Caltech Division of the Humanities and Social Sciences

Expository Notes on the Kolmogorov Extension Problem*

KC Border

Fall 1997, Revised May 1998

Harsanyi [8] introduced the notion of "types" of players as a way of reducing games of incomplete information to games of imperfect information. The Kolmogorov extension theorem has been used to represent types as "infinite hierarchies of beliefs" with varying degrees of explicitness by Böge and Eisele [4], Brandenburger and Dekel [5], and Mertens and Zamir [11]. Recent work by Heifetz and Samet [9, 10] shows the extension theorem is crucial for this interpretation. These notes describe the Kolmogorov extension problem and present an important counterexample (Proposition 17) due to Andersen[1] and Jessen [2].

1 The Kolmogorov extension problem

The usual statement of the Kolmogorov extension theorem is that if the joint distributions of the random variables $X_1, \dots, X_n$, $n = 1, 2, \dots$ are consistent, then it is possible to define a probability measure on the infinite product space $R^{\mathbb{N}}$ that agrees with all of the finite dimensional distributions. But it is easier (as is so often the case) to start with a more abstract problem.

The abstract version of the Kolmogorov extension problem is this. Consider a set X and a net $\{\Sigma_\nu : \nu \in \mathfrak{I}\}$ of σ-algebras of subsets of X, directed upward by inclusion $\supset$. That is, for any ν and μ in $\mathfrak{I}$ there is some γ belonging to $\mathfrak{I}$ such that $\Sigma_\gamma \supset \Sigma_\nu$ and $\Sigma_\gamma \supset \Sigma_\mu$.

*In Sections 5 through 8 I borrow heavily from Halmos's *Measure Theory* [7, pp. 68–70], but so does everyone. In particular, this same material may also be found in Dudley [6, pp. 80–81]. I have rewritten the material to avoid the mention of inner measure, but this is essentially just a detailed working out of the exercises in Halmos [7, p. 214] and Dudley [6, p. 353], making use of their generous hints. I can't remember when or from whom I picked up the abstract formulation of the Kolmogorov extension problem, but it was probably Tom Armstrong circa 1975. I have subsequently discovered from Rao [13] that Bochner [3, §5.1] uses directed families of probability measures in his version of the Kolmogorov theorem. The proof of the abstract extension theorem is based on Neveu [12, Proposition I.6.2, p. 27]. Its relation to the standard extension theorem is adapted from Neveu[12, pp. 82–83] and Aliprantis and Border [1, Section 12.6].

[1] Halmos uses the surname Sparre Andersen when alphabetizing this reference.

For each ν let P_ν be a probability measure on Σ_ν. The family $\{P_\nu : \nu \in \mathcal{I}\}$ is **Kolmogorov consistent** if

$$\Sigma_\nu \supset \Sigma_\mu \quad \text{implies} \quad P_\nu|_{\Sigma_\mu} = P_\mu,$$

where $P_\nu|_{\Sigma_\mu}$ is the restriction of P_ν to the σ-subalgebra Σ_μ of Σ_ν. In this case, let us call $\{(\Sigma_\nu, P_\nu) : \nu \in \mathcal{I}\}$ a **Kolmogorov net** on X.

As we shall see, the union $\mathcal{A} = \bigcup_{\nu \in \mathcal{I}} \Sigma_\nu$ of these σ-algebras is itself an algebra of subsets of X. The extension problem is this:

> *Is there a probability measure P on the algebra $\mathcal{A}$ that simultaneously extends every P_ν?*

Clearly the only candidate for an extension is the set function P on $\mathcal{A}$ defined by

$$P(E) = P_\nu(E) \quad (E \in \Sigma_\nu).$$

Kolmogorov consistency is exactly the condition that this is well defined, so the question reduces to whether P is a probability measure. Now P is clearly nonnegative and $P(X) = P_\nu(X) = 1$ for all ν. In addition, P is finitely additive, for if a finite collection of sets belongs to $\mathcal{A}$, then since $\{\Sigma_\nu\}$ is a net, there is some ν for which every member of the collection belongs to Σ_ν. Consequently their union belongs to Σ_ν and hence to $\mathcal{A}$. This proves the assertion above that $\mathcal{A}$ is an algebra. The finite additivity of P is then guaranteed by that of P_ν. *So the only nontrivial issue is whether P is countably additive on $\mathcal{A}$.* If P is countably additive, we call it the **Kolmogorov extension** of the family $\{P_\nu\}$ to $\mathcal{A}$. We can then use the Carathéodory extension procedure (see, e.g., Aliprantis–Border [1, Section 8.6]) to extend it uniquely to $\sigma(\mathcal{A})$, the σ-algebra generated by $\mathcal{A}$.

So the abstract Kolmogorov extension problem boils down in the end to the question of whether a finitely additive probability on an algebra is countably additive.

We digress for a moment to consider this general question.

2 A digression on countable additivity

Recall that a family of sets has the **finite intersection property** if the intersection of every finite subfamily is nonempty. Following Neveu [12, Definition I.6.2, p. 26], let us call a family $\mathcal{C}$ of subsets of X a **compact class** if every sequence $C_1, C_2, \dots$ in $\mathcal{C}$ with the finite intersection property has a nonempty intersection. (For instance, the family of compact sets in a Hausdorff topological space is a compact class.) An equivalent restatement is that $\mathcal{C}$ is a compact class if $C_1 \cap C_2 \cap \cdots = \varnothing$ implies there is some N for which $C_1 \cap \cdots \cap C_N = \varnothing$.

The next result (Neveu [12, Proposition I.6.2, p. 27]) states that if a finitely additive set function on an algebra is "tight" relative to a compact class, then it is countably additive.

1 Proposition *Let P be a finitely additive probability set function on the algebra $\mathcal{A}$ of subsets of X. Let $\mathcal{C}$ be a compact subclass of $\mathcal{A}$, and suppose*

$$P(E) = \sup\{P(C) : C \subset E,\ C \in \mathcal{C}\}$$

for every $E \in \mathcal{A}$. Then P is countably additive on $\mathcal{A}$.

Proof: Let $E_n \downarrow \varnothing$, where each E_n belongs to $\mathcal{A}$. It suffices to show that $P(E_n) \downarrow 0$ (see, e.g., Aliprantis–Border [1, Lemma 8.32, p. 285]).

Let $\varepsilon > 0$ be given, and for each n choose $C_n \in \mathcal{C}$ satisfying $C_n \subset E_n$ and $P(E) \leqslant P(C_n) + \frac{\varepsilon}{2^n}$. Observe that

$$\Big(\bigcap_{m=1}^{n} E_m\Big) \setminus \Big(\bigcap_{m=1}^{n} C_m\Big) \subset \bigcup_{m=1}^{n} \Big(E_m \setminus C_m\Big). \tag{1}$$

Now let $\mathcal{C}'$ be the collection of all finite intersections of sets from $\mathcal{C}$. That is, $\mathcal{C}' = \{C_1 \cap \cdots \cap C_n : C_i \in \mathcal{C},\ i = 1, \dots, n,\ n \in \mathbb{N}\}$. Obviously $\mathcal{C}'$ is also a compact class. Now write $K_n = \cap_{m=1}^{n} C_m$, which belongs to $\mathcal{C}' \cap \mathcal{A}$. Then $\bigcap_{m=1}^{n} K_m \subset \bigcap_{m=1}^{n} E_m$, so $K_n \downarrow \varnothing$. By the compactness of $\mathcal{C}'$, there is some N for which $K_N = \varnothing$.

Since the E_ns are nested, for $n \geqslant N$ equation (1) reduces to

$$E_n \subset \bigcup_{m=1}^{n} \Big(E_m \setminus C_m\Big),$$

so

$$P(E_n) \leqslant \sum_{m=1}^{n} P\Big(E_m \setminus C_m\Big) \leqslant \varepsilon, \quad n \geqslant N.$$

This proves that $\lim_{n\to\infty} P(E_n) = 0$. ∎

3 An abstract Kolmogorov extension theorem

We now state a positive result on the existence of a Kolmogorov extension. Its usefulness is somewhat limited, as it is not obvious when the hypotheses are satisfied. The next section remedies this defect by providing a more tractable sufficient condition.

2 Theorem (Abstract Kolmogorov Extension Theorem) *Let $\{(\Sigma_\nu, P_\nu) : \nu \in \mathcal{I}\}$ be a Kolmogorov net on X. Suppose that there is a compact class $\mathcal{C}$ having the property that*

for each $\nu \in \mathcal{I}$, and for each $E \in \Sigma_\nu$, $P_\nu(E) = \sup\{P_\nu(C) : C \subset E,\ C \in \mathcal{C} \cap \Sigma_\nu\}$.

Then there is a unique Kolmogorov extension to the σ-algebra $\sigma\Big(\bigcup_{\nu\in\mathcal{I}} \Sigma_\nu\Big)$.

Proof: It is clear from the discussion above that all we need to do is verify that the set function P is countably additive on $\mathcal{A} = \bigcup_{\nu\in\mathcal{I}} \Sigma_\nu$, where P is defined on $\mathcal{A}$ by $P(E) = P_\nu(E)$ for $E \in \Sigma_\nu$. But this is immediate from Proposition 1. We may then use the Carathéodory extension procedure (see, e.g., Aliprantis–Border [1, Section 8.6]) to extend it to $\sigma(\mathcal{A})$. ∎

4 The standard Kolmogorov extension problem

The abstract formulation of the Kolmogorov extension may look alien, but I find it simplifies the description of Kolmogorov consistency, and Theorem 2 captures the essence of the problem. I now show why this really is an abstraction of the standard approach.

In the standard approach we are given a family $\{(X_t, \Sigma_t) : t \in T\}$ of measurable spaces. The index set T is generally interpreted as a set of time periods or dates. For any subset H of T define

$$X_H = \prod_{t \in H} X_t \quad \text{and} \quad X_{-H} = \prod_{t \notin H} X_t.$$

To ease notation write X_{-t} for $X_{-\{t\}}$. When $H \subset G \subset T$, let f_{GH} denote the natural projection of X_G on X_H.

For each finite subset F of T we are given a probability measure P_F on X_F with its product σ-algebra $\mathcal{S}_F$. The measure P_F is called the finite-dimensional distribution on X_F. The family $\{P_F\}$ is consistent if the distribution of f_{GF} under P_G on X_F is P_F, that is, if $P_G \circ f_{GF}^{-1} = P_F$. The Kolmogorov extension is a probability P on the infinite product X_T with its product σ-algebra that extends each finite dimensional distribution in the sense that $P \circ f_{TF}^{-1} = P_F$.

To translate this into the abstract framework, identify each $\mathcal{S}_F$ with the collection $\widehat{\mathcal{S}}_F$ of F-cylinder sets in X_T. That is, $\widehat{\mathcal{S}}_F$ consists of all sets of the form $A \times X_{-F}$, where A belongs to $\mathcal{S}_F$. By definition the product σ-algebra $\bigotimes_{t \in T} \Sigma_t$ is the σ-algebra generated by $\left\{\widehat{\mathcal{S}}_F : F \text{ is a finite subset of } T\right\}\Big)$. Define $\widehat{P}_F$ by $\widehat{P}_F(A \times X_{-F}) = P_F(A)$. Regard the family of finite subsets of T as a net directed upward by inclusion, and we are back in the abstract framework, and the two notions of consistency coincide.

The real work of proving a standard version of the extension theorem is verifying the existence of a compact class with the desired property. The following version of the Kolmogorov extension theorem is taken from Neveu [12, unnumbered theorem on p. 82].

3 Kolmogorov Extension Theorem *Let $\{(X_t, \Sigma_t) : t \in T\}$ be a family of measurable spaces, and for each finite subset F of T let P_F be a probability measure on $X_F = \prod_{t \in F} X_t$ with its product σ-algebra $\mathcal{S}_F$. Assume the measures $\{\widehat{P}_F\}$ are Kolmogorov consistent. Suppose that for each t there is a compact class $\mathcal{C}_t \subset \Sigma_t$ such that*

$$\text{for each } A \in \Sigma_t,\ P_t(A) = \sup\{P_t(C) : C \subset A \text{ and } C \in \mathcal{C}_t\}.$$

Then there is a unique probability measure on the infinite product σ-algebra $\Sigma_T = \bigotimes_{t \in T} \Sigma_t$ that extends each $\widehat{P}_F$.

Before we can prove this theorem we need some additional results on compact classes. The next result on compact families is taken from Neveu [12, Lemma I.6.1, p. 26]. [Since I wrote these notes I received an e-mail in 2008 from David Epstein at Warwick expressing doubts about the validity of the next result. I should try to prove it myself, but have not yet done so.]

4 Lemma *If $\mathcal{C}$ is a compact class, then the smallest family including $\mathcal{C}$ and closed under finite unions and countable intersections is also a compact class.*

Proof: ****** To be added. ****** ■

The next result, Neveu [12, Proposition I.6.2, p. 27], is a generalization of Proposition 1 to semialgebras

5 Proposition *Let P be a finitely additive probability set function on the semialgebra $\mathcal{S}$ of subsets of X. Let $\mathcal{C}$ be a compact subclass of $\mathcal{S}$, and suppose*

$$P(E) = \sup\{P(C) : C \subset E,\ C \in \mathcal{C}\}$$

for every $E \in \mathcal{S}$. Then P is countably additive on $\mathcal{S}$.

Proof: By Lemma 4 the class $\mathcal{C}_u$ of finite unions of members of $\mathcal{C}$ is a compact class. Further, it is included in the algebra $\mathcal{A}$ generated by $\mathcal{S}$. Extend P to the set function P' on $\mathcal{A}$ by $P'(\cup_{n=1}^m E_n) = \sum_{n=1}^m P(E_n)$ whenever $E_1, \ldots, E_m$ are pairwise disjoint elements of $\mathcal{S}$. (The algebra generated by $\mathcal{S}$ consists precisely of such sets, Aliprantis–Border [1, Lemma 8.5, p. 274].) I leave it as an exercise to show that this well defined and additive on $\mathcal{A}$, and that P' and $\mathcal{C}_u$ satisfy the hypotheses of Proposition 1. ■

Proof of Theorem 3: Recall that the collection of measurable rectangles $\widehat{E_t} = E_t \times X_{-t}$, as t ranges over T and E_t ranges over Σ_t is enough to generate the product σ-algebra. Define $\mathcal{C}^0$ to be the collection of rectangles of the form $\widehat{C_t} = C_t \times X_{-t}$ in Σ_T, where C_t belongs to $\mathcal{C}_t$. I claim that $\mathcal{C}^0$ is a compact class. For every intersection $\bigcap_{n=1}^\infty \widehat{C_{t_n}} = \bigcap_{n=1}^\infty C_{t_n} \times X_{-t_n}$ is actually a product set $\prod_{t\in T} E_t$. (But only countably many E_t differ from X_t.) The only way a product set can be empty is if some factor, say E_s, is empty. But $\mathcal{C}_s$ is a compact class, so there is some N for which $\bigcap_{n=1}^N \widehat{C_{t_n}} = \varnothing$. This shows that $\mathcal{C}^0$ is a compact class. Consequently the collection $\mathcal{C}$ closed under finite unions and countable intersections generated by $\mathcal{C}^0$ is also a compact class.

Now consider the finite-dimensional measurable rectangle $E = \prod_{t\in F} E_t \times X_{-F}$, where F is a finite subset of T and each E_t belongs to Σ_t. The sets of this form constitute a semialgebra that generates the product σ-algebra. Let $\varepsilon > 0$, and choose $C_t \in \mathcal{C}_t$ so that

$$C_t \subset E_t, \text{ and } P_t(E_t) < P_t(C_t) + \frac{\varepsilon}{n}, \quad t \in F,$$

where n is the cardinality of F. Then $C = \prod_{t\in F} C_t \times X_{-F} = \bigcap_{t\in F} C_t \times X_{-t}$ belongs to $\mathcal{C}$, and

$$E \setminus C \subset \bigcup_{t\in F} \{(E_t \setminus C_t) \times X_{-t}\}.$$

Since P is finitely additive and nonnegative, it is finitely subadditive, so

$$P(E) - P(C) \leqslant \sum_{t\in F} \Big(P_t(E_t) - P(C_t)\Big) < \varepsilon.$$

This shows that $P(E) = \sup\{P(C) : C \subset E,\ C \in \mathcal{C}\}$, and the theorem now follows from Proposition 5. ■

This proposition gives rise to a number of corollaries based on the regularity of Borel measures on topological spaces.

6 Corollary *Let $\{X_t : t \in T\}$ be a family of Polish spaces equipped with their Borel σ-algebras. For each finite subset F of T let P_F be a Borel probability measure on $X_F = \prod_{t\in F} X_t$ with its product (Borel) σ-algebra $\mathcal{S}_F$. Assume the distributions $\{\widehat{P}_F\}$ are Kolmogorov consistent. Then there is a unique probability measure on the infinite product σ-algebra $\Sigma_T = \bigotimes_{t\in T} \Sigma_t$ that extends each $\widehat{P}_F$.*

Proof: By Aliprantis–Border [1, Theorem 11.20, p. 382] the hypotheses of Theorem 3 are satisfied for the class $\mathcal{C}_t$ of compact subsets of X_t. ∎

7 Corollary *Corollary 6 remains true for universally measurable spaces X_t.*

Proof: By Dudley [6, Theorem 11.5.1, p. 315] the hypotheses of Theorem 3 are satisfied for the class $\mathcal{C}_t$ of compact subsets of X_t. ∎

8 Corollary *Corollary 6 remains true if each X_t is Hausdorff, and each P_F is tight (in the sense of Aliprantis–Border [1, Definition 11.14, p. 379]).*

Proof: In a Hausdorff space, every compact set is closed, so it follows that the class $\mathcal{C}_t$ of compact subsets of X_t is a compact class, and hence satisfies the hypotheses of Theorem 3. ∎

Section 8 presents a moderately well known example of a consistent family that has no Kolmogorov extension. But a number of preliminary results are presented first.

5 Two obscure facts regarding Lebesgue measure

Throughout the remainder of these notes λ denotes Lebesgue measure on the real line. You may recall that Lebesgue measure is the Carathéodory extension of length from the semiring of half-open intervals to the σ-algebra of Lebesgue measurable sets. See, for instance, Aliprantis–Border [1, Section 8.8]. Also $A + B$ denotes the algebraic sum of A and B, namely

$$A + B = \{x + y : x \in A \text{ and } y \in B\}$$

and for real α, the set αA is defined by

$$\alpha A = \{\alpha x : x \in A\}.$$

A Lebesgue measurable subset E of the real line cannot be "evenly distributed." By this I mean that there is never any fraction $0 < \alpha < 1$ so that for every interval J the set E occupies a fraction α of J, or $\lambda(E \cap J) = \alpha\lambda(J)$. There is always some interval in which E is highly concentrated. The next result is similar to one pointed out by van Vleck [14] and may be found in Halmos [7, Theorem A, p. 69] or Dudley [6, Proposition 3.4.2, p. 80].

9 Lemma *Let E be a Lebesgue measurable set with $\lambda(E) > 0$. Then for any $0 \leqslant \alpha < 1$ there is some open interval J, so $\lambda(J) > 0$, such that*

$$\lambda(J) \geqslant \lambda(E \cap J) > \alpha\lambda(J).$$

Proof: By the outer regularity of Lebesgue measure there is an open set U with $E \subset U$ and

$$\lambda(E) > \alpha\lambda(U).$$

But every open subset of the line is a countable union of disjoint open intervals, so $U = \bigcup_n J_n$, where the open intervals J_n are pairwise disjoint. Thus $E = \bigcup_n (E \cap J_n)$. It follows that

$$\lambda(E) = \sum_n \lambda(E \cap J_n) > \alpha \sum_n \lambda(J_n),$$

so some J_n must satisfy the conclusion of the lemma. ■

A corollary of this result is a rather remarkable result about Lebesgue measurable sets of positive measure. It is due to Steinhaus, and also appears Halmos [7, Theorem B, p. 69] or Dudley [6, Proposition 3.4.3, p. 80].

10 Corollary *If E is a Lebesgue measurable set with $\lambda(E) > 0$, then the algebraic difference $E - E$ is a neighborhood of zero.*

Proof: By Lemma 9 there is some interval J with $\lambda(J \cap E) > \frac{3}{4}\lambda(J) > 0$. I claim that $E - E$ includes the open interval $\left(-\frac{1}{2}\lambda(J), \frac{1}{2}\lambda(J)\right)$, which makes it a neighborhood of zero. For let $-\frac{1}{2}\lambda(J) < x < \frac{1}{2}\lambda(J)$. Then

$$F_x = (E \cap J) \cup \Big[(E \cap J) + x\Big] \subset J \cup (J + x).$$

Now $J \cup (J + x)$ is an interval whose length is less than $\frac{3}{2}\lambda(J)$. This implies that $(E \cap J)$ and $(E \cap J) + x$ are not disjoint. Otherwise $\lambda(F_x) = 2\lambda(E \cap J) > \frac{3}{2}\lambda(J)$, a contradiction.

Thus there is some y belonging to both $(E \cap J)$ and $(E \cap J) + x$. This means in particular that y belongs to E and $y = z + x$ where z belongs to E. Thus $x = y - z$ belongs to $E - E$. Since x is an arbitrary element of $\left(-\frac{1}{2}\lambda(J), \frac{1}{2}\lambda(J)\right)$, this interval is included in $E - E$, as claimed. ■

6 A family of nonmeasurable sets

For this section fix an irrational number α. Let A be the set of numbers of the form $n + m\alpha$, where n and m are integers (positive, negative, or zero). Thus

$$A = \mathbb{Z} + \alpha\mathbb{Z}.$$

Note that the set A is closed under addition and subtraction, and under multiplication by integers. Since α is irrational, the elements of A have the following **distinctness property**:

> *If $n + m\alpha = n' + m'\alpha$, then $n = n'$ and $m = m'$.*
>
> (If $m - m' = 0$, then $n - n' = 0$ too. Otherwise solve for $\alpha = \frac{n'-n}{m-m'}$, which is decidedly rational.)

The distinctness property would not be satisfied if α were rational. It also follows from the irrationality of α that A is dense in $\boldsymbol{R}$. In fact, we shall prove an even stronger result. For each nonnegative integer k, define the sets A_k, B_k, and C_k by

$$\begin{aligned} A_k &= \{n + m\alpha : n, m \in \mathbb{Z} \text{ and } |n| \geqslant k\} \\ B_k &= \{n + m\alpha \in A_k : n \text{ is even}\} \\ C_k &= \{n + m\alpha \in A_k : n \text{ is odd}\} \end{aligned}$$

Clearly

$$A_k = B_k \cup C_k$$

for each $k = 0, 1, 2, \ldots$, and it follows from the distinctness property that

$$B_k \cap C_k = \varnothing,$$

for each k. By construction $A = A_0 \supset A_1 \supset A_2 \supset \cdots$, and the distinctness property implies $\bigcap_{k=1}^{\infty} A_k = \varnothing$. In other words,

$$A_k \downarrow \varnothing.$$

Likewise, $B_k \downarrow \varnothing$ and $C_k \downarrow \varnothing$.

The case $k = 0$ of the next result may be found in Halmos [7, Theorem C, p. 70] or in the proof of Dudley [6, Theorem 3.4.4, p. 81].

11 Proposition *For each nonnegative integer k, the sets A_k, B_k, and C_k defined above are dense.*

Proof: I shall prove the denseness of B_k, which implies that of A_k. The case of C_k is similar. But first I prove the denseness of $A = A_0$.

Begin by noting that for each integer m, there is a unique even integer n_m (not necessarily positive) such that $0 \leqslant n_m + m\alpha < 2$. (In fact, $-n_m$ is the largest even integer less than or equal to $m\alpha$.) Set $x_m = n_m + m\alpha$, and let $X = \{x_m : m \in \mathbb{Z}\}$. I now claim:

> *For each $\varepsilon > 0$ the set X contains a point x_{m^*} satisfying $0 < x_{m^*} < \varepsilon$.*
>
> *Proof of the claim:* Choose m so that $0 < \frac{1}{m} < \varepsilon$. Consider the $2m + 1$ distinct points $x_1, \ldots, x_{2m+1}$. These all lie in $[0, 2)$ so there must be at least two of them, say x_i and x_j, with $0 < x_i - x_j < \frac{1}{m}$, otherwise not all of them would fit in $[0, 2)$. But $x_i - x_j$ belongs to A, and the claim is proved. ∎

Given the claim, it is clear that every real is within ε of nx_{m^*} for some n. But A contains each nx_{m^*}, so A is dense.

Indeed there are infinitely many such points $0 < x_m < \varepsilon$ in X. (Once you have found one, let it replace ε and find another, etc.) This implies that each set B_k also includes a point $x_{m^{**}}$ satisfying $0 < x_{m^{**}} < \frac{\varepsilon}{2}$. For the only way a point x_m in X can fail to belong to B_k is if $|n_m| < k$. But this can happen for only finitely many m, namely those m for which $|m\alpha| < k$ if k is even or $|m\alpha| < k+1$ if k is odd.

Now every real is within ε of $x_{m^{**}}$ itself or of some nonzero even multiple of $x_{m^{**}}$. But B_k is closed under multiplication by nonzero even integers, so it is dense. ∎

Now define the relation $\sim$ on $\boldsymbol{R}$ by

$$x \sim y \text{ if and only if } x - y \in A.$$

This is an equivalence relation: it is reflexive since $x - x = 0 + 0\alpha$; it is symmetric since $x - y = n + m\alpha$ implies $y - x = -n + (-m)\alpha$; and it is transitive since if $x - y = n + m\alpha$ and $y - z = n' + m'\alpha$, then $x - z = x - y + (y - z) = n + n' + (m + m')\alpha$. Note that the equivalence class of a number x is the countable set $x + A$.

By the Axiom of Choice

> *there is a set V consisting of exactly one element from each $\sim$-equivalence class.*[2]

It follows that

> *$V - V$ contains no nonzero element of A.*
>
> For if $x, y \in V$ satisfy $x - y \in A$, i.e., $x \sim y$, we must have $x = y$.

Moreover, if x and y are distinct elements of A, then $V + x$ and $V + y$ are disjoint. For suppose $v + x = v' + y$. Then $v - v' = y - x \in A$, so $v \sim v'$. But by construction V has only one element from each $\sim$-equivalence class, so $v = v'$ and $x = y$. Thus

> *$\boldsymbol{R} = V + A$, a countable union of the pairwise disjoint sets $V + x$ for $x \in A$.*

The next result may be found in Halmos [7, Theorem D, p. 69] or Dudley [6, Theorem 3.4.4, p. 80].

12 Proposition *The set V is not Lebesgue measurable.*

[2] Halmos [7, p. 70] describes this construction of V as "well known." Dudley [6, p. 84] attributes it to van Vleck [14]. One can rewrite van Vleck's construction in terms of the equivalence relation defined by $x \sim y$ if $x - y = \frac{m}{2^n}$ for some integers m and n, which is not quite the same, but still interesting.

Proof: Suppose by way of contradiction that V is Lebesgue measurable, and suppose first that $\lambda(V) > 0$. Then $V - V$ includes an open interval about zero. Since the set A is dense, there is a nonzero point belonging to A that lies in this interval and hence in $V - V$, a contradiction. Therefore if V is measurable, $\lambda(V) = 0$. But Lebesgue measure is translation invariant so $\lambda(V) = \lambda(V + x)$ for any x, so

$$\lambda(\boldsymbol{R}) = \lambda\Big(\bigcup_{x \in A} V + x\Big) = \sum_{x \in A} \lambda(V) = 0,$$

another contradiction. Therefore V cannot be Lebesgue measurable. ∎

In fact, there are even more pathological sets. The idea is this. We saw in Lemma 9 that there is no such thing as an evenly distributed Lebesgue measurable set, but what about an evenly distributed nonmeasurable set? We remarked above that $\boldsymbol{R} = V + A$, and in a very real sense A can be decomposed into two "halves" B_0 and C_0. What are they like? We might hope that they have the property that $\lambda(E \cap (V + B_0)) = \frac{1}{2}\lambda(E)$ for every interval E. Oops, I forgot that $\lambda(E \cap (V + B_0))$ is probably not defined if $V + B_0$ is not Lebesgue measurable, so use the outer measure λ^* instead. In this case we get an even more surprising result: For any Lebesgue measurable set E, we have $\lambda^*(E \cap (V + B_0)) = \lambda^*(E \cap (V + C_0))$, so $V + B_0$ is in a sense evenly distributed, but the outer measure of each "half" of E is not $\frac{1}{2}\lambda(E)$—it is the whole measure $\lambda(E)$.[3]

The next result is proven for the case $k = 0$ in Halmos [7, Theorem E, p. 70] or Dudley [6, Theorem 3.4.4, p. 80].

13 Proposition *For each nonnegative integer k, the set $M_k = V + B_k$ has the property that for any Lebesgue measurable set E,*

$$\lambda^*(E \cap M_k) = \lambda(E) = \lambda^*(E \cap M_k^c),$$

where λ^* is the outer measure induced by λ.

Proof: Suppose by way of contradiction that $\lambda^*(E \cap M_k^c) < \lambda(E)$. Then there is a Borel set B satisfying $E \cap M_k^c \subset B \subset E$ with

$$\lambda^*(E \cap M_k^c) \leqslant \lambda(B) < \lambda(E).$$

From this it follows that the measurable set $F := E \setminus B \subset E \cap M_k \subset M_k$ has $\lambda(F) > 0$. This implies that $F - F$ includes a neighborhood of zero, so that $M_k - M_k$ also includes a neighborhood of zero, and hence there is some nonzero x belonging to $M_k - M_k$ and also belonging to the dense set C_k. Let us write $x = v + a - (v' + a') = v - v' + (a - a')$, where v and v' belong to V and a and a' belong to B_k. Then $a' - a$ belongs to A, so $v \sim v'$, which in turn implies $v = v'$ by the construction of V. Thus $x = a - a'$, which belongs to B_0, but x also belongs to C_k, a contradiction. Therefore

$$\lambda(E) = \lambda^*(E \cap M_k^c).$$

A similar argument interchanging the roles of B_k and C_k shows $\lambda(E) = \lambda^*(E \cap M_k)$. ∎

[3]The construction in van Vleck [14] also has this property.

7 A digression on thick sets

Consider the probability space (I, Σ, λ), where I is the closed unit interval, Σ is the collection of Lebesgue measurable subsets of I, and λ is Lebesgue measure. Let us say that a subset X of I is **thick**[4] if $\lambda^*(X) = 1$. Two immediate consequences of thickness are that if E is a Lebesgue measurable subset of I and $E \supset X$, then $\lambda(E) = 1$; and if $E \cap X = \varnothing$, then $\lambda(E) = 0$.

All the sets $X_k = M_k \cap I$, where M_k is described above, are thick. Furthermore $X_k \downarrow \varnothing$. This shows that λ^* is not countably additive, by the way. The next theorem appears (in a more general form) in Halmos [7, Theorem A, p. 75].

14 Proposition *Let X be a thick subset of I, and set $\Sigma_X = \{E \cap X : E \in \Sigma\}$. Define the set function λ_X on Σ_X by $\lambda_X(E \cap X) = \lambda(E)$. Then λ_X is a well defined probability measure on the σ-algebra Σ_X.*

Proof: It is easy to see that Σ_X is a σ-algebra. To see that λ_X is well defined we have to show that if $E \cap X = F \cap X$, where E and F belong to Σ, then $\lambda(E) = \lambda(F)$. In this case, $(E \setminus F) \cap X = \varnothing$ and $(F \setminus E) \cap X = \varnothing$. Since X is thick we have $\lambda(E \setminus F) = \lambda(F \setminus E) = 0$, so $\lambda(E) = \lambda(F)$.

Next we have to verify that λ_X is countably additive. So let $\{A_n\}$ be a sequence of pairwise disjoint subsets in Σ_X, and let $\{E_n\}$ be a sequence of subsets in Σ satisfying $A_n = E_n \cap X$ for each n. We can "disjoin" the sets E_n by the usual trick of setting $\tilde{E}_n = E_n \setminus \bigcup_{i=1}^{n-1} E_i$. Then

$$(E_n \setminus \tilde{E}_n) \cap X = A_n \setminus \bigcup_{i=1}^{n-1} E_i \cap X = A_n \setminus \bigcup_{i=1}^{n-1} A_i = \varnothing,$$

where the last equality follows from the pairwise disjointness of the sets A_n. Since $\tilde{E} \subset E$, it follows that $\lambda(\tilde{E}) = \lambda(E)$. Thus

$$\begin{aligned}
\sum_{n=1}^{\infty} \lambda_X(A_n) &= \sum_{n=1}^{\infty} \lambda(E_n) \\
&= \sum_{n=1}^{\infty} \lambda(\tilde{E}_n) \\
&= \lambda\Big(\bigcup_{n=1}^{\infty} \tilde{E}_n\Big) \\
&= \lambda\Big(\bigcup_{n=1}^{\infty} E_n\Big) \\
&= \lambda_X\Big(\bigcup_{n=1}^{\infty} A_n\Big).
\end{aligned}$$

[4] Halmos defines thickness in terms of inner measure, but this definition is equivalent for totally finite measures, such as Lebesgue measure on the unit interval.

Thus λ_X is a measure. ∎

The next proposition is immediate from the definitions.

15 Proposition *Let X be a thick subset of I and let $f\colon I \to \boldsymbol{R}$ be Lebesgue measurable. Then $f|_X$ is Σ_X-measurable.*

The next one is also immediate from the definitions, provided you remember the definition of the product σ-algebra.

16 Proposition *Let $X_1, \ldots, X_n$ be thick subsets of I. Then $X = X_1 \times \cdots \times X_n$ is a thick subset of I^n with n-dimensional Lebesgue measure λ^n. Furthermore, a subset A belongs to the product σ-algebra $\bigotimes_{i=1}^n \Sigma_{X_i}$ if and only if there is a λ^n-measurable subset E of I^n with $A = E \cap X$.*

8 The Andersen–Jessen construction

Let the sets M_k, $k = 0, 1, 2 \ldots$, be defined as in Proposition 13. Set $X_k = M_k \cap [0, 1]$, let $\Sigma_k = \{E \cap X_k : E \text{ is Lebesgue measurable}\}$, and define the measure λ_k on Σ_k by $\lambda_k(E \cap X_k) = \lambda(E)$. Each X_k is a thick subset of $I = [0, 1]$, so by Proposition 14, these are well defined probability spaces. Also recall that $X_k \downarrow \varnothing$.

Now let $X = \prod_{k=0}^\infty X_k$ endowed with the product σ-algebra $\mathcal{S} = \bigotimes_{k=0}^\infty \Sigma_k$. For each nonnegative integer n, let $\mathcal{S}_n$ denote the σ-subalgebra of $\mathcal{S}$ consisting of all the cylinder sets of the form $A \times \prod_{k=n+1}^\infty X_k$, where A belongs to the product σ-algebra $\bigotimes_{k=0}^n \Sigma_k$.

Consider the functions $f_n\colon X_n \to X_0 \times \cdots \times X_n$, $n = 0, 1, 2, \ldots$, defined by

$$f_n(x) = (\underbrace{x, \ldots, x}_{n+1 \text{ times}}).$$

I claim these are measurable functions. To see this, extend f_n to $\hat{f}_n\colon I \to I^{n+1}$, where $\hat{f}_n(x) = (x, \ldots, x)$. This is a Lebesgue measurable function. Further for any A belonging to $\bigotimes_{k=0}^n \Sigma_k$, let E be a Lebesgue measurable subset of I^{n+1} satisfying $E \bigcap \prod_{k=0}^n X_k = A$. Then $f_n^{-1}(A) = \hat{f}_n^{-1}(E) \cap X_n$, so f_n is measurable.

Define the probability measure P_n on $\mathcal{S}_n$ by

$$\begin{aligned} P_n\Big(A \times \prod_{k=n+1}^\infty X_k\Big) &= \lambda_n(\{x \in X_n : (\underbrace{x, \ldots, x}_{n+1 \text{ times}}) \in A\}) \\ &= \lambda_n\Big(f_n^{-1}(A)\Big) \end{aligned}$$

The remarkable fact about this construction is that the probability measures P_n are Kolmogorov consistent, that is, if $n > m$, then $\mathcal{S}_m \subset \mathcal{S}_n$ and

$$P_n|_{\mathcal{S}_m} = P_m.$$

To see this, let $B = A \times \prod_{k=m+1}^{\infty} X_k$ belong to $\mathcal{S}_m$ and let E be a Lebesgue measurable set satisfying

$$E \cap X_m = \{x \in X_m : (\underbrace{x, \dots, x}_{m+1 \text{ times}}) \in A\}.$$

Such a set E exists since A belongs to $\bigotimes_{k=1}^{m} \Sigma_k$. Then $P_m(B) = \lambda(E)$. But since $X_n \subset X_m$, we also have

$$E \cap X_n = \{x \in X_n : (\underbrace{x, \dots, x}_{m+1 \text{ times}}) \in A\} = \{x \in X_n : (\underbrace{x, \dots, x}_{n+1 \text{ times}}) \in A \times \prod_{k=m+1}^{n} X_k\},$$

so $P_n(B) = \lambda(E) = P_m(B)$.

17 Proposition *There is no probability measure on* $\mathcal{S}$ *that extends each* P_n*.*

Proof: Suppose the set function P on $\mathcal{S}$ extends every P_n. Consider the sets of the form

$$D_n = \{(\underbrace{x, \dots, x}_{n+1 \text{ times}}) \times \prod_{k=n+1}^{\infty} X_k : x \in X_n\}.$$

Since the diagonal of I^{n+1} is Lebesgue measurable, Proposition 16 implies that each D_n belongs to $\mathcal{S}_n$, so we have

$$P(D_n) = P_n(D_n) = \lambda_n(X_n) = 1$$

for all n. But since $X_n \downarrow \varnothing$, we also have

$$D_n \downarrow \varnothing,$$

so P cannot be countably additive. ∎

9 The Heifetz–Samet refinement

Heifetz and Samet [9] refine the Andersen–Jessen construction by constructing a product $\prod_{n=1}^{\infty} X$ of infinitely many copies of the same space, instead of different X_k spaces. They need to do this to interpret their spaces as a set of hierarchies of beliefs.

Let X_k, $k = 0, 1, 2, \dots$, and X be the measurable spaces defined in Section 8 above. For the Heifetz–Samet construction, the underlying set is

$$S = \prod_{n=1}^{\infty} X = \prod_{n=1}^{\infty} \Big(\prod_{k=0}^{\infty} X_k \Big)$$

The σ-algebra $\mathcal{F}_n$ is the collection of cylinder sets in S determined by the first $n+1$ coordinates. I shall now proceed to abuse notation and identify the cylinder sets and their finite dimensional bases, and similarly for probability measures on the cylinder algebras.

The measure on $\mathcal{F}_0$ (identified with the measure on X) is just the product measure $\lambda_0 \times \lambda_1 \times \cdots$.

For $n > 0$, the description of the probability measures Q_n on $\mathcal{F}_n$ is a little involved. For $n > m$ define the measurable functions $f_{m,n} \colon X_n \to \prod_{k=m}^n X_k$ via $f_{m,n}(x) = (x, \dots, x)$ and define $\lambda_{m,n}$ on $\prod_{k=m}^n X_k$ by $\lambda_{m,n} = \lambda_n \circ f_{m,n}^{-1}$. The measure $\lambda_{0,n}$ is an analogue of P_n from the previous construction.

Now write

$$X^{n+1} = \prod_{m=1}^{n+1} \Big(\prod_{k=0}^{\infty} X_k \Big)$$

as a one-dimensional array as follows:

$$X^{n+1} = X_0 \times (X_0 \times X_1) \times \cdots \times (X_0 \times \cdots \times X_n) \times (X_1 \times \cdots \times X_{n+1}) \times (X_2 \times \cdots \times X_{n+2}) \cdots$$

The measure Q_{n+1} on $\mathcal{F}_{n+1}$ (identified with X^{n+1}) is written corresponding to the one-dimensional array as

$$Q_{n+1} = \lambda_0 \times \lambda_{0,1} \times \cdots \times \lambda_{0,n} \times \lambda_{1,n+1} \times \lambda_{2,n+2} \cdots$$

******************* TO BE COMPLETED *******************

[I have run out of energy and have not yet finished the exposition of their construction. Part of the problem is finding a manageable system of notation.]

References

[1] C. D. Aliprantis and K. C. Border. 1994. *Infinite dimensional analysis: A hitchhiker's guide*. Number 4 in Studies in Economic Theory. Berlin: Springer–Verlag.

[2] E. S. Andersen and B. Jessen. 1948. On the introduction of measures in infinite product sets. *Det Kongelige Danske Videskabernes Selskab Mathematik-Fysike Meddelelser* 25(4).

[3] S. Bochner. 1960. *Harmonic analysis and the theory of probability*. Berkeley: University of California Press. Second printing.

[4] W. Böge and T. Eisele. 1979. On solutions of Bayesian games. *International Journal of Game Theory* 8(4):193–215. DOI: 10.1007/BF01766706

[5] A. Brandenburger and E. Dekel. 1993. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory* 59(1):189–198. DOI: 10.1006/jeth.1993.1012

[6] R. M. Dudley. 1989. *Real analysis and probability*. Pacific Grove, California: Wadsworth & Brooks/Cole.

[7] P. R. Halmos. 1974. *Measure theory*. Graduate Texts in Mathematics. New York: Springer–Verlag. Reprint of the edition published by Van Nostrand, 1950.

[8] J. C. Harsanyi. 1967. Games with incomplete information played by "Bayesian" players, I–III. Part I. The basic model. *Management Science* 14:159–182. http://www.jstor.org/stable/2628393

[9] A. Heifetz and D. Samet. 1996. Coherent beliefs are not always types. Manuscript, School of Mathematical Sciences and Faculty of Management,Tel Aviv University.

[10] ——— . 1996. Topology-free typology of beliefs. Manuscript, School of Mathematical Sciences and Faculty of Management,Tel Aviv University.

[11] J.-F. Mertens and S. Zamir. 1985. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory* 14(1):1–29. DOI: 10.1007/BF01770224

[12] J. Neveu. 1965. *Mathematical foundations of the calculus of probability*. Holden–Day Series in Probability and Statistics. San Francisco: Holden–Day.

[13] M. M. Rao. 1971. Projective limits of probability spaces. *Journal of Multivariate Analysis* 1(1):28–57. DOI: 10.1016/0047-259X(71)90028-5

[14] E. B. van Vleck. 1908. On non-measurable sets of points, with an example. *Transactions of the American Mathematical Society* 9(2):237–244. http://www.jstor.org/stable/1988654