Discussion Papers of the
Max Planck Institute for
Research on Collective Goods
2017/6

Probability Measures on Product Spaces with Uniform Metrics

Martin F. Hellwig

MAX PLANCK
SOCIETY

Discussion Papers of the
Max Planck Institute
for Research on Collective Goods 2017/6

Probability Measures on Product Spaces with Uniform Metrics

Martin F. Hellwig

May 2017

This version December 2020

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
https://www.coll.mpg.de

Probability Measures on Product Spaces with Uniform Metrics*

Martin F. Hellwig
Max Planck Institute for Research on Collective Goods
Kurt Schumacher-Str. 10
D-53113 Bonn, Germany
hellwig@coll.mpg.de

December 23, 2020

Abstract

For a countable product of complete separable metric spaces, with a topology induced by a uniform metric, the σ-algebra generated by the open balls, which was introduced by Dudley (1966), coincides with the product σ-algebra. Any probability measure on the product space with this σ-algebra is *quasi-separable* in the sense that, for any union of open balls that has full measure, there is a countable sub-union that also has full measure. With suitably adapted definitions, the topology of weak convergence on the space of such measures is equivalent to the topology induced by the Prohorov metric. The projection mapping from such measures to sequences of measures on the first ℓ factors, $\ell = 1, 2, ...$, is a homeomorphism if the range of this mapping is also given a uniform metric. These findings are relevant for the theory of games of incomplete information, where a topology on the space of belief hierarchies that is based on a uniform metric has been proposed as being more appropriate for capturing the continuity properties of strategic behaviour.

Key Words: Product spaces with uniform metrics, weak convergence of non-Borel measures, σ-algebras generated by the open balls, quasi-separable measures, Prohorov metric.

MSC Classification 60B05

JEL Classification: C02, C72

*Without implicating them, I thank Eduardo Faingold and Alia Gizatulina for helpful discussions.

1 Introduction

Let $X_1, X_2, ...$ be non-singleton complete separable metric spaces with metrics $\rho_1, \rho_2, ...$. Suppose that the product

$$X = \prod_{k=1}^{\infty} X_k \tag{1.1}$$

has the topology induced by the uniform metric ρ^u where, for any x and $\hat{x}$ in X,

$$\rho^u(x, \hat{x}) = \sup_k \rho_k(\pi_k(x), \pi_k(\hat{x})) \tag{1.2}$$

and π_k is the projection from X to X_k. I use the notation X^u to indicate that X has the topology induced by ρ^u. Let $\mathcal{B}_0(X^u)$ be the σ-algebra that is generated by the ρ^u-open balls and let $\mathcal{M}_0(X^u)$ be the space of probability measures on $(X^u, \mathcal{B}_0(X^u))$. Let $CB_0(X^u)$ be the space of bounded, ρ^u-continuous, and $\mathcal{B}_0(X^u)$-measurable real-valued functions on X. Say that a sequence $\{\mu^r\}$ of measures in $\mathcal{M}_0(X^u)$ converges weakly to a measure $\mu \in \mathcal{M}_0(X^u)$ if and only if

$$\int_{X^u} f(x) d\mu^r(x) \to \int_{X^u} f(x) d\mu(x) \tag{1.3}$$

for all $f \in CB_0(X^u)$.

This paper shows that, even though X^u is non-separable and $\mathcal{B}_0(X^u)$ is not a Borel σ-algebra, yet, under the assumption that the cardinal $\mathbf{c}$ of the continuum is not atomlessly measurable, the topology of weak convergence on $\mathcal{M}_0(X^u)$ is metrizable by a suitably adapted version of the Prohorov metric. The argument is similar to the argument for Borel measures on possibly non-separable metric spaces.[1] However, whereas the latter argument relies on the fact that, under the given assumption about the cardinal of the continuum, any Borel measure on a metric space is concentrated on a separable set,[2] I show that, under the same assumption, every measure in $\mathcal{M}_0(X^u)$ is concentrated on what I call a *quasi-separable* set, i.e., a set with the property that any covering of the set by open balls contains a countable subcovering. The proof of the latter result uses the fact that X has a product structure and combines ideas from Banach (1930) and Billingsley (1968).

[1] See, e.g., Theorem 5, p. 238, in Billingsley (1968).

[2] See Theorem III in Marczewski and Sikorski (1948) or Theorem 2, p. 235, in Billingsley (1968).

The σ-algebra $\mathcal{B}_0(X^u)$, the space $\mathcal{M}_0(X^u)$, and the topology of weak convergence on $\mathcal{M}_0(X^u)$ were introduced by Dudley (1966, 1967) in order to avoid certain inconveniences associated with the Borel σ-algebra $\mathcal{B}(X^u)$ that are induced by the non-separability of X^u.[3] Dudley was interested in the convergence properties of sequences of stochastic processes when the space of sample paths of the processes has the uniform topology. By using the coarser σ-algebra induced by the open balls, he avoided the difficulties arising from the large size of the Borel σ-algebra. Dudley's approach provides an alternative to the use of the Skorokhod topology and associated Borel σ-algebra on the space of the sample paths.

Dudley did not actually study the topology that is induced by the concept of weak convergence on $\mathcal{M}_0(X^u)$. Given his interest in the convergence properties of sequences of stochastic processes, he merely considered the convergence behaviour of sequences of integrals of bounded continuous functions, including upper and lower integrals for functions that are not measurable with respect to the smaller σ-algebra induced by the open balls. In this analysis, he assumed that the limit measure of such a sequence is concentrated on a separable set; this assumption presumes that the limit measure can be extended to the Borel σ-algebra for the uniform topology so that the result of Marczewski and Sikorski could be appealed to.[4] Dudley's approach thus involves an implicit asymmetry between measures that can be so extended and measures that cannot be so extended.

For Dudley's reasearch programme, this asymmetry did not matter because the limit measures in his convergence theorems satisfied extendability condition anyway. In other contexts, the asymmetry is problematic. In the final section of this paper, I discuss recent developments in game theory that have provided the motivation for the research presented here.

2 Quasi-Separability of Measures in $\mathcal{M}_0(X^u)$

Because each the spaces $X_1, X_2, X_3, ...$ has more than one element, the cardinal of X is at least $2^\omega = \mathbf{c}$ where ω is the cardinal of the natural numbers

[3] See also Wichura (1970), Dudley (1978), and Pollard (1979).

[4] Pollard (1979) implicitly makes the same assumption. He uses the separability of the support of the limit measure in order to define a topology that is smaller than the topology of weak convergence on the space of measures on the σ-algebra generated by the open balls, but sufficiently large to provide for weak convergence in a neighbourhood of the limit measure. Wichura (1970) considers the random variables associated with the measures on the σ-algebra generated by the open balls.

and $\mathbf{c}$ is the cardinal of the continuum. It is also no greater than $\mathbf{c}$ because, as a product of complete separable metric spaces, with the product topology, X itself is a complete separable metric space so the cardinal of X is either finite or ω or $\mathbf{c}$.[5]

The cardinal of the σ-algebra $\mathcal{B}_0(X^u)$ is also equal to $\mathbf{c}$. To see this, note that, by a result of Dudley (1967, p. 449), the claim is true if $\mathbf{c}$ is the smallest cardinal of a dense set in X^u. To see that X^u satisfies this condition, for $k = 1, 2, \ldots$, let x_k^1, x_k^2 be two distinct elements of X_k and consider the set $\hat{X} = \{x = (x_1, x_2, \ldots) \in X | x_k \in \{x_k^1, x_k^2\} \text{ for all } k\}$. Clearly, $\hat{X} \subset X$ has the cardinal $\mathbf{c}$, and so does any subset of $\hat{X}$, or of X, that is ρ^u-dense.

Throughout the analysis, I impose the following assumption.

Assumption 1 The cardinal $\mathbf{c}$ of the continuum is not atomlessly measurable: no set of cardinality $\mathbf{c}$ or less admits a nontrivial atomless measure that is defined on all subsets of the set.

Definition 2.1 *A measure $\mu \in \mathcal{M}_0(X^u)$ is called quasi-separable if, for any family $\mathcal{G}$ of ρ^u-open balls that covers X, there exists a countable subfamily $\mathcal{G}^*$ such that the union G^* of the sets in $\mathcal{G}^*$ satisfies $\mu(G^*) = 1$.*

Proposition 2.2 *Under Assumption 1, any measure $\mu \in \mathcal{M}_0(X^u)$ is quasi-separable.*

The proof of Proposition 2.2 makes use of the outer measure that a measure $\mu \in \mathcal{M}_0(X^u)$ induces on every subset of X. For a given set $U \subset X$, the value $\mu^*(U)$ of the outer measure μ^* that is induced by μ is defined as[6]

$$\mu^*(U) = \inf \sum_{i=1}^{\infty} \mu(B_i) \quad \text{s.t.} \quad U \subset \bigcup_{i=1}^{\infty} B_i \quad \text{and} \quad B_i \subset \mathcal{B}_0(X^u) \quad \text{for all } i. \tag{2.1}$$

The following preliminary lemma is of interest in its own right.

Lemma 2.3 *Let μ be a finite measure on $(X^u, \mathcal{B}_0(X^u))$. Let I be a set of indices and let $\{B_i\}_{i \in I}$ be a family of open ρ^u-spheres such that $\mu(B_i) = 0$ for all $i \in I$. Then the set*

$$U_I := \bigcup_{i \in I} B_i \tag{2.2}$$

has the outer measure $\mu^*(U_I) = 0$.

[5] Theorem 13.1.1 in Dudley (2002).

[6] See Dudley (2002), p. 89.

Proof. For any $i \in I$ and any natural number k, let

$$\Pi^k(B_i) := \pi^k(B_i) \times X^{k+1} \times X^{k+2} \times ... \tag{2.3}$$

where $\pi^k(B_i)$ is the projection of B_i to the finite product $X_1 \times \ldots \times X_k$. For any set $\hat{I} \subset I$ and any k, let

$$U^k_{\hat{I}} := \bigcup_{i \in \hat{I}} \Pi^k(B_i). \tag{2.4}$$

For any i and k, $\pi^k(B_i)$ is a finite product of open spheres. Therefore, $\Pi^k(B_i)$ is an open set in the product topology on X. For any $\hat{I} \subset I$, therefore, $U^k_{\hat{I}}$ is also an open set in the product topology on X . Because the product σ-algebra on X is coextensive with the σ-algebra $\mathcal{B}_0(X^u)$, it follows that $\mu(U^k_{\hat{I}})$ is well defined for all $\hat{I} \subset I$ and all k. For any k, therefore, the formula

$$\nu^k(\hat{I}) := \mu(U^k_{\hat{I}}) \tag{2.5}$$

defines a set function on the subsets of I.

From (2.3), one finds that, for any i, the sequence $\{\Pi^k(B_i)\}_{k=1}^{\infty}$ is nonincreasing. For any $\hat{I} \subset I$, therefore, the sequence $\{U^k_{\hat{I}}\}_{k=1}^{\infty}$ is also nondecreasing. For any $\hat{I} \subset I$, therefore, the limit

$$\nu^\infty(\hat{I}) := \lim_{k \to \infty} \nu^k(\hat{I}) \tag{2.6}$$

is well defined. I will show that the set function ν^∞ is atomless as well as countably additive.

I first show that ν^∞ is atomless, i.e., that

$$\nu^\infty(\{i\}) = 0 \tag{2.7}$$

for all $i \in I$. For any $i \in I$, obviously,

$$\bigcap_{k=1}^{\infty} \Pi^k(B_i) = B_i, \tag{2.8}$$

and therefore

$$\nu^\infty(\{i\}) := \lim_{k \to \infty} \nu^k(\{i\}) = \lim_{k \to \infty} \mu(\Pi^k(B_i)) = \mu(B_i), \tag{2.9}$$

so (2.7) follows from the assumption that $\mu(B_i) = 0$ for all $i \in I$.

I next show that ν^∞ is countably additive. Let $I_j, j = 1, 2, ...,$ be a sequence of disjoint subsets of I. For any k, obviously,

$$\nu^k\left(\bigcup_{j=1}^{\infty} I_j\right) = \mu\left(\bigcup_{j=1}^{\infty} U_{I_j}^k\right) = \sum_{j=1}^{\infty} \mu(U_{I_j}^k) = \sum_{j=1}^{\infty} \nu^k(I_j). \tag{2.10}$$

The desired result,

$$\nu^\infty\left(\bigcup_{j=1}^{\infty} I_j\right) = \sum_{j=1}^{\infty} \nu^\infty(I_j),$$

follows by a standard monotone-convergence argument: Since

$$\nu^\infty\left(\bigcup_{j=1}^{\infty} I_j\right) = \lim_{k\to\to\infty} \nu^k\left(\bigcup_{j=1}^{\infty} I_j\right),$$

(2.10) implies that it suffices to show that

$$\sum_{j=1}^{\infty} \nu^\infty(I_j) = \lim_{k\to\infty} \sum_{j=1}^{\infty} \nu^k(I_j). \tag{2.11}$$

Because, for any k, the infinite series $\sum_{j=1}^{\infty} \nu^k(I_j)$ converges and, moreover, the terms $\nu^k(I_j)$ are non-increasing in k, it follows that, for each $\varepsilon > 0$, there exists a finite set J_ε, with complement J_ε^c , such that

$$\sum_{j\in J_\varepsilon^c} \nu^k(I_j) < \varepsilon$$

for all k and, hence, by monotonicity,

$$\sum_{j\in J_\varepsilon^c} \nu^\infty(I_j) < \varepsilon.$$

For fixed ε, since J_ε is finite,

$$\sum_{j\in J_\varepsilon} \nu^\infty(I_j) = \lim_{k\to\infty} \sum_{j\in J_\varepsilon} \nu^k(I_j).$$

Hence

$$\left|\sum_{j=1}^{\infty} \nu^\infty(I_j) - \lim_{k\to\infty} \sum_{j=1}^{\infty} \nu^k(I_j)\right| \leq \left|\sum_{j\in J_\varepsilon^c} \nu^\infty(I_j) - \lim_{k\to\infty} \sum_{j\in J_\varepsilon^c} \nu^k(I_j)\right| \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, (2.11) follows.

Because the cardinal of $\mathcal{B}_0(X^u)$ is $\mathbf{c}$, the cardinal of I is at most $\mathbf{c}$. By Assumption 1, it follows that the atomless measure ν^∞, which is defined on all subsets $\hat{I} \subset I$, must be trivial, i.e., that $\nu^\infty(\hat{I}) = 0$ for all $\hat{I} \subset I$.[7] In particular, $\nu^\infty(I) = 0$. By (2.5) and (2.6), therefore, $\lim_{k\to\infty} \mu(U_I^k) = 0$. By the monotonicity of the sequence $\{U_I^k\}_{k=1}^\infty$, one also has $\bigcap_{k=1}^\infty U_I^k \in \mathcal{B}_0(X^u)$ and therefore

$$\mu^* \left(\bigcap_{k=1}^\infty U_I^k \right) = \mu \left(\bigcap_{k=1}^\infty U_I^k \right) = \lim_{k\infty} \mu(U_I^k) = 0 \tag{2.12}$$

To complete the proof, I note that, for any $x \in U_I = \bigcup_{i\in I} B_i$, there exists $i \in I$ such that $\pi^k(x) \in \pi^k(B_i)$ for all k, hence $x \in U_I^k$ for all k . Thus $U_I \subset \bigcap_{k=1}^\infty U_I^k$. By the monotonicity of outer measure, it follows that $\mu^*(U_I) = 0$. ■

Proposition 2.4 *For any measure $\mu \in \mathcal{M}_0(X^u)$ and any family $\mathcal{G}$ of ρ^u-open balls covering X, there exists a countable subfamily $\mathcal{G}^* \subset \mathcal{G}$ such that $\mu(G^*) = 1$, where G^* is the union of the sets in $\mathcal{G}^*$.*

Proof. The proof follows along similar lines as the proof of Theorem 2, p. 235, in Billingsley (1968). Let μ and $\mathcal{G}$ be as specified in the proposition. Because X^u is a metric space, Theorem 4.21, p. 129, in Kelley (1955), implies that $\mathcal{G}$ has a σ-discrete refinement, i.e., there exists a family $\mathcal{H}$ of ρ^u-open sets (not necessarily balls) covering X such that, for every $H \in \mathcal{H}$, there exists $G(H) \in \mathcal{G}$ such that $H \subset G(H)$ and, moreover, $\mathcal{H}$ can be written as a countable union

$$\mathcal{H} = \bigcup_{t=1}^\infty \mathcal{H}_t$$

where, for any t, any two sets H_{ti}, H_{tj} in $\mathcal{H}_t$, there exists $\varepsilon_t > 0$ such that $\rho^u(x^i, x^j) \geq \varepsilon_t$ for all $x^i \in H_{ti}$ and all $x^j \in H_{tj}$.

For any t, let I_t be the set of indices i such that $H_{ti} \in \mathcal{H}_t$. Moreover, let I_t^* be the set of indices $i \in I_t$ such that $\mu(B_{ti}) > 0$ for some ρ^u-open

[7] The underlying idea of this proof is due to Banach (1930), see also the proof of Theorem I in Marczewski and Sikorski (1948).

ball $B_{ti} \subset H_{ti}$, and let $\mathcal{H}_t^*$ be the family of sets H_{ti}, $i \in I_t^*$. I claim that, for any t, $\mathcal{H}_t^*$ has at most countably many elements. To prove this claim, I note that $\mathcal{H}_t^*$ can be written as a countable union,

$$\mathcal{H}_t^* = \bigcup_{n=1}^{\infty} \mathcal{H}_{tn}^*,$$

where for any n,

$$\mathcal{H}_{tn}^* = \{H_{ti} \in \mathcal{H}_t^* | \mu(B_{ti}) > \frac{1}{n} \text{ for some } \rho^u\text{-open ball } B_{ti} \subset H_{ti}\}.$$

I claim that, for any n, the set $\mathcal{H}_{tn}^*$ has no more than n elements. For suppose that, contrary to the claim, $\mathcal{H}_{tn}^*$ has $n' > n$ elements, $H_{ti_1}, H_{ti_2}, ..., H_{ti_{n'}}$, and let $B_{ti_1}, B_{ti_2}, ..., B_{ti_{n'}}$ be the associated ρ^u-open balls. Because the sets $H_{ti_1}, H_{ti_2}, ..., H_{ti_{n'}}$ are disjoint, so are the associated ρ^u-open balls $B_{ti_1}, B_{ti_2}, ..., B_{ti_{n'}}$. The measure of the union of these balls is therefore equal to the sum $\sum_{j=1}^{n'} \mu(B_{ti_j}) > n' \cdot \frac{1}{n} > 1$. This conclusion is incompatible with the fact that this measure cannot exceed $\mu(X) = 1$. The assumption that $\mathcal{H}_{tn}^*$ has $n' > n$ elements has thus led to a contradiction and must be false. Hence $\mathcal{H}_{tn}^*$ has no more than n elements, and the countable union $\mathcal{H}_t^* = \bigcup_{n=1}^{\infty} \mathcal{H}_{tn}^*$ also has no more than countably many elements.

Now consider the union $\mathcal{H}^* = \bigcup_{t=1}^{\infty} \mathcal{H}_t^*$. As a countable union of countable sets, $\mathcal{H}^*$ is also countable. Recalling that, for each $H \in \mathcal{H}$, there exists $G(H) \in \mathcal{G}$ such that $H \subset G(H)$, define $\mathcal{G}^* \subset \mathcal{G}$ so that $G \in \mathcal{G}^*$ if and only if $G = G(H)$ for some $H \in \mathcal{H}^*$. Since $\mathcal{H}^*$ is countable, so is $\mathcal{G}^*$. Moreover, since the elements of $\mathcal{G}^*$ belong to $\mathcal{B}_0(X^u)$, so does the countable union

$$G^* = \bigcup_{H \in \mathcal{H}^*} G(H).$$

I claim that $\mu(G^*) = 1$. For suppose that $\mu(G^*) < 1$. Then also $\mu(X \backslash G^*) > 0$. (Since G^* belongs to $\mathcal{B}_0(X^u)$, so does $X \backslash G^*$.) Hence also $\mu^*(X \backslash G^*) > 0$, where μ^* is the outer measure defined by μ , in accordance with (2.1) above. Since $H \subset G(H)$ for all H, we also have $X \backslash G^* \subset X \backslash H^*$, where H^* is the union of the sets H_{ti} over all t and all $i \in I_t^*$. Thus, $\mu^*(X \backslash G^*) > 0$ implies $\mu^*(X \backslash H^*) > 0$. Because the refinement $\mathcal{H}$ of the family $\mathcal{G}$ covers X , the set $X\backslash H^*$ is a subset of the union $\hat{H}$ of the sets H_{ti} over all t and all $i \in \hat{I}_t$ where, for any t, $\hat{I}_t := I_t \backslash I_t^*$. By the monotonicity of outer measure, it follows that $\mu^*(X\backslash H^*) > 0$ implies $\mu^*(\hat{H}) > 0$.

I claim that $\mu^*(\hat{H}) > 0$ is incompatible with Lemma 2.3. For this purpose, for any t, let $\hat{H}_t$ be the union of the sets H_{ti} over all $i \in \hat{I}_t$. Because the sets H_{ti} for $i \in \hat{I}_t$ are at ρ^u-distance ε_t from each other, it follows that any ρ^u-open ball $\hat{B}_t$ that is contained in $\hat{H}_t$ is in fact contained in H_{ti} for some $i \in \hat{I}_t$. By the definition of $\hat{I}_t$, however, any ρ^u-open ball $\hat{B}_t$ that is contained in H_{ti} for some $i \in \hat{I}_t$.satisfies $\mu(\hat{B}_{ti}) = 0$. By Lemma 2.3, it follows that $\mu^*(\hat{H}_t) = 0$ for all t. By the countable subadditivity of outer measure,[8] therefore, $\mu^*(\hat{H}) > 0$.

The assumption that $\mu(G^*) < 1$ has thus led to a contradiction and must be false. ■

Proposition 2.2, i.e., the quasi-separability of the measures in $\mathcal{M}_0(X^u)$, follows immediately.

3 Metrizability of the Topology of Weak Convergence on $\mathcal{M}_0(X^u)$

Given the quasi-separability of measures in $\mathcal{M}_0(X^u)$, the proof that the topology of weak convergence on $\mathcal{M}_0(X^u)$ is metrizable is basically the same as the proof of the analogous result for Borel measures, e.g., Theorem 5, p. 238, in Billingsley (1968). Some modifications are needed to take account of the fact that $\mathcal{B}_0(X^u)$ is not a Borel σ-algebra and the measures in $\mathcal{M}_0(X^u)$ are quasi-separable, rather than separable.

The following lemma is fundamental to the argument.

Lemma 3.1 *If a set $B \subset X$ belongs to $\mathcal{B}_0(X^u)$, the set*

$$B^\varepsilon := \{x' \in X | \rho^u(x', x) < \varepsilon \text{ for some } x \in B\} \tag{3.1}$$

also belongs to $\mathcal{B}_0(X^u)$.

Proof. Consider the class $\mathcal{C} \subset \mathcal{B}_0(X^u)$ of sets for which the lemma is true. It is easy to see that $\mathcal{C}$ contains the finite-dimensional cylinder sets.

Moreover, $\mathcal{C}$ is closed under countable unions: If $B_r, r = 1, 2, ...,$ is any countable family of sets in $\mathcal{C}$, a point x belongs to the ε-neighbourhood of

[8] See Lemma 3.1.5, p. 89, in Dudley (2003).

$\cup_r B_r$ if and only if it belongs to B_r^ε for some r. The ε-neighbourhood of $\cup_r B_r$ is therefore equal to the union $\cup_r B_r^\varepsilon$. Since $B_r \in \mathcal{C}$ implies $B_r^\varepsilon \in \mathcal{B}_0(X^u)$ and $\mathcal{B}_0(X^u)$ is closed under countable unions, it follows that $\cup_r B_r^\varepsilon \in \mathcal{B}_0(X^u)$ and hence that $\cup_r B_r \in \mathcal{C}$.

Finally, $\mathcal{C}$ is also closed under countable intersectons: If $B_r, r = 1, 2, ...,$ is any countable family of sets in $\mathcal{C}$, a point x belongs to the ε-neighbourhood of $\cap_r B_r$ if and only if it belongs to B_r^ε for all r. The ε-neighbourhood of $\cap_r B_r$ is therefore equal to the intersection $\cap_r B_r^\varepsilon$. Since $B_r \in \mathcal{C}$ implies $B_r^\varepsilon \in \mathcal{B}_0(X^u)$ and $\mathcal{B}_0(X^u)$ is closed under countable intersections, it follows that $\cap_r B_r^\varepsilon \in \mathcal{B}_0(X^u)$ and hence that $\cap_r B_r \in \mathcal{C}$.

Since $\mathcal{B}_0(X^u)$ is the smallest σ-algebra that is closed under countable unions and countable intersections and that contains the ρ^u-open balls in X, it follows that $\mathcal{C} = \mathcal{B}_0(X^u)$. ■

Lemma 3.1 implies that, for any $\mu \in \mathcal{M}_0(X^u)$ and $B \in \mathcal{B}_0(X^u)$, not only $\mu(B)$ but also $\mu(B^\varepsilon)$ is well defined for any $\varepsilon > 0$. One can therefore define the Prohorov metric on $\mathcal{M}_0(X^u)$ by specifying the distance between any two measures μ and $\hat{\mu}$ in $\mathcal{M}_0(X^u)$ as the greatest lower bound on the set of $\varepsilon > 0$ such that

$$\mu(B) \leq \hat{\mu}(B^\varepsilon) + \varepsilon \text{ and } \hat{\mu}(B) \leq \mu(B^\varepsilon) + \varepsilon \tag{3.2}$$

for all sets $B \in \mathcal{B}_0(X^u)$.

Proposition 3.2 *Under Assumption 1, the topology of weak convergence on $\mathcal{M}_0(X^u)$ and the topology on $\mathcal{M}_0(X^u)$ that is induced by the Prohorov metric are equivalent.*

The proof of the analogous result in Billingsley (1968) comes in two distinct steps. The first step (Theorem 3, p. 236) specifies several families of sets of measures and shows that each family is a base for the topology of weak convergence. The second step uses this finding to prove the equivalence of the topology of weak convergence with the topology generated by the Prohorov metric.

The argument here has the same structure. For the first step of the argument, I note that the family $\mathcal{F}_0$ of sets taking the form

$$\left\{ \nu \in \mathcal{M}_0(X^u) | \left| \int_{X^u} f_i(x) d\nu(x) - \int_{X^u} f_i(x) d\mu(x) \right| < \varepsilon,\ i = 1, ..., k \right\} \tag{3.3}$$

for some $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and $f_1, ..., f_k$ in $CB_0(X^u)$ is a base for the topology of weak convergence on $\mathcal{M}_0(X^u)$. I also consider families $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ of sets taking the forms

$$\{\nu \in \mathcal{M}_0(X^u)|\ \nu(F_i) < \mu(F_i) + \varepsilon,\ i = 1, ..., k\} \tag{3.4}$$

for some $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and ρ^u-closed sets $F_1, ..., F_k$ belonging to $\mathcal{B}_0(X^u)$ in the case of $\mathcal{F}_1$,

$$\{\nu \in \mathcal{M}_0(X^u)|\ \nu(G_i) < \mu(G_i) + \varepsilon,\ i = 1, ..., k\} \tag{3.5}$$

for some $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and ρ^u-open sets $G_1, ..., G_k$ belonging to $\mathcal{B}_0(X^u)$ in the case of $\mathcal{F}_2$, and

$$\{\nu \in \mathcal{M}_0(X^u)|\ |\nu(A_i) - \mu(A_i)| < \varepsilon,\ i = 1, ..., k\} \tag{3.6}$$

for some $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and μ-continuity sets $A_1, ..., A_k$ belonging to $\mathcal{B}_0(X^u)$ in the case of $\mathcal{F}_3$. Each one of the families $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ is the base for a topology on $\mathcal{M}_0(X^u)$. The following result provides an analogue of Theorem 3, p. 236, in Billingsley (1968).

Proposition 3.3 *Each of the families $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ is a base for the topology of weak convergence on $\mathcal{M}_0(X^u)$.*

Proof. The proof has two parts. The first part shows that the topologies induced by the families $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ are equivalent to each other. The second part shows that the topology induced by the family $\mathcal{F}_1$ is equivalent to the topology induced by the family $\mathcal{F}_0$. For the first part, I refer to the argument of Billingsley (1968, p. 237), which goes through without any change. Because of Lemma 3.1, the requirement that the sets F_i, G_i, A_i must all belong to $\mathcal{B}_0(X^u)$ plays no role in the argument.

For the second part, the requirement that the sets F_i in $\mathcal{F}_1$ must belong to $\mathcal{B}_0(X^u)$ does play a role. Therefore I give the adapted argument in detail.

I first show that any element of $\mathcal{F}_1$ contains an element of $\mathcal{F}_0$ as a subset. Any set $N \in \mathcal{F}_1$ takes the form (3.4) for some $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and ρ^u-closed sets $F_1, ..., F_k$ in $\mathcal{B}_0(X^u)$. By Lemma 3.1, for any $\delta > 0$, the sets $F_i^\delta = \{x \in X^u | \rho^u(x, F_i) < \delta\}$, $i = 1, ..., k$, also belong to $\mathcal{B}_0(X^u)$. Let $\delta > 0$ be such that, for $i = 1, ..., k$, the set F_i^δ is a μ-continuity set and, moreover,

$$\mu(F_i^\delta) < \mu(F_i) + \frac{\varepsilon}{2}. \tag{3.7}$$

Next, let $\varphi : \mathbb{R}_+ \to [0,1]$ be a continuous function such that, for any $t \in \mathbb{R}$,

$$\varphi(t) = 1 - t \text{ if } t \in [0,1) \text{ and } \varphi(t) = 1 \text{ if } t \geq 1.$$

For any i, define a function $f_i : X^u \to [0,1]$ by setting

$$f_i(x) = \varphi\left(\frac{1}{\delta} \cdot \rho^u(x, F_i)\right), \tag{3.8}$$

where again $\rho^u(x, F_i) = \min_{\hat{x} \in F_i} \rho^u(x, \hat{x})$. The function f_i is obviously bounded and ρ^u-continuous. Since $F_i \in \mathcal{B}_0(X^u)$, By Lemma 3.1, f_i is also measurable with respect to $\mathcal{B}_0(X^u)$.[9] We also have $f_i(x) = 0$ for $x \in X^u \backslash F_i^\delta$ and $f_i(x) = 1$ for $x \in F_i$. For any $\nu \in \mathcal{M}_0(X^u)$, therefore,

$$\int_{X^u} f_i(x) d\nu(x) < \int_{X^u} f_i(x) d\mu(x) + \frac{\varepsilon}{2}$$

implies

$$\nu(F_i) \leq \int_{X^u} f_i(x) d\nu(x) < \int_{X^u} f_i(x) d\mu(x) + \frac{\varepsilon}{2} \leq \mu(F_i^\delta) + \frac{\varepsilon}{2} < \mu(F_i) + \varepsilon$$

For any $\nu \in \mathcal{M}_0(X^u)$ satisfying

$$\left| \int_{X^u} f_i(x) d\nu(x) - \int_{X^u} f_i(x) d\mu(x) \right| < \frac{\varepsilon}{2}$$

for $i = 1, ..., k$, we therefore have

$$\nu(F_i) < \mu(F_i) + \varepsilon$$

for $i = 1, ..., k$. Thus the set $N \in \mathcal{F}_1$ of measures $\nu \in \mathcal{M}_0(X^u)$ that satisfy (3.4) for the given $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and $F_1, ..., F_k$ contains a set measures $\nu \in \mathcal{M}_0(X^u)$ that satisfy (3.3) for the same $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and the specified functions $f_i, i = 1, ..., k$. The latter set an element of $\mathcal{F}_0$. Thus every element of $\mathcal{F}_1$ contains an element of $\mathcal{F}_0$.

For the claim that every element of $\mathcal{F}_0$ also contains an element of $\mathcal{F}_1$, the argument in Billingsley (1968) applies with hardly any change. For example, let $N \in \mathcal{F}_0$ be the set of measures $\nu \in \mathcal{M}_0(X^u)$ such that, for given $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and $f \in CB_0(X^u)$,

$$\left| \int_{X^u} f(x) d\nu(x) - \int_{X^u} f(x) d\mu(x) \right| < \varepsilon. \tag{3.9}$$

[9] $f_i(x) > f_i(\hat{x})$ implies $\rho^u(x, F_i) < \rho^u(\hat{x}, F_i)$ so that, for some $\eta \in (\rho^u(x, F_i), \rho^u(\hat{x}, F_i))$, $x \in F_i^\eta$ and $\hat{x} \in X^u \backslash F_i^\eta$.

Without loss of generality, assume that f takes values in the unit interval. Choose k' so that $\frac{1}{k'} < \varepsilon$ and, for $i = 1, ..., k'$, let $F_i = \{x \in X^u | \frac{i}{k'} \leq f(x)\}$. Then, for any i, F_i is ρ^u-closed. Moreover, since f is $\mathcal{B}_0(X^u)$-measurable, F_i is in $\mathcal{B}_0(X^u)$. By standard arguments, therefore, $\nu(F_i) < \mu(F_i) + \varepsilon$ implies

$$\int_{X^u} f(x)d\nu(x) < \frac{1}{k'}+\frac{1}{k'}\sum_{i=1}^{k'}\nu(F_i) < \frac{1}{k'}+\frac{1}{k'}\sum_{i=1}^{k'}\mu(F_i)+\varepsilon < \int_{X^u} f(x)d\mu(x)+2\varepsilon.$$

By a parallel argument for the function $x \to 1 - f(x)$, there also exist ρ^u-closed, $\mathcal{B}_0(X^u)$-measurable sets F_i, $i = k' + 1, ..., 2k'$, such that, for $i = k' + 1, ..., 2k'$, $\nu(F_i) < \mu(F_i) + \varepsilon$ implies

$$\int_{X^u} (1 - f(x))d\nu(x) < \int_{X^u} (1 - f(x))d\mu(x) + 2\varepsilon,$$

or, equivalently,

$$\int_{X^u} f(x)d\mu(x) < \int_{X^u} f(x)d\nu(x) + 2\varepsilon.$$

Upon combining these arguments, one finds that the set N of measures $\nu \in \mathcal{M}_0(X^u)$ that satisfy (3.9) for some given $\mu \in \mathcal{M}_0(X^u), \varepsilon > 0$, and $f \in CB_0(X^u)$, an element of $\mathcal{F}_0$, contains a subset consisting of measures that satisfy $\nu(F_i) < \mu(F_i) + \varepsilon$ for the given $\mu, \varepsilon, k = 2k'$, and ρ^u-closed, $\mathcal{B}_0(X^u)$-measurable sets $F_i, i = 1, ..., k$. Thus N has a subset that belongs to $\mathcal{F}_1$.

By taking intersections of sets like N, with different functions f, the argument can be generalized to all sets in $\mathcal{F}_0$. Thus every element of $\mathcal{F}_0$ contains an element of $\mathcal{F}_1$ as a subset. This completes the proof of the claim that the topologies induced by $\mathcal{F}_0$ and $\mathcal{F}_1$ are equivalent. ■

Proposition 3.4 *For any measure $\mu \in \mathcal{M}_0(X^u)$ that is quasi-separable, the topology induced by the Prohorov metric and the topology of weak convergence on $\mathcal{M}_0(X^u)$ are equivalent at μ.*

Proof. Let $N(\mu)$ be any $\mathcal{F}_1$-neighbourhood of μ, characterized by $\varepsilon > 0$ and ρ^u-closed, $\mathcal{B}_0(X^u)$-measurable sets $F_i, i = 1, ..., k$, such that $\nu \in N(\mu)$ if and only if ν satisfies (3.4) for μ, ε, and $F_1, ..., F_k$. Let $\delta < \varepsilon$ be such that, for $i = 1, ..., k$, the set $F_i^\delta = \{x \in X^u | \rho^u(x, F_i) < \delta\}$ is a μ-continuity set and, moreover, $\mu(F_i^\delta) < \mu(F_i) + \frac{\varepsilon}{2}$. For any $\nu \in \mathcal{M}_0(X^u)$ such that $p(\nu, \mu) < \delta$, one has

$$\nu(F_i) < \mu(F_i^\delta) + \delta < \mu(F_i) + \varepsilon,$$

so the p-open ball with radius δ around μ is a subset of $N(\mu)$. Therefore the topology on $\mathcal{M}_0(X^u)$ that is induced by the Prohorov metric is at least as large at μ as the topology of weak convergence.

Next, suppose that μ is quasi-separable. I will show that for any $\varepsilon > 0$, the p-open ball with radius ε around μ contains an $\mathcal{F}_3$-neighbourhood $N(\mu)$ of μ. Fix a cover of X^u by ρ^u-open, μ-continuity balls with diameters less than δ, where $\delta < \frac{\varepsilon}{3}$. Appealing to quasi-separability, pass to a countable subfamily $\{B_i\}_{i=1}^{\infty}$ such that $\mu(\cup_i B_i) = 1$. Construct disjoint μ-continuity stes $A_1, A_2, ...$ by setting $A_1 = B_1$ and, for $i > 1$, $A_i = B_i \backslash \cup_{j<i} A_j$. Choose k so that

$$\mu(\cup_{i=1}^{k} A_i) > 1 - \delta \tag{3.10}$$

and let $\mathcal{A}$ be the set of unions of the sets A_i over subsets of the indices $i = 1, ..., k$. Then each $A \in \mathcal{A}$ is a μ-continuity set, and, by Proposition 3.3, there is a neighbourhood $N(\mu) \in \mathcal{F}_3$ of μ such that, for any $\nu \in N(\mu)$,

$$|\nu(A) - \mu(A)| < \delta \quad \text{for all} \quad A \in \mathcal{A}. \tag{3.11}$$

Iclaim that $N(\mu)$ is contained in the p-open ball with radius ε around μ. To prove this claim, consider any $B \in \mathcal{B}_0(X^u)$. Let I_B be the set of indices i such that $A_i \cap B \neq \emptyset$ and let $A_B := \cup_{i \in I_B} A_i$. Then $B \subset A_B \cup (X^u \backslash \cup_{i=1}^{k} A_i)$. Moreover, $A_B \in \mathcal{A}$ and, because the sets A_i all have diameters less than δ, $A_B \subset B^\delta$. Using (3.10) and (3.11), one obtains

$$\mu(B) \leq \mu(A_B) + 1 - \mu(\cup_{i=1}^{k} A_i) < \nu(A_B) + 2\delta \leq \nu(B^\delta) + 2\delta < \nu(B^\varepsilon) + \varepsilon.$$

Similiarly, taking account of the fact that (3.10) and (3.11) imply

$$\nu(\cup_{i=1}^{k} A_i) > 1 - 2\delta,$$

one also obtains

$$\nu(B) \leq \nu(A_B) + 1 - \nu(\cup_{i=1}^{k} A_i) < \mu(A_B) + 3\delta \leq \mu(B^\delta) + 3\delta < \mu(B^\varepsilon) + \varepsilon.$$

Thus, for any $\nu \in N(\mu)$, the Prohorov distance between ν and μ is less than ε. The specified set $N(\mu) \in \mathcal{F}_3$ is contained in the p-open ball with radius ε around μ. At μ, therefore the topology of weak convergence on $\mathcal{M}_0(X^u)$ is at least as large as the topology that is induced by the Prohorov metric. Given that the topology induced by the Prohorov metric is also at least as large at μ as the topology of weak convergence, it follows that the two topologies are equivalent at μ. ■

Now Proposition 3.2 is a straightforward corollary to Propositions 2.2 and 3.4.

4 Why Bother about $\mathcal{B}_0(X^u)$ and $\mathcal{M}_0(X^u)$?

Dudley's initiative in introducing the σ-algebra $\mathcal{B}_0(X^u)$ and the space of measures $\mathcal{M}_0(X^u)$ was not much followed. For the purpose that he was interested in, the σ-algebra induced by the Skorokhod topology and the associated space of Borel measures served the same purpose. Moreover, whereas the uniform topology, as well as $\mathcal{B}_0(X^u)$ and the space of measures $\mathcal{M}_0(X^u)$, are conceptually simpler, they suffer from the inconveniences of having to deal with non-Borel measures.

Recently, however, the issue that the Borel σ-algebra $\mathcal{B}(X^u)$ that is induced by the ρ^u-topology on X is "too large" has come up in another context. Recent developments in the theory of strategic games of incomplete information suggest that, for the product spaces under consideration there, a reliance on the σ-algebra $\mathcal{B}_0(X^u)$ and the space of measures $\mathcal{M}_0(X^u)$ may be necessary.

Incompleteness of information in a strategic game concerns not only an agent's information (probabilistic beliefs) about the exogenous parameters of the game, but also the agent's information about the other agents' information about the exogenous parameters of the game, the agent's information about the other agents' infomation about everybody else' information about the exogenous parameters of the game, and so on, in a seemingly intractable infinite regress.

To avoid having to deal with the seemingly intractable hierarchy of different orders of probabilistic beliefs about exogenous parameters and other agents' beliefs of different orders, Harsanyi (1967/68) introduced the notion of an agent's *abstract type*, together with a *belief function* that maps the agent's abstract types into probability measures over exogenous parameters and other agents' abstract types. This formalism provides a workable apparatus for the analysis of strategic interdependence with incomplete information in many areas of applied work. However, it suffers from the fact that the notion of an abstract type space is a black box and the notion of beliefs as type-determined probability measures over exogenous data and other agents' abstract types involves an element of circularity.

These shortcomings were remedied by Mertens and Zamir (1985), Brandenburger and Dekel (1993), and Heifetz (1993). These authors showed that one can actually interpret any hierarchy of beliefs of different orders of an agent as a "type" in the Harsanyi sense. Moreover, Kolmogorov's extension theorem implies that any belief hierarchy in which the beliefs of different orders are mutually consistent determines a unique measure on the product σ-algebra for space of exogenous parameters and of other agents' belief hierarchies. The infinite regress in belief formation is shown to be harmless when the model is formulated in terms of infinite-dimensional products of spaces of beliefs of different orders.

If the spaces of consistent belief hierarchies are given the product topology and if the space of probability measures on exogenous parameters and other agents' belief hierarchies is given the associated topology of weak convergence, the mapping from the space of an agent's belief hierarchies to probability measures over exogenous parameters and other agents' belief hierarchies that is given by Kolmogorov's theorem is a homeomorphism. Moreover, for any abstract type space model à la Harsanyi, if belief functions are continuous, there is a natural embedding of the different agents' abstract type spaces in the corresponding spaces of belief hierarchies with the product topology.

The product topology on the space of belief hierarchies is well suited to capturing upper hemi-continuity properties of optimal choices, e.g., best responses to other agents' strategies; it is also well suited to capturing upper hemi-continuity properties of outcomes under various equilibrium concepts (Bayes-Nash or rationalizable). But it is not well suited to capturing lower hemi-continuity properties of strictly ε-rationalizable outcomes. The reason is that the product topology gives decreasing weight to beliefs of very high order in the hierarchy of beliefs, but in some games, optimizing or ε-optimizing behaviour can be sensitive to changes in beliefs of arbitrarily high orders in the hierarchy.[10]

This shortcoming of the product topology has induced Dekel et al. (2006) and Chen et al. (2010, 2017) to propose finer topologies for the spaces of belief hierarchies. Chen et al. treat the space of an agent's consistent belief hierarchies as a subspace of the product of the spaces of beliefs of different orders and impose a uniform metric on this space. Dekel et al. (2006) specify finer topologies directly in terms of the desired continuity properties of strategic behaviour, but Chen et al. (2010, 2017) show that these topologies can be reinterpreted in terms of their own *uniform weak* topology.

What then becomes of the Mertens-Zamir reconciliation of an approach based on belief hierarchies with the Harsanyi abstract-type-space approach? In particular, what becomes of the proposition that Kolmogorov's extension theorem can be understood as defining a homeomorphism between the space of belief hierarchies and the space of probability measures on exgenous parameters and other agents' belief hierarchies?

[10]See Dekel et al. (2006, 2007), Chen et al. (2010, 2017).

This is not just a question about the impact of having a finer topology for the domain of the mapping. It is also a question about the impact of having a finer topology on the range; indeed, it is even a question about the specification of the range itself. Kolmogorov's extension theorem gives measures on the product σ-algebra for the space of exogenous parameters and other agents' belief hierarchies or, equivalently measures on the Borel σ-algebra for the product topology. The Borel σ-algebra for the uniform weak topology is strictly larger.

How then does one deal with the discrepancy between the domain of the measures defined by Kolmogorov's theorem? One approach might be to try and extend the measures to the larger σ-algebra. Because the infinite-dimensional product space with a uniform metric is non-separable, the first approach does not work. A measure on the product σ-algebra can always be extended to a measure on the Borel σ-algebra for the finer topology but the extension need not be unique and it need not be countably additive.[11]

An alternative approach restricts the measures in question to a σ-algebra on X that is coextensive with the product σ-algebra. The σ-algebra $\mathcal{B}_0(X^u)$ that is generated by the ρ^u-open spheres in X serves this purpose very well because it is in fact coextensive with the product σ-algebra $\mathcal{B}(X)$. The space $\mathcal{M}_0(X^u)$ of probability measures on $(X^u, \mathcal{B}_0(X^u))$ is coextensive with the space $\mathcal{M}(X)$ of probability measures on $(X, \mathcal{B}(X))$. To make use of this observation in studying the relation between the space of belief hierarchies of an agent and the space of probability measures over exogenous parameters and other agents' belief hierarchies, one needs to understand the topology of weak convergence on $\mathcal{M}_0(X^u)$.

[11] For the existence of a finitely additive extension, see Bachman and Sultan (1980). The sufficient conditions they give for uniqueness and countable additivity are not satisfied here. To see the problem, set $X_k = \{0, 1\}$ in (1), and let $\mu \in \mathcal{M}(X)$ be the product measure that assigns probability $\frac{1}{2}$ to each of the events $\{x_k = 0\}$ and $\{x_k = 1\}$, for all k. By Theorem 2.1, p. 290, of Bachman and Sultan (1980), this measure can be extended to a finitely additive measure on the algebra generated by the lattice of ρ^u-closed subsets of X. However, the values of the extension for the set A_0 of sequences $x \in X$ such that $x_k = 1$ no more than finitely many times and for the set A_1 of sequences $x \in X$ such that $x_k = 0$ no more than finitely many times are to some extent arbitrary because, in the product topology on X, the empty set is the largest closed set that is contained in A_0 or A_1 and X itself is the smallest closed set that contains A_0 or A_1 . Thus the extended measure μ^* might have $\mu^*(A_0) = 0$ and $\mu^*(A_1) = 1$ or $\mu^*(A_0) = 1$ and $\mu^*(A_1) = 0$. As for the possibility that μ^* might be countably additive, notice that the ρ^u-topology on $\{0, 1\}^\infty$ is equivalent to the discrete topology, so the associated Borel σ-algebra is the power set. Moreover, μ^* assigns probability zero to the singletons $\{x\} \subset \{0, 1\}^\infty$. Countable additivity of μ^* would be incompatible with the axiom that the cardinal $\mathbf{c}$ of the continuum is not atomlessly measurable.

This is where Propositions 2.2 and 3.2 are useful. In Hellwig (2016/2020), I use these propositions to show that, with the uniform weak topology on the spaces of belief hierarchies, Kolmogorov's extension theorem still provides a homeomorphism between the space of an agent's belief hierarchies and the space of probability measures over exogenous parameters and other agents' belief hierarchies (with the topology of weak convergence). Without going into the game theoretic details, the argument boils down down to the following.

For $\ell = 1, 2, ...$, let π^ℓ be the projection from X to the finite product

$$X^\ell = \prod_{k=1}^{\ell} X_k. \tag{4.1}$$

For any $\mu \in \mathcal{M}_0(X^u)$ and any ℓ, let

$$\Pi^\ell(\mu) = \mu \circ (\pi^\ell)^{-1} \in \mathcal{M}\left(X^\ell\right), \tag{4.2}$$

and let

$$\Pi^\infty(\mu) = (\Pi^1(\mu), \Pi^2(\mu), ...). \tag{4.3}$$

Proposition 4.1 *Suppose that each of the spaces $\mathcal{M}\left(X^\ell\right)$, $\ell = 1, 2, ...,$is endowed with the Prohorov metric p_ℓ, and that the product space $\prod_{\ell=1}^{\infty} \mathcal{M}\left(X^\ell\right)$ is endowed with the uniform metric d^u such that, for any two sequences $\{\mu^\ell\}_{\ell=1}^\infty, \{\hat{\mu}^\ell\}_{\ell=1}^\infty$,*

$$d^u(\{\mu^\ell\}_{\ell=1}^\infty, \{\hat{\mu}^\ell\}_{\ell=1}^\infty) = \sup_\ell d_\ell(\mu^\ell, \hat{\mu}^\ell). \tag{4.4}$$

If $\mathcal{M}_0(X^u)$ has the topology of weak convergence, then, under Assumption 1, the mapping Π^∞ is a homeomorphism between $\mathcal{M}_0(X^u)$ and the subspace $H^u \subset \prod_{\ell=1}^{\infty} \mathcal{M}\left(X^\ell\right)$ that consists of those sequences $\{\mu^\ell\}_{\ell=1}^\infty$ that are mutually consistent in that $\mu^{\ell+1} \circ (\pi^\ell)^{-1} = \mu^\ell$ for all ℓ.

Proof. The mapping Π^∞ from $\mathcal{M}_0(X^u)$ to $H^u \subset \prod_{\ell=1}^{\infty} \mathcal{M}\left(X^\ell\right)$ is obviously injective and onto. By Kolmogorov's extension theorem, the inverse $(\Pi^\infty)^{-1}$ from H^u to $\mathcal{M}_0(X^u)$ is also injective and onto. Therefore it suffices to show that Π^∞ and $(\Pi^\infty)^{-1}$ are both continuous.

Continuity of Π^∞ is straightforward: For any ℓ and any set $W^\ell \in \mathcal{B}(X^\ell)$, the cylinder set

$$\hat{W}^\ell = W^\ell \times X_{\ell+1} \times X_{\ell+2} \times ...$$

belongs to $\mathcal{B}_0(X^u)$, and, for any $\varepsilon > 0$, the cylinder set

$$\hat{W}^{\ell\varepsilon} = (W^\ell)^\varepsilon \times X_{\ell+1} \times X_{\ell+2} \times ...$$

that is defined by the ε-neighbourhood $(W^\ell)^\varepsilon$ of W^ℓ in X^ℓ is an ε-neighbourhood of $\hat{W}^\ell$ in X^u. If the Prohorov distance between two measures $\mu, \hat{\mu}$ in $\mathcal{M}_0(X^u)$ is less than ε, we have

$$\mu(\hat{W}^\ell) < \hat{\mu}(\hat{W}^{\ell\varepsilon}) + \varepsilon$$

and

$$\hat{\mu}(\hat{W}^\ell) < \mu(\hat{W}^{\ell\varepsilon}) + \varepsilon.$$

By the definition of the marginal distributions $\mu^\ell = \Pi^\ell(\mu)$ and $\hat{\mu}^\ell = \Pi^\ell(\hat{\mu})$, it follows that

$$\mu^\ell(W^\ell) < \hat{\mu}^\ell(W^{\ell\varepsilon}) + \varepsilon$$

and

$$\hat{\mu}^\ell(W^\ell) < \mu^\ell(W^{\ell\varepsilon}) + \varepsilon.$$

Because the choice of $W^\ell \in \mathcal{B}(X^\ell)$ was arbitrary, it follows that the Prohorov distance $d_\ell(\mu^\ell, \hat{\mu}^\ell)$ between $\mu^\ell = \Pi^\ell(\mu)$ and $\hat{\mu}^\ell = \Pi^\ell(\hat{\mu})$ in $\Pi^\ell(\mathcal{M}_0(X^u))$ is no greater than ε. Since ε may be taken to be arbitrarily close to the Prohorov distance $p(\mu, \hat{\mu})$ between μ and $\hat{\mu}$ in $\mathcal{M}_0(X^u)$, it follows that the Prohorov distance between $\mu^\ell = \Pi^\ell(\mu)$ and $\hat{\mu}^\ell = \Pi^\ell(\hat{\mu})$ in $\Pi^\ell(\mathcal{M}_0(X^u))$ is no greater than $p(\mu, \hat{\mu})$. Since this is true for all ℓ, it follows that

$$d^u(\Pi^\infty(\mu), \Pi^\infty(\hat{\mu})) = \sup_\ell d_\ell(\Pi^\ell(\mu), \Pi^\ell(\hat{\mu})) \leq p(\mu, \hat{\mu}).$$

Continuity of the map Π^∞ from $\mathcal{M}_0(X^u)$ to H^u follows immediately.

Next, consider the map $\beta := (\Pi^\infty)^{-1}$ from H^u to $\mathcal{M}_0(X^u)$ that is given by Kolmogorov's extension theorem. Proceeding indirectly, suppose that β is not continuous. Then there exist sequences $h^r = \{\mu^{\ell r}\}_{\ell=1}^\infty, r = 1, ..., \infty$, and $h = \{\mu^\ell\}_{\ell=1}^\infty$ in H^u such that h^r converges to $h \in H^u$ but $\beta(h^r)$ does not converge to $\beta(h)$ in $\mathcal{M}_0(X^u)$. Convergence of h^r to h implies that

$$\lim_{r\to\infty} \sup_\ell d_\ell(\mu^{\ell r}, \hat{\mu}^\ell) = 0. \tag{4.5}$$

Non-convergence of $\beta(h^r)$ to $\beta(h)$ implies that, for some $\varepsilon > 0$ and some subsequence $\{h^{r'}\}$ of $\{h^r\}$,

$$p(\beta(h^r), \beta(h)) \geq \varepsilon$$

for all r'. Thus, for every r', there exists a set $B_{r'} \in \mathcal{B}_0(X^u)$ such that

$$\beta(B_{r'}|h^r) > \beta(B_{r'}^\varepsilon|h) + \varepsilon \tag{4.6}$$

or

$$\beta(B_{r'}|h) > \beta(B_{r'}^\varepsilon|h^r) + \varepsilon, \tag{4.7}$$

where $B_{r'}^\varepsilon \in \mathcal{B}_0(X^u)$ is the ε-neighbourhood of $B_{r'}$ in X^u.

For any ℓ, let $B_{r'}^\ell = \pi^\ell(B_{r'})$ be the projection of $B_{r'}$ to $X^\ell = \pi^\ell(X^u)$ and let $(B_{r'}^\ell)^\varepsilon$ be an ε-neighbourhood of $B_{r'}^\ell$ in X^ℓ. Let

$$\hat{B}_{r'}^\ell := B_{r'}^\ell \times X_{\ell+1} \times X_{\ell+2} \times ...$$

and

$$(\hat{B}_{r'}^\ell)^\varepsilon := (B_{r'}^\ell)^\varepsilon \times X_{\ell+1} \times X_{\ell+2} \times ...$$

be the cylinder sets in X^u that are defined by $B_{r'}^\ell$ and $(B_{r'}^\ell)^\varepsilon$. One easily verifies that the sequences $\{\hat{B}_{r'}^\ell\}$ and $\{(\hat{B}_{r'}^\ell)^\varepsilon\}$ are nonincreasing and that

$$B_{r'} = \bigcap_{\ell=1}^{\infty} \hat{B}_{r'}^\ell \quad \text{and} \quad B_{r'}^\varepsilon = \bigcap_{\ell=1}^{\infty} (\hat{B}_{r'}^\ell)^\varepsilon \tag{4.8}$$

for all r'. By elementary measure theory,[12] for any r' and any $\delta > 0$, there exists an integer $L^{r'}(\delta)$ such that, for $\ell > L^{r'}(\delta)$,

$$\beta(B_{r'}^\varepsilon|h) \geq \beta((\hat{B}_{r'}^\ell)^\varepsilon|h) - \delta \tag{4.9}$$

and

$$\beta(B_{r'}^\varepsilon|h^{r'}) \geq \beta((\hat{B}_{r'}^\ell)^\varepsilon|h^{r'}) - \delta. \tag{4.10}$$

Moreover, by (4.8),

$$\beta(B_{r'}|h) \leq \beta(\hat{B}_{r'}^\ell|h) \tag{4.11}$$

and

$$\beta(B_{r'}|h^{r'}) \leq \beta(\hat{B}_{r'}^\ell|h^{r'}). \tag{4.12}$$

Set $\delta = \frac{\varepsilon}{2}$ and combine (4.9) - (4.12) with (4.6) and (4.7). Thereby one finds that, for all r', all $\delta > 0$, and all $\ell > L^{r'}(\delta)$, either

$$\beta(\hat{B}_{r'}^\ell|h^{r'}) > \beta((\hat{B}_{r'}^\ell)^\varepsilon|h) + \frac{\varepsilon}{2} \tag{4.13}$$

or

$$\beta(\hat{B}_{r'}^\ell|h) > \beta((\hat{B}_{r'}^\ell)^\varepsilon|h^{r'}) + \frac{\varepsilon}{2}. \tag{4.14}$$

[12] Theorem 3.1.1, p. 86, in Dudley (2002).

Since $(\hat{B}^{\ell}_{r'})^{\frac{\varepsilon}{2}} \subset (\hat{B}^{\ell}_{r'})^{\varepsilon}$, it follows that, for all r', all $\delta > 0$, and all $\ell > L^{r'}(\delta)$, either

$$\beta(\hat{B}^{\ell}_{r'}|h^{r'}) > \beta((\hat{B}^{\ell}_{r'})^{\frac{\varepsilon}{2}}|h) + \frac{\varepsilon}{2} \tag{4.15}$$

or

$$\beta(\hat{B}^{\ell}_{r'}|h) > \beta((\hat{B}^{\ell}_{r'})^{\frac{\varepsilon}{2}}|h^{r'}) + \frac{\varepsilon}{2}. \tag{4.16}$$

By the definition of β as the inverse of $\Pi^{\infty} = (\Pi^1, \Pi^2, ...)$, and the cylinder nature of the sets $\hat{B}_{r'}$ and $(\hat{B}^{\ell}_{r'})^{\frac{\varepsilon}{2}}$, we also have

$$\beta(\hat{B}^{\ell}_{r'}|h^{r'}) = \Pi^{\ell}(B^{\ell}_{r'}|\beta(h^{r'})) = \mu^{\ell r'}(B^{\ell}_{r'}), \tag{4.17}$$

$$\beta(\hat{B}^{\ell}_{r'}|h) = \Pi^{\ell}(B^{\ell}_{r'}|\beta(h)) = \mu^{\ell}(B^{\ell}_{r'}), \tag{4.18}$$

$$\beta((\hat{B}^{\ell}_{r'})^{\frac{\varepsilon}{2}}|h^{r'}) = \Pi^{\ell}((B^{\ell}_{r'})^{\frac{\varepsilon}{2}}|\beta(h^{r'})) = \mu^{\ell r'}((B^{\ell}_{r'})^{\frac{\varepsilon}{2}}), \tag{4.19}$$

$$\beta((\hat{B}^{\ell}_{r'})^{\frac{\varepsilon}{2}}|h) = \Pi^{\ell}((B^{\ell}_{r'})^{\frac{\varepsilon}{2}}|\beta(h)) = \mu^{\ell}((B^{\ell}_{r'})^{\frac{\varepsilon}{2}}). \tag{4.20}$$

Thus, (4.15) and **??**) can be rewritten as

$$\mu^{\ell r'}(B^{\ell}_{r'}) > \mu^{\ell}((B^{\ell}_{r'})^{\frac{\varepsilon}{2}}) + \frac{\varepsilon}{2} \tag{4.21}$$

and

$$\mu^{\ell}(B^{\ell}_{r'}) > \mu^{\ell r'}((B^{\ell}_{r'})^{\frac{\varepsilon}{2}}) + \frac{\varepsilon}{2}, \tag{4.22}$$

and one of (4.21), (4.22) must hold if r', $\delta = \frac{\varepsilon}{2}$, and $\ell > L^{r'}(\delta)$. But then, for such r', δ, and ℓ, $d_{\ell}(\mu^{\ell r'}, \hat{\mu}^{\ell}) \geq \frac{\varepsilon}{2}$, contrary to (4.5). The assumption that $\beta = \Pi^{\infty}$ is not continuous has thus led to a contradiction and must be false.
■

References

[1] Bachman, G., and A. Sultan (1980), On Regular Extensions of Measures, *Pacific Journal of Mathematics* 86 (2), 389-395.

[2] Banach, S. (1930), Über additive Maßfunktionen in abstrakten Mengen, *Fundamenta Mathematicae* 15, 97-101.

[3] Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

[4] Brandenburger, A., and E. Dekel (1993), Hierarchies of Beliefs and Common Knowledge, *Journal of Economic Theory* 59, 189 - 198.

[5] Chen, Y., A. Di Tillio, E. Faingold, and S. Xiong (2010), Uniform Topologies on Types, *Theoretical Economics* 5, 445 -478.

[6] Chen, Y., A. Di Tillio, E. Faingold, and S. Xiong (2017), Characterizing the Strategic Impcat of Misspecified Beliefs, *Review of Economic Studies* 84 (4), 1424-1471.

[7] Dekel, E., D. Fudenberg, and S. Morris (2006), Topologies on Types, *Theoretical Economics* 1, 275 - 309.

[8] Dekel, E., D. Fudenberg, and S. Morris (2007), Interim Correlated Rationalizability, *Theoretical Economics* 2, 15 -40.

[9] Dudley, R.M. (1966), Weak Convergence of Probability Measures on Nonseparable Metric Spaces and Empirical Measures on Euclidean Spaces, *Illinois Journal of Mathematics* 10, 109-126.

[10] Dudley, R.M. (1967), Measures on Non-Separable Metric Spaces, *Illinois Journal of Mathematics* 11, 449-453.

[11] Dudley, R.M. (1978), Central Limit Theorems for Empirical Measures, *Annals of Probability* 6, 899-929.

[12] Dudley, R.M. (2002), *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics 74, Cambridge University Press, Cambridge, UK.

[13] Harsanyi, J.C. (1967–68), Games with incomplete information played by 'Bayesian' players, I–III. *Management Science* 14, 159–182, 320–334, 486–502.

[14] Hellwig, M.F. (2016/20), A Homeomorphism Theorem for the Universal Type Space with the Uniform Weak Topology, Preprint 17/2016, Max Planck Institute for Research on Collective Goods, Bonn, revised December 2020. http://www.coll.mpg.de/pdf_dat/2016_17online.pdf.

[15] Kelley, J.L. (1955), *General Topology*, Van Nostrand, New York.

[16] Marczewski, E., and R. Sikorski (1948), Measures in Non-Separable Metric Spaces, *Colloquium Mathematicum 1, 133-139.*

[17] Mertens, J.F., and S. Zamir (1985), Formulation of Bayesian Analysis for Games with Incomplete Information, *International Journal of Game Theory* 14, 1–29.

[18] Pollard, D. (1979), Weak Convergence on Non-Separable Metric Spaces, Journal of the Australian Mathematical Society (Series A) 28, 197-204.

[19] Wichura, M.J. (1970), On the Construction of Almost Uniformly Convergent Random Variables with Given Weakly Convergent Image Laws, *Annals of Mathematical Statistics* 41 (1), 284-291.