(\Omega, \mathcal{F})$, to M. The preferences can be represented by $X \succsim Y$ iff

$$(5) \qquad \int_{\Omega} u(X(\omega)) \, dP(\omega) \geq \int_{\Omega} u(Y(\omega)) \, dP(\omega),$$

where $P \in \Delta_{\mathcal{F}}$ is unique and the $u \in C_b(M)$ 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 each side of (4) and (5) 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. Sko(b) says that we can analyze continuity either using $\mu_n \to \mu_0$ or $X_n \to X_0$ a.e.

Over the last several decades, the systematic inability of either approach to explain behavior in the face of ambiguity, modeled as unknown probabilities, has led many to replace Savage's single prior, P, with a set of priors, $\mathbf{\Pi}$ (see e.g. the monograph [7]). One of the most general forms of preferences is given by $X \succsim Y$

iff

$$(6) \qquad \alpha \min_{p \in \mathbf{\Pi}} \int u(X)\, dp(\omega) + (1-\alpha) \max_{q \in \mathbf{\Pi}} \int u(X)\, dq(\omega) \geq$$

$$\alpha \min_{p \in \mathbf{\Pi}} \int u(Y)\, dp(\omega) + (1-\alpha) \max_{q \in \mathbf{\Pi}} \int u(Y)\, dq(\omega).$$

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 this paper was "What conditions on $\mathbf{\Pi}$ yield both domain equivalence and continuity?"

Many models of choice in the presence of ambiguity use what are called *belief functions*. These are the conjugates of Choquet capacities called the hitting functionals of a random closed set, defined by $\nu(A) = P(X \cap A \neq \emptyset)$ for compact $A \subset M$ where $\omega \mapsto X(\omega)$ is a measurable mapping from Ω to the class of closed subsets of M. Of particular interest are the cores of these capacities, $\mathrm{core}(\nu) := \{\mu \in \Delta_{\mathcal{M}} : \mu(A) \leq \nu(A)\}$. and the associated infima and suprema of the integrals against these sets as in (6). When each $X(\omega)$ is a compact set, the core is closed [4, Corollary 3.4] so that the infimum and supremum are achieved. Continuity questions arise as follows: Terán [11] shows that if the random closed sets converge in distribution, then the closure of the cores converge.

3.3. **The Role of Completeness.** Skorohod's construction has been generalized to metric spaces that are separable but not complete, and to non-separable spaces if the limit, μ_0, has separable support (see [6, §11.7] and the accompanying references). Let (X, d) be a separable metric space. If we can find a jointly measurable $h : [0, 1] \times [0, 1] \to X$ such that $\{h(r, \lambda) : r \in [0, 1]\} = \Delta_X$, then we can: replace the metric on Δ_X with an ℓ^2 type of metric topologizing the weak* topology; replace continuous selection proof of Lemma with an ℓ^2-minimum distance approach. Such an $h(\cdot, \cdot)$ will exist if (X, d) is measurably isomorphic to a Borel subset of a Polish space, but we suspect that such a proof strategy will not work for more general separable metric spaces. The essential difficulty is that the range of the mapping $r \mapsto h(r, \lambda)$ is the measurable image of a measurable set, and to the extent that Δ_X inherits the properties of X (see e.g. [5, Theorem III.60]), this covers a strict, albeit highly interesting, subset of the separable metric spaces.

3.4. **The Role of Compactness.** There are several different approaches to extending the proof strategy for Theorem 1(b) from sequences of compact sets $K_n \to K_0$ to sequences of closed sets $F_n \to F_0$ depending on the definition of convergence that one uses on the class of closed sets: convergence in the the Hausdorff metric does not work; convergence in the Wijsman topology does work; and neither convergence in the Vietoris topology nor Painlevé-Kuratowski convergence are satisfactory for applications.

3.4.1. *Hausdorff metrics on closed sets.* Let $Cl(\Delta_{\mathcal{M}})$ denote the set of closed subsets of $\Delta_{\mathcal{M}}$. On $Cl(\Delta_{\mathcal{M}})$, the metrics $d_H^\rho(\cdot)$ and $d_H^{\rho'}(\cdot)$ are equivalent iff ρ and ρ' induced the same uniformity on $\Delta_{\mathcal{M}}$ [1, Theorem 3.3.2, p. 92]. However, for the simultaneous retract Lemma to work, we need $(Cl(\Delta_{\mathcal{M}}), d_H^\rho)$ to be separable, which requires that $\Delta_{\mathcal{M}}$ be totally bounded in the ρ metric, and we need $(Cl(\Delta_{\mathcal{M}}), d_H^\rho)$ to be complete, which requires that $\Delta_{\mathcal{M}}$ be complete in the ρ metric. $(\Delta_{\mathcal{M}}, \rho)$ being totally bounded and complete is equivalent to it being compact, and $\Delta_{\mathcal{M}}$ being compact in the weak* topology is equivalent to (M, d) being compact.

One might hope to replace the proof of Lemma 1 by having $f_A(\mu)$ be the nearest point to μ in the convex set A. However, for many well-known metrics on $\Delta_\mathcal{M}$, the following example shows that such a nearest point may not exist, or the set of nearest points may be wildly discontinuous as a function of A.

Example 2. *Let (M, d) be a separable Hilbert space, $\{e_n : n \in \mathbb{N}\}$ an orthnormal basis, and give $\Delta_\mathcal{M}$ the Prokhorov metric, $\rho(\mu, \nu) = \inf\{\epsilon \geq 0 : \forall F \text{ closed}, \ \mu(F) \leq \nu(F^\epsilon) + \epsilon, \nu(F) \leq \mu(F^\epsilon) + \epsilon\}$ where $F^\epsilon := \cup_{x \in F} B_\epsilon(x)$ is the ϵ-ball around the set F. For $x \in M$, let $\delta_x \in \Delta_\mathcal{M}$ be point mass on x, i.e. $\delta_x(E) = 1_E(x)$. Let $(r_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{R}$ with $|r_n| \leq 1/10$ and $r_n \to 0$, let A be the closed convex hull of $\{\delta_{(\frac{1}{2} + r_n)e_n} : n \in \mathbb{N}\}$, and let μ be δ_0.*

No solution. *If r_n is a strictly decreasing positive sequence, then there is no solution to the problem $\min_{\nu \in A} \rho(\mu, \nu)$.*

Discontinuous solutions. *For each finite $N \subset \mathbb{N}$, set $r_n^N = -\epsilon \cdot 1_N(n)$, $\epsilon > 0$. The set of solutions to $\min_{\nu \in A^N} \rho(\mu, \nu)$, where A^N is the closed convex hull of probabilities associated with the sequence r^N, is the convex hull of the point masses of the form $\delta_{(\frac{1}{2} + r_n)e_n}$, $n \in N$. For finite disjoint $N_1, N_2 \subset \mathbb{N}$, the distance between the solution sets is at least $\frac{1}{2} - \epsilon$, even though the distance between the associated A^{N_1} and A^{N_2} is $\epsilon \cdot \sqrt{\#(N_1 \cup N_2)}$.*

3.4.2. *Wijsman topologies.* Suppose now that the metric ρ on $\Delta_\mathcal{M}$ not only gives the weak* topology, but it also makes $(\Delta_\mathcal{M}, \rho)$ into a Polish space (which is possible iff there is a metric on M making it Polish). The Wijsman topology, τ_ρ, on $Cl(\Delta_\mathcal{M})$ is the topology of pointwise convergence of the distance functionals $\rho(\mu, \cdot)$, that is, a net F_α τ_ρ-converges to F_0 iff for all $\mu \in \Delta_\mathcal{M}$, $d(\mu, F_\alpha) \to d(\mu, F_0)$. As the distance functionals have Lipschitz constant 1, this is not as weak a topology as it might seem, and $(Cl(\Delta_\mathcal{M}), \tau_\rho)$ admits a metric making it Polish (see [1, Theorem 2.5.4 (p. 71)]).

Corollary 1.1. *If $(\Delta_\mathcal{M}, \rho)$ is a Polish space, $(\Omega, \mathcal{F})$ is a standard measure space, $\mathbf{\Pi} \subset \Delta_\mathcal{F}$ is a measurable, uncountable, and strongly zero-one set of non-atomic probabilities, then*

(b′): *there exists a jointly measurable $H : Cl(\Delta_\mathcal{M}) \times \Omega \to M$ such that for all closed $F \subset \Delta_\mathcal{M}$, $H(F, \mathbf{\Pi}) = F$, and if $F_n \to_{\tau_\rho} F_0$, then for all $p \in \mathbf{\Pi}$, $p(\{\omega : H(F_n, \omega) \to_{\tau_\rho} H(F_0, \omega)\}) = 1$.*

Proof. The proof of the simultaneous retract Lemma is the same because the Wijsman topology is also Polish. This in turn implies that the compact convex sets with the Hausdorff topology can be replaced by closed convex sets with the Wijsman topology in the part of the proof of Theorem 1(b) dealing with convex sets. The extension of this result to sequences of closed sets requires replacing the various Hausdorff metric topologies by the corresponding Wijsman topologies, and then checking that $\Delta(F_n)$ Wijsman converges to $\Delta(F_0)$ in the class of closed sets of probabilities on probabilities iff F_n Wijsman converges to F_0 in the class of closed sets of probabilities on M, which is immediate. □

3.4.3. *Other topologies on the class of closed sets.* Two other well-known options for the sense of convergence of the closed sets of probabilities are the Vietoris topology and Painlevé-Kuratowski convergence. We chose not to work with the Vietoris topology on closed sets because it is typically too strong for applications

— taking $M = \mathbb{R}^2$, if $F_n = \Delta(\{1/n\} \times \mathbb{R})$ and $F_0 = \Delta(\{0\} \times \mathbb{R})$, then F_n Wijsman converges to F_0 (if the metric ρ on $\Delta_{\mathcal{M}}$ is e.g. the Prokhorov metric from Example 2), but does not Vietoris converge. Painlevé-Kuratowski convergence for closed, non-compact sets is a sensible kind of convergence if one is working with sequences of closed subsets of a locally compact space, essentially because one works in the one-point compactification of the space and uses the Hausdorff topology. However, in $\Delta_{\mathcal{M}}$ with the weak* topology, local compactness is equivalent to compactness.[1]

3.5. **The Standardness Assumption.** The assumption that $(\Omega, \mathcal{F})$ is a standard space can be significantly relaxed provided that one is willing to deal with classes of probabilities defined on a sub-σ-field of $\mathcal{F}$. Suppose that there exists a countably additive non-atomic $p \in \Delta_{\mathcal{F}}$. Let $X : \Omega \to [0,1]$ be any measurable function having the property that $p(X \le t) = t$ for $t \in [0,1]$ and let $\mathcal{X} \subset \mathcal{F}$ be the smallest σ-field making X measurable. One works with sets of probabilities $\mathbf{\Pi} \subset \Delta_{\mathcal{X}}$, and this is a standard space: as $\mathcal{X}$ is countably generated, we can define the Blackwell pseudo-metric, $d_B(\omega, \omega') = 1/\min\{n : 1_{E_n}(\omega) \ne 1_{E_n}(\omega')\}$ where $\{E_n : n \in \mathbb{N}\}$ is a countable set generating $\mathcal{X}$; the pseudo-metric space (Ω, d_B) has $\mathcal{X}$ as its σ-field; and after passing to d_B-equivalence classes, $(\Omega, \mathcal{X})$ is a standard space, so that all of the work above applies to it.

3.6. **The Minimality of $\mathbf{\Pi}$.** The set $\mathbf{\Pi}$ can be either too large or too small to satisfy Theorem 1(a): $\mathbf{\Pi} = \Delta(\mathcal{F})$ is too large, for any $H : \Omega \to M$, $H(\Delta(\mathcal{F}))$ contains all of the distributions on the range of H; $\mathbf{\Pi} = \{\alpha p + (1 - \alpha)q : \alpha \in [0,1]\}$ is too small, for any $H : \Omega \to M$, $H(\mathbf{\Pi})$ is at most 1-dimensional. The question remains, among the uncountable sets of non-atomic probabilities, are there minimal sets, $\mathbf{\Pi}_m$, satisfying the conclusion of Theorem 1(a), that for all measurable $A \subset \Delta_{\mathcal{M}}$, there exists a measurable $H_A : \Omega \to M$ such that $H_A(\mathbf{\Pi}_m) = A$?

Minimality cannot be in terms of set inclusion, if $\mathbf{\Pi}$ satisfies the conditions of Theorem 1, then so does any uncountable measurable subset of $\mathbf{\Pi}$. For the same reason, minimality also cannot be in terms of mutual orthogonality.

Example 3. *Let* $\mathbf{\Pi}' = \{p_r : r \in [0, a] \cup \{1\}\}$ *where* $0 < a < 1$ *and* p_r *is the uniform distribution on* $\{r\} \times [0,1] \subset [0,1]^2$, *and let* $\mathbf{\Pi} = \mathbf{\Pi}' \cup \{\frac{1}{2}p_a + \frac{1}{2}p_1\}$. $\mathbf{\Pi}$ *clearly violates strong zero-one condition but satisfies the conclusion of Theorem 1(a) — pick an* H_A *that covers A with the set* $\{p_r : r \in [0, a)\}$, *pick* $\mu \in A$, *and define* H_A *on* $\{a\} \times [0,1]$ *and* $\{1\} \times [0,1]$ *so that* $H_A(p_a) = H_A(p_1) = \mu$.

However, if $\mathbf{\Pi}$ satisfies the conclusions of Theorem 1(a), then by composition of measurable functions, it satisfies the conclusions of 1(b) — simply let $f_\circ : \Omega \to \Omega'$ have the property that $\mathbf{\Pi}' := f_\circ(\mathbf{\Pi})$ satisfies all of the conditions in Theorem 1.

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. 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)

---

[1]If $\Delta_{\mathcal{M}}$ is locally compact, then for any $p \in \Delta_{\mathcal{M}}$ and any neighborhood G_p of p with compact closure, for sufficiently small $r > 0$, $(1 - r)p + r\Delta_{\mathcal{M}} \subset G_p$, but $(1 - r)p + r\Delta_{\mathcal{M}}$ has compact closure iff $\Delta_{\mathcal{M}}$ is compact.

[4] Adriana Castaldo, Fabio Maccheroni, and Massimo Marinacci, *Random correspondences as bundles of random variables*, Sankhyā **66** (2004), no. 3, 409–427. MR2108198 (2005i:28030)

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

[6] 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)

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

[8] Ernest Michael, *Continuous selections. I*, Ann. of Math. (2) **63** (1956), 361–382. MR0077107 (17,990e)

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

[10] A. V. Skorohod, *Limit theorems for stochastic processes*, 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. 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, University of Texas, Austin, TX 78712-0301 USA, e-mail: mdumav@gmail.com, max.stinchcombe@gmail.com