LEARNING FINITELY ADDITIVE PROBABILITIES: AN IMPOSSIBILITY THEOREM

MAXWELL B. STINCHCOMBE

Abstract. If $(X, \mathcal{X})$ is a measure space and $\mathcal{F}^\circ \subset \mathcal{X}$ is a field generated either by a countable class of sets or by a Vapnik-Červonenkis class, then if μ is purely finitely additive, there exist uncountably many μ' agreeing with μ on $\mathcal{F}^\circ$ and having $|\mu(A) - \mu'(A)| = 1$ for uncountably many A. If μ is also non-atomic, then for any $r \in (0, 1]$, $|\mu(A_r) - \mu'(A_r)| = r$ for uncountably many A_r. Al-Najjar's [1] unlearnability result is: if $(X, \mathcal{X})$ belongs to a class of measure spaces not supporting countably additive nonatomic distributions and $\mathcal{C}$ is a Vapnik-Červonenkis class, then there exist purely finitely additive nonatomic probabilities, μ and μ', agreeing on $\mathcal{C}$ and having $|\mu(A_\alpha) - \mu'(A_\alpha)| = \alpha$ for any $\alpha \in (0, \frac{1}{2}]$ for uncountably many A_α.

Alice laughed. "There's no use trying," she said: "one *ca'n't* believe impossible things."

"I daresay you haven't had much practice," said the Queen. "When I was your age, I always did it for half-an-hour a day. Why, sometimes I've believed as many as six impossible things before breakfast." (Lewis Carroll, *Through the Looking Glass*)

1. Introduction

The learnable classes of events, $\mathcal{C} \subset \mathcal{X}$, $\mathcal{X}$ a σ-field of subsets of a non-empty set X, are the Vapnik-Červonenkis (VČ) classes. They are characterized by the property that the empirical distributions of iid draws distributed according to any countably additive probability P satisfy both a strong law of large numbers and a central limit theorem. The useful VČ classes $\mathcal{C}$ determine countably additive probabilities so that after learning each $P(C)$, $C \in \mathcal{C}$, one has learned P.

There is no representation in X for the mass of a purely finitely additive probability, μ, on $(X, \mathcal{X})$. This lack of representation leads to a fundamental indeterminacy that makes learning purely finitely additive

Date: February 18, 2010

Key words and phrases. Learning, Vapnik-Červonenkis classes, finitely additive probabilities.

probabilities impossible. Theorem 1 shows that for any purely finitely additive μ, the set of μ' that agree with μ on $\mathcal{F}^\circ(\mathcal{C})$, the field generated either by a countable class of sets or by a VČ class, contains uncountably many μ' having support disjoint from the support of μ.

1.1. **No Representation.** Let X be an infinite set; $\mathcal{X}$ a σ-field of subsets of X that separates points; [1] Δ^{fa} the set of finitely additive probabilities on $\mathcal{X}$; and $\Delta^{ca} \subset \Delta^{fa}$ the set of countably additive probabilities. Following Kingman [14], the **deficiency** of a $\mu \in \Delta^{fa}$ is

$$\sup\{\delta \geq 0 : \exists E_n \downarrow \varnothing,\ \mu(E_n) \geq \delta\}, \tag{1}$$

where the supremum is taken over all nested sequences of sets E_n in $\mathcal{X}$ that decrease to the empty set.[2] If $\mu \in \Delta^{ca}$, then its deficiency is 0, and μ is **purely finitely additive** if its deficiency is 1.

Every finitely additive probability on a measure space $(X, \mathcal{X})$ is the trace of a unique countably additive probability on a larger measure space $(\widehat{X}, \widehat{\mathcal{X}})$. If the larger space is chosen to be compact, then it is determined by the requirement that X can be imbedded as a dense subset of $\widehat{X}$ in such a fashion that every bounded measurable $\mathcal{X}$-measurable function has a unique continuous extension to $\widehat{X}$. For such spaces, every $\mu \in \Delta^{fa}(\mathcal{X})$ has a unique extension $\widehat{\mu} \in \Delta^{ca}(\widehat{\mathcal{X}})$ determined by $\widehat{\mu}(\widehat{E}) = \mu(E)$.

The space $\widehat{X}$ contains a representation of the mass in μ that has no representation in X in the following sense: if μ has deficiency δ, then $\widehat{\mu}(\widehat{X} \setminus X) = \delta$. This result contains the intuition for many of the results for finitely additive probabilities. For example, when $\delta = 1$, i.e. when μ is purely finitely additive, the intuition for the unlearnability of μ is that all of the mass of μ is carried on $\widehat{X} \setminus X$, and since points in $\widehat{X} \setminus X$ are not observed, μ can not be learned.

1.2. **Vapnik-Červonenkis Classes and Learnability.** If $P \in \Delta^{ca}$, and ξ_n is a sequence of independent, identically distributed (iid) random variables taking values in X and having distribution P, then a sequence of empirical measures $\{P_n : n \in \mathbb{N}\}$ can be defined by $P_n = \frac{1}{n}(\delta_{\xi_1} + \cdots + \delta_{\xi_n})$ where δ_x is point mass on x. Vapnik and Červonenkis [22] gave a combinatorial condition on class $\mathcal{C} \subset \mathcal{X}$ that is sufficient to guarantee

[1] For $x \neq x'$, there exist $E \in \mathcal{X}$ such that $1_E(x) \neq 1_E(x')$. This is satisfied if e.g. singleton sets are measurable, or if the σ-field is Hausdorff.

[2] The basics of finitely additive measures were originally laid out in Yosida and Hewitt [23], and their Theorem 1.22 implies that the fact that the supremum is achieved in (1). See also Dunford and Schwarz [13, Ch. III].

that for all $P \in \Delta^{ca}$,

$$\sup_{E \in \mathcal{C}} |P_n(C) - P(C)| \stackrel{a.e.}{\to} 0. \tag{2}$$

Under measurability conditions (involving being able to take the supremum over an uncountable class of events), these conditions are also necessary, and a class satisfying their conditions is called a **VČ class**.

Dudley, Giné, and Zinn [12, Proposition 11] give the uniform version of this result, showing that, subject to the measurability conditions, $\mathcal{C}$ is a VČ class if and only if

$$\sup_{P \in \Delta^{ca},\ E \in \mathcal{C}} |P_n(C) - P(C)| \stackrel{a.e.}{\to} 0. \tag{3}$$

Thus, if we ask for learnability of a class of events for any $P \in \Delta^{ca}$, then only VČ classes can be learned, and they can be learned uniformly.

Being central to the theory of learnability, VČ classes have been extensively studied (Dudley's monograph [11] is a good source). The crucial result for present purposes is that they are, very nearly, compact metric spaces for many metrics — $\mathcal{C}$ is a VČ classes if and only if it is totally bounded,[3] in a uniform sense, for *every* $L^2(p)$ metric on the indicator functions of the $C \in \mathcal{C}$, p finitely supported [11, Theorems 3.7.2 and 10.1.7]. In particular, this implies that they are separable in the following sense:[4] for any $P \in \Delta^{ca}$ and any VČ class $\mathcal{C}$, there is a countable $\mathcal{C}_0 \subset \mathcal{C}$ such that, outside of a null set of realizations of the sequence of random variables, for every $n \in \mathbb{N}$ and every $E \in \mathcal{X}$,

$$P_n(E) = \lim_i P_n(E_i) \text{ for some } \{E_i : i \in \mathbb{N}\} \subset \mathcal{C}_0 \text{ with } P(E_i \Delta E) \to 0. \tag{4}$$

It is possible (see Lemma 4.4 below) to use this result for finitely additive μ by passing to the associated countably additive $\widehat{\mu}$.

The following gives the classic example of a VČ class as well as several related classes of sets that fail to be VČ classes. For any $\mathcal{A} \subset \mathcal{X}$, $\mathcal{F}^\circ(\mathcal{A})$ denotes the smallest field, not σ-field, containing $\mathcal{A}$.

Example 1.A. *When* $(X, \mathcal{X}) = ([0,1], \mathcal{B})$*, the unit interval with the Borel* σ*-field,* $\mathcal{C} = \{[0,r] : r \in \mathbb{R}\}$ *is a VČ class — by the Glivenko-Cantelli theorem,* (3) *holds. One countable subclass of* $\mathcal{C}$ *that is dense in the sense of* (4) *is* $\mathcal{C}_0 = \{[0,q] : q \in \mathbb{Q}\}$ *where* $\mathbb{Q}$ *is the rational numbers. Note that neither* $\mathcal{F}^\circ(\mathcal{C})$ *nor the countable class of sets* $\mathcal{F}^\circ(\mathcal{C}_0)$ *is a VČ class because the uniform convergence of* (3) *does not hold.*

---

[3] A metric space M is totally bounded iff for every $\epsilon > 0$ there is a finite set M_ϵ with $d(x, M_\epsilon) < \epsilon$ for every $x \in M$, and it is compact iff it is both totally bounded and complete. See e.g. [8, Theorem 4.7.15].

[4] See also Pollard [17, §5] for this.

The following gives some VČ classes for discrete spaces (see Dudley [11, Theorem 4.2.6, and preceding remarks]), as well as related classes of sets that fail to be VČ classes. $\#A$ denotes the cardinality of A.

Example 1.B. *If* $(X, \mathcal{X}) = (\mathbb{N}, \mathcal{N})$ *where* $\mathcal{N} = 2^{\mathbb{N}}$ *is the class of all subsets of* $\mathbb{N}$*, then the following classes of sets are VČ classes:* $\mathcal{C}_{[} = \{\{n, n+1, \dots\} : n \in \mathbb{N}\}$*;* $\mathcal{C}_{]} = \{\{1, \dots, n\} : n \in \mathbb{N}\}$*; and for any* $r \in \mathbb{N}$*, the class* $\mathcal{C}_r = \{C \in \mathcal{N} : \#C \leq r\}$*. The countably infinite fields,* $\mathcal{F}^{\circ}(\mathcal{C}_{[})$*,* $\mathcal{F}^{\circ}(\mathcal{C}_{]})$ *and* $\mathcal{F}^{\circ}(\mathcal{C}_r)$*, are not VČ classes because the uniform convergence of* (3) *does not hold.*

1.3. **VČ Classes and Indeterminacy.** In Example 1.A and in the first three parts of Example 1.B, the VČ classes **determine** countably additive probabilities, i.e. if $P, P' \in \Delta^{ca}(\mathcal{X})$, then $P = P'$ iff $P(E) = P'(E)$ for all $E \in \mathcal{C}$. Most of the well-studied VČ classes determine countably additive probabilities because the point of learning $P(E)$ for all $E \in \mathcal{C}$ is that one wants to learn P. Countable classes and VČ classes of sets leave purely finitely additive probabilities indeterminate in the strong sense that not even their support sets/carriers are identified.

Stating the result requires some preliminaries.

(1) The weak* topology on $\Delta^{fa}(\mathcal{X})$ is defined by $\mu_\alpha \to_{w^*} \mu$ iff $\mu_\alpha(E) \to \mu(E)$ for all $E \in \mathcal{X}$.
(2) $\Delta^{fa}(\mathcal{X})$ is compact in the weak* topology, so that any set defined by equalities of the form $\mu(E) = r$ is compact and convex.
(3) ω denotes the cardinality of $\mathbb{N}$ so that 2^ω is the cardinality of the continuum, and 2^{2^ω} is a strictly larger cardinality.

Theorem 1. *If* $\mathcal{C} \subset \mathcal{X}$ *is countable or is a VČ class and* μ *is purely finitely additive, then* $M(\mu) := \{\mu' \in \Delta^{fa}(\mathcal{X}) : (\forall E \in \mathcal{F}^{\circ}(\mathcal{C}))[\mu'(E) = \mu(E)]\}$ *is compact, convex, and contains at least* 2^{2^ω} *many* μ' *having* $|\mu(A) - \mu'(A)| = 1$ *for uncountably many* A*.*

The proof is in §4. The first main step is Lemma 4.2, which shows that if μ is a point mass, then there are 2^{2^ω} distinct point masses μ' that agree with μ on $\mathcal{F}^{\circ}(\mathcal{C})$. Thus, even in the set of point masses, the failure of countable additivity leads to indeterminacy of an order strictly larger than the continuum.

The only way to have $|\mu(A) - \mu'(A)| = 1$ is to have $\mu(A) = 1$ or $\mu(A^c) = 1$. If μ is non-atomic and $\mu(B) = 1$, there for every $r \in (0, 1]$, there is a $B_r \subset B$ such that $\mu(B_r) = 1$. These two observations deliver the following.

Corollary 1.1. *If the conditions of Theorem* 1 *hold and* μ *is non-atomic, then there exist at least* 2^{2^ω} *purely finitely additive* μ' *that agree* *with μ on $\mathcal{F}^\circ(\mathcal{C})$ and satisfy the following condition: for any $r \in (0,1]$, $|\mu(A_r) - \mu'(A_r)| = r$ for uncountably many A_r.*

From Yosida and Hewitt [23, Theorem 1.24], every probability can be decomposed into a countably additive part and a purely finitely additive part. This yields the following, which shows that the size of the failure of countable additivity is the same as the size of the indeterminacy.

Corollary 1.2. *If the conditions of Theorem 1 hold except that μ has a deficiency $\delta \in (0,1)$, and $\mathcal{F}^\circ(\mathcal{C})$ determines countably additive probabilities, then there exist at least 2^{2^ω} μ' that agree with μ on $\mathcal{C}$, $\sup_{A\in\mathcal{X}} |\mu(A) - \mu'(A)| = \max_{A\in\mathcal{X}} |\mu(A) - \mu'(A)| = \delta$, and the maximum is achieved for uncountably many A.*

Al-Najjar [1] shows that if $\mathcal{X}$ contains all the subsets of X, a condition that rules out the existence of countably additive nonatomic probabilities, then for any VČ class $\mathcal{C}$ and any $\alpha \in (0, \frac{1}{2}]$, there exist a pair μ, μ' of purely finitely additive nonatomic probabilities that agree on $\mathcal{C}$ and have $|\mu(A) - \mu'(A)| = \alpha$ for uncountably many A. By contrast, the results here

(a) give at least 2^{2^ω} different μ' for any μ;
(b) apply to all measure spaces that separate points;
(c) apply to countable classes as well as VČ classes;
(d) tie the deficiency of μ to the maximal difference for $|\mu(A) - \mu'(A)|$; and
(e) show that if $\mathcal{C}$ determines countably additive probabilities and the purely finitely additive part of μ is non-atomic, then $|\mu(A) - \mu'(A)|$ can be any number in the interval $(0, \delta]$ where δ is the deficiency of μ.

The last item comes from combining Corollaries 1.1 and 1.2.

2. Examples of Finitely Additive Probabilities

A probability μ is purely finitely additive if there exists a sequence $E_n \downarrow \varnothing$ in $\mathcal{X}$ with $\mu(E_n) \equiv 1$. Define a random variable $f : X \to \mathbb{N}$ by $f(x) = \sum_n n \cdot 1_{E_{n+1}\setminus E_n}(x)$, so that it takes values only in $\mathbb{N}$, yet has $\mu(\{f > n\}) = 1$ for all $n \in \mathbb{N}$. It is the possibility of constructions such as this that lead to difficulties in the use of finitely additive probabilities.

This first part of this section contains a pair of examples of purely finitely additive probabilities that agree on rich classes of sets yet have different carriers. The claim in Theorem 1 is that these kinds of examples will always occur, and the examples give good intuitions for this result. The second part examines how the deficiencies of finitely additive probabilities cause deep problems in Bayesian statistics, choice under uncertainty, and stochastic process theory.

2.1. **Indeterminacy Examples.** The constructions begin with two purely finitely additive distributions, ρ and λ, on $(\mathbb{N}, \mathcal{N})$ where $\mathcal{N} = 2^{\mathbb{N}}$ is the class of all subsets of $\mathbb{N}$. Note that any $\mu \in \Delta^{fa}(\mathcal{N})$ is a point in the compact product space $Y(\mathcal{N}) := [0,1]^{\mathcal{N}}$, and that Δ^{fa} is a weak* closed, hence compact, subset of $Y(\mathcal{N})$.

•Let ρ be an accumulation point of the set $\{\delta_n : n \in \mathbb{N}\}$ where δ_n is the probability defined by $\delta_n(E) := 1_E(n)$, i.e. is point mass on n. Since $\delta_n(E) \in \{0,1\}$ for all E, the same must be true of ρ, i.e. ρ is a point mass. Since $\delta_n(F) \to 0$ for all finite F, $\rho(F) = 0$ and ρ is purely finitely additive.

•Let λ be an accumulation point of $\{U_n : n \in \mathbb{N}\}$ where U_n is the uniform distribution on the set $\{1, \ldots, n\}$. λ is a non-atomic, purely finitely additive probability.

One sees most of Theorem 1 in the following continuation of Example 1.A.

Example 2.A. *Let $(X, \mathcal{X}) = ([0,1], \mathcal{B})$, the unit interval with the Borel σ-field, let $\mathcal{C}$ to be the Glivenko-Cantelli class, $\{[0,r] : r \in [0,1]\}$, and let $\mathcal{F}^\circ(\mathcal{C})$ to be the smallest field containing $\mathcal{C}$.*

Let $S \subset \mathbb{R}_{++}$ be an uncountable set of strictly positive irrational numbers with the property that for all $s \neq s' \in S$, s/s' is irrational. For any $r > 0$, consider the infinite set of sequences $x_n^s := r - \frac{s}{n} \uparrow r$, $s \in S$. For $s \neq s'$, the sets $E^s = \{x_n^s : n \in \mathbb{N}\}$ $E^{s'} = \{x_n^{s'} : n \in \mathbb{N}\}$ are disjoint. For each $s \in S$, define two probabilities on $(X, \mathcal{X})$ by $\rho^s(E) = \rho(\{n \in \mathbb{N} : x_n^s \in E\})$ and $\lambda^s(E) = \lambda(\{n \in \mathbb{N} : x_n^s \in E\})$. Since $\rho^s((r-\epsilon, r)) = \lambda^s((r-\epsilon, r)) = 1$ for every $\epsilon > 0$, these probabilities are purely finitely additive. Further, for each $E \in \mathcal{F}^\circ(\mathcal{C})$, $\rho^s(E) = \lambda^s(E) = \delta_r(E)$, that is, we have infinitely many probabilities that all agree with each other on the $\mathcal{F}^\circ(\mathcal{C})$, and which all agree with the countably additive point mass δ_r. For any three distinct $s, s', s'' \in S$,

$$|\gamma^s(E^s) - \gamma^{s'}(E^s)| = |\gamma^s(E^s \cup E^{s''}) - \gamma^{s'}(E^s \cup E^{s''})| = 1 \tag{5}$$

for either $\gamma = \rho$ or $\gamma = \lambda$. More generally, all of the purely finitely additive probability on any of the uncountably many sets E^s agree on $\mathcal{F}^\circ(\mathcal{C})$, and since each E^s is countable, the set of purely finitely additive point masses on E^s has the same cardinality as the Stone-Čech compactification of the integers, that is, 2^{2^ω}.

The following continues Example 1.B, and demonstrates how finitely additive probabilities on discrete spaces also suffer from strong indeterminacy.

Example 2.B. *$\mathcal{F}^\circ(\mathcal{C}_{[}) = \mathcal{F}^\circ(\mathcal{C}_{]}) = \mathcal{F}^\circ(\mathcal{C}_r)$, and all three are equal to the field, $\mathcal{F}_{cof}$, of sets that are either finite or co-finite (have finite complement), a field that determines countably additive probabilities. Partition $\mathbb{N}$ into infinitely many disjoint, infinite subsets, S. Enumerate each $S \subset \mathbb{N}$ as $\{x_{S,n} : n \in \mathbb{N}\}$, define $\rho^S(E) = \rho(\{n : x_{S,n} \in E\})$, and define $\lambda^S(E) = \lambda(\{n : x_{S,n} \in E\})$. Each ρ_S and λ_S agree on $\mathcal{F}_0$, and for distinct S, S', S'',*

$$|\gamma^S(S) - \gamma^{S'}(S)| = |\gamma^S(S \cup S'') - \gamma^{S'}(S \cup S'')| = 1 \tag{6}$$

for either $\gamma = \rho$ or $\gamma = \lambda$. Much as in the previous example, all of the purely finitely additive probability on any of the sets S agree on $\mathcal{F}_{cof}$, and since each S is countably infinite, the set of purely finitely additive point masses on S has cardinality 2^{2^ω}.

2.2. **Deficiency Examples.** We now turn to examples of the problems caused by deficiencies in Bayesian statistics, choice under uncertainty, and stochastic process theory. In the first two examples, deficiencies make Bayes' Law unapplicable because partitions of the state space may omit substantial parts of the mass of the probability.

2.2.1. *Bayesian Statistics.* A probability $\mu \in \Delta^{fa}(\mathcal{X})$ fails **conglomerability** if there exists a countable partition $\pi = \{E_1, E_2, \ldots\}$ of X, an event $E \in \mathcal{X}$, and constants $k_1 \leq k_2$ such that $k_1 \leq \mu(E|E_n) \leq k_2$ for each $E_n \in \pi$, yet $\mu(E) < k_1$ or $\mu(E) > k_2$. The **conglomerability** of a probability is equivalent to it having no deficiency, that is, it is equivalent to countable additivity.[5]

The following is loosely based on Dubins [10, Theorem 2].

Example 2.C. *Let Θ be the union of two copies of the integers, indexed by $j = 0$ or $j = 1$, $\Theta = \cup\{(i, j) : i \in \mathbb{N},\ j = 0, 1\}$. The statistician's prior is $\nu = (1-\delta)Q + \delta\mu$ where: μ is purely finitely additive and Q is countably additive so that δ is the deficiency of ν; $\mu(\{j = 1\}) = 1$; and $Q(\{(i, j)\}) > 0$ for all $(i, j) \in \Theta$. The statistician observes $i = n$ and is interested in the probability that $j = 1$. This is given by $\nu(j = 1 \big| i = n) = Q(j = 1 \big| i = n)$ because $\mu(\{i = n\}) \equiv 0$. The events*

[5]For studies of conglomerability and its properties, see de Finetti's [9] discussions of coherence for fields of sets, Dubins [10] discussion of disintegration for σ-fields, Schervish et. al.'s [19] demonstration that the deficiency of a finitely additive probability determines the size of the failure of conglomerability, and Armstrong's [2] extension of Schervish et. al.

$\{\{i = n\} : n \in \mathbb{N}\}$ partition Θ, and after learning that any event in the partition has occured, the statistician ignores the deficient part of their prior in forming their posterior.

2.2.2. *Choice Under Uncertainty.* Seidenfeld and Schervish [20] showed that the failure of conglomerability leads to money pumps.

Example 2.D (Seidenfeld and Schervish)**.** *Suppose that an expected utility maximizer's prior beliefs, μ, on a state space $(X, \mathcal{X})$ is a non-atomic, purely finitely additive probability (as in Savage* [18]*). Then there exists a vector-valued random variable $(i, j)(x)$, $i \in \mathbb{N}$, $j \in \{0, 1\}$, such that: $\mu(\{j = 1\}) = \mu(\{j = 0\}) = \frac{1}{2}$; $\mu(\{i = n\}) > 0$ for all $n \in \mathbb{N}$; and $\mu(\{n, 0\}) = 2\mu(\{n, 1\})$ for all $n \in \mathbb{N}$. This failure of conglomerability guarantees that $\mu(j = 1 \big| i = n) = \frac{1}{3}$ for all $n \in \mathbb{N}$.*

Suppose that the act $\boldsymbol{a}$ is constant, always delivering a consequence c with $u(c) = 35$ while the act $\boldsymbol{b}$ delivers a consequence c' with $u(c') = 60$ if $j = 1$ and a consequence c'' with $u(c'') = 0$ if $j = 0$. Because $35 > \frac{1}{2}(0) + \frac{1}{2}(60) = 30$, $\boldsymbol{a} \succ \boldsymbol{b}$. However, the events $\{\{i = n\} : n \in \mathbb{N}\}$ partition X, and conditional on any event $\{i = n\}$ in this partition, $\boldsymbol{b} \succ \boldsymbol{a}$ because $\frac{1}{3}(0) + \frac{2}{3}(60) > 35$. A (Savage) expected utility maximizer with these preferences would pay a strictly positive amount to move from $\boldsymbol{b}$ to $\boldsymbol{a}$, and then, conditional on each and every element of a countable partition of the state space, pay a strictly positive amount to reverse the decision.

Allowing failures of countable additivity in Savage's theory of decisions under uncertainty requires giving up countable partitions of state spaces or it requires accepting a model of decisions under uncertainty in which there are money pumps. Countable partitions arise as soon as there are continuously distributed random variables, and money pumps are a form of "free lunch," so either choice carries an unacceptably high price for applications. Stinchcombe [21] shows how the points in $\widehat{X} \setminus X$ resolve the money pumps in Savage's decision theory.

2.2.3. *Stochastic Process Theory.* The previous two examples started with finitely additive probabilities on a smaller space that can be usefully identified with countably additive probabilities on a larger space. An alternative approach is to start with countably additive probabilities on a larger space and restrict them to smaller spaces.[6] This is the approach taken by Kingman [14] in his study of finitely additive probabilities in stochastic process theory.

[6]A classic example is the purely finitely additive μ on the set of all subsets of $\mathbb{Q} \cap (0, 1]$ determined by $\mu((a, b] \cap \mathbb{Q}) := b - a$.

Let $\widehat{X} = D([0,\infty))$ denote the (larger) space of right-continuous left-limit functions from $[0,\infty)$ to $\mathbb{R}$. With the Skorohod metric, this is a complete separable metric space, and $\widehat{\mathcal{X}}$, the Borel σ-field, is the smallest one making all of the evaluation mappings $f \mapsto f(t)$ measurable, $t \in [0,\infty)$, $f \in D([0,\infty))$. Further, $\widehat{\mathcal{X}}$ determines countably additive probabilities (see Billingsley [5, Ch. 3]).

For any finite set $\{t_1,\dots,t_K\} \subset [0,1]$, the mapping $f \mapsto (f(t_1),\dots,f(t_K))$ from $\widehat{X}$ to $\mathbb{R}^K$ is measurable. For a given $P \in \Delta^{ca}(\mathcal{X})$, the induced distribution is called a **finite dimensional distribution (fidi) of** P, and knowing all of the fidi's determines any $P \in \Delta^{ca}(\widehat{\mathcal{X}})$. Intuitively, the value of a continuous time stochastic process can be observed at any finite set of times, and the distributions of these observations determine the distribution of the process.

Let $X \subset D([0,\infty))$ denote the (smaller) space of polynomials, and let $\mathcal{X}$ denote the trace of $\widehat{\mathcal{X}}$ on X, i.e. the class of sets of the form $E \cap X$, $E \in \widehat{\mathcal{X}}$. Kingman shows that, given any $\widehat{\mu} \in \Delta^{ca}(\widehat{\mathcal{X}})$, there is a unique finitely additive $\mu \in \Delta^{fa}(\mathcal{X})$ that has exactly the same fidi's as $\widehat{\mu}$. This means that drawing a polynomial according to μ gives rise to observations indistinguishable from e.g. a point process (or a Brownian motion, or a combination of the two). This leaves one in the uncomfortable position of asking "So, which polynomial was that?" after one have observed the outcome of a pure jump process. Here, we see directly that $\widehat{X} \setminus X$ provides the appropriate model for the observed outcomes.

3. Discussion

Al-Najjar [1] compares the possibility of learning the probability of events in a VČ class of events in two contexts: first, when the probability is a countably additive and defined on the Borel σ-field of a complete separable metric space; and second, when it is purely finitely additive and defined on the σ-field of all subsets of an infinite space. He argues that, because the latter class of models do not have a metric structure, they provide a better model for very large finite sets, and from this he argues that the learnability arises from metric structures.

There are at least three reasons to doubt these arguments. First, the work here shows that the difference in learnability arises entirely because of the failure of countable additivity: Theorem 1 shows that nothing about metrizability or nonmetrizability of $(X,\mathcal{X})$ matters for the failure of learnability; and Example 2.A gives $2^{2^{\omega}}$ purely finitely additive probabilities on the compact metric space $[0, 1]$ that are completely indistinguishable on the classic Glivenko-Cantelli class of sets. Second, Lemma 4.4 shows that even with purely finitely additive probabilities, learnable classes of events are totally bounded metric spaces. It is learnability that delivers metrizability, rather than the other way around. Third, there is nothing about random draws in large finite spaces that resembles the paradoxes that arise from purely finitely additive probabilities.

However, purely finitely additive probabilities do have at least three advantages: the Bayes optimality of certain 'natural' statistical procedures; uncovering the basic structures of probabilistic limit arguments in the law of large numbers, the central limit theorem, and the 0-1 laws; and being able to dispense with measurability considerations.

There are many settings in which natural, non-Bayesian statistical procedures are not admissible unless one allows for uniform priors on 'arbitrarily large intervals in $\mathbb{R}$,' e.g. an accumulation point of the sequence of uniform distributions on $[-n, +n] \subset \mathbb{R}$ (see the postscript to Blackwell and Diaconis [6] for a brief discussion and some references). In these contexts, the natural procedures are the limits, as $n \to \infty$, of what a Bayesian would do if one's true prior distribution were uniform on $[-n, +n]$. Because the optimal actions at n converge to the optimal actions in the purely finitely additive limit, the deficiency of the limit has no effect.

The usual proofs of the basic limit theorems of probability, the strong law of large numbers, the central limit theorem, the various 0-1 laws, depend in a crucial fashion on countable additivity. By replacing countable additivity by succesively stronger restrictions on finitely additive models of probability, one can recover the limit theorems (see Berti et. al. [3] and Berti and Rigo [4] for an overview and recent progress in this area). This has the potential to make clearer the structure of the fundamental arguments behind the limit results.

As to measurability, by the Hahn-Banach theorem, it is always possible to extend a probability from a field or σ-field to the collection of all subsets of a space provided that one is willing to forego both the uniqueness and the countable additivity of the extension. This means that one can, if one wishes, work with probabilities in a setting where all sets are measurable if one is willing to work with sets of finitely additive probabilities.[7]

[7] In order to extend a probability to the class of all subsets, one must make uncountably many choices, and it is the Axiom of Choice that allows this. There is a symmetry here, the existence of the problematic, non-measurable sets depends on the

However, at the level of pragmatic modeling choices for economics, this seems like a bad trade: in the theory of choice under uncertainty, the cost of avoiding measurability issues is either ruling out random variables with infinite supports or accepting money pumps; in the theory of Bayesian statistics, the cost is either ruling out random variables with infinite supports or accepting that prior knowledge need not be reflected when using Bayes law; and, as this paper has shown, in the theory of learning, it means accepting complete failures of learnability from iid samples.

4. Proof

The outline of the proof is as follows.

4.1 Construct $(\widehat{X}, \widehat{\mathcal{X}})$ with $\widehat{X}$ compact and give its basic properties.
4.2 Identify points in $\widehat{X}$ with ultrafilters.
4.3 Show that for any purely finitely additive point mass, there are at least 2^{2^ω} many that agree on a countable $\mathcal{F}^\circ(\mathcal{C})$.
4.4 Show that for any purely finitely probability, there are at least 2^{2^ω} many that agree on a countable $\mathcal{F}^\circ(\mathcal{C})$.
4.5 Reduce the VČ class case to the countable case by observing that VČ classes are totally bounded metric spaces.

4.1. The Space $(\widehat{X}, \widehat{\mathcal{X}})$ and its Properties.

For the measure space $(X, \mathcal{X})$, $Y(\mathcal{X}) = [0,1]^{\mathcal{X}}$ denotes the compact product space of all functions from $\mathcal{X}$ to $[0,1]$. $\Delta^{fa}(\mathcal{X}) \subset Y(\mathcal{X})$ is a closed, hence compact, subset of $Y(\mathcal{X})$. We take $\Delta^{fa}(\mathcal{X})$ to have the relative topology, i.e. $\mu_\alpha \to_{w^*} \mu$ iff $\mu_\alpha(E) \to \mu(E)$ for all $E \in \mathcal{X}$, equivalently, iff $\int f\, d\mu_\alpha \to \int f\, d\mu$ for all bounded measurable f. The σ-field on $\Delta^{fa}(\mathcal{X})$ is denoted $\widetilde{\mathcal{X}}$ and defined as the trace on $\Delta^{fa}(\mathcal{X})$ of the Borel σ-field on $Y(\mathcal{X})$.

Definition 1. *$\rho \in \Delta^{fa}(\mathcal{X})$ is a **point mass** if $\rho(E) \in \{0,1\}$ for all $E \in \mathcal{X}$, and it is an X-**point mass** if $\rho(E) = 1_E(x)$ for some $x \in X$.*

Following Yosida and Hewitt [23, §4], we take the space $\widehat{X}$ to be the set of point masses in $\Delta^{fa}(\mathcal{X})$ and take $\widehat{\mathcal{X}}$ to be the trace of $\widetilde{\mathcal{X}}$ on $\widehat{X}$.[8] To every set $E \in \mathcal{X}$, the associated set of limit points in $\widehat{X}$ is $\widehat{E} = \{\rho : \rho(E) = 1\}$. In particular, $\widehat{\{x\}}$ is the (set containing the) X-point mass defined by $\rho_x(E) = 1_E(x)$.

Axiom of Choice, as does the existence of a solution, finitely additive probabilities. See Lauwers [15] for this.

[8]For an alternative construction of $\widehat{X}$, see Dunford and Schwarz [13, Theorem V.8.11].

The following basic properties of $(\widehat{X}, \widehat{\mathcal{X}})$ arise directly from the definitions.

(1) Being a closed subset of $\Delta^{fa}(\mathcal{X})$, $\widehat{X}$ is compact.
(2) For each $E \in \mathcal{X}$, $\widehat{E}$ is closed, hence compact, because it the intersection of $\widehat{X}$ and the closed subset of $Y(\mathcal{X})$, $\text{proj}_E^{-1}(1)$.
(3) For each $E \in \mathcal{X}$, $\widehat{E}$ is open because it the intersection of $\widehat{X}$ and the open subset of $Y(\mathcal{X})$, $\text{proj}_E^{-1}((1-\epsilon, 1+\epsilon))$, $\epsilon \in (0,1)$.
(4) For $E, F \in \mathcal{X}$, $\widehat{E \cup F} = \widehat{E} \cup \widehat{F}$, $\widehat{E \cap F} = \widehat{E} \cap \widehat{F}$, and $\widehat{E^c} = \widehat{E}^c$, from which we conclude that the class of sets $\{\widehat{E} : E \in \mathcal{X}\}$ is a field of subsets of $\widehat{X}$.
(5) The X-point masses are dense in $\widehat{X}$.[9]

Lemma 4.1. *For each $\mu \in \Delta^{fa}(\mathcal{X})$, there is a unique $\widehat{\mu}$ in $\Delta^{ca}(\widehat{\mathcal{X}})$ defined by $\widehat{\mu}(\widehat{E}) = \mu(E)$ for $E \in \mathcal{X}$. Further, the mapping $\mu \leftrightarrow \widehat{\mu}$ is one-to-one and onto.*

Proof. Let $\widehat{\mathcal{X}}^\circ$ denote the field $\{\widehat{E} : E \in \mathcal{X}\}$ and note that $\widehat{\mu}$ is finitely additive on $\widehat{\mathcal{X}}^\circ$. By Carathéodory's extension theorem, uniqueness follows if $\widehat{\mu}$ is countably additive on $\widehat{\mathcal{X}}^\circ$. Let $\widehat{E}_n \downarrow \varnothing$ be a sequence in $\widehat{\mathcal{X}}^\circ$. Since the $\widehat{E}_n$ are compact, the finite intersection property implies that $\cap_n \widehat{E}_n = \varnothing$ iff $\widehat{E}_m = \varnothing$ for all $m \geq N$ for some N, so that $\lim_n \widehat{\mu}(\widehat{E}_n) = 0$. The one-to-one and onto properties are immediate. □

Here is why the space $\widehat{X}$ contains representations of the points that carry a finitely additive probability: if $\mu(E_n) \equiv 1$ while $E_n \downarrow \varnothing$, i.e. if μ is purely finitely additive, then $\widehat{\mu}(K) = 1$ where K is the non-empty, compact set $\cap_n \widehat{E}_n$; and, because $\{x\} \in \mathcal{X}$ for all $x \in X$, $K \cap \{x\} = \varnothing$, i.e. $K \subset \widehat{X} \setminus X$.

4.2. **Points in $\widehat{X}$ as Ultrafilters.** Points in $\widehat{X}$ are identified with $\mathcal{X}$-ultrafilters.

Definition 2. *For $\mathcal{G} \subset \mathcal{X}$ a field (or σ-field), an $\mathcal{G}$-**ultrafilter** is a class of non-empty sets, $\mathbb{U} \subset \mathcal{G}$ that is closed under finite intersection, $[A, B \in \mathbb{U}] \Rightarrow [A \cap B \in \mathbb{U}]$, contains all supersets of its elements in $\mathcal{G}$, $[A \in \mathbb{U},\ B \in \mathcal{G},\ A \subset B] \Rightarrow [B \in \mathbb{U}]$, and has the property that for*

[9]Sets of the form $G(\rho) := \{\rho' : |\rho'(E_i) - \rho(E_i)| < \epsilon_i,\ i = 1, \ldots, I\}$ form a neighborhood basis for a point mass ρ. Since $\rho(E) \in \{0,1\}$ and ρ is a probability, there is no loss in changing E_i for E_i^c so that each $\rho(E_i) = 1$ in forming these open neighborhoods. For any finite collection of sets E_i, $i = 1, \ldots, I$ with $\rho(E_i) = 1$, let E denote the necessarily non-empty set $\cap_{i=1}^I E_i$, pick $x \in E$ and note that the X-point mass defined by $\rho'(E) = 1_E(x)$ belongs to $G(\rho)$.

*all $G \in \mathcal{G}$, either $G \in \mathbb{U}$ or $G^c \in \mathbb{U}$. An $\mathcal{X}$-ultrafilter is **fixed** if its intersection is non-empty, otherwise it is **free**.*

Because $\mathcal{X}$ separates points, the intersection of any fixed $\mathcal{X}$-ultrafilter is of the form $\{x\}$ for some $x \in X$. From Zorn's lemma, we know that any collection of sets in $\mathcal{X}$ with the finite intersection property is contained in an $\mathcal{X}$-ultrafilter. In particular, every $\mathcal{G}$-ultrafilter is contained in at least one $\mathcal{X}$-ultrafilter.

The following observations are immediate from properties of point masses: to any point mass ρ, there corresponds a unique $\mathcal{X}$-ultrafilter $\mathbb{U}_\rho := \{E \in \mathcal{X} : \rho(E) = 1\}$; to any $\mathcal{X}$-ultrafilter $\mathbb{U}$, there corresponds a unique point mass $\rho_{\mathbb{U}}(E) = 1_{\mathbb{U}}(E)$, i.e. $\rho_{\mathbb{U}}(E) = 1$ iff $E \in \mathbb{U}$; the mappings $\rho \mapsto \mathbb{U}_\rho$ and $\mathbb{U} \mapsto \rho_{\mathbb{U}}$ are both are one-to-one and onto and are inverses of each other. Further, the points in $\widehat{X} \setminus X$ correspond to the free $\mathcal{X}$-ultrafilters, and the points in X correspond to the fixed ultrafilters with $\{x\}$ being the smallest element.

4.3. **Countable Class Point Mass Indeterminacy.** ω denotes the cardinality of $\mathbb{N}$ so that 2^ω is the cardinality of the continuum, and 2^{2^ω} is a strictly larger cardinality.

Lemma 4.2. *If $\mathcal{C} \subset \mathcal{X}$ is countable and ρ is a purely finitely additive point mass, then the set $\{\rho' : (\forall C \in \mathcal{F}^\circ(\mathcal{C}))[\rho(C) = \rho'(C)]\}$ contains at least 2^{2^ω} many distinct point masses.*

Proof. As $\mathcal{C}$ is countably infinite, so is $\mathcal{F}^\circ(\mathcal{C})$. Let $\mathcal{S} = \{C \in \mathcal{F}^\circ(\mathcal{C}) : \rho(C) = 1\}$ and enumerate $\mathcal{S}$ as $\{S_m : m \in \mathbb{N}\}$. Pick a sequence $E_n \downarrow \varnothing$ in $\mathcal{X}$ with $\rho(E_n) \equiv 1$ and define $D_n = E_n \cap S_n$. For each D_n, $\rho(D_n) = 1$, and since ρ is purely finitely additive, each D_n is infinite, and the collection D_n has the finite intersection property. The result is now a Corollary to the disjoint refinement lemma for ultrafilters [7, Lemma 7.7 and Cor. 7.9], or alternately from the observation that $\widehat{D}_n$ must contain a set homeomorphic to the Stone-Čech compactification of the integers, which has cardinality 2^{2^ω}. $\square$

4.4. **Countable Class Indeterminacy.** If $\mu(E_n) \equiv 1$ and $E_n \downarrow \varnothing$, then $\widehat{\mu}$ is supported on the non-empty compact set $K := \cap_n \widehat{E}_n$. By the Choquet-Bishop-de Leeuw theorem for compact non-metrizable spaces (e.g. Phelps [16, §4]), $\widehat{\mu}$ has a unique representation as an integral over the points masses in K. Since each point mass has at least 2^{2^ω} other point masses that agree on $\mathcal{C}$, a little bit of rearrangement proves the following, which is the part of Theorem 1 that deals with countable classes of sets.

Lemma 4.3. *If $\mathcal{C} \subset \mathcal{X}$ is countable and μ is purely finitely additive, then there exist at least 2^{2^ω} purely finitely additive μ' that agree with μ on $\mathcal{C}$ and have $|\mu(A) - \mu'(A)| = 1$ for uncountably many A.*

Proof. Let $E_n \downarrow \varnothing$ be a sequence in $\mathcal{X}$ with $\mu(E_n) \equiv 1$. Enumerate $\mathcal{C}$ as $\{T_n : n \in \mathbb{N}\}$ and let $\{C_{n,m} : m = 1, \ldots M(n)\}$ denote the partition of X generated by $T_1, \ldots, T_n$. For each $n \in \mathbb{N}$, pick all of the $m(n) \in \{1, \ldots M(n)\}$ such that $C_{n,m(n)} \cap E_n$ forms a chain with $\mu(C_{n,m(n)} \cap E_n) > 0$ for each $n \in \mathbb{N}$. Because μ is purely finitely additive, each $C_{n,m(n)} \cap E_n$ is infinite. As in Lemma 4.2, each set $\cap_n(\widehat{C}_{n,m(n)} \cap \widehat{E}_n)$ has cardinality at least 2^{2^ω}, and it can be partitioned into uncountably many disjoint measurable sets. For each pair $n, m(n)$, pick a point mass, $\widehat{\rho}_{n,m(n)}$, on one of the partition elements of $\cap_n(\widehat{C}_{n,m(n)} \cap \widehat{E}_n)$. Set $\beta_{n,m(n)} = \mu(C_{n,m(n)} \cap E_n)$. The set of accumulation points of all the probabilities $\widehat{\mu}'_n := \sum \beta_{n,m(n)} \widehat{\rho}_{n,m(n)}$ that arise from different sequences of choices of point masses $\widehat{\rho}_{n,m(n)}$ has cardinality at least 2^{2^ω}, and the partitions of the sets $\cap_n(\widehat{C}_{n,m(n)} \cap \widehat{E}_n)$ give the uncountably many A with $|\mu(A) - \mu'(A)| = 1$. $\square$

4.5. **VČ Class Indeterminacy.** For a pseudo-metric space (S, d) and $\epsilon > 0$, $D(\epsilon, S, d)$ is defined as the maximum number of points in S that are all more than ϵ apart. For a probability Q and measurable sets E, F, we have the pseudo-metric $d_{2,Q}(E, F) = Q(E\Delta F)^{1/2} = \left(\int (1_{E_1} - 1_{E_2})^2 \, dp\right)^{1/2}$ where $E\Delta F$ is the symmetric difference of E and F. For a class of sets $\mathcal{C}$, $D^{(2)}(\epsilon, \mathcal{C})$ is defined as the supremum of $D(\epsilon, \mathcal{C}, d_{2,Q})$ where the supremum is taken over finitely supported Q. From Dudley [11, Theorem 10.1.7], if $\mathcal{C}$ is a VČ class, then

$$\log D^{(2)}(\epsilon, \mathcal{C}) = O(\epsilon^{-2}) \text{ as } \epsilon \downarrow 0. \tag{7}$$

We now argue that:

(1) $\mathcal{C}$ satisfies (7) iff $\widehat{\mathcal{C}}$ satisfies (7) where $\widehat{\mathcal{C}} = \{\widehat{C} : C \in \mathcal{C}\}$;
(2) that the supremum in $D^{(2)}(\epsilon, \widehat{\mathcal{C}})$ can be taken over all countably additive probabilities on $\widehat{\mathcal{X}}$; and that
(3) this in turn implies that for any purely finitely additive μ, $\mathcal{C}$ is totally bounded in the $d_\mu(C_1, C_2) := \mu(C_1 \Delta C_2)^{1/2}$ pseudo-metric.

Letting $\mathcal{C}_0$ be a countable d_μ-dense subset of $\mathcal{C}$ and applying Lemma 4.3 will then complete the proof of Theorem 1.

Lemma 4.4. *$\mathcal{C}$ satisfies* (7) *iff $\widehat{\mathcal{C}}$ satisfies* (7)*.*

Proof. It is sufficient to show that $\widehat{X}$, $\sup_p D(\epsilon, \mathcal{C}, d_{2,p}) = \sup_q D(\epsilon, \widehat{\mathcal{C}}, d_{2,q})$ where the first supremum is taken over finitely supported probabilities in X and the second is taken over finitely supported probabilities in $\widehat{X}$. This follows from the openness of each $\widehat{C}$ and denseness of X in $\widehat{X}$. □

From the usual approximation of all countably additive probabilities on a compact Hausdorff space by finitely supported probabilities on dense subsets, this means that the supremum $\sup_q D(\epsilon, \widehat{\mathcal{C}}, d_{2,q})$ could just as well be taken over all $q \in \Delta^{ca}(\widehat{\mathcal{X}})$. Therefore, for any purely finitely additive μ, we know that

$$\log D(\epsilon, \widehat{\mathcal{C}}, d_{2,\widehat{\mu}}) = \log D(\epsilon, \mathcal{C}, d_{2,\mu}) \leq O(\epsilon^{-2}) \text{ as } \epsilon \downarrow 0. \tag{8}$$

Since $\widehat{\mu}(\widehat{C}) = \mu(C)$ for all $C \in \mathcal{C}$, this means that $\mathcal{C}$ is totally bounded in the pseudo-metric $d_{2,\mu}$.

Lemma 4.5. *If $\mathcal{C} \subset \mathcal{X}$ is a VČ class and μ is purely finitely additive, then there exist at least 2^{2^ω} purely finitely additive μ' that agree with μ on $\mathcal{C}$ and have $|\mu(A) - \mu'(A)| = 1$ for uncountably many A.*

Proof. The foregoing establishes that there exists a countable $\mathcal{C}_0 \subset \mathcal{C}$ that is $d_{2,\mu}$-dense in $\mathcal{C}$. From Lemma 4.3, there are at least 2^{2^ω} purely finitely additive probabilities μ' agreeing with μ on $\mathcal{C}_0$ and satisfying $|\mu(A) - \mu'(A)| = 1$ for uncountably many $A \in \mathcal{X}$. For any of these μ', $\widehat{\mu}'$ agrees with $\widehat{\mu}$ on $\widehat{\mathcal{C}}_0$. Since $\widehat{\mu}'$ and $\widehat{\mu}$ are countably additive and $\widehat{\mathcal{C}}_0$ is $d_{2,\widehat{\mu}}$ dense in $\widehat{\mathcal{C}}$, agreement of μ' and μ on $\mathcal{C}_0$ implies agreement on $\mathcal{C}$. □

Since compactness and convexity of $M(\mu)$ in Theorem 1 are immediate, Lemmas 4.3 and 4.5 complete the proof.

References

[1] Al-Najjar, N. Decision makers as statisticians: diversity, ambiguity, and learning. *Econometrica 77*, 5 (2009), 1371–1401.

[2] Armstrong, T. E. Conglomerability of probability measures on Boolean algebras. *J. Math. Anal. Appl. 150*, 2 (1990), 335–358.

[3] Berti, P., Regazzini, E., and Rigo, P. Modes of convergence in the coherent framework. *Sankhyā 69*, 2 (2007), 314–329.

[4] Berti, P., and Rigo, P. Finitely additive uniform limit theorems. *Sankhyā 68*, 1 (2006), 24–44.

[5] Billingsley, P. *Convergence of probability measures*, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1999. A Wiley-Interscience Publication.

[6] Blackwell, D., and Diaconis, P. A non-measurable tail set. In *Statistics, probability and game theory*, vol. 30 of *IMS Lecture Notes Monogr. Ser.* Inst. Math. Statist., Hayward, CA, 1996, pp. 1–5.

[7] Comfort, W. W., and Negrepontis, S. *The theory of ultrafilters.* Springer-Verlag, New York, 1974. Die Grundlehren der mathematischen Wissenschaften, Band 211.

[8] Corbae, D., Stinchcombe, M. B., and Zeman, J. *An introduction to mathematical analysis for economic theory and econometrics.* Princeton University Press, Princeton, NJ, 2009.

[9] de Finetti, B. *Probability, induction and statistics. The art of guessing.* John Wiley & Sons, London-New York-Sydney, 1972. Wiley Series in Probability and Mathematical Statistics.

[10] Dubins, L. E. Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probability 3* (1975), 89–99.

[11] Dudley, R. M. *Uniform central limit theorems*, vol. 63 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1999.

[12] Dudley, R. M., Giné, E., and Zinn, J. Uniform and universal Glivenko-Cantelli classes. *J. Theoret. Probab. 4*, 3 (1991), 485–510.

[13] Dunford, N., and Schwartz, J. T. *Linear operators. Part I.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[14] Kingman, J. F. C. Additive set functions and the theory of probability. *Proc. Cambridge Philos. Soc. 63* (1967), 767–775.

[15] Lauwers, L. Purely finitely additive measures are non-constructible objects. Working Paper, Center for Economic Studies, K.U.Leuven, 2010.

[16] Phelps, R. R. *Lectures on Choquet's theorem*, second ed., vol. 1757 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2001.

[17] Pollard, D. Limit theorems for empirical processes. *Z. Wahrsch. Verw. Gebiete 57*, 2 (1981), 181–195.

[18] Savage, L. J. *The foundations of statistics*, revised ed. Dover Publications Inc., New York, 1972.

[19] Schervish, M. J., Seidenfeld, T., and Kadane, J. B. The extent of non-conglomerability of finitely additive probabilities. *Z. Wahrsch. Verw. Gebiete 66*, 2 (1984), 205–226.

[20] Seidenfeld, T., and Schervish, M. J. A conflict between finite additivity and avoiding Dutch book. *Philos. Sci. 50*, 3 (1983), 398–412.

[21] Stinchcombe, M. B. Countably additive subjective probabilities. *Rev. Econom. Stud. 64*, 1 (1997), 125–146.

[22] Vapnik, V. N., and Červonenkis, A. J. On the uniform convergence of frequencies of the appearance of events to their probabilities. *Theor. Probability Appl. 16* (1971), 264–280.

[23] Yosida, K., and Hewitt, E. Finitely additive measures. *Trans. Amer. Math. Soc. 72* (1952), 46–66.

Department of Economics, University of Texas, Austin, TX 78712-0301 USA, e-mail: maxwell@eco.utexas.edu