OBJECTIVE AND SUBJECTIVE FOUNDATIONS FOR MULTIPLE PRIORS

MAXWELL B. STINCHCOMBE

Abstract. Foundations for priors can be grouped in two broad categories: objective, deriving probabilities from observations of similar instances; and subjective, deriving probabilities from the internal consistency of choices. Partial observations of similar instances and the Savage-de Finetti extensions of subjective priors yield objective and subjective *sets* of priors suitable for modeling choice under ambiguity. These sets are best suited to such modeling when the distribution of the observables, or the prior to be extended, is non-atomic. In this case, the sets can be used to model choices between elements of the closed convex hull of the faces in the set of distributions over outcomes, equivalently, all sets bracketed by the upper and lower probabilities induced by correspondences.

1. Introduction

An uncertain decision problem is one where chance intervenes between the decision and the outcome. When chance intervenes according to a known probability distribution, it called a risky decision problem, when the probability distribution is unknown or only partially known, it is called an ambiguous problem. Introspection and experiments have motivated research into the systematic differences between ambiguous decisions and risky decisions in single agent problems, general equilibrium, finance theory, and games.

The model of risky choice most frequently extended to ambiguous choice involves preferences between state-contingent bundles, that is, between functions from a state space to utility relevant consequences. The model covers risky choice problems by assuming that there is a single prior probability distribution over states, it covers ambiguous choice problems by replacing the single prior by a set of priors. In modeling either risky or ambiguous choice problems, the states are, by assumption, utility neutral. For risky problems, the neutrality assumption implies that preferences between functions depend only on the distributions they induce, for ambiguous problems, it implies that preferences depend only on the set of distributions they induce. The Skorokhod question for a set of priors is what sets they can induce.

1.1. **Foundations and the Skorokhod Question.**

For models of risky choice, approaches to the foundations of probability and its meaning are grouped in two broad categories. The objective approaches derive the single prior from observations of similar instances, the subjective approaches derive the single prior from an internal consistency of choices that reveals prior knowledge. This paper develops two observational learning models and one subjective model as foundations for the sets of priors that replace the single prior in ambiguous choice models, and then answers the pragmatic, Skorokhod question, "Which problems can and which cannot be modeled using these sets of priors?"

For single prior models of risky choice, the answer to the corresponding pragmatic question turns on the non-atomicity of the prior. If a decision maker's prior has an atom of some given size, then every distribution induced by a function from states to consequences has an atom of at least that size. By contrast, if the prior is non-atomic, then for a broad range of models of utility relevant consequences, Skorokhod's representation theorem [33, Thm. 3.1.1] implies that *every* distribution on consequences arises as the image law of the prior under some function. The

---

Date: April 22, 2016. Many thanks to two anonymous referees, Ed Green, Urmee Khan, Renou Ludovic, Patrick O'Callaghan, and Marcin Peski for perspective on and help with this project.

Key words and phrases: multiple prior models and ambiguous choice; partial observability and partially identified models; finitely additive learning models; Savage-de Finetti indeterminacy; Dempster compatible sets of probabilities; difficulty of learning problems.

answer for the sets of priors developed here is not correspondingly inclusive, though the descriptive possibilities are largest in the presence of non-atomicity.

The objective models developed here derive sets of priors from repeated partial observations, and the partial nature of the information leads to the indeterminacy needed to model ambiguous choice problems. The subjective model supposes that there is internal consistency on a subset of possible choices, a subset that only partially determines a prior. In both classes of models, the sets of priors are defined as all probabilities consistent with partial information, equivalently, as the set of all extensions of a probability defined on a small field or σ-field of sets to a larger one. Such sets of extensions have an integral representation, and it is the integral representation that leads to the answer to the Skorokhod question. The following is the main result of the paper.

> If the distribution of observables or the prior to be extended is non-atomic, then in both classes of models, the descriptive range of the associated sets of probabilities is the closed convex hull of the set of closed faces.

We turn now to short discussions of the three models and the class of decision problems the associated sets of priors can model, suppressing many details.

1.2. **Partially Observable Econometric Models.** In this class of models, repeated observations of a random vector (Y, X) lead, in the limit, to knowledge of the joint distribution, denoted μ. From this joint distribution one can answer questions of the form "How would the expected value of Y change if we observed $X = x'$ rather than x?" However, economic interest centers on the *cæteris paribus* value of the effect of changing X from x to x', and knowledge of the joint distribution of Y and X is not sufficient to identify this in the presence of partial observability.

If there is an unobserved random Z with the property that $E(Y|X)$ is not equal to $E(Y|X, Z)$, then we have a partially observable model. Examples include selection biases, errors-in-variables and other forms of endogeneity. Selection biases arise when agents are described by (e.g. demographic) variables X but use the unobserved Z (e.g. abilities and/or resources) to make choices that affect the value of Y (e.g. levels of education, income or savings). Errors-in-variables arise when the observed values of X are an error-perturbed version of the true causal variable, and the perturbations, Z, are stochastically related to the outcome, Y.

In this context, the learned set of priors is the set of joint distributions of the vector $((Y, X), Z)$ having marginal equal to μ. Let $\mathcal{C}$ denote the σ-field for the observations of (Y, X) and $\mathcal{X}$ the σ-field for $((Y, X), Z)$. The set of extensions of interest is the **partially identified set**, here defined as the set of countably additive ν on $\mathcal{X}$ such that $\nu(E) = \mu(E)$ for all $E \in \mathcal{C}$, and denoted $\Pi(\mu)$.

In models with utility neutral states, if a decision maker must choose between the measurable functions, say $((y, x), z) \mapsto f((y, x), z)$ and $((y, x), z) \mapsto g((y, x), z)$, then the choice depends only on preferences between the sets of distributions on consequences induced by the functions, that is, depends only on preferences between $f(\Pi(\mu))$ and $g(\Pi(\mu))$. The **descriptive range** of $\Pi(\mu)$, is the class of all sets of the form $f(\Pi(\mu))$. Integrating conditional probabilities provides an answer to the Skorokhod question, "How large is the descriptive range of $\Pi(\mu)$?"

For every (y, x), let $q_{y,x}$ be a distribution for Z. The integrals of this mapping, $\nu(E \times E') = \int_E q_{y,x}(E')\, d\mu(y, x)$, gives rise to a ν in the partially identified set, E a measurable set of observables E' a measurable set of unobservables. Further, all partially identified ν arise from some measurable $(y, x) \mapsto q_{y,x}$ under this mapping.[1]

For a function $((y, x), z) \mapsto f((y, x), z)$, $F(y, x)$ denotes the range of $f((y, x), \cdot)$ and $\Delta(F(y, x))$ denotes the set of distributions putting mass 1 on $F(y, x)$. Conditional on any given (y, x), the set of image measures for the partially identified set is $\Delta(F(y, x))$ because there are no restrictions on the distribution $q_{y,x}$. From the definition of the integral of a correspondence, this yields

$$f(\Pi(\mu)) = \int \Delta(F(y, x))\, d\mu(y, x). \tag{1}$$

Thus, $f(\Pi(\mu))$ is a convex combination of sets of the form $\Delta(F)$ with convex weights given by $\mu(\cdot, \cdot)$.

[1]For a quick development of these results, see Dellacherie and Meyer [12, III.72-4].

The set of distributions putting mass 1 on a set F is denoted $\Delta(F)$ and called a **face** in the simplex of distributions. The integral representation in (1) gives the answer to the Skorokhod question — the sets of distributions over consequences that decision maker can conceive of choosing between are convex combinations of faces. Further, we will see that if μ is non-atomic, then all convex combinations of faces arise this way.

1.3. **Purely Finitely Additive Partial Observability.** In this class of models, repeated observations allow the decision maker to learn the probability, $\mu(E)$, of every set E in a field of observable events, $\mathcal{C}$ where $\mathcal{C}$ is a strict subset of the set of all events, $\mathcal{X}$. Again, we define $\Pi(\mu)$ as the set of all probabilities ν that agree with μ on $\mathcal{C}$. There are two differences between this model and the previous one: first, $\mathcal{C}$ generates the σ-field, $\mathcal{X}$, on which μ is defined; and second, μ is, by assumption, purely finitely additive (pfa).

If μ is countably additive on $\mathcal{C}$, then Carathéodory's extension theorem (e.g. [6, Thm. 3.1, p. 36]) guarantees that the set of countably additive extensions of μ from $\mathcal{C}$ to $\mathcal{X}$, is a singleton set. It is the failure of countable additivity that makes this class of models indeterminate: Lemma 4 (below) shows that if the smallest field containing $\mathcal{C}$ is a strict subset of $\mathcal{X}$, then for some probability μ, $\Pi(\mu)$ contains a probability ν with $\sup_{E\in\mathcal{X}} |\mu(E)-\nu(E)| = 1$. Under an additional condition on $\mathcal{C}$, Theorem 2 shows that this diameter 1 result holds for all purely finitely additive μ.

We can see how the indeterminacy of $\Pi(\mu)$ arises in the context of the simplest infinite state space: let $\mathbb{X}$ be the complete separable metric space of integers, $\mathbb{N}$, with the metric $d(n,m) = |n-m|$; and let the σ-field, $\mathcal{X}$, be the associated Borel σ-field. A probability, μ on $\mathcal{X}$, is finitely additive if for all disjoint sets E and F, $\mu(E\cup F) = \mu(E)+\mu(F)$. A probability μ on $\mathcal{X}$ is **countably additive (ca)** on a field $\mathcal{C}\subset\mathcal{X}$ if for every decreasing sequence of sets E_n in $\mathcal{C}$ with empty intersection, $\mu(E_n)\downarrow 0$, and it is **purely finitely additive (pfa)** on $\mathcal{C}$ if $\mathcal{C}$ contains a decreasing sequence of sets, E_n with empty intersection and $\mu(E_n)\equiv 1$.

Every observation, $X_n\in\mathbb{X}$, is recorded as the vector $(1_C(X_n))_{C\in\mathcal{C}}$. We suppose that the law of large numbers holds, that is, that $\frac{1}{N}\sum_{n=1}^{N} 1_C(X_n)\to\mu(C)$ for every $C\in\mathcal{C}$. Thus, asymptotically, each $\mu(C)$ is known, and $\Pi(\mu)$ is the set of finitely additive probabilities consistent with that knowledge.

Using the field $\mathcal{C}$ consisting of all finite sets and their complements yields an example in which the analysis is more transparent.[2] The value of any probability μ on the field $\mathcal{C}$ are determined by the values of the cumulative distribution function (cdf), $F_\mu(N) := \mu(\{1,\ldots,N\})$. A probability μ is pfa if and only if $F_\mu(N)\equiv 0$, equivalently, if and only if $\mu(\{N+1,N+2,\ldots\})\equiv 1$.

The cdf being identically equal to 0 gives one sense the partial observability — one never observes an integer when drawing from a pfa because all the mass is far out "to the right of $\mathbb{N}$." To get a sharper sense of this, let $S=\{0,1\}^{\mathcal{C}}$ denote the space of possible records and let $s^\infty\in S$ denote the vector with $s^\infty_C = 0$ for all finite $C\in\mathcal{C}$ and $s^\infty_C = 1$ for $C\in\mathcal{C}$ with finite complement. A probability on $\mathbb{X}$ is pfa if and only if it is a probability 1 event that the record is s^∞. Here is a more precise sense of the partial observability — the state space $\mathbb{X}$ is not observable because there is no point in $\mathbb{X}$ that yields the record s^∞ and s^∞ happens with probability 1.

In this example, the probability induced by the observations is always the same point mass, which implies that $\Pi(\mu)$ is equal to the set of all pfa probabilities. Therefore, for any function f, $f(\Pi(\mu)) = \Delta(F)$ where F is the set of utility relevant consequences, w for which $f^{-1}(w)$ is infinite. In this, the simplest version of the pfa observational model, the decision maker can conceive of choosing between faces, but cannot conceive of choosing between any other sets of probabilities on utility relevant consequences. Here we see that it is the interaction between the probability μ and the class of observable events, $\mathcal{C}$, that determines how useful the indeterminacy is — if μ is non-atomic on $\mathcal{C}$, the descriptive range of $\Pi(\mu)$ will contain all convex combinations of faces.

1.4. **The Savage-De Finetti Extension Set.** In this class of models, we assume that the decision maker can, in principle at least, specify for each event E in a class $\mathcal{C}$ and each $r\in[0,1]$, whether they would prefer to take the bet $f_E^r := 1_E - r$ or to take the other side of the bet,

[2]The following analysis is a slight elaboration of [43, Note 1.8a].

$g_E^r := r - 1_E$. Define r_E as the number making the bets f_E^r and g_E^r indifferent. Ramsey and de Finetti independently showed that a preference ordering over this collection of bets is internally consistent if and only if the mapping $E \mapsto r_E$ is a finitely additive probability.[3]

The proposed subjective model for sets of priors supposes that the mapping $E \mapsto r_E$ determines a prior on $\mathcal{C}$ but that the full set of events is the much larger class $\mathcal{X} = \mathcal{P}(\mathbb{X})$, the class of all subsets of $\mathbb{X}$. Use of the measure space $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ imposes very strong restrictions on the structure of the set of probabilities: all countably additive probabilities are either finite or countably infinite convex combinations of point masses; all non-atomic probabilities are pfa. Both Savage and de Finetti argued these restrictions are acceptable, even necessary, because the existence of non-measurable sets represents a challenge, perhaps insuperable, to the meaning of probability (e.g. [9, p. 124], [31, Ch. 3, §3]).

The **Savage-de Finetti extension set**, $\Pi(\mu)$, is the set of probabilities on $\mathcal{X} = \mathcal{P}(\mathbb{X})$ such that $\nu(C) = \mu(C)$ for all $C \in \mathcal{C}$. De Finetti argued that the indeterminacy of $\Pi(\mu)$ captures something fundamental about probability (see esp. [9, pp. 229-231]). We will see that, under quite general conditions, there exists a measurable class, $\mathcal{E}$, of disjoint, uncountable, measurable sets of states and a probability η_μ on $\mathcal{E}$ such that, for any f,

$$f(\Pi(\mu)) = \int_{\mathcal{E}} \Delta(f(E))\, d\eta_\mu(E) \tag{2}$$

where $\Delta(f(E))$ is the set of probabilities putting mass 1 on $f(E)$. The integral representation in (2) yields an answer to the Skorokhod question: each $f(\Pi(\mu))$ is a convex combination of faces in the space of distributions over utility relevant consequences; and if η_μ is non-atomic, then all convex combinations of faces are of the form $f(\Pi(\mu))$ for some state-contingent bundle f.

1.5. **The Skorokhod Question.** The descriptive range of a set of priors is the class of sets of distributions over consequences for which it provides a model. As noted, if the distribution of observables or the prior to be extended is non-atomic, then in all three classes of models, the descriptive range of the associated $\Pi(\mu)$ is the closed convex hull of the set of closed faces. In the case that there three utility relevant consequences, $\mathbb{W} = \{a, b, c\}$, it is easy to see that this result implies that the sets of priors developed here can model a limited, but non-trivial, class of ambiguous choice problems.

The set of closed convex subsets of $\Delta(\{a, b, c\})$ is infinite dimensional, but the convex hull of the faces is the closed, seven-dimensional class of sets spanned by $\Delta(\{a\})$, $\Delta(\{b\})$, $\Delta(\{c\})$, $\Delta(\{a, b\})$, $\Delta(\{a, c\})$, $\Delta(\{b, c\})$, and $\Delta(\{a, b, c\})$. Convex combinations of the first three sets are risky outcomes, the more general elements of this class are also known as Dempster's [13, §2] class of "compatible measures," that is, they are sets of probabilities that are bracketed between the upper and lower probabilities induced by a correspondence.

A widely used subclass of the convex combinations of faces are the upper/lower probability intervals. For $0 \leq r \leq s \leq 1$, these are defined by

$$A = (s - r)\Delta(\{w, w'\}) + r\Delta(\{w\}) + (1 - s)\Delta(\{w'\}) = \{\beta\delta_w + (1 - \beta)\delta_{w'} : \beta \in [r, s]\} \tag{3}$$

where δ_w and $\delta_{w'}$ are point masses on the consequences w and w'.

The need for non-atomicity can be seen in the version of the pfa observational model sketched in §1.3 — because the observables had a point mass distribution, no convexification was possible. A decision maker with the associated set of priors, the set of all pfa probabilities, can only conceive of choices between faces, a class that includes no non-degenerate risky choices.

1.6. **Outline.** Before turning to the three models, §2 gives more detail about the descriptive ranges of a sets of priors. §3 covers the countably additive partially observable econometrics models while §4 covers the two purely finitely additive models. §5 contains comparisons between the countably additive and the purely finitely additive models, as well as comparisons of the present approach to Dempster's [13, §2] class of "compatible measures," and to interval-valued probabilities. §6 contains a summary, sketches of some extensions and open questions, as well as a discussion of where the present work fits into the literatures on related topics.

[3]We will cover this argument and a generalization of it due to Smith [34] below.

Throughout: all probabilities are assumed to be finitely additive, when used, the assumption of countable additivity will be explicitly invoked; we also assume that events determine states, that is, that for states $x \neq x'$, there exists an event E in the σ-field $\mathcal{X}$ such that $1_E(x) \neq 1_E(x')$; to avoid complete triviality, we always assume that the space of utility relevant consequences, $\mathbb{W}$ contains two or more points; to avoid some distracting complications, we also assume that $\mathbb{W}$ is a compact metric space and that preferences over constant functions (those inducing point mass distributions) are continuous in the topology on $\mathbb{W}$.

Finally, "Theorem" is reserved for the main results of the paper, other results are "Lemmas," and all proofs are relegated to the appendix.

2. The Descriptive Range of a Set of Priors

The von Neumann-Morgenstern [42] approach to risky choice posits preferences over distributions p, q in $\Delta(\mathbb{W})$, the distributions on utility relevant consequences. By contrast, the Savage [31] approach posits preferences over measurable functions taking values in $\mathbb{W}$ that depend only on the distributions the pfa prior μ induces. For a countably additive prior, μ, Skorokhod's representation theorem [33, Thm. 3.1.1] tells us that every countably additive $p \in \Delta(\mathbb{W})$ can be induced by some measurable function if and only if μ is non-atomic. The following extends this result, so far as possible, to finitely additive priors.

Lemma 1. *A probability $\mu \in \Delta^{fa}(\mathcal{X})$ is non-atomic if and only if for every probability p on $\mathbb{W}$, there exists a measurable $f_p : \mathbb{X} \to \mathbb{W}$ such that for all continuous $v : \mathbb{W} \to \mathbb{R}$, $\int_{\mathbb{X}} v(f_p(x))\, d\mu(x) = \int_{\mathbb{W}} v(w)\, dp(w)$.*

The descriptive range of a set of priors, Π, is the class of sets of distributions that can be induced on $\mathbb{W}$ using measurable functions. For the sets of priors developed in this paper, the descriptive ranges are the closed convex hull of the faces in $\Delta(\mathbb{W})$. This class of descriptive ranges is Dempster's [13, §2] class of "compatible measures," those induced by correspondences.

2.1. **Notation.** The space $\mathbb{W}$ is equipped with its Borel σ-field, $\mathcal{W}$. Two (finitely additive) probabilities, p, p' on $\mathcal{W}$ are **equivalent**, $p \sim p'$, if $\int_{\mathbb{W}} v(w)\, dp(w) = \int_{\mathbb{W}} v(w)\, dp'(w)$ for every continuous $v : \mathbb{W} \to \mathbb{R}$. Integration against finitely additive probabilities defines a continuous linear functional on $C(\mathbb{W})$, the continuous functions on $\mathbb{W}$. Therefore the Riesz representation theorem (e.g. [8, Thm. 9.6.4]) implies that every equivalence class of finitely additive probabilities contains a countably additive probability.

The set of measurable functions $f : \mathbb{X} \to \mathbb{W}$ is denoted $M(\mathbb{X}; \mathbb{W})$. For $\mu \in \Delta^{fa}(\mathcal{X})$ and $f \in M(\mathbb{X}; \mathbb{W})$, $f(\mu)$ denotes the (equivalence class of) of the induced probability on $\mathbb{W}$, that is, the probability assigning $\mu(f^{-1}(E))$ to each measurable $E \in \mathcal{W}$. The following class of sets of distributions on $\mathbb{W}$ will play a crucial role below.

Definition 1. *A set D of probabilities on $\mathbb{W}$ is a **face** associated with a closed $F \subset \mathbb{W}$ if $p \in D$ iff $p \sim p'$ for some countably additive p' putting mass 1 on F. Any such face is denoted $\Delta(F)$.*

We will see that the descriptive range of the sets of priors developed here are the class of convex combinations of faces.

Definition 2. *The **descriptive range of a single prior,** μ is $\mathcal{R}(\mu) := \{f(\mu) : f \in M(\mathbb{X}; \mathbb{W})\}$. The **descriptive range of a set of priors,** Π is $\mathcal{R}(\Pi) := \{f(\mu) : f \in M(\mathbb{X}; \mathbb{W}), \mu \in \Pi\}$. A convex set of priors is **descriptively complete** if $\mathcal{R}(\Pi)$ contains, up to equivalence, every closed convex subset of $\Delta(\mathbb{W})$ for every compact metric space $\mathbb{W}$.*

The convex combinations in the descriptive range of the sets of priors developed here will contain sets with integral representations of the following form,

$$B = \int_{\mathbb{X}} \Delta(F_x)\, d\mu(x) \tag{4}$$

where $x \mapsto F_x$ is a measurable correspondence. A special case of (4) are the finite convex combinations, those of the form $\sum_{i=1}^{I} \Delta(F_i)\mu(E_i)$ where the correspondence $x \mapsto F_x$ is constant on sets E_i, $i = 1, \ldots, I$, forming a measurable partition of $\mathbb{X}$. More generally, the integral of a correspondence is the set of integrals of almost everywhere selections, that is, $p \in B$ if and only if there exists a measurable $x \mapsto q_x$ such that $\mu(\{x : q_x \in \Delta(F_x)\}) = 1$ and $p = \int q_x \, d\mu(x)$. It is when μ is non-atomic that all convex combinations of faces are in the descriptive range.

2.2. **Examples.** For multiple prior models of preferences over measurable functions, one replaces the single prior, μ, with a set of priors Π. Suppose that f and g are elements of $M(\mathbb{X}; \mathbb{W})$ representing e.g. insurance policies with different coverages and deductibles. *State independence* is the assumption that all that matters for preferences over functions are the induced sets of distributions, the assumption that the comparison between f and g is the comparison between the sets of distributions over utility relevant consequences, $f(\Pi)$ and $g(\Pi)$.

Another way to express state independence is that one can use the change of variable theorem to do the analysis.

Example 1. *The Hurwicz criterion or α-Minmax Expected Utility function (α-MEU,* [22, §6]*) ranks elements of $M(\mathbb{X}; \mathbb{W})$ using a weighted average of the worst and the best outcomes. These preferences can be represented by the utility function $f \mapsto U(f)$ given by*

$$U(f) := \alpha \cdot \min_{\mu \in \Pi} \int_{\mathbb{X}} v(f(x)) \, d\mu(x) + (1 - \alpha) \cdot \max_{\nu \in \Pi} \int_{\mathbb{X}} v(f(x)) \, d\nu(x) \tag{5}$$

where $0 \leq \alpha \leq 1$ and $v : \mathbb{W} \to [0, 1]$ is a Bernoulli utility function. Letting $B = f(\Pi) \subset \Delta(\mathbb{W})$, the utility function in (5) *can be re-written, after a change of variable, as*

$$U(B) = \alpha \cdot \min_{p \in B} \int_{\mathbb{W}} v(c) \, dp(c) + (1 - \alpha) \cdot \max_{q \in B} \int_{\mathbb{W}} v(c) \, dq(c). \tag{6}$$

All functions with $f(\Pi) = B$ are indistinguishable for α-MEU preferences and for all other state independent preferences over $M(\mathbb{X}; \mathbb{W})$.

Any set of distributions that is *not* in the descriptive range of Π *cannot* be modeled as a choice using the set priors Π and state-independent preferences over functions. It is tempting, but mistaken, to conclude that a larger set of probabilities has a larger descriptive range.

Example 2. *If Π is $\Delta^{fa}(\mathcal{X})$, the set of all probabilities on $(\mathbb{X}, \mathcal{X})$, then for any $f \in M(\mathbb{X}; \mathbb{W})$, $f(\Pi) = \Delta(F)$ where F is the closure of $f(\mathbb{X})$, the range the function f. A decision maker with $\Delta^{fa}(\mathcal{X})$ as their set of priors cannot conceive of any non-degenerate risky choices — $\{p\} \in \mathcal{R}(\Delta^{fa}(\mathcal{X}))$ if and only if p is point mass on some $w \in \mathbb{W}$. A decision maker with $\Delta^{fa}(\mathcal{X})$ as their set of priors can only conceive of upper/lower probability intervals* (3) *of the form $A = \Delta(\{w, w'\})$, that is, $A = \{\beta\delta_w + (1 - \beta)\delta_{w'} : \beta \in [0, 1]\}$.*

A **capacity** on $(\mathbb{X}, \mathcal{X})$ is a function $c : \mathcal{X} \to [0, 1]$ that is monotonic, $[E \subset E'] \Rightarrow [c(E) \leq c(E')]$, and normalized, $c(\emptyset) = 0$ and $c(\mathbb{X}) = 1$. A capacity is **convex** (or supermodular) if $c(E) + c(E') \leq c(E \cup E') + c(E \cap E')$ for all $E, E' \in \mathcal{X}$. The **core** of a convex capacity c is defined as the set of all $\mu \in \Delta^{fa}(\mathcal{X})$ such that $\mu(E) \geq c(E)$ for all $E \in \mathcal{X}$ and denoted $\mathbb{P}_c$. The **ignorance capacity** is defined by $c(\mathbb{X}) = 1$ and $c(E) = 0$ if $E^c \neq \emptyset$. Here $\mathbb{P}_c = \Delta^{fa}(\mathcal{X})$ and the descriptive range of $\mathbb{P}_c$ is the same as in Example 2.

It is easy to visualize the descriptive ranges of the various sets of priors when restricting to two utility relevant outcomes.

Example 3. *We can represent closed convex subsets of $\Delta(\{w, w'\})$ as upper-lower probability intervals, $[r, s]$, $0 \leq r \leq s \leq 1$, equivalently, as points in the triangle $T = \{(r, s) : 0 \leq r \leq s \leq 1\}$ as in Figure* 1.

- *The descriptive range of the three sets of priors developed here is all of T if the distribution of observables or the prior to be extended is non-atomic.*
- *The descriptive range of $\Delta^{fa}(\mathcal{X})$, the core of the ignorance capacity, consists of the three extreme points of T.*

*Let $\tau \in \Delta^{fa}(\mathcal{X})$ be non-atomic, let $\varphi : [0, 1] \to [0, 1]$ be a strictly increasing, onto, strictly convex function, define the φ-**distortion capacity** $c_\varphi : \mathcal{X} \to [0, 1]$ by $c_\varphi(E) = \varphi(\tau(E))$ for $E \in \mathcal{X}$, and let $\mathbb{P}_{c_\varphi}$ be its core.*

• *The curved line joining* $(0,0)$ *to* $(1,1)$ *in* T *represents* $\mathcal{R}(\mathcal{P}_{c_\varphi}) = \{[\varphi(t), 1-\varphi(1-t)] : t \in [0,1]\}$*.*

• *If* μ *and* μ' *have bounded densities with respect to each other and* Π *is the convex hull of* μ *and* μ'*, then* $\mathcal{R}(\Pi)$ *consists of line segments in* T *with bounds on the ratio of the probabilities of* w *and* w'*.*

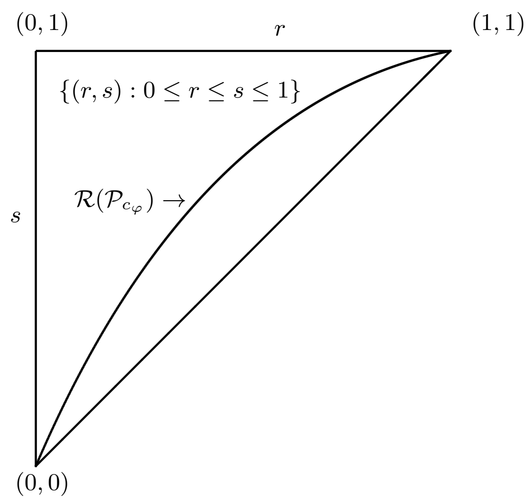

Figure 1

2.3. **Descriptive Completeness.** The following is a useful representation property for sets of countably additive probabilities. A metric space is **Polish** if there is an equivalent metric making it both complete and separable.

Definition 3. *A (convex) set of countably additive probabilities* Π *on* $(\mathbb{X}, \mathcal{X})$ *is* ***descriptively complete*** *if for every Polish* $\mathbb{W}$ *and every (convex) measurable set* B *of countably additive probabilities on* $\mathbb{W}$*, there exists a measurable* $f_B : \mathbb{X} \to \mathbb{W}$ *such that* $f_B(\Pi) = B$*.*

From [18], $\Pi \subset \Delta^{ca}(\mathcal{X})$ is descriptively complete if and only if it is

(a) measurably mutual orthogonal, there exists a measurable, onto $d : \Pi \to [0,1]$ and a measurable, onto $\varphi : \Omega \to [0,1]$ such that for all $r \in [0,1]$, for all $\mu \in d^{-1}(r)$, $\mu(\varphi^{-1}(r)) = 1$, and

(b) simultaneously Skorokhod, for every Polish space $\mathbb{W}$ and every countably additive probability p on consequences, there exists a measurable $f : \mathbb{X} \to \mathbb{W}$ such that for all $\mu \in \Pi$, $f(\mu) = p$.

Condition (a) is a measurable version of the requirement that Π can be partitioned into sets probabilities with disjoint carriers: if $\mu_r \in d^{-1}(r)$ and $\mu_s \in d^{-1}(s)$, $r \neq s$, then μ_r and μ_s are have disjoint carriers because $\varphi^{-1}(r)$ and $\varphi^{-1}(s)$ are disjoint.

In decision theoretic terms, (b) is the requirement that the descriptive range of Π contain all risky choices, a condition that requires that each $\mu \in \Pi$ be non-atomic and entails the existence of a rich class of unambiguous events. Because each μ is non-atomic, for every countably additive probability p on $\mathbb{W}$, there exists a measurable $f_{p,\mu} : \mathbb{X} \to \mathbb{W}$ with $f_{p,\mu}(\mu) = p$. The simultaneity condition is that a single measurable f_p serve for all of the $\mu \in \Pi$.

A simple descriptively complete set is $\Pi^\circ = \{\lambda_r : r \in [0,1]\}$ where λ_r is the uniform distribution on the slice $\{r\} \times [0,1]$ in $[0,1] \times [0,1]$. Because this is a collection of disjointly supported probabilities with no probabilistic structure on the indexing set of r's, it seems difficult to design a learning model delivering knowledge of Π°.

3. Countably Additive Learning Models

The theory of countably additive learning models for risky problems has been extensively studied. We begin with a short coverage of a simple case in which naive empiricism yields a viable and effective learning and optimization strategy, and then turn to a quick analysis of the comparative difficulty of estimation problems. With these in hand, we give the partially observable econometrics models and indicate the development of the parallel results for partially observable econometric models.

In this section, all probabilities are countably additive.

3.1. **Naive Empiricism.** Objective answers to the question "Where do single priors come from?" are of the form, "From observing a great deal of relevant data." We begin with a learning model from statistical decision theory gives an observational foundation for a single prior and a corresponding decision theory. The model involves a sequence of problems, and a limit problem,

$$\max_{a\in A} \int u(a,x)\,dP_n(x),\ n\in\mathbb{N},\ \text{and}\ \max_{a\in A}\int u(a,x)\,d\mu(x), \tag{7}$$

where the P_n are a stochastic sequence of probabilities converging to μ.

Call the problems in (7) well-behaved if A is a compact metric space, $u(\cdot,\cdot)$ is bounded, without loss $0 \le u(a,x) \le 1$ for all (a,x), each $u(\cdot,x) : A \to [0,1]$ is continuous, and each $u(a,\cdot) : \mathbb{X} \to [0,1]$ is $\mathcal{X}$-measurable. The next result, the uniform convergence of the value functions in (7) fails without the assumption of countable additivity.

Lemma 2. *If μ and each P_n, $n \in \mathbb{N}$, is countably additive and $P_n(E) \to \mu(E)$ for every $E \in \mathcal{X}$, then for any well-behaved $(a,x) \mapsto u(a,x)$,*

$$\left|\max_{a\in A}\int u(a,x)\,dP_n(x) - \max_{b\in A}\int u(b,x)\,d\mu(x)\right| \to 0, \tag{8}$$

and the solution set at P_n converges upper hemicontinuously to the solution set at μ.

In the independent and identically distributed (iid) version of this model, the P_n are the empirical distributions of n iid observations on $\mathbb{X}$ that have distribution μ. In the iid case, and in much more general ergodic models, the condition that $P_n(E) \to \mu(E)$ for every $E \in \mathcal{X}$ is satisfied for a probability 1 set of sequences of P_n's.

In statistical problems, the a might be an estimator in a set of parameters A, and the expected value of the utility function increases if the parameter more closely fits the data, which is encapsulated in the stochastic P_n. In economics, the $a \in A$ might be insurance or investment policies, the expected value of the utility function increases as the distribution of income is increased and/or smoothed, and experience with this problem is encapsulated in the P_n. Observations pin down the probability μ asymptotically, and Lemma 2 tells us that the naive empiricist strategy of best responding to the interim probabilities, P_n, leads to solutions to the risky limit problem.

Data about μ is relevant to related problems and the same naive empiricist strategy also solves these. Let $(\mathbb{X}', \mathcal{X}')$ be another state space. If $P_n(E) \to \mu(E)$ for every measurable E, then for any measurable function $g : \mathbb{X} \to \mathbb{X}'$, $Q_n(E') := g(P_n)(E') \to \nu(E') := g(\mu)(E')$ for every measurable $E' \subset \mathbb{X}'$. Therefore, if A' is a different set of actions and $v : A' \times \mathbb{X}' \to [0,1]$ is well-behaved, we again have value convergence and upper hemicontinuous argmax convergence for the problems

$$\max_{a'\in A'}\int v(a',x')\,dQ_n(x'),\ n\in\mathbb{N},\ \text{and}\ \max_{a'\in A'}\int v(a',x')\,d\nu(x'). \tag{9}$$

3.2. **The Difficulty of Risky Decision Problems.** Study of the comparative difficulty of countably additive problems starts with value function $V(\nu) := \max_{a\in A}\int u(a,x)\,d\nu(x),\ \nu \in \Delta^{ca}(\mathcal{X})$. As observed, for countably additive probabilities, if $P_n(E) \to \mu(E)$ for every measurable E, then $V(P_n) \to V(\mu)$. A well-studied measure of the difficulty of the problems is the rate and which $|V(P_n) - V(\mu)|$ converges to 0, and slower convergence is a marker of a more difficult problem.

In the simplest case, the convergence is at a square root of n rate, that is, $|V(P_n) - V(\mu)| = \mathcal{O}(\frac{1}{\sqrt{n}})$ with probability 1. This arises in the iid model if the set of slices, $\mathcal{U}_A := \{u(a,\cdot) : a \in A\}$, is a VC class of functions.[4] Every finite class of bounded functions is a VC class, so

[4]The question that Vapnik and Červonenkis [41] answered is "How large can a class of measurable sets $\mathcal{C}$ (resp. class of measurable functions $\mathcal{F}$) be and still have the convergence of $P_n(C)$ to $\mu(C)$ (resp. $\int f\,dP_n$ to $\int f\,d\mu$)) be

finite problems, however large, fall into the same difficulty class as finite dimensional parametric estimation problems with iid data, however high the dimensionality.

Slower convergence of the value differences indicates that the learning problem is more difficult. This directly parallels the slower convergence of optimal estimators for complicated estimation problems. In high dimensional non-parametric regression problems, the values can converge very slowly, see [37] for the genericity of arbitrarily slow convergence in infinite dimensional problems.

3.3. **Partially Observable Econometric Models.** To get at the essentials of the partially observable econometric models, let $(M_1, \mathcal{M}_1)$ and $(M_2, \mathcal{M}_2)$ be uncountable standard measure spaces,[5] let $(\mathbb{X}, \mathcal{X}) = (M_1 \times M_2, \mathcal{M}_1 \otimes \mathcal{M}_2)$ be the product space with the product σ-field, let $\mathcal{C}$ be the sub-σ-field $\mathcal{M}_1 \otimes \{\emptyset, M_2\}$, and let μ be a distribution in $\Delta^{ca}(\mathcal{X})$. The following generalizes the partially identified set introduced §1.2 — set $M_1 = \mathbb{R}^{1+k}$ and $M_2 = \mathbb{R}^{\ell}$.

Definition 4. *The **partially identified set** is denoted $\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ and defined as the set of all $\nu \in \Delta^{ca}(\mathcal{X})$ such that $\nu(C) = \mu(C)$ for all $C \in \mathcal{C}$.*

We state the next result for finite spaces of consequences. With sufficient care about closure and measurability assumptions, it extends to Polish spaces of outcomes.

Theorem 1. *For finite $\mathbb{W}$, every element of the descriptive range of $\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ is a convex combination of faces in $\Delta(\mathbb{W})$, and if the marginal of μ on M_1 is non-atomic, then the descriptive range is the class of all convex combinations.*

The proof shows that $\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ can be expressed as the integral of a measurable set of disjointly supported probabilities: each observable x_1 can arise from any point in $\{x_1\} \times M_2$ so that $\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}}) = \int_{M_1} \Delta(\{x_1\} \times M_2)\, d\mu_1(x_1)$ where μ_1 is the marginal of μ on M_1; therefore, if f is a measurable function from $M_1 \times M_2$ to $\mathbb{W}$, then $f(\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}})) = \int_{M_1} \Delta(F_{x_1})\, d\mu_1(x_1)$ where $F_{x_1} := f(\{x_1\} \times M_2)$ is the range of $f(x_1, \cdot)$.

3.4. **Toward a Decision Theory for Partial Observability.** Suppose that $P_n(C) \to \mu(C)$ for all C in the class of sets $\mathcal{C} = \mathcal{M}_1 \otimes \{\emptyset, M_2\}$ given above. Let Π_n be the set of probabilities on $\mathcal{X}$ that agree with P_n on $\mathcal{C}$, and let Π be the set of probabilities that agree with μ. The sequence of sets Π_n converges to Π: for every $\nu \in \Pi$, there exists $\nu_n \in \Pi_n$ such that $\nu_n(E) \to \nu(E)$ for every $E \in \mathcal{X}$; and if $\nu_n \in \Pi_n$ satisfies $\nu_n(E) \to \nu'(E)$ for all $E \in \mathcal{X}$, then $\nu' \in \Pi$. In parallel with the model in (7), consider the problems

$$\max\nolimits_{a \in A} U(a, \Pi_n),\ n \in \mathbb{N},\ \text{and}\ \max\nolimits_{a \in A} U(a, \Pi) \tag{10}$$

where each $U(a, \cdot)$ is an extension of the expected utility preferences on singleton sets, $\{\nu\} \mapsto \int u(a, x)\, d\nu(x)$, to non-singleton sets of probabilities.

There are two basic parts to the the theory of countably additive learning models for risky problems. The first is the uniform convergence of the value functions and the related upper hemicontinuous convergence of the argmax set, the second is the study of conditions under which the value functions converge at a $\sqrt{n}$ or a slower rate. We begin with an outline of sufficient conditions for uniform convergence.

Suppose that for each $a \in A$, the mapping $\Pi' \mapsto U(a, \Pi')$ is linear and continuous in a topology that makes the closed convex sets of probabilities on $(\mathbb{X}, \mathcal{X})$ into a Frechet space. From Dieudonne's version of the Banach-Steinhaus (aka uniform boundedness) theorem [14, Ch. XII, §6], the uniform boundedness of the linear mappings $\{U(a, \cdot) : a \in A\}$ implies that the class of mappings is equicontinuous. In particular, this means that if $\Pi_n \to \Pi$, then

$$|\max_{a \in A} U(a, \Pi_n) - \max_{b \in A} U(b, \Pi)| \to 0. \tag{11}$$

With this and the compactness of A, upper hemicontinuous convergence of the argmax sets follows.

---

uniform over $\mathcal{C}$ (resp. over $\mathcal{F}$) and uniform over countably additive μ's?" Such classes of sets or functions are now called VC classes. Dudley's monograph [15] thoroughly covers VC classes and the uniform central limit theorems they satisfy.

[5] A measure space is standard if it is measurably isomorphic to a measurable subset of a Polish metric space.

The set of continuous linear functionals on the class of closed convex sets of probabilities resists an easy characterization as soon as there are three or more points in the probability space. There is, in [19], a representation result for a dense class of the functionals. It involves signed measures on the continuous, bounded functions on $\mathbb{X}$. It seems that conditions on the supports of those signed measures, e.g. to uniformly equicontinuous or other VC classes of functions, should deliver rate of convergence results in (11).

4. Two Purely Finitely Additive Models

The first purely finitely additive model supposes that internal consistency of choices, in the form of a Dutch book argument, delivers a probability on a field $\mathcal{C}$ where $\mathcal{X}$ strictly contains $\sigma(\mathcal{C})$. This model applies if the internal consistency arises e.g. from a sequence of choices or a sequence of elicited preferences. The second model is observational, it gives rise to the set of extensions of a purely finitely additive probability from a field $\mathcal{C}$ to the smallest σ-field containing $\mathcal{C}$, that is, to $\mathcal{X} = \sigma(\mathcal{C})$.

4.1. **The Savage-De Finetti Extension Set.** A Dutch book argument due separately to Ramsey and to de Finetti involves decisions between different possible bets, and gives a subjective foundation for a single prior on a field of events $\mathcal{C}$. Given an event $E \in \mathcal{C}$, the two numerical consequences 0 and 1, and a number r, consider the two functions (state-contingent outcomes), $f_E^r(x) = 1_E(x) - r$ and $g_E^r(x) = r - 1_E(x)$. The first function models paying r to buy a bet on E that pays off 1 if E happens and pays off 0 if E does not happen, while the second function models the opposite side of that bet.

Assume that for each E, the decision maker must name a price, r_E, and that after they have named these prices, they can be forced to take any finite sum of the bets, $\gamma(x) = \sum_E h_E(x)$ where h_E can be either f_E^r or g_E^r at the chosen price r_E. If the decision maker chooses r_E's with the property that some γ satisfies $\gamma(x) < 0$ for all x, then we say that we can make a **Dutch book** against them. Avoiding Dutch book is a form of rationality that is equivalent to choosing a pricing function, $E \mapsto r_E$, interpretable as a probabilistic prior belief.

Lemma 3 (Ramsey, de Finetti)**.** *There is no finite sum with $\gamma(x) < 0$ for all x if and only if $E \mapsto r_E$ is a finitely additive probability on $\mathcal{C}$.*

Both Savage and de Finetti argued that the existence of non-measurable sets represents a challenge, perhaps insuperable, to the meaning of probability (e.g. [9, p. 124], [31, Ch. 3, §3]). To avoid this, both advocated using the measure space $(\mathbb{X}, \mathcal{P}(\mathbb{X}))$ as the model of randomness ($\mathcal{P}(\mathbb{X})$ is the power set of $\mathbb{X}$, that is, the class of all subsets of $\mathbb{X}$). The choice to use this measure space imposes strong restrictions on the probabilities: all countably additive probabilities are either finite or countably infinite convex combinations of point masses; all non-atomic probabilities are purely finitely additive.

The **Savage-de Finetti extension set** is the set of probabilities on $\mathcal{P}(\mathbb{X})$ such that $\nu(C) = \mu(C)$ for all $C \in \mathcal{C}$, denoted $\Pi^{fa}_{\mathcal{P}(X)}(\mu_{|\mathcal{C}})$. Typically, $\mu_{|\mathcal{C}}$ will have many Hahn-Banach extensions from $\mathcal{C}$ to $\mathcal{P}(\mathbb{X})$, and the Savage-de Finetti extension set is that indeterminate set. De Finetti argued that this indeterminacy captures something fundamental about probability (see esp. [9, pp. 229-231]).

4.2. **Purely Finitely Additive Observational Models.** An **observational learning model** for a class of events $\mathcal{C} \subset \mathcal{X}$, $\mathcal{X} = \sigma(\mathcal{C})$, and a probability μ is a stochastic sequence of probabilities, $P_n : \mathcal{C} \to [0,1]$, with the property that, with probability 1, $P_n(C) \to \mu(C)$ for all C in $\mathcal{F}^\circ(\mathcal{C})$, the smallest field containing $\mathcal{C}$. The iid learning model is the special case in which $P_n(C) = \frac{1}{n}\sum_{m \leq n} 1_C(Y_n)$ is a sequence of empirical distributions and the vectors $(1_C(Y_n))_{C \in \mathcal{C}} \in \{0,1\}^{\mathcal{C}}$ are iid with distribution given by $E\, 1_C(Y_n) = \mu(C)$. The **partially identified set** for a learning model, $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$, is defined as the set of all probabilities $\nu \in \Delta^{fa}(\mathcal{X})$ with $\nu(C) = \mu(C)$ for all $C \in \mathcal{C}$.

The **observation space** is $S := \{0,1\}^{\mathcal{C}}$ with its product σ-field $\mathcal{S}$, and the **observation mapping** from $\mathbb{X}$ to S is $\varphi_{\mathcal{C}}(x) := (1_C(x))_{C \in \mathcal{C}}$. A learning model for $(\mathcal{X}, \mathcal{C}, \mu)$ is **determinate** if the partially identified set is completely identified, $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}}) = \{\mu\}$, otherwise it is **indeterminate**. Indeterminacy arises if the mapping $\mu \mapsto \varphi_{\mathcal{C}}(\mu)$ is many to one.

Carathéodory's extension theorem tells us that observational models for countably additive probabilities are determinate. In the following example, observations are consistent with only one countably additive μ, and if μ is the true distribution, then it can be learned at a $\sqrt{n}$ rate in the iid model. By contrast, if we assume that μ is purely finitely additive, then observations cannot determine its support set.

Example 4. *Suppose that* $\mathbb{X} = [0,1]$*, that* $\mathcal{C}$ *is the field generated by the intervals* $\{[0,r] : r \in [0,1]\}$*, that* $\mathcal{X} = \sigma(\mathcal{C})$ *is the usual Borel* σ*-field, and that* $\mu([0,r]) = r$ *for all* $r \in [0,1]$*. If* μ *is countably additive, the learning model is determinate. By contrast, there are uncountably many disjoint, uncountable sets dense subsets of* $[0,1]$*, each containing uncountably many disjoint countable dense subsets supporting a purely finitely additive probability that agrees with with* μ *on* $\mathcal{C}$*.*[6]

The advantage of the indeterminacy is that the set of probabilities consistent with μ can be used as a set of priors. The following shows how the non-atomicity of μ on the observable field $\mathcal{C}$ leads to arbitrary convex combinations of faces being in the descriptive range of $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$.

Example 5. *Suppose that* $f : \mathbb{X} \to \mathbb{W} = \{a, b, c\}$ *has the folowing properties:* $f : [0, \beta] \to \{a, b\}$*;* $f : (\beta, 1] \to \{b, c\}$*;* $f^{-1}(a)$ *and* $f^{-1}(b)$ *are dense subsets of* $[0, \beta]$*; while* $f^{-1}(b)$ *and* $f^{-1}(c)$ *are dense subsets of* $(\beta, 1]$*. Using the claim in the previous Example,* $f(\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}}))$ *is the convex combination of faces* $\beta\Delta(\{a, b\}) + (1-\beta)\Delta(\{b, c\})$*.*

4.3. **Structural Results.** The determinacy of countably additive learning models is a restatement of Carathéodory's extension theorem. The first structural result gives the extent to which the Carathéodory's theorem depends on countable additivity — either the field $\mathcal{C}$ is equal to the σ-field $\mathcal{X}$, or the diameter of $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ is 1 for some $\mu \in \Delta^{fa}(\mathcal{X})$. The second result gives a condition on $\mathcal{C}$, essential separability, that guarantees that for any purely finitely additive μ, $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ can be expressed as an integral of simplexes with uncountable, disjoint support sets. This guarantees that the diameter is 1 and leads directly to the descriptive range results.

4.3.1. *The Diameter Dichotomy.* The **norm distance** between $\nu, \nu' \in \Delta^{fa}(\mathcal{X})$ is

$$\|\nu - \nu'\| := \sup_{E \in \mathcal{X}} |\nu(E) - \nu'(E)|. \tag{12}$$

Two probabilities are **orthogonal** if $\nu(E) = 1$ and $\nu'(E) = 0$ for some $E \in \mathcal{X}$, and orthogonal probabilities are at norm distance 1 from each other.[7] The **diameter** of a set of probabilities Π is $Diam(\Pi) = \sup\{\|\nu - \nu'\| : \nu, \nu' \in \Pi\}$. The Hahn-Banach extension theorem yields

Lemma 4 (The diameter dichotomy)**.** *If* $\mathcal{C} \subset \mathcal{X}$ *is a field, then either* $\mathcal{C} = \mathcal{X}$ *so that* $Diam(\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})) = 0$ *for all* $\mu \in \Delta^{fa}(\mathcal{X})$*, or for some* $\mu \in \Delta^{fa}(\mathcal{X})$*,* $Diam(\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})) = 1$*.*

4.3.2. *Essentially Separable Classes.* It will matter that the countable class $\mathcal{C}'$ in the following may depend on μ.

Definition 5. *A class* $\mathcal{C}$ *is* ***essentially separable*** *if for any* $\mu \in \Delta^{fa}(\mathcal{X})$*, there is a countable* $\mathcal{C}' \subset \mathcal{C}$ *such that* $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}'}) = \Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$*.*

The smallest field containing any countable collection of sets is countable, and any countable class of sets is essentially separable. For any class of sets $\mathcal{E}$, $\mathcal{F}^\circ(\mathcal{E})$ denotes the smallest field of sets containing $\mathcal{E}$. If $\mathcal{I}$ is the class of intervals (q, q'), $(q, q']$, $[q, q')$ or $[q, q']$, q, q' rational in $\mathbb{R}^\ell$, then $\mathcal{F}^\circ(\mathcal{I})$ is essentially separable. If $(\mathbb{X}, d)$ is a metric space with a countable dense set, E, and $\mathcal{E}$ is the class of open balls with rational radius around points in E, then $\mathcal{F}^\circ(\mathcal{E})$ is essentially separable. There are also many uncountable, essentially separable sets.

[6] The proof of this claim is in the appendix.

[7] Countably additive ν, ν' at norm distance 1 from each other are orthogonal by the Hahn decomposition theorem ([20, Theorem III.4.10]). Finitely additive probabilities at norm distance 1 may not be orthogonal to each other, see [3] for a study of the implications.

If $\mathcal{B}$ is the class of open balls with finite radius in a separable Banach space or a separable Frechet space, then $\mathcal{F}^\circ(\mathcal{B})$ is essentially separable. These spaces cover those parts of continuous time stochastic process theory that use the Banach space $C([0,1])$ with the sup norm topology or the Frechet space $C([0,\infty))$ with the topology of uniform convergence on compacta.

If $\mathcal{G}$ is the class of open sets with compact closure in a locally compact metric space, then $\mathcal{F}^\circ(\mathcal{G})$ is essentially separable. These spaces include $\mathbb{R}^k$ and the space of closed subsets of a compact metric space with the Hausdorff metric. One can also look for uniform compactness conditions on classes of sets that deliver essential separability. The following is one such condition.

Lemma 5. *If $\mathcal{C} = \mathcal{F}^\circ(\mathcal{V})$ and $\mathcal{V}$ is a VC class, then $\mathcal{C}$ is essentially separable.*

In econometrics, VC classes of functions, or the closely related μ-Donsker classes, are often called uniformly equi-continuous classes [2]. VC classes have combinatorial and metric entropy characterizations that are often easier to verify, and the proof of Lemma 5 uses a basic metric entropy result for VC classes. Dudley, Giné, and Zinn [16, Proposition 11] show the equivalence between the VC class property and the uniform strong law of large numbers property. For example, the uncountable field $\mathcal{F}^\circ(\{(-\infty,\boldsymbol{x}] : \boldsymbol{x} \in \mathbb{R}^k\})$ is not a VC class (except in the finite case, no field of events can be a VC class), but it is essentially separable because multi-dimensional Glivenko-Cantelli theorem implies that the given class of intervals is a VC class.[8]

4.3.3. *The Descriptive Range Result.* There are two differences between the next result and Theorem 1. First, because we are here using purely finitely additive probabilities, there is no extra difficulty in replacing finite spaces of consequences with general compact spaces. Second, we here need a condition on $\mathcal{C}$. Recall the maintained assumption that $\mathbb{W}$ is a compact metric space.

Theorem 2. *If μ is purely finitely additive and $\mathcal{C}$ is essentially separable, then every element of the descriptive range of $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ is a convex combination of faces in $\Delta(\mathbb{W})$, and if μ is non-atomic on $\mathcal{C}$, then the descriptive range is the class of all convex combinations.*

The proof uses properties of the Stone space for L_∞, a construction that contains the essential intuition about the non-observability of the state space when the probability is purely finitely additive.

5. Comparisons

The parallels between the partially observed econometric models and the Savage-de Finetti extension set are immediate and direct: the indeterminacy arises because the value of a probability is only observed for the sets in a sub-field $\mathcal{C}$ with $\sigma(\mathcal{C})$ a strict subset of $\mathcal{X}$. We begin this section by showing that the finitely additive observational model has the same structure. We then turn to the relation between the representation results given here in the presence of non-atomicity and Dempster's [13] method of arriving at convex combinations of faces. We end this section with comparisons between the present approach and interval-valued probabilities and between countably and finitely additive stochastic search.

5.1. **Domains for Purely Finitely Additive Probabilities.** The examples of observational models with a purely finitely additive μ given here have had the property that they put mass 1 on observations that cannot arise from states in the carrier of μ. This is not an accident, and this sort of phenomenom has been long recognized. For an example, consider the following summary of the mathematics taken from Uhl's [40] review of Rao and Rao's definitive monograph on finitely additive measures [5].

> There are essentially three ways to prove theorems about finitely additive measures. The easiest is usually proof via the Stone representation theorem which allows a direct transfer of the finitely additive case to the countably additive case. ... The second is Drewnowski's principle ... both show that a finitely additive measure is just a countably additive measure that was unfortunate enough

[8]Dudley's monograph [15] contains a systematic development of this and the associated uniform central limit theorems in empirical process theory.

to have been cheated on its domain. The third approach (followed by Rao and Rao) is to prove everything directly with absolutely no reference to the countably additive case.

In many cases, having a probability that has been 'cheated on its domain' is a tremendously useful device, and here it delivers indeterminacy that can be used for modeling ambiguous choice problems. The proof of Theorem 2 replaces the state space $(\mathbb{X}, \mathcal{X}, \mu)$ with the mentioned Stone representation. A version of the Stone representation taken from Yosida and Hewitt [43, §4] allows us to see how partial observability arises from a projection mapping, just as occurs in the partially observed econometrics models.[9]

(i) Let $(\widehat{Y}, \widehat{\mathcal{Y}})$ be the product space $\{0,1\}^{\mathcal{X}}$ with the product topology and the associated Borel σ-field. Because events determine states, the mapping $\psi(x) := (1_E(x))_{E\in\mathcal{X}}$ is one-to-one. Identify each $x \in \mathbb{X}$ with $\psi(x)$ and identify $\mathbb{X}$ with $\psi(\mathbb{X})$.

(ii) Define $\widehat{\mathbb{X}}$ as the closure of $\psi(\mathbb{X})$ in $\widehat{Y}$ so that $\mathbb{X}$ is dense in $\widehat{\mathbb{X}}$, and let $\widehat{\mathcal{X}}$ denote the Borel measurable subsets of the compact set $\widehat{\mathbb{X}}$.

(iii) For any $E \in \mathcal{X}$ and all $x \in \mathbb{X}$, $1_E(x) = \text{proj}_E(\psi(x))$ where proj_E is the canonical projection map for the product space $\{0,1\}^{\mathcal{X}}$. The topology on $\widehat{\mathbb{X}}$ is defined by the requirement that these projections be continuous.

(iv) Every simple measurable $h(\cdot) = \sum_i \beta_i 1_{E_i}(\cdot)$ on $\mathbb{X}$ has a unique continuous extension, $\widehat{h}(\cdot) = \sum_i \beta_i \,\text{proj}_{E_i}(\cdot)$ on $\widehat{\mathbb{X}}$. Taking uniform limits delivers a sup norm isomorphism $h \leftrightarrow \widehat{h}$ between the bounded measurable functions on $\mathbb{X}$ and the continuous functions on the compact Hausdorff space $\widehat{\mathbb{X}}$.

(v) Integration of measurable functions on $\mathbb{X}$ becomes integration of continuous functions on $\widehat{\mathbb{X}}$, and this in turn induces a one-to-one, onto variation norm isometry between $\Delta^{fa}(\mathcal{X})$ and $\Delta^{ca}(\widehat{\mathcal{X}})$ defined by $\int_{\mathbb{X}} h\, d\mu = \int_{\widehat{\mathbb{X}}} \widehat{h}\, d\widehat{\mu}$ for the bounded measurable h.

(vi) Finally, the observation mapping $\varphi_{\mathcal{C}}(x) := (1_C(x))_{C\in\mathcal{C}}$ is the many-to-one projection from $\widehat{\mathbb{X}} \subset \{0,1\}^{\mathcal{X}}$ to $\{0,1\}^{\mathcal{C}}$.

The **corona** of $\mathbb{X}$ is $\widehat{\mathbb{X}} \setminus \mathbb{X}$. It is easy to show that μ is purely finitely additive if and only if $\widehat{\mu}$ puts mass 1 on the corona.[10] It is in this sense that μ has been 'cheated on its domain,' and this contains an intuition about the indeterminacy of observational models: learning requires observations, but pfa probabilities put their mass on the unobserved corona. In §1.3, all of the points in the corona led to the same observation point in $S = \{0,1\}^{\mathcal{C}}$, and that point does not arise from any point in $\mathbb{X}$. In Example 4, the points in the corona gave rise to observations just like those arising from a countably additive model even when the support sets in $\mathbb{X}$ are disjoint.

There is also a cardinality based intuition: there are at least $2^{\mathfrak{c}}$ pfa point masses where $\mathfrak{c}$ is the cardinality of $\mathbb{R}$; every pfa μ is a convex combination of those point masses; when $\mathcal{C}$ is countable, the cardinality of the observation space, $S = \{0,1\}^{\mathcal{C}}$, is the strictly smaller $\mathfrak{c}$. One cannot determine $2^{\mathfrak{c}}$ points using an injection into a space having only $\mathfrak{c}$ points.

5.2. **Dempster's Compatible Sets of Probabilities.** For simplicity, we consider only the there are only finitely many consequences, $\#\mathbb{W} < \infty$. Suppose that $\mu \in \Delta^{fa}(\mathcal{X})$ is non-atomic and that $\Psi : \mathbb{X} \rightrightarrows \mathbb{W}$ is a non-empty valued and measurable correspondence. A measurable function $f : \mathbb{X} \to \mathbb{W}$ is **an almost everywhere selection from** Ψ if $\mu(\{x : f(x) \in \Psi(x)\}) = 1$.

With a non-atomic μ, Dempster's [13, §2] set of probabilities compatible with Ψ is

$$\Psi(\mu) = \{f(\mu) : f \text{ is an almost everywhere selection from } \Psi\}. \tag{13}$$

Using the definition of the integral of a correspondence taking values in the convex set $\Delta(\mathbb{W})$, it is nearly immediate that $\Psi(\mu) = \sum_F \Delta(F)\mu(\Psi = F)$ where the sum is over non-empty $F \subset \mathbb{W}$.

[9]The properties given below follow directly from known results, for completeness the arguments are sketched in the Appendix. When $\mathbb{X} = \mathbb{N}$ and $\mathcal{X} = \mathcal{P}(\mathbb{N})$, the $\widehat{\mathbb{X}}$ constructed below is the Stone-Čech compactification of the integers, denoted $\beta\mathbb{N}$. Maharam [29] was an early study of properties of $\beta\mathbb{N}$ that made use of the isometry isometry between finitely additive measures on $\mathbb{N}$ and the countably additive measures on $\beta\mathbb{N}$ given in (v) below.

[10]Let $h_n = 1_{E_n}$ where $E_n \downarrow \emptyset$ and $\mu(E_n) \equiv 1$. Because $\int h_n\, d\mu \equiv \int \widehat{h}_n\, d\widehat{\mu}$, $\widehat{\mu}$ is carried on the subset of the corona given by $\cap_N \{\widehat{x} \in \widehat{\mathbb{X}} : \widehat{h}_n(\widehat{x}) = 1\}$.

The choice theory presented in this paper posits choices between $f(\Pi(\mu))$ and $g(\Pi(\mu))$, and in the presence of non-atomicity, these are choices between convex combinations of closed faces. The Dempster alternative is to model choices as being between $\Psi(\mu)$ and $\Phi(\mu)$ where Ψ and Φ are correspondences and μ is non-atomic. Once again, these are choices between convex combinations of faces.

5.3. **Interval-Valued Probabilities.** An **interval-valued probability** is a mapping $E \mapsto P(E) = [\underline{r}_E, \overline{r}_E] \subset [0,1]$. In the early 1900's, Keynes [25] derived interval-valued probabilities and their properties from epistemic considerations. They were systematically investigated by de Finetti [10, Appendix §19.3] who assumed that the interval-valued probabilities contain a probability, that is, that there exists a probability $\mu : \mathcal{X} \to [0,1]$ such that $\mu(E) \in P(E)$ for all E. Suppes and Zanotti [39] give conditions under which de Finetti's interval-valued probabilities contain at least one probability.

As noted by Savage and de Finetti in their discussion of how to choose a prior [11], Smith's work [34] provides a decision theoretic foundation for interval-valued probabilities. Smith gives the following variant of the Ramsey and de Finetti derivation of subjective probabilities: suppose that the decision maker can specify for each event E and each $r \in [0,1]$, whether they would take the bet $f_E^r = 1_E - r$, take the other side of the bet, $g_E^r = r - 1_E$, or would prefer not to bet either way on the event E a the price r. Assuming that E is not null, we expect the decision maker to take $1_E - r$ for sufficiently low r, to take the other side for sufficiently high r, and if decision maker prefers not to bet on E at a price r, then this should happen for all r in an interval $[\underline{r}_E, \overline{r}_E]$. The mapping $E \mapsto [\underline{r}_E, \overline{r}_E]$ is the interval-valued probability.

Given an interval-valued probability, the set of probabilities that it contains is closed and convex. However, it need not belong to the closed convex hull of the faces (see [13, Fig. 1] for an example). The difference arises because Smith's work delivers an interval of possible values for each event $E \in \mathcal{X}$ and asks for the set of μ compatible with those values, while the present approach focuses on the set of μ compatible with exact knowledge of the probability of all E is a sub-field $\mathcal{C} \subset \mathcal{X}$.

5.4. **Recoupling Learning and Decisions.** The objective foundation for a single prior given in §3.1 consists of a sequence of problems and a limit problem,

$$\max_{a \in A} \int u(a,x)\, dP_n(x) \text{ and } \max_{a \in A} \int u(a,x)\, d\mu(x) \tag{14}$$

where P_n converges to μ. This model captures the idea that the reward structure, $u(\cdot,\cdot)$, is known, but the stochastic structure, μ, is not known. There are two potential sources of information about μ: first, observing the stochastic sequence $1_C(Y_n)$ for all $C \in \mathcal{F}^\circ(\mathcal{C})$ where $E\,1_C(Y_n) = \mu(C)$; second, observing the random rewards, $R(a_n)$, to taking an action a_n where the probability that $R(a_n) \in [0,r]$ is equal to $\mu(\{x : u(a_n,x) \in [0,r]\})$ for $r \in [0,1]$.

The analysis so far has been decoupled from the second source of information. This section develops a condition on the interaction of the stochastic structure and the reward structure, which, if satisfied, guarantee that simple stochastic search makes the second source of information sufficient to find the solution to the limit problem. An example failing the condition demonstrates that having both sources of information need not be sufficient to reduce ambiguity and re-inforces the domain intuitions about finitely additive priors.

From above, the limit problem is **well-behaved** if A is compact, $0 \le u(a,x) \le 1$, for each $x \in \mathbb{X}$ the function $u(\cdot,x)$ is continuous on A, and for each $a \in A$, $u(a,\cdot)$ is measurable on $\mathbb{X}$. $V_0^1(A)$ denotes the set of continuous functions on A bounded between 0 and 1 with the sup norm, and $\mathcal{V}$ denotes its Borel σ-field. $V_0^1(A)$ is a Polish metric space, and it is compact if and only if A is finite.

A finitely additive probability p on the Borel σ-field of a Polish space M is **tight** if for every $\epsilon > 0$, there exists a compact K_ϵ such that for every continuous $f : M \to [0,1]$ with $f(x) = 1$ for all $x \in K_\epsilon$, $\int f\, dp > (1-\epsilon)$. Finitely additive probabilities, p, q, on M are **Prohorov equivalent** if $\int f\, d(p-q) = 0$ for all bounded continuous f. It can be shown that p is Prohorov equivalent to a countably additive probability if and only if it is tight, and when M is compact, all probabilities are Prohorov equivalent to a countably additive probability.

For a well-behaved $u : A \times \mathbb{X} \to [0,1]$, we say that $\mu \in \Delta^{fa}(\mathcal{X})$ is **tight for** $u(\cdot,\cdot)$ if the distribution $p \in \Delta^{fa}(\mathcal{V})$ defined by $p(E) = \mu(\{x : u(\cdot,x) \in E\})$, $E \in \mathcal{V}$, is tight. Any countably additive μ is tight for all $u(\cdot,\cdot)$ because $V_0^1(A)$ is Polish and all countably additive probabilities on Polish spaces are tight. If A is finite, then $V_0^1(A)$ is compact, and every $\mu \in \Delta^{fa}(\mathcal{X})$ is tight for all $u(\cdot,\cdot)$. If the set $\{u(\cdot,x) : x \in \mathbb{X}\}$, is equi-continuous, then every $\mu \in \Delta^{fa}(\mathcal{X})$ is tight for $u(\cdot,\cdot)$ by the Arzelà-Ascoli theorem (e.g. [8, Thm. 6.2.61]).

The most naive form of stochastic search takes repeated independent draws according to a single distribution. The next result shows that this can be effective in the presence of tightness. For a countably additive probability Q on A, Q^∞ denotes the countable product measure on $A^{\mathbb{N}}$ having all marginals equal to Q. Recall that $V(\mu) := \max_{a\in A} \int u(a,x)\,d\mu(x))$.

Lemma 6. *If $\mu \in \Delta^{fa}(\mathcal{X})$ is tight for $u(\cdot,\cdot)$, then $a \mapsto \int u(a,x)\,d\mu(x)$ is continuous, and for any full support, countably additive Q on A and for any $\epsilon > 0$,*

$$Q^\infty(\#\{(a_n)_{n\in\mathbb{N}} : E\,R(a_n) > V(\mu) - \epsilon\} = \infty) = 1. \tag{15}$$

Thus, in the presence of tightness, naive stochastic search will, with probability 1, yield infinitely many observations with expected payoffs at least ϵ-close to the maximum achievable. In the following example, μ is not tight for $u(\cdot,\cdot)$ and knowing both the asymptotic results of a stochastic search and $\mu(C)$ for all C in a rich field $\mathcal{C}$ does not reduce the residual ambiguity.

Example 6. *Let $A = [0,1]$. We describe a class of probabilities on $\mathbb{X}$ that have the property that $a \mapsto E\,R(a)$ is equal to 0 except for a single $r \in [0,1]$. Unless $Q(\{r\}) > 0$, naive stochastic search will, with probability 1 only see the realizations $R(a_n) = 0$.*

Suppose that $u(\cdot,\cdot)$ satisfies the following richness condition — $\{u(\cdot,x) : x \in \mathbb{X}\} = V_0^1(A)$. For $r \in [0,1]$ and $n \in \mathbb{N}$, let $v_n \in V_0^1(A)$ be the function $\max\{0, 1 - n|r-a|\}$, and pick x_n such that $u(\cdot,x_n) = v_n(\cdot)$. For any pfa probability γ on the integers, define $\mu_\gamma(E) = \gamma(\{n : x_n \in E\})$. For $a \neq r$, $\int u(a,x)\,d\mu_\gamma(x) = 0$ while $\int u(r,x)\,d\mu_\gamma(x) = 1$.

Unless For each sup norm open ball, $B_\epsilon(v) \subset V_0^1(A)$, $v \in V_0^1(A)$ and $\epsilon < \frac{1}{2}$, let $E_\epsilon(v) = \{x \in \mathbb{X} : u(\cdot,x) \in B_\epsilon(v)\}$. Let $\mathcal{C}$ be the field generated by the class of sets $\{E_\epsilon(v) : \epsilon < \frac{1}{2}, v \in V_0^1(A)\}$. Because $\epsilon < \frac{1}{2}$, $\mu(E_\epsilon(v)) = 0$ for all v and $\epsilon < \frac{1}{2}$, which implies that $\mu(C) = 0$ or $\mu(C) = 1$ for all $C \in \mathcal{C}$.

This barely scratches the set of mappings $a \mapsto E\,R(a)$ that arise from different $\mu \in \Delta^{fa}(\mathcal{X})$. Using Kingman [26, Thm. 2 and §4], one can show the following: if $\{u(\cdot,x) : x \in \mathbb{X}\} = V_0^1(A)$, then for any $g : A \to [0,1]$, measurable or not, there exists a non-empty, convex set $\mathbb{P}_g \subset \Delta^{fa}(\mathcal{X})$ having diameter 1 and having $E\,R(a) \equiv g(a)$. Whether or not descriptive range of sets such as $\mathbb{P}_g$ have better properties than the decoupled sets investigated in this paper is an open question, but it seems reasonable to expect that their properties are the same.

6. Summary, Extensions, and Relations with the Literature

This paper has analyzed three possible foundations for sets of priors useful in multiple prior theories of choice under ambiguity. The first, data-based approach involves the econometrics of partial observability, the second involves observational learning models where the data has a pfa distribution. The third approach is the Savage-de Finetti set of extensions of a subjective prior.

The three approaches have the same weakness, the indeterminate sets contain uncountably many non-atomic probabilities with mutually disjoint supports, and strength, if the distribution of observables or the prior to be extended is non-atomic, then all convex combinations of faces can be modeled in the space of distributions over consequences. The mutually disjoint support property makes the three sets unsuitable as models of diversity of opinion. The strength allows a small, albeit interesting, class of ambiguous choice problems to be modeled, one that encompasses Dempster's compatible sets, in particular, includes all two outcome upper-lower probability sets.

6.1. **Generalizations.** Interest centers on preferences between sets of distributions induced by different functions applied to a set of priors Π. After choosing a function, say f, the rewards contain information about the induced distribution $f(\mu)$. It is easy to give conditions under which the descriptive range of the set of priors is unchanged after incorporating the extra information. Generalizations to the space of utility relevant consequences, $\mathbb{W}$, are also available, as are generalizations to probabilities μ that are neither countably additive nor purely finitely additive.

6.1.1. *Incorporating extra information.* The analysis of learning models has decoupled decisions and observations. In §5.4, we have seen that for tight pfa probabilities, incorporating what is learned by making decisions reduces ambiguity, for decision purposes, to 0, and that without tightness, there may be no reduction. Here we look at the descriptive range after information from a simple decision is incorporated in the partially observable econometric models.

Let $\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ be the partially identified set. Suppose that one has learned Π and is repeatedly faced with the choice between the functions $f = 1_F$ and $g = 1_G$, F, G measurable subsets of $M_1 \times M_2$. Experimentation will soon give information on $\mu(F)$ and $\mu(G)$. Consider the new partially observed model where μ_1, the marginal of μ on M_1, and, say, $\mu(F)$ has been learned. One question of interest is the descriptive range of this new model.

For each $x_1 \in M_1$, let $F(x_1)$ be the corresponding slice of F, $F(x_1) = \{x_2 \in M_2 : (x_1, x_2) \in F\}$, and let $F' \subset M_1$ be the set of x_1 for which $F(x_1)$ is uncountable. If $\mu_1(F') = 1$ or if $\mu(F') = \mu(\text{proj}_{M_1}(F))$, then the arguments for Theorem 1 can be applied to show that the descriptive range of the new partially observed model has the same properties as the original set (for the measurability of F', $\text{proj}_{M_1}(F)$, and the existence and measurability of the requisite selections from F, see e.g. [38]).

6.1.2. *General spaces of utility relevant consequences.* The descriptive range for the partially observable econometric models was given only for finite spaces of consequences. To generalize to compact metric spaces of consequences requires the use of compact-valued correspondences and some care with measurability and closure considerations both in the statements and proofs of the results. However, the requisite techniques for this extension are well-known.

To generalize the descriptive range results to Polish spaces of consequences, e.g. $\mathbb{W} = \mathbb{R}$, requires yet more care. This is mostly because the pfa probabilities on non-compact spaces of consequences can be at arbitrary Prohorov distances from the set of countably additive probabilities. Again, the requisite techniques are well-known, in this case they involve homeomorphically imbedding $\mathbb{W}$ as a $\mathcal{G}_\delta$ subset of $[0,1]^{\mathbb{N}}$.

6.1.3. *General finitely additive probabilities.* From the Yosida-Hewitt decomposition [43, Theorem 1.23], any $\mu \in \Delta^{fa}(\mathcal{X})$ has a unique expression as a convex combination of a purely finitely additive and a countably additive probability, $\mu = \delta\mu^{pfa} + (1-\delta)\mu^{ca}$. Taking $E_n \downarrow \emptyset$ with $\mu^{pfa}(E_n) \equiv 1$ delivers $\mu(E_n) \downarrow \delta$. All results developed here for pfa and for countably additive probabilities can be combined using this decomposition. For example, if $\mu = \delta\mu^{pfa} + (1-\delta)\mu^{ca}$, then the orthogonality of the probabilities in asymptotic set for an observational learning model is correspondingly modified, $(1-\delta)$ of the mass is identified and δ of the mass is spread between uncountably many disjointly supported measures. The descriptive range becomes the class of sets of the form $(1-\delta)\{p\} + \delta A$, p a countably additive probability on $\mathbb{W}$ and A an element of the closed convex hull of the set of faces.

6.2. **Open Questions.** The are open questions about: the generality of the state space assumptions; the dependence of the asymptotic set on the class of learned events $\mathcal{C}$; the existence of variants of learning or extension models with indeterminate sets having a more complete descriptive range; and completing the comparison with the cores of convex capacities.

6.2.1. *General state space questions.* The arguments for Theorem 2 used the assumption that events identify states, and used essential separability. The assumptions are weak for learning models, and cover the Savage-de Finetti extension sets when the class $\mathcal{C}$ need only be used in models with spaces where sequences are an adequate tool for studying continuity. Still, one might want a more general extension theory.

Let $\mathcal{X} \subset \mathcal{P}(\mathbb{X})$ be a σ-field, let $\mathcal{C} \subset \mathcal{X}$ be a subclass of sets, and let $L(\mathcal{C})$ be the span of the indicators of sets belonging to $\mathcal{F}^\circ(\mathcal{C})$, the field generated by $\mathcal{C}$. A satisfactory extension of the present work involves characterizing triples $(\mathcal{X}, \mathcal{C}, \mu)$ with the property that $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$, viewed as the set of Hahn-Banach extensions from $L(\mathcal{C})$ to the bounded measurable functions, has an integral representation as a convex combination of disjointly supported probability simplexes. This is a stronger property than the existence of a μ with the diameter of $\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ being 1, a property characterized in Lemma 4 by the sup norm closure of $L(\mathcal{C})$ not being the class of all bounded measurable functions.

6.2.2. *Dependence on the class of learned events.* For ambiguous problems, we replace the risky decision problems $v(\mu) = \max_{a\in A} \int u(a,x)\,d\mu(x)$ by $V(\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}})) = \max_{a\in A} U(a, \Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}}))$ where each $U(a,\cdot)$ is an extension of expected utility preferences on singleton sets to non-singletons sets of probabilities (as in [19]). The set $\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}})$ contains all probabilities consistent with knowing the probability of every event in the class $\mathcal{F}^\circ(\mathcal{C})$. Much is known about the continuity of value and the upperhemicontinuity of behavior in the risky problems as $\mathcal{C}_n \to \mathcal{C}$ (see [36] for a Bayesian approach). Similar information about the behavior of the value function and the best response sets for ambiguous problems would be a useful first step in studies of the effects of changing information in general equilibrium models and Nash equilibrium models with ambiguity.

6.2.3. *Learning sets of priors with larger descriptive ranges.* A non-atomic prior, μ, is *descriptively complete* in models of risky choice because any distribution over $\mathbb{W}$ can be modeled as a measurable function. As discussed above, this is a consequence of Skorokhod's representation theorem, for any $p \in \Delta(\mathbb{W})$, there exists a measurable $f_p : \mathbb{X} \to \mathbb{W}$ such that $f_p(\mu)$ is at Prohorov distance 0 from p. In a similar fashion, a (convex) set of priors, Π, is **descriptively complete** if for every non-empty (convex) measurable $A \subset \Delta(\mathbb{W})$, there exists a measurable function, $f_A : \mathbb{X} \to \mathbb{W}$, such that $f_A(\Pi) = A$ up to equivalence.

We have seen how small the class of closed faces is, and this can matter in substantive ways — one cannot model ambiguous choices over, say, the three outcomes $\mathbb{W} = \{a, b, c\}$ answering to the description "a is at least as likely as b but I have no information about the relative likelihood of c." Such limitations can be avoided. There are descriptively complete sets of countably additive priors, and their characterization [18] shows that they must be an uncountable collection of non-atomic probabilities, must satisfy a measurable form of each probability having a support set disjoint from the support sets of all the other probabilities, and must have a rich class of unambiguous events. It seems difficult to construct a learning model for such a class of priors. However, if we replace learning the values of μ for every $C \in \mathcal{C}$ with learning an interval containing $\mu(C)$ as in Smith [34], then we may increase the descriptive range in important ways.

6.2.4. *Completing the comparisons with capacities.* This paper has characterized the descriptive ranges of sets of priors arising from three different kinds of models. To finish the comparisons with the cores of convex capacities begun in §2 requires a characterization of those cores. The results in [21] seem likely to be key, and it can be shown that no convex capacity has a core that is descriptively complete.

6.3. **Literature.** The ability to model two outcome upper/lower probability intervals is intimately related to the interval-valued probabilities introduced by Keynes [25], given a decision theoretic foundation by Smith [34], and studied by de Finetti [10, Appendix §19.3]. However, there seems to be no previous detailed examination of the properties Savage-de Finetti set of extensions. The key is to show that the Savage-de Finetti set can be expressed as a convex combination/integral of disjointly supported probability simplexes.

Suppes and Zanotti [39] give sufficient conditions for de Finetti's interval valued probabilities to contain a probability. In the present context, this is not an issue because we start with the probability μ. Rather, Theorem 2 contains an answer to a dual question: when will the interval valued probability $E \mapsto \{\nu(E) : \nu \in \Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})\}$ provide a model for all upper/lower probability intervals over pairs of outcomes?

There are vast literatures on partially observable econometrics models and countably additive learning models. In graduate economics curricula, the partially observable models make their first appearance in the part of the econometrics sequence known affectionately as "The Linear Model with Diseases." Much of the recent theoretical research in econometrics concerns the very hard problem of making inferences in the presence of different kinds of partial observability. This typically involves postulating some extra structure, hopefully minimal and having a sound theoretical basis, on the joint distribution of the observables and the unobservables and then deriving good estimators of the *cæteris paribus* quantities of interest.

The literature on countably additive learning models and the associated decision theory as well as the approximation of difficult problems by simpler, is comprehensively developed in [27] or [28]. There is also a well-developed theory of Bayesian statistical problems with pfa priors e.g. [24, Ch. 12, 15]. There is nearly nothing about learning pfa probabilities, and still less about the interactions of learning and decision theory. Perhaps the present results and the difficulties of recoupling learning and observations given in §5.4 explains this lack. A notable exception is a study of purely finitely additive, decoupled observational learning models in Al-Najjar [1] (AN below).

For the unit interval $[0,1]$, the class of intervals $[0,a]$ are a VC class generating the Borel σ-field. AN notes that if $(\mathbb{X},\mathcal{X})$ is a Polish space, then it can be imbedded as a measurable subset of $[0,1]$, hence the inverse images of the $[0,a]$ are a VC class in $\mathbb{X}$ that generates $\mathcal{X}$. From this, AN's Theorem 3 states that countably additive observational models are determinate. By contrast, AN's Theorem 4 shows that for the state space $(\mathbb{N},\mathcal{P}(\mathbb{N}))$ and any VC class $\mathcal{C}$, there exists a pfa μ with $Diam(\Pi_{\mathcal{X}}^{fa}(\mu_{|\mathcal{C}})) \geq \frac{1}{2}$. AN argues (p. 1373) that his Theorem 3, the determinacy result for Polish spaces with VC classes indicates that "asymptotic learning is easy" for Polish spaces (p. 1383), and that this "is a consequence of implicit structural restrictions these (Polish) spaces impose." He also argues that the indeterminacy in his Theorem 4 arises because the space $(\mathbb{N},\mathcal{P}(\mathbb{N}))$ "is free from any inductive biases involving notions of distance, ordering, or similarity." These claims seem questionable and/or misleading on several levels. Recall that in observational learning models, $\mathcal{X} = \sigma(\mathcal{C})$.

- If $\mathcal{X} = \sigma(\mathcal{C})$, then AN's Theorem 3, the determinacy of countably additive learning models, is Carathéodory's extension theorem, a result that holds for all measure spaces, not just the Polish ones, and does not depend on $\mathcal{C}$ being a VC class.
- The state space $(\mathbb{N},\mathcal{P}(\mathbb{N}))$ used in AN's Theorem 4 is also Polish, and the offered proof of the existence of a purely finitely additive μ on the integers for which $Diam(\Pi_{\mathcal{X}}^{fa}(\mu_{|\mathcal{C}})) \geq \frac{1}{2}$, applies to all infinite measure spaces.
- If $\mathcal{X} = \sigma(\mathcal{C})$, then the only way to avoid the existence of a diameter 1 asymptotic set in a finitely additive observational model is to have $\mathcal{F}^\circ(\mathcal{C})$ equal to the class of all events, $\mathcal{X}$, in which case there is no ambiguity in any asymptotic set.
- The claim that asymptotic learning is "easy" in all Polish settings with countably additive probabilities is belied by the entire theory of non-parametric regression with Polish spaces of functions. The difficulties in this area cannot be solved by using the suggested wildly irregular inverse images of the intervals $[0,a]$ in spaces of functions.
- Except in "emancipatory" mathematics [35, p. 231], the correctness of mathematical arguments is independent of human inductive biases.

AN §4 describes three examples of multiple prior preferences: Bewley's better than the status quo preferences [4]; the Gilboa-Schmeidler preferences [23], that is, the α-MEU preferences given above with $\alpha = 1$; and "Bayesianism," selecting and using as priors some sequence of probabilities from sets $\Pi_n(\mu)$ that have the property that $Diam(\cap_n\Pi_n(\mu)) = 1$ (which implies that the selected sequence of $\mu_n \in \Pi_n(\mu)$ can have disjoint carrier sets).

AN then makes the fascinating suggestion that the partial indeterminacy of at least one purely finitely additive asymptotic set with $\mathcal{C}$ being a VC class provides a learning justification for the sets of priors used in the first two classes of models. Unfortunately, AN contains no supporting arguments for this suggestion. Lemma 5 and Theorem 2 show that for purely finitely additive priors and VC classes, the indeterminacy is total, all diameters are 1. Further, Theorem 2 gives the class of problems that can be modeled using these asymptotic sets: if the observations support the hypothesis that μ is non-atomic, then the asymptotic set provides a model for the closed convex hull of the set of closed faces.

References

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

[2] Donald W. K. Andrews. Empirical process methods in econometrics. In *Handbook of Econometrics, Vol. IV*, volume 2 of *Handbooks in Economics*, pages 2247–2294. North-Holland, Amsterdam, 1994.

[3] Thomas E. Armstrong. Strong singularity, disjointness, and strong finite additivity of finitely additive measures. *J. Math. Anal. Appl.*, 131(2):565–587, 1988.

[4] Truman F. Bewley. Knightian decision theory. I. *Decis. Econ. Finance*, 25(2):79–110, 2002.

[5] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges: A Study of Finitely Additive Measures*, volume 109 of *Pure and Applied Mathematics*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1983.

[6] Patrick Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

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

[8] Dean Corbae, Maxwell B. Stinchcombe, and Juraj Zeman. *An Introduction to Mathematical Analysis for Economic Theory and Econometrics*. Princeton University Press, Princeton, NJ, 2009.

[9] Bruno de Finetti. *Theory of Probability: A Critical Introductory Treatment. Vol. 1*. John Wiley & Sons, London-New York-Sydney, 1974.

[10] Bruno de Finetti. *Theory of Probability: A Critical Introductory Treatment. Vol. 2*. John Wiley & Sons, London-New York-Sydney, 1975. Translated from the Italian by Antonio Machì and Adrian Smith, Wiley Series in Probability and Mathematical Statistics.

[11] Bruno de Finetti and Leonard J. Savage. Sul modo di scegliere le probabilità iniziali. *Biblioteca del Metron, Ser. C*, I:81–154, 1962.

[12] Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[13] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339, 1967.

[14] Jean Dieudonné. *Treatise on Analysis. Vol. II*. Translated from the French by I. G. Macdonald. Pure and Applied Mathematics, Vol. 10-II. Academic Press, New York-London, 1970.

[15] Richard M. Dudley. *Uniform Central Limit Theorems*, volume 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.

[16] Richard M. Dudley, Evarist Giné, and Joel Zinn. Uniform and universal Glivenko-Cantelli classes. *J. Theoret. Probab.*, 4(3):485–510, 1991.

[17] James Dugundji. *Topology*. Allyn and Bacon, Inc., Boston, Mass., 1966.

[18] Martin Dumav and Maxwell B. Stinchcombe. Skorokhod's representation theorem for sets of probabilities. *Trans. Amer. Math. Soc.*, 2015.

[19] Martin Dumav and Maxwell B. Stinchcombe. The von Neumann/Morgenstern approach to ambiguity. Working Paper, Department of Economics, U.T. Austin, 2015.

[20] Nelson Dunford and Jacob T. Schwartz. *Linear Operators. Part I: General Theory*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1988.

[21] Aziz El Kaabouchi. Mesures dominées par une capacité alternée d'ordre 2. *Proc. Amer. Math. Soc.*, 121(3):823–832, 1994.

[22] Paolo Ghirardato, Fabio Maccheroni, and Massimo Marinacci. Differentiating ambiguity and ambiguity attitude. *J. Econom. Theory*, 118(2):133–173, 2004.

[23] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with nonunique prior. *J. Math. Econom.*, 18(2):141–153, 1989.

[24] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003.

[25] John Maynard Keynes. *A Treatise on Probability. Vol VIII in The collected writings of John Maynard Keynes*. Macmillan Press Ltd., London; Cambridge University Press, New York, 1988.

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

[27] Lucien Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986.

[28] Friedrich Liese and Klaus-J. Miescke. *Statistical Decision Theory*. Springer Series in Statistics. Springer, New York, 2008.

[29] Dorothy Maharam. Finitely additive measures on the integers. *Sankhyā Ser. A*, 38(1):44–59, 1976.

[30] Robert R. Phelps. *Lectures on Choquet's Theorem*, volume 1757 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2001.

[31] Leonard J. Savage. *The Foundations of Statistics*. Dover Publications, Inc., New York, revised edition, 1972.

[32] Roman Sikorski. *Boolean Algebras.* Second edition. Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue-Folge, Band 25. Academic Press Inc., New York; Springer-Verlag, Berlin-New York, 1964.
[33] Anatoliĭ V. Skorohod. Limit theorems for stochastic processes. *Teor. Veroyatnost. i Primenen.*, 1:289–319, 1956.
[34] Cedric A. B. Smith. Consistency in statistical inference and decision. *J. Roy. Statist. Soc. Ser. B*, 23:1–37, 1961.
[35] Alan D Sokal. Transgressing the boundaries: Toward a transformative hermeneutics of quantum gravity. *Social text*, 46/47:217–252, 1996.
[36] Maxwell B. Stinchcombe. Bayesian information topologies. *J. Math. Econom.*, 19(3):233–253, 1990.
[37] Maxwell B. Stinchcombe. The gap between probability and prevalence: loneliness in vector spaces. *Proc. Amer. Math. Soc.*, 129(2):451–457, 2001.
[38] Maxwell B. Stinchcombe and Halbert White. Some measurability results for extrema of random functions over random sets. *Rev. Econom. Stud.*, 59(3):495–514, 1992.
[39] Patrick Suppes and Mario Zanotti. Conditions on upper and lower probabilities to imply probabilities. *Erkenntnis*, 31(2):323–345, 1989.
[40] J. Jerry Uhl. Review of Rao and Rao (1983). *Bull. Lond. Math. Soc.*, 16(4):431–432, 1984.
[41] Vladimir N. Vapnik and A. Ja. Červonenkis. On the uniform convergence of frequencies of the appearance of events to their probabilities. *Theor. Probability Appl.*, 16:264–280, 1971.
[42] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, NJ, anniversary edition, 2007.
[43] Kôsaku Yosida and Edwin Hewitt. Finitely additive measures. *Trans. Amer. Math. Soc.*, 72:46–66, 1952.

Appendix A. Proofs

Proof of Lemma 1. If μ has an atom of size q, then the minimum size of the atoms in an image measure, $f(\mu)$, is of size at least q. As $\mathbb{W}$ contains two or more points, the set of distributions with smaller mass points is non-empty, and all such probabilities are ruled out. Suppose now that μ is non-atomic. Construct $h : \mathbb{X} \to [0, 1]$ such that $\mu(h^{-1}((a, b])) = (b - a)$ for all $0 \leq a \leq b \leq 1$. By Skorokhod's representation theorem, there exists $g_p : [0, 1] \to \mathbb{W}$ such that $g_p(\lambda) = p$ where λ is the uniform distribution on $[0, 1]$. Set $f_p(x) = g_p(h(x))$. The verification that $\int_{\mathbb{X}} v(f_p(x))\, d\mu(x) = \int_{\mathbb{W}} v(w)\, dp(w)$ for continuous v is routine. □

Proof of Lemma 2. Because $\left|\max_{a\in A} \int u(a, x)\, dP_n(x) - \max_{b\in A} \int u(b, x)\, d\mu(x)\right|$ is less than or equal to $v_n := \max_{a\in A} \left|\int u(a, x)\, dP_n(x) - \int u(a, x)\, d\mu(x)\right|$, it is sufficient to show $v_n \to 0$. For this, it is in turn sufficient to show that every subsequence, $v_{n'}$ has a further subsequence, $v_{n''}$, with $v_{n''} \to 0$. Along the subsequence n', let $a_{n'}$ maximize $v_{n'}$. Because A is compact, there exists an $a^* \in A$ and a further subsequence, $a_{n''} \to a^*$. Relabeling the subsubsequence $a_{n''}$ as a_m, we have $a_m \to a^*$, and we must show that $v_m = | \int u(a_m, x)\, dP_m(x) - \int u(a_m, x)\, d\mu(x)| \to 0$.

Adding and subtracting $\int u(a^*, x)\, dP_m(x)$ and $\int u(a^*, x)\, d\mu(x)$ to v_m, we have

$$\begin{aligned} v_m &= \left|\int u(a_m, x)\, dP_m(x) - \int u(a_m, x)\, d\mu(x)\right| && (16)\\ &\leq \left|\int u(a_m, x)\, dP_m(x) - \int u(a^*, x)\, dP_m(x)\right| && (17)\\ &\quad + \left|\int u(a^*, x)\, dP_m(x) - \int u(a^*, x)\, d\mu(x)\right| && (18)\\ &\quad\quad + \left|\int u(a^*, x)\, d\mu(x) - \int u(a_m, x)\, d\mu(x)\right|. && (19)\end{aligned}$$

The term in (18) goes to 0 because $x \mapsto u(a^*, x)$ is measurable and bounded and $P_n(E) \to \mu(E)$ for all $E \in \mathcal{X}$. The term in (19) goes to 0 by Lebesgue's dominated convergence theorem. The term in (17) can be integrated separately over any set E and its complement, which yields

$$\begin{aligned} &\left|\int_E u(a_m, x)\, dP_m(x) - \int_E u(a^*, x)\, dP_m(x)\right| + && (20)\\ &\quad \left|\int_{E^c} u(a_m, x)\, dP_m(x) - \int_{E^c} u(a^*, x)\, dP_m(x)\right|. && (21)\end{aligned}$$

Because $u(a_m, \cdot) \to u(a^*, \cdot)$ pointwise, there exists a set E an integer M_1 with $\mu(E) > 1 - (\epsilon/2)$ and $|u(a_m, x) - u(a^*, x)| < \epsilon$ for all $x \in E$, and $|P_m(E) - \mu(E)| < \epsilon/2$ for $m \geq M_1$. Thus, for $m \geq M_1$, the term in (20) is less than ϵ. Taking $f = 1_{E^c}$, we can pick M_2 such that for all $m \geq M_2$, $| \int 1_{E^c}\, dP_m - \int 1_{E^c}\, d\mu| < \epsilon$. Because $0 \leq u(a, x) \leq 1$ for all a and x and $\mu(E^c) < \epsilon$, we have the following bounds in equation (21),

$$\left|\int_{E^c} u(a_m,x)\,dP_m(x) - \int_{E^c} u(a^*,x)\,dP_m(x)\right| \tag{22}$$

$$\leq \left|\int_{E^c} u(a_m,x)\,dP_m(x)\right| + \left|\int_{E^c} u(a^*,x)\,dP_m(x)\right| \tag{23}$$

$$\leq \left|\int_{E^c} 1\,dP_m(x)\right| + \left|\int_{E^c} 1\,dP_m(x)\right| < 2\cdot\epsilon + 2\cdot\epsilon. \tag{24}$$

Thus, for all $m \geq \max\{M_1, M_2\}$, the term in (17) is less than $5\cdot\epsilon$. □

Proof of Theorem 1. From the theory of regular conditional probabilities for standard measure spaces (e.g. [12, III.72-4]), ν belongs to the partially identified set if and only if there is a measurable $x_1 \mapsto q_{x_1} \in \Delta(M_2)$ such that $\nu(E_1 \times E_2) = \int_{E_1} q_{x_1}(E_2)\,d\mu_1(x_1)$. By the definition of the descriptive range, if $B \in \mathcal{R}(\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}}))$, then there exists a measurable $f : M_1 \times M_2 \to \mathbb{W}$ such that $B = f(\Pi^{ca}_{\mathcal{X}}(\mu_{|\mathcal{C}}))$. Define $F_{x_1} = f(x_1, M_2)$ and note that $B = \int \Delta(F_{x_1})\,d\mu_1(x_1)$. For given measurable, non-empty valued $x_1 \mapsto F_{x_1} \subset \mathbb{W}$, let $f : M_1 \times M_2 \to \mathbb{W}$ be any measurable function with $f(x_1, M_2) = F_{x_1}$. □

Proof of Claims in Example 4. For $0 < s \leq s' < 1$, let $I_{s,s'}$ be the set of $x \in [0,1]$ with the lim inf of the frequency of 1's in their binary expansions equal to s and the lim sup equal to s'. If μ is countably additive, then by Borel's normal number theorem [6, Ch. 1, §1], $\mu(I_{\frac{1}{2},\frac{1}{2}}) = 1$.

For each non-empty, open $G \subset [0,1]$ and each $I_{s,s'}$, $G \cap I_{s,s'}$ is uncountable. Therefore, each $I_{s,s'}$ contains uncountably many countable disjoint, dense sets. For any countable dense $D \subset [0,1]$ and $0 \leq a < b \leq 1$, define a ν satisfying $\nu(D) = 1$ by $\nu(D \cap (a,b]) = (b-a)$. By the Hahn-Banach extension theorem, ν has a pfa extension to the class of all subsets of $[0,1]$. □

Proof of Lemma 3. Suppose that there is no sum γ that is everywhere negative. Taking $\gamma(x) = 1_E(x) - r_E$ shows that $r_E \leq 1$, taking $\gamma(x) = r - 1_E(x)$ shows that $0 \leq r_E$. Combining, each r_E belongs to the interval $[0,1]$. For disjoint E and E', the sum of three bets, $\gamma = (1_{E\cup E'} - r_{E\cup E'}) + (r_E - 1_E) + (r_{E'} - 1_{E'})$, is everywhere negative if $r_{E\cup E'} > r_E + r_{E'}$. The sum of the the other side of the three bets is everywhere negative if $r_{E\cup E'} < r_E + r_{E'}$. Combining, $r_{E\cup E'} = r_E + r_{E'}$ for disjoint sets.

Suppose now that $r_E = \mu(E)$ for some probability μ and let $E^\mu(\cdot)$ be the expectation operator for μ. We have $E^\mu(1_E - r_E) = E^\mu(r_E - 1_E) = 0$ for all E so that $E^\mu \sum_E h_E = 0$. No mean 0 finite sum of simple functions can be everywhere negative. □

Proof of Lemma 4. The field $\mathcal{C}$ is a strict subset of $\mathcal{X}$ iff the closure of $M_S(\mathcal{C})$, the class of simple $\mathcal{C}$-measurable functions, is a strict subset of $M_b(\mathcal{X})$, the class of $\mathcal{X}$-measurable, bounded functions. By the Hahn-Banach theorem, there exists a non-zero ϕ in the dual space of $M_b(\mathcal{X})$, the finite signed measures on $\mathcal{X}$ [20, Theorem IV.5.1], such that $\int f\,d\phi = 0$ for all $f \in M_S(\mathcal{C})$. The Jordan decomposition for finitely additive signed measures [20, Theorem III.1.8] expresses ϕ as the sum $\phi = r\nu - s\nu'$ where $\nu, \nu' \in \Delta^{fa}(\mathcal{X})$ are probabilities at norm distance 1 from each other and $r, s \geq 0$. Because $M_S(\mathcal{C})$ contains the constant functions, $r = s$, because ϕ is non-zero, $r, s > 0$. Rescaling r and s to 1 and setting $\mu = \nu$ yields $Diam(\Pi^{fa}_{\mathcal{X}}(\mu_{|\mathcal{C}})) = 1$. □

Sketches for the properies of $(\widehat{\mathbb{X}}, \widehat{\mathcal{X}})$. The class of cylinder sets in $\widehat{Y}$ are the product field, $\mathcal{Y}^\circ$, every cylinder set is compact, so $F_n \downarrow \emptyset$ in $\mathcal{Y}^\circ$ implies that $F_N = \emptyset$ for some N, hence Carathéodory's extension theorem implies that any probability on $\mathcal{Y}^\circ$ has a unique continuous extension to the product σ-field, and the extension to the Borel σ-field is routine e.g. [32, p. 202]. The existence of continuous extensions follows from the proof of [17, Theorem 8.2(1)]. (v) arises from the previous point and the bijective representation of elements of $\Delta^{fa}(\mathcal{X})$ as elements of the dual space of the class of bounded measurable functions [20, Theorem IV.5.1] and the bijective representation of countably additive probabilities on $\widehat{\mathbb{X}}$ as elements of the dual space of the continuous functions on $\widehat{\mathbb{X}}$ [20, Theorem IV.6.3].

Proof of Theorem 2. We use the construction of $\widehat{\mathbb{X}}$ and $\widehat{\mathcal{X}}$ given in §5.1. A $\nu \in \Delta^{fa}(\mathcal{X})$ is a **point mass** if for all $E \in \mathcal{X}$, $\nu(E)$ is either equal to 0 or to 1. Point masses can be identified with the ultrafilters $\{E \in \mathcal{X} : \nu(E) = 1\}$ and pfa point masses can be identified with the free ultrafilters, that is, with point masses on the corona, $\widehat{\mathbb{X}} \setminus \mathbb{X}$. There are at least 2^{2^ω} pfa point masses where ω is the cardinality of $\mathbb{N}$ [7, Cor. 7.4].

Suppose first that μ is a pfa point mass. Without loss, assume that $\mathcal{C}$ is a countably infinite field of sets. Let $\widehat{x} \in \widehat{\mathbb{X}}$ be the support point for the of countably additive point mass $\widehat{\mu}$, and define $x_S = \mathrm{proj}_{\mathcal{C}}(\widehat{x}) \in S$ where S is the observation space $\{0,1\}^{\mathcal{C}}$. Because $\mathrm{proj}_{\mathcal{C}}^{-1}(x_S)$ is a compact subset of the corona of $\mathbb{X}$, it is sufficient to show that $\mathrm{proj}_{\mathcal{C}}^{-1}(x_S)$ has cardinality at least $2^{2^{\omega}}$ where ω is the cardinality of $\mathbb{N}$. Enumerate $\mathcal{C}_1 = \{C \in \mathcal{C} : \mu(C) = 1\}$ as $\{C_n : n \in \mathbb{N}\}$. Let E_n be a sequence in $\mathcal{X}$ with $E_n \downarrow \emptyset$ and $\mu(E_n) \equiv 1$. Define $D_n = E_n \cap \bigcap_{m=1}^{n} C_m$ so that $D_n \downarrow \emptyset$ and $\mu(D_n) \equiv 1$. Define $\widehat{D} = \cap_n \widehat{D}_n$, and note that $\widehat{D}$ is a subset of $\mathrm{proj}_{\mathcal{C}}^{-1}(x_S)$. We now show that there are $2^{2^{\omega}}$ different point masses supported on the compact set $\widehat{D}$.

Define $F_0 = D_1^c$, $F_1 = D_1 \setminus D_2$, ..., $F_n = D_n \setminus D_{n+1}$. Renumbering if necessary, there is no loss in assuming that each F_n is non-empty. Let $\mathcal{G} = \sigma(\mathcal{C}, \{F_n : n \in \mathbb{N}\})$. For each pfa point mass γ on the integers, define $\nu_\gamma \in \Delta^{fa}(\mathcal{G})$ by $\nu_\gamma(G) = \gamma(\{n \in \mathbb{N} : F_n \subset G\})$. Each ν_γ is pfa and agrees with μ on the class $\mathcal{C}$. Each ν_γ has a non-empty, compact, convex set of Hahn-Banach extensions to $\mathcal{X}$, all of them necessarily pfa, and all of them agreeing with μ on $\mathcal{C}$. By the Krein-Milman theorem [20, Theorem V.8.4], the set of extensions is the closed convex hull of its extreme points. The extreme points are again point masses, hence each ν_γ has a point mass extension, also denoted ν_γ. By construction, $\widehat{\nu}_\gamma$ is point mass on a point in $\widehat{D}$. Since the pfa point masses on the integers correspond to points $\beta\mathbb{N} \setminus \mathbb{N}$ where $\beta\mathbb{N}$ is the Stone-Čech compactification of the integers, there are $2^{2^{\omega}}$ of them.

Now pick arbitrary pfa μ, assume (again without loss) that $\mathcal{C}$ is a countably infinite field of sets, and let $\mu_S = \mathrm{proj}_{\mathcal{C}}(\widehat{\mu})$. By the Choquet-Bishop-de Leeuw theorem for countably additive probabilities on compact non-metrizable spaces (e.g. [30, §4]), any pfa $\widehat{\nu}$ has a unique expression as an integral of point masses on the corona of $\mathbb{X}$. Since $\Pi_{\mathcal{X}}^{fa}(\mu_{|\mathcal{C}}) = \{\nu \in \Delta^{fa}(\mathcal{X}) : \mathrm{proj}_{\mathcal{C}}(\widehat{\nu}) = \mu_S\}$, the result for point masses implies that

$$\Pi_{\mathcal{X}}^{fa}(\mu_{|\mathcal{C}}) = \int_S \Delta(\mathrm{proj}_{\mathcal{C}}^{-1}(x_s))\, d\mu_S(x_s) \tag{25}$$

where $\Delta(\mathrm{proj}_{\mathcal{C}}^{-1}(x_s))$ is the set of probabilities on the compact set $\mathrm{proj}_{\mathcal{C}}^{-1}(x_s)$. Finally, $\mathrm{proj}^{-1}(\mu_S)$ is non-atomic if μ_S is. □

The essential fact about VC classes is that they are simultaneously totally bounded for a wide range of metrics. Recall that totally bounded and complete subsets of metric spaces are compact. Thus, VC classes are, up to completion, compact in a variety of metrics.

Proof of Lemma 5. We use notation from the proof of Theorem 2. For a pseudo-metric space (S,d) and $\epsilon > 0$, define $D(\epsilon, S, d)$ 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 $d_{2,Q}(E,F) = Q(E\Delta F)^{1/2} = \left(\int (1_{E_1} - 1_{E_2})^2\, dQ\right)^{1/2}$ 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 [15, Theorem 10.1.7], if $\mathcal{C}$ is a VC class, we have the following metric entropy result,

$$\log D^{(2)}(\epsilon, \mathcal{C}) = O(\epsilon^{-2}) \text{ as } \epsilon \downarrow 0. \tag{26}$$

We now show the following.

1. $\mathcal{C}$ satisfies (26) iff $\widehat{\mathcal{C}}$ satisfies (26) where $\widehat{\mathcal{C}} = \{\widehat{C} : C \in \mathcal{C}\}$.
2. The supremum in $D^{(2)}(\epsilon, \widehat{\mathcal{C}})$ can be taken over all countably additive probabilities on $\widehat{\mathcal{X}}$.
3. The previous 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; and
4. if $\mathcal{C}_0$ is a countable d_μ-dense subset of $\mathcal{C}$, then $\Pi(\mu_{|\mathcal{C}_0}, \mathcal{X}) = \Pi_{\mathcal{X}}^{fa}(\mu_{|\mathcal{C}})$.

To show that $\mathcal{C}$ satisfies (26) iff $\widehat{\mathcal{C}}$ satisfies (26), it is sufficient to show that $\sup_p D(\epsilon, \mathcal{C}, d_{2,p}) = \sup_q D(\epsilon, \widehat{\mathcal{C}}, d_{2,q})$ where the first supremum is taken over finitely supported probabilities in $\mathbb{X}$ and the second is taken over finitely supported probabilities in $\widehat{\mathbb{X}}$. This follows from the openness of each $\widehat{C}$ in $\widehat{\mathbb{X}}$ and denseness of $\mathbb{X}$ in $\widehat{\mathbb{X}}$.

From the usual weak* approximation of all countably additive probabilities on a compact Hausdorff space by finitely supported probabilities on dense subsets, this means that the supremum

$\sup_Q D(\epsilon, \widehat{\mathcal{C}}, d_{2,Q})$ is, without loss, taken over all $Q \in \Delta^{ca}(\widehat{\mathcal{X}})$. Therefore, for any pfa μ, we know that

$$\log D(\epsilon, \widehat{\mathcal{C}}, d_{2,\widehat{\mu}}) = \log D(\epsilon, \mathcal{C}, d_{2,\mu}) \leq O(\epsilon^{-2}) \text{ as } \epsilon \downarrow 0. \tag{27}$$

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 $d_{2,\mu}$. Let $\mathcal{C}_0$ be a countable d_μ-dense subset of $\mathcal{C}$. It is immediate that for any countable, d_μ-dense $\mathcal{C}_0 \subset \mathcal{C}$, $\Pi(\widehat{\mu}_{|\widehat{\mathcal{C}}_0}, \widehat{\mathcal{X}}) = \Pi(\widehat{\mu}_{|\widehat{\mathcal{C}}}, \widehat{\mathcal{X}})$. □

Proof of Lemma 6. If $a \mapsto E\,R(a)$ is continuous, then the set of a such that $E\,R(a) > s - \epsilon$ is non-empty and open. Since Q is full support, it puts strictly positive mass on this set, and the Borel-Cantelli lemma delivers $Q^\infty(A_\epsilon) = 1$.

To show continuity, pick arbitrary $\epsilon > 0$. Note that the evaluation mapping $e : A \times V_0^1(A) \to [0,1]$ defined by $e(a,v) = v(a)$, is jointly continuous, hence uniformly continuous on $A \times K$ for any compact $K \subset V_0^1(A)$. Since μ is tight for $u(\cdot,\cdot)$, there exists a countably additive probability $p \in \Delta(\mathcal{V})$ at Prohorov distance 0 from the distribution $p' := u(\cdot, \mu)$. For every $\epsilon > 0$, there exists a compact $K_\epsilon \subset V_0^1(A)$ such that $p(K_\epsilon) > (1-\epsilon)$. Because K_ϵ is compact, it is equi-continuous. Therefore, there exists $\delta > 0$ such that $d(a,a') < \delta$ implies that $|v(a) - v(a')| < \epsilon$ for all $v \in K_\epsilon$. Since p is at Prohorov distance 0 from p', $\int e(a,v)\,dp(v) = \int e(a,v)\,dp'(v)$ for all $a \in A$. For $d(a,a') < \delta$, $|\int e(a,v)\,dp(v) - \int e(a',v)\,dp(v)| < 2\epsilon$. □

Dept. of Economics, University of Texas, Austin, e-mail: max.stinchcombe@gmail.com