Statistical Methods & Applications
https://doi.org/10.1007/s10260-019-00486-6

ORIGINAL PAPER

What finite-additivity can add to decision theory

Mark J. Schervish[1] · Teddy Seidenfeld[2] · Rafael B. Stern[3] · Joseph B. Kadane[1]

Accepted: 9 August 2019

Abstract

We examine general decision problems with loss functions that are bounded below. We allow the loss function to assume the value ∞. No other assumptions are made about the action space, the types of data available, the types of non-randomized decision rules allowed, or the parameter space. By allowing prior distributions and the randomizations in randomized rules to be finitely-additive, we prove very general complete class and minimax theorems. Specifically, under the sole assumption that the loss function is bounded below, we show that every decision problem has a minimal complete class and all admissible rules are Bayes rules. We also show that every decision problem has a minimax rule and a least-favorable distribution and that every minimax rule is Bayes with respect to the least-favorable distribution. Some special care is required to deal properly with infinite-valued risk functions and integrals taking infinite values.

Keywords Admissible rule · Bayes rule · Complete class · Least-favorable distribution · Minimax rule

Mathematics Subject Classification Primary 62C07; Secondary 62C20

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10260-019-00486-6) contains supplementary material, which is available to authorized users.

✉ Mark J. Schervish
mark@cmu.edu

Teddy Seidenfeld
teddy@stat.cmu.edu

Rafael B. Stern
rbstern@gmail.com

Joseph B. Kadane
kadane@stat.cmu.edu

1 Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

2 Departments of Philosophy and Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

3 Statistics Department, Federal University of São Carlos, São Carlos 13565-905, Brazil

Published online: 22 August 2019

1 Introduction

1.1 Motivation

The following example, adapted from Example 3 of Schervish et al. (2009), is a case in which countably-additive randomized rules do not contain a minimal complete class. It involves a discontinuous version of squared-error loss in which a penalty is added if the prediction and the event being predicted are on opposite sides of a critical cutoff.

Example 1 A decision maker is going to offer predictions for an event B and its complement. The parameter space is $\Theta = \{B, B^C\}$ while the action space is $\mathcal{A} = [0, 1]^2$, pairs of probability predictions. The decision maker suffers the sum of two losses (one for each prediction) each of which equals the usual squared-error loss (square of the difference between indicator of event and corresponding prediction) plus a penalty of 0.5 if the prediction is on the opposite side of 1/2 from the indicator of the event. In symbols, the loss function equals

$$\begin{aligned} L(\theta, (a_1, a_2)) &= (I_B - a_1)^2 + (I_{B^C} - a_2)^2 \\ &+ \frac{1}{2}\begin{cases} I_{[0,1/2]}(a_1) + I_{(1/2,1]}(a_2) & \text{if } \theta = B, \\ I_{(1/2,1]}(a_1) + I_{[0,1/2]}(a_2) & \text{if } \theta = B^C. \end{cases} \end{aligned}$$

To keep matters simple, we assume that no data are available, but one could rework the example with potential data at the cost of more complicated calculations. Figure 1 is a plot of all of the pairs $(L(B, (a_1, a_2)), L(B^C, (a_1, a_2)))$, which shows all of the possible risk functions of non-randomized rules (pure strategies when there are no data available). The admissible non-randomized strategies are the pairs $(p, 1 - p)$ for $p \in [0, 1]$. The corresponding points in Fig. 1 are

(i) from but not including (1.5,0.5) to (3,0) in the lower right, corresponding to $p \in [0, 1/2)$,
(ii) (1,1) in the middle section, corresponding to $p = 1/2$, and
(iii) from (0,3) up to but not including (0.5,1.5) in the upper left, corresponding to $p \in (1/2, 1]$.

The countably-additive randomized rules have risk functions in the convex hull of the set plotted in Fig. 1. The resulting set is not closed, and its lower boundary is missing all points on the closed line segment from (0.5, 1.5) to (1.5, 0.5) except (1, 1). One consequence of these points being missing is that there are many inadmissible rules, corresponding to points just above that line segment, that are dominated by other inadmissible rules, but not dominated by an admissible rule. In other words, the admissible rules do not form a complete class in this problem.

If, however, one is willing to introduce finitely-additive randomizations, all of the missing risk functions of admissible rules become available. For example, there are finitely-additive probabilities P_- and P_+ on the power set of [0, 1] such that $P_-(A) = 1$ for every set of the form $A = (1/2 - \epsilon, 1/2)$ and $P_+(C) = 1$ for every set of the form $C = (1/2, 1/2 + \epsilon)$. Every missing point on the line segment between (0.5, 1.5) and (1.5, 0.5) is the risk function of a randomized rule that gives a_1 the distribution

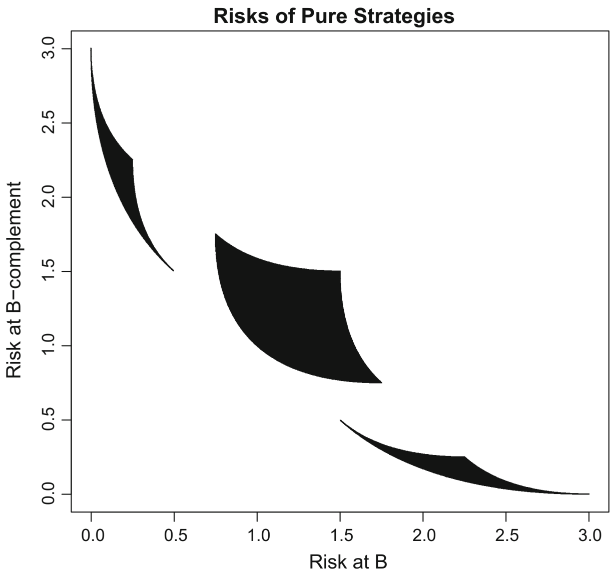

Fig. 1 The risk functions of non-randomized rules in Example 1

$\alpha P_{-} + (1-\alpha)P_{+}$ for some $\alpha \in [0, 1/2) \cup (1/2, 1]$ and sets $a_2 = 1 - a_1$. For example, the risk function for $\alpha = 0$ is the point (0.5, 1.5) while the risk function for $\alpha = 1$ is the point (1.5, 0.5).

One result (Theorem 1) that we prove in this paper is that every decision problem with a loss function bounded below has a minimal complete class consisting of Bayes rules if finitely-additive randomizations are allowed. There are several complete class theorems in the countably-additive literature that make additional assumptions about the loss function (e.g., continuous, convex) and about the distributions of the available data (e.g., exponential family distributions). Some of these results can be found in Berger and Srinivasan (1978), Brown (1971), and Lehmann and Casella (1998, Section 5.7). Our other main theorem is a general minimax theorem (Theorem 2) stating that every decision problem with a loss function bounded below has a minimax rule and least-favorable prior such that all minimax rules are Bayes with respect to that least-favorable prior. A finitely-additive minimax theorem when risk functions are bounded is proven in Heath and Sudderth (1972). Cases in which a Bayesian analysis will be performed and the joint distribution of data and parameter can be computed by integration in both orders (a property not possessed by all finitely-additive distributions) is the subject of Heath and Sudderth (1978), which presents results about Bayes rules and extended admissible rules.

Our results also cover cases in which the loss function is allowed to assume the value ∞. An example of such a loss function is the logarithmic loss for predicting events. In the notation of Example 1, replace $(I_B - a_1)^2 + (I_{B^C} - a_2)^2$ by $-\log(a_1[1 - a_2])I_B - \log(a_2[1 - a_1])I_{B^C}$. Dealing with loss functions that assume the value ∞ requires special care. In particular, the set of functions that need to be integrated is not a linear space. In order to get ∞ into the range of our loss and risk functions, we use $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ as the range, and endow $\widehat{\mathbb{R}}$ with the topology that is generated by the open intervals together with sets of the form $(c, \infty) \cup \{\infty\}$ for all real c. This is similar to the usual two-point compactification of the reals, but we leave out all the sets containing $-\infty$. As a result, $\widehat{\mathbb{R}}$ is not compact, but each "interval" of the form $[c, \infty]$ (for finite c) is compact. Our topology also differs from the one-point compactification of Alexandroff (see Kelley 1955, Theorem 21, p. 150) whose "intervals" around ∞ include both positive and negative numbers with arbitrarily large absolute value.

Throughout the paper, we adhere to the convention that $\infty \times 0 = 0$. We take arithemetic that includes $\infty - \infty$ to be undefined. For example, if f and g are both functions on the same space $\mathcal{Z}$, and $f(z) = g(z) = \infty$ for some z, then $f(z) - 2g(z)$ is not defined. In particular, $f - 2g$ does not have the same domain as f and g.

1.2 Organization

The paper is organized as follows. We start with an introduction to finitely-additive expectations in Sect. 2. Section 3 contains a general overview of decision theory. We then prove some facts about the structure of the risk set in Sect. 4. The first of the main theorems is the complete class theorem (Theorem 1) that appears in Sect. 5. It states that every decision problem has a minimal complete class of rules consisting of admissible Bayes rules. The second theorem (Theorem 2 in Sect. 6) says that every decision problem has a minimax rule and a least-favorable prior. Also, each minimax rule is Bayes with respect to the least-favorable prior. Section 7 gives some results on the existence and non-existence of Bayes rules with respect to specific types of priors.

We place a large amount of mathematical background needed to prove the main theorems in an appendix and in the Online Resource. Section A of the appendix gives an overview of the theory of finitely-additive expectations. Section B lists some results which are cited in this paper and whose proofs appear in the Online Resource. Section 1 of the Online Resource includes background on topology and convergence. Section 2 of the Online Resource gives more detail on finitely-additive expectations. Section 3 of the Online Resource gives some results on separation of convex sets of unbounded functions. The randomized rules that we construct in Sect. 3 of this paper are seemingly less general than the randomized rules that typically arise in the countably-additive decision theory. Section 4 in the Online Resource demonstrates that the risk set that we construct from our finitely-additive randomizations includes the risk functions for the more general randomizations when all probabilities are countably-additive. Section 5 of the Online Resource presents an interesting property of pointwise convergence that helps to understand why some prior distributions (both finitely-additive and countably-additive) might not have Bayes rules.

2 Finitely-additive expectations

Because risk functions are expected values of loss functions with respect to various probability distributions, it is necessary to understand what we mean by finitely-additive expectations. Our brief introduction here consists of two parts:

– Important features of the countably-additive theory of expectations.
– How we generalize those features in the finitely-additive setting.

More details are provided in Sect. A of the appendix and in the Online Resource.

Let P be a countably-additive probability defined on a σ-field Σ of subsets of a set $\mathcal{Z}$. If f is a bounded real-valued Σ-measurable function on $\mathcal{Z}$, there is a unique definition of the integral $\int_{\mathcal{Z}} f(z)P(dz)$ in terms of the function f and the probability P. (A similar unique definition exists for the integral of a function that is bounded below. Since functions that are bounded below do not form a linear space, special care is needed to deal with them in general.) The integral is also called the expectation of f.

Temporarily identify each set $A \in \Sigma$ with its indicator function

$$I_A(z) = \begin{cases} 1 & \text{if } z \in A, \\ 0 & \text{if } z \in A^C. \end{cases}$$

Then, we can think of Σ as a subset of the linear space $\mathcal{F}$ of bounded Σ-measurable functions on $\mathcal{Z}$. In this way, the integral $\int_{\mathcal{Z}} f(z)P(dz)$ can be thought of as a function $P(f)$ that is a natural extension of the probability P. The domain of P was originally Σ, but we have now extended the domain to $\mathcal{F}$. As a function on $\mathcal{F}$, $P(\cdot)$ is a linear functional with two other key properties:

(i) $P(1) = 1$ (*normalized*), and
(ii) $f \leq g$ implies $P(f) \leq P(g)$ (*monotone*).

Every normalized monotone linear functional can be interpreted as a finitely-additive expectation that extends the domain of a probability P to a linear space of functions. (See Sect. A of the appendix for more details.) What distinguishes countably-additive P from merely finitely-additive P is a third property, which is a form of continuity. (See Royden 1968, Chapter 13 on Daniell integrals for more detail, if desired.) The countably-additive theory allows us to extend the domain of a countably-additive expectation to include functions that are bounded below. The same extension is available for finitely-additive expectations. Since all of our loss and risk functions are bounded below, we will not pursue possible ways to further extend the domain of an expectation to include functions that are unbounded in both directions.

Countably-additive expectations are typically defined on sets of functions that are measurable with respect to σ-fields of subsets of their domain space $\mathcal{Z}$. Except in special cases, they cannot be extended to countably-additive expectations on the set of all functions in $\mathbb{R}^{\mathcal{Z}}$. In contrast, finitely-additive expectations can be extended to arbitrary sets of real-valued functions, although the linearity property must be modified to accommodate some unbounded functions. For example, the set $\mathcal{M}_{\mathcal{Z}}$ of $\widehat{\mathbb{R}}$-valued functions that are bounded below is not a linear space, so not all linear combinations of its elements need to be in the domain of a finitely-additive expectation. Furthermore, two functions f and g might both have infinite expectation. If their difference is well-defined and bounded below, the expectation of $f - g$ can be defined, but we will not need it.

Section A.1 (starting with Definition 8) contains the details on how we handle the fact that $\mathcal{M}_{\mathcal{Z}}$ is not a linear space. We assume that our finitely-additive probabilities on $\mathcal{Z}$ are defined for all elements of $2^{\mathcal{Z}}$, and we assume that our finitely-additive expectations are defined on all of $\mathcal{M}_{\mathcal{Z}}$. These may seem like strong assumptions, but the domain of every finitely-additive probability, no matter how small, can be extended to all of $2^{\mathcal{Z}}$. Similarly every finitely-additive expectation on a set of bounded functions can be extended to $\mathcal{M}_{\mathcal{Z}}$.

In addition, Sect. A of the appendix gives details on how we extend the concept of finitely-additive expectation to domains that include functions that assume the value $+\infty$, producing a meaningful definition of expectation and/or integral with respect to a finitely-additive probability. One feature that distinguishes the countably-additive theory from the finitely-additive theory that we adopt in this paper is the acceptability of non-uniqueness in the extension of a finitely-additive expectation from the domain consisting of bounded functions to the domain that consists of all functions that are bounded below. In the countably-additive theory, the expectation of a function f that is bounded below is the supremum of the expectations of all measurable bounded functions g such that $g \leq f$. The same idea can be applied in the finitely-additive theory to define what we call the minimum extension in Definition 9 in Sect. A.1. However, for each finitely-additive expectation P on the set of bounded functions, there are multiple extensions to the domain of functions that are bounded below. One of them is the minimum extension but there are others. De Finetti (1974, Section 6.54) embraces this non-uniqueness, but does not pursue its consequences very far. Because of this non-uniqueness, we distinguish between what we call a finitely-additive integral and the more general finitely-additive expectations that extend the same finitely-additive probability P.

Throughout the rest of the paper, when we write an integral of a function $f : \mathcal{Z} \to \widehat{\mathbb{R}}$ with respect to a finitely-additive probability P, $\int_{\mathcal{Z}} f(z)P(dz)$, we are referring to the minimum extension of P. General finitely-additive expectations will be denoted with notation such as $Q(f)$. The restriction of each such general expectation to the domain of indicators of subsets $\mathcal{Z}$ is a finitely-additive probability (call it P_Q). If Q happens to be the minimum extension of P_Q, then $P_Q(f) = Q(f) = \int_{\mathcal{Z}} f(z)Q(dz)$ for all f that are bounded below. If Q is not the minimum extension of P_Q, then $Q(f) \geq P_Q(f)$ for all f that are bounded below, and there exists at least one f such that $Q(f) > P_Q(f)$. (See Lemma 4 in Sect. 4.2.)

Finitely-additive expectations will arise in the following places in this paper:

- To calculate the risk functions of non-randomized rules, we use expectations of $\widehat{\mathbb{R}}$-valued functions defined on a data space $\mathcal{X}$. As in the familiar countably-additive decision theory, these expectations, denoted P_θ, are indexed by a parameter θ from a parameter space Θ. Our theory is general enough to allow an arbitrary parameter space, and the P_θ can be either countably-additive or merely finitely-additive.
- To calculate the risk functions of randomized rules, we use expectations of $\widehat{\mathbb{R}}$-valued functions of non-randomized rules. Our theory is general enough to allow an arbitrary action space $\mathcal{A}$, and an arbitrary subset $\mathcal{H}_0 \subseteq \mathcal{A}^{\mathcal{X}}$ of non-randomized rules. The randomized rules are expectations of $\widehat{\mathbb{R}}$-valued functions defined on $\mathcal{H}_0$. These expectations can be either countably-additive or merely finitely-additive.
- To calculate the Bayes risks of decision rules, we use expectations (called priors) of $\widehat{\mathbb{R}}$-valued functions defined on the parameter space Θ. These expectations can be either countably-additive or merely finitely-additive.

Table 1 Notation used throughout this document

Symbol	Meaning
$\mathcal{A}$	The action space
$\mathcal{X}$	The space where available data take their values
Θ	The parameter space that indexes the set of data distributions
$\widehat{\mathbb{R}}$	$\mathbb{R} \cup \{\infty\}$ endowed with the topology generated by all open subsets of $\mathbb{R}$ together with all sets of the form $(c, \infty) \cup \{\infty\}$
$\mathcal{Z}$	A generic set that serves as the domain of a class of $\widehat{\mathbb{R}}$-valued functions
$\mathcal{M}_{\mathcal{Z}}$	The subset of $\widehat{\mathbb{R}}^{\mathcal{Z}}$ consisting of those elements that are bounded below
$L(\cdot, \cdot)$	The loss function mapping $\Theta \times \mathcal{A}$ to $\widehat{\mathbb{R}}$ and bounded below
P_θ	For each $\theta \in \Theta$, P_θ is an expectation on $\mathcal{M}_{\mathcal{X}}$
$\mathcal{P}_{\mathcal{Z}}$	The set of all finitely-additive expectations (see Sect. A of the appendix) on $\mathcal{M}_{\mathcal{Z}}$
$\Lambda_{\mathcal{Z}}$	The subset of $\mathcal{P}_{\mathcal{Z}}$ consisting of those finitely-additive expectations that are minimum extensions (see Definition 9)
$\mathcal{S}_{\mathcal{Z}}$	The subset of $\Lambda_{\mathcal{Z}}$ corresponding to extensions of all simple probability measures on $2^{\mathcal{Z}}$, i.e., those supported on a finite set
$\mathcal{H}_0$	The subset of $\mathcal{A}^{\mathcal{X}}$ consisting of all non-randomized rules available to the decision maker
$R_0(\cdot, \gamma)$	The risk function of a non-randomized rule γ as defined in Definition 1
$R(\cdot, \delta)$	The risk function of a rule δ as defined in Definition 2
$\mathcal{R}$	The set of risk functions for all finitely-additive randomized rules
$\mathcal{R}_1$	The subset of $\mathcal{R}$ whose randomizations are minimum extensions (see Definition 9)

The subscript $\mathcal{Z}$ on symbols like $\mathcal{M}_{\mathcal{Z}}$, $\Lambda_{\mathcal{Z}}$, etc. will vary when referring to a specific set like Θ or $\mathcal{H}_0$

3 General decision theory

Table 1 contains some notation that gets used frequently.

3.1 Decision rules

Definition 1 A *non-randomized rule* is a function $\gamma : \mathcal{X} \to \mathcal{A}$. We denote by $\mathcal{H}_0$ the set of all non-randomized rules that the decision maker has available. The *risk function* of a non-randomized rule γ is the function $R_0(\cdot, \gamma) : \Theta \to \widehat{\mathbb{R}}$ defined by

$$R_0(\theta, \gamma) = P_\theta \left[L(\theta, \gamma(\cdot)) \right], \tag{1}$$

which can be written $\int_{\mathcal{X}} L(\theta, \gamma(x)) P_\theta(dx)$, if P_θ is a minimum extension.

To be clear, for each non-randomized rule γ, each loss function L, and each $\theta \in \Theta$ there is a function $f_{\gamma,L,\theta} : \mathcal{X} \to \widehat{\mathbb{R}}$ defined by

$$f_{\gamma,L,\theta}(x) = L(\theta, \gamma(x)).$$

The risk function of γ is $R_0(\theta, \gamma) = P_\theta(f_{\gamma,L,\theta})$. We see that $R_0(\theta, \gamma)$ is an element of $\mathcal{M}_\Theta$ as a function of θ for fixed γ, and it is an element of $\mathcal{M}_{\mathcal{H}_0}$ as a function of γ for fixed θ.

Definition 2 A *randomized rule* δ is a finitely-additive expectation on $\mathcal{M}_{\mathcal{H}_0}$. In general, its *risk function* is

$$R(\theta, \delta) = \delta[R_0(\theta, \cdot)], \tag{2}$$

which can be written $\int_{\mathcal{H}_0} R_0(\theta, \gamma)\delta(d\gamma)$, if δ is a minimum extension. A randomized rule $\delta \in \mathcal{S}_{\mathcal{M}_{\mathcal{H}_0}}$ is called *simple*.

To be clear, for each loss function L, and each $\theta \in \Theta$ there is a function $g_{L,\theta} : \mathcal{H}_0 \to \widehat{\mathbb{R}}$ defined by

$$g_{L,\theta}(\delta) = R_0(\theta, \delta).$$

The risk function of δ is $R(\theta, \delta) = \delta(g_{L,\theta})$, and is an element of $\mathcal{M}_\Theta$ for fixed δ.

The operational understanding of a randomized rule δ is that a decision maker can select a γ from $\mathcal{H}_0$ using an auxiliary randomization with distribution δ.

We can think of a non-randomized rule γ as the special case of a randomized rule δ with $\delta(f) = f(\gamma)$, for $f \in \mathcal{M}_{\mathcal{H}_0}$. With this understanding, (2) is a generalization of (1). In this case, $R_0(\theta, \gamma) = R(\theta, \delta)$ for all θ.

We are deliberately vague about which elements of $\mathcal{A}^{\mathcal{X}}$ constitute the set $\mathcal{H}_0$. The reason is that our results are general enough to cover whatever set $\mathcal{H}_0$ one wishes to contemplate. One could choose all of $\mathcal{A}^{\mathcal{X}}$, or just those rules that are formal Bayes rules with respect to various priors, or whatever subset of $\mathcal{A}^{\mathcal{X}}$ one wishes. The rules can arise from fixed-sample-size experiments or from sequential experiments. The only structure we assume for the decision problem is that the loss L is bounded below and that all finitely-additive randomizations (as defined in Definition 2) are allowed.

Two types of randomized rules are described by Wald and Wolfowitz (1951). That article refers to the randomizations we defined in Definition 2 as *special randomizations*. A different kind of randomization, that is common in countably-additive decision theory, is called a *general randomization*. A general randomization is a mapping δ from $\mathcal{X}$ to probability distributions over $\mathcal{A}$. In this manner, the randomization performed after observing a particular x can be virtually unrelated to the randomization that is performed after observing each other x. The only requirement is that the following integral be defined and be a measurable function of x:

$$L(\theta, \delta(x)) = \int_{\mathcal{A}} L(\theta, a)\delta(x)(da). \tag{3}$$

Then (3) is inserted directly into (1) to define the risk function of a general randomization δ as

$$R(\theta, \delta) = P_\theta \left[\int_{\mathcal{A}} L(\theta, a)\delta(\cdot)(da) \right]. \tag{4}$$

Special randomizations are special cases of general randomizations. For the case in which all of the P_θ probabilities and all of the general randomizations are countably additive, Section 4 of the Online Resource contains the proof that the risk functions that result from general randomizations are included in the set of risk functions that result from finitely-additive special randomizations. Aside from the restriction to countably additive probabilities, the major difference between general and special randomizations involves the orders in which the expectations are performed. For general randomizations, the inner expectation is over the randomization with the outer expectation over the data distribution P_θ. For special randomizations, the expectation over the data distribution is inside, and the outer expectation is over the randomization. For finitely-additive expectations, the order in which these expectations is computed matters to a larger extent than it does in the countably-additive case. (See Section 2.4 in the Online Resource for more detail.)

3.2 Admissibility, risk sets, Bayes rules, and minimax rules

Dominance is defined in Definition 3 to match the idea that "smaller is better" when comparing risk functions in decision theory.

Definition 3 Let $\mathcal{Z}$ be a set, and let $f, g \in \widehat{\mathbb{R}}^{\mathcal{Z}}$. We say that g *dominates* f if (i) $g(z) \leq f(z)$ for all $z \in \mathcal{Z}$ and (ii) there is z such that $g(z) < f(z)$. The *lower boundary* of a subset A of $\widehat{\mathbb{R}}^{\mathcal{Z}}$, denoted $\partial_L A$, is the set of all functions $f \in \overline{A}$ such that there is no $g \in \overline{A}$ that dominates f. (Here, $\overline{A}$ denotes the closure of A.)

The lower boundary of a set A of functions is defined in terms of the closure of A. If more than one topology is available, one needs to be clear about which closure one is using. The topology of pointwise convergence of functions is always available for a function space since it is the product topology (see Definition 1 in Section 1 of the Online Resource) obtained by identifying each function from a space S_1 to S_2 as an element of $S_2^{S_1}$. If the functions are all bounded, the topology of uniform convergence is also available. Since our results will apply to general function spaces, we will use only the topology of pointwise convergence.

Definition 4 The *risk set* is the set of all risk functions of decision rules. A decision rule δ *dominates* another decision rule δ' if $R(\cdot, \delta)$ dominates $R(\cdot, \delta')$. A rule δ is *admissible* if no other rule dominates δ. A subset $\mathcal{C}$ of the set of all decision rules is called a *complete class* if, for every $\delta' \notin \mathcal{C}$, there exists a $\delta \in \mathcal{C}$ such that δ dominates δ'. A set $\mathcal{C}$ is an *essentially complete class* if, for every $\delta' \notin \mathcal{C}$, there exists $\delta \in \mathcal{C}$ such that $R(\theta, \delta) \leq R(\theta, \delta')$ for all θ. A(n essentially) complete class is *minimal* if no proper subset is (essentially) complete.

In addition to (as well as related to) the risk function, the Bayes risk of a decision rule with respect to one or several prior distributions is important to decision making.

A prior distribution λ over Θ can also be extended (as in Sect. A) to a finitely-additive expectation on $\mathcal{M}_\Theta$, hence we can treat such λ as elements of $\mathcal{P}_\Theta$.

Definition 5 If δ is a randomized rule, the *Bayes risk* of δ with respect to λ is the value

$$r(\lambda, \delta) = \lambda[R(\cdot, \delta)] \tag{5}$$

Let $r_0(\lambda) = \inf_\delta r(\lambda, \delta)$. If

$$r(\lambda, \delta_0) = r_0(\lambda), \tag{6}$$

then δ_0 is called a *Bayes rule* with respect to λ.

An alternative to choosing a rule to minimize the Bayes risk is to choose a rule to minimize the supremum of the risk function.

Definition 6 A rule δ is called a *minimax rule* if

$$\sup_\theta R(\theta, \delta) = \inf_\gamma \sup_\theta R_0(\theta, \gamma). \tag{7}$$

A finitely-additive expectation λ_0 on $\mathcal{M}_\Theta$ is called a *least-favorable prior* if $r_0(\lambda_0) = \sup_\lambda r_0(\lambda)$. The right-hand side of (7) is called the *minimax value* of the decision problem.

4 The risk set

Consider a decision problem with loss function that is bounded below. We assume that the set of all randomized rules, as defined in Definition 2, are available to the decision maker. We denote the risk set $\mathcal{R}$. The set of risk functions for randomized rules with minimum extension randomizations is the smaller set $\mathcal{R}_1$. The following two results, whose proofs are included in this section, are important to the proofs of the main theorems of this paper.

Lemma 1 *The set $\mathcal{R}$ is a compact Hausdorff space.*

Lemma 2 *The lower boundary of $\mathcal{R}$ is the same as the lower boundary of $\mathcal{R}_1$.*

The reason that we care that $\mathcal{R}$ is a Hausdorff space is so that, when we have a convergent net, we know that the limit is unique:

Proposition 1 (Willard 1970: Theorem 13.7, p. 86) *For a convergent net in a Hausdorff space, there is a unique limit that is also the unique cluster point.*

4.1 Proof of Lemma 1

The proof relies on a number of topological results. The first is straightforward and stated without proof.

Proposition 2 *Let $x = \{x_\eta\}_{\eta\in\mathcal{D}}$ be a net in a topological space $\mathcal{T}$ with a subnet $\{y_\gamma\}_{\gamma\in\mathcal{D}'}$. Let $f : \mathcal{T} \to \mathcal{V}$ be a function to another topological space $\mathcal{V}$. Then $\{f(x_\eta)\}_{\eta\in\mathcal{D}}$ is a net in $\mathcal{T}$ with a subnet $\{f(y_\gamma)\}_{\gamma\in\mathcal{D}'}$.*

The following results contain references to well-known sources for their proofs.

Proposition 3 (Dunford and Schwartz 1957: I.5.7(a), p. 17) *A closed subset of a compact space is compact.*

Proposition 4 (Dunford and Schwartz 1957: I.7.2, p. 27 or Kelley 1955: Theorem 2(b), p. 66) *The closure of a set C is the set of all limits of nets in C that converge.*

Proposition 5 (Dunford and Schwartz 1957: I.7.9, p. 29 or Kelley 1955: Theorem 2, p. 136) *A topological space is compact if and only if every net has a cluster point.*

Proposition 6 (Willard 1970: Theorem 11.5 and Example 11.4(e), p. 75) *A net has p as a cluster point if and only if there is a subnet that converges to p. A net converges if and only if every subnet converges to the same limit.*

Proposition 7 (Dunford and Schwartz 1957: I.8.2, p. 32 or Kelley 1955: Theorem 5, p. 92) *The product of a collection of Hausdorff spaces is Hausdorff in the product topology.*

Proposition 8 (Dunford and Schwartz 1957: I.8.5, p. 32 or Kelley 1955: Theorem 13, p. 143) *The product of a collection of compact spaces is compact in the product topology.*

In addition, we need a result that concerns the way that we have defined $\widehat{\mathbb{R}}$.

Lemma 3 *For finite c_0, $[c_0, \infty]$ is a compact Hausdorff subset of $\widehat{\mathbb{R}}$.*

Proof First, we prove that $\widehat{\mathbb{R}}$ is Hausdorff, hence so is every subset. Let $x \neq y \in \widehat{\mathbb{R}}$. If both are finite then the following open intervals $(x-\epsilon, x+\epsilon)$ and $(y-\epsilon, y+\epsilon)$ are disjoint, where $\epsilon < |x-y|/2$. If $x < \infty = y$, then $(x-1, x+1)$ and $(x+2, \infty]$ are disjoint open intervals. Next, we prove that $[c_0, \infty]$ is compact. Let $\mathcal{A} = \{A_\alpha\}_{\alpha\in C}$ be an open cover of $[c_0, \infty]$. Since $\infty \in \widehat{\mathbb{R}}$, there exists $\alpha_0 \in C$ such that A_{α_0} contains a set of the form $(c, \infty]$. For the other elements of $\mathcal{A}$, let $B_\alpha = A_\alpha \backslash \{\infty\}$. Then $\{B_\alpha\}_{\alpha\neq\alpha_0}$ is an open cover of $[c_0, c]$, which has a finite subcover $\{B_{\alpha_1}, \ldots, B_{\alpha_k}\}$, because $[c_0, c]$ is a compact subset of $\mathbb{R}$. Then $\{A_{\alpha_0}, \ldots, A_{\alpha_k}\}$ is a finite subcover of $[c_0, \infty]$ from $\mathcal{A}$. □

Finally, we can prove Lemma 1.

Proof *(Lemma 1)* Let c_0 be a finite lower bound for the loss function. Recall that each element of $\mathcal{R}$ has the form $R(\theta, \delta) = \delta\,[R_0(\theta, \cdot)]$ for some $\delta \in \mathcal{P}_{\mathcal{H}_0}$. Propositions 7 and 8 and Lemma 3 tell us that $[0, \infty]^\Theta$ is a compact Hausdorff space, where each net has at least one cluster point by Proposition 5. Since $\mathcal{R} \subseteq [0, \infty]^\Theta$, we have that $\mathcal{R}$ a Hausdorff space. Using Proposition 3, we need only show that $\mathcal{R}$ is closed. By Proposition 4, we can do this by showing that $\mathcal{R}$ contains the limit of every convergent net in $\mathcal{R}$. Each net in $\mathcal{R}$ can be expressed as $\mathbf{R} = \{R(\cdot, \delta_\eta)\}_{\eta\in\mathcal{D}}$ for some net $\mathbf{D} =$

$\{\delta_\eta\}_{\eta\in\mathcal{D}}$ in $\mathcal{P}_{\mathcal{H}_0}$. Suppose that such a net $\mathbf{R}$ converges to a limit $r(\cdot) \in \mathcal{M}_\Theta$. We need to show that $r \in \mathcal{R}$. Proposition 10 (in Sect. B) says that $\mathcal{P}_{\mathcal{H}_0}$ is a compact set. Hence, the net $\mathbf{D}$ has a cluster point $\delta \in \mathcal{P}_{\mathcal{H}_0}$, which is the limit of a convergent subnet $\mathbf{D}'$. Let $\mathbf{R}' = \{R(\cdot, \beta_\gamma)\}_{\gamma\in\mathcal{D}'}$ be the subnet of $\mathbf{R}$ corresponding to the convergent subnet $\mathbf{D}' = \{\beta_\gamma\}_{\gamma\in\mathcal{D}'}$ of $\mathbf{D}$ (Proposition 2.) Then $\mathbf{R}'$ also converges to r (Proposition 6.) That is, $r(\theta) = \lim_\gamma \beta_\gamma[R_0(\theta, \cdot)]$. Since $\lim_\gamma \beta_\gamma = \delta$, we have that, for all θ, $r(\theta) = \delta[R_0(\theta, \cdot)] = R(\theta, \delta)$, so r is an element of $\mathcal{R}$. □

4.2 Proof of Lemma 2

The proof of Lemma 2 relies on the following result.

Lemma 4 *Let $\mathcal{Z}$ be a set. Let P be a finitely-additive expectation on the bounded elements of $\mathcal{M}_\mathcal{Z}$. Let Q and L be the minimum and maximum extensions of P respectively. (See Definition 9 in Sect. A.1.) Let S be a general extension of P to $\mathcal{M}_\mathcal{Z}$. Then, for all $f \in \mathcal{M}_\mathcal{Z}$, $Q(f) \le S(f) \le L(f)$. If $S \ne Q$, there exists $f \in \mathcal{M}_\mathcal{Z}$ such that $S(f) > Q(f)$.*

Proof To see that $Q(f) \le S(f)$, we know that $S(g) \le S(f)$ for every bounded $g \le f$. Since $S(g) = Q(g) = P(g)$ for every bounded g, the definition of Q implies that $Q(f) \le S(g)$. That $S(f) \le L(f)$ follows from the fact that $L(f) = \infty$ for all f such that $S(f) \ne L(f)$. The final claim is obvious. □

Now, we prove Lemma 2.

Proof *(Lemma 2)* Let $f \in \partial_L\mathcal{R}$. Since $\mathcal{R}$ is closed, f is not dominated by another element of $\mathcal{R}$. Suppose that f is the risk function of a randomized rule δ that is *not* a minimum extension. Let δ' be the minimum extension of the restriction of δ to the bounded elements of $\mathcal{M}_{\mathcal{H}_0}$, and let f' be the corresponding risk function. By Lemma 4, $f' \le f$. Since f is not dominated by f' we must have $f = f'$. Since f is not dominated by another element of $\mathcal{R}$, it is not dominated by another element of $\overline{\mathcal{R}_1}$ (a subset of $\mathcal{R}$). Hence $f \in \partial_L\mathcal{R}_1$.

Finally, let $g \in \partial_L\mathcal{R}_1$. Suppose, to the contrary, that $g \notin \partial_L\mathcal{R}$. Then, there is $g' \in \mathcal{R}$ that dominates g. By the argument in the previous paragraph, there is $h \in \mathcal{R}_1$ such that $h \le g$, hence h dominates g. But this contradicts $g \in \partial_L\mathcal{R}_1$. □

5 A complete class theorem

In this section, we state and prove a general complete class theorem. The proof refers to several results that appear in Sect. B of the appendix (with appropriate forward references) or in Sect. 4.

Theorem 1 (Complete Class Theorem) *The decision rules whose risk functions are on $\partial_L\mathcal{R}$ form a minimal complete class. Each admissible rule is a Bayes rule.*

Proof Let $\mathcal{H}_L$ be the set of decision rules whose risk functions are on $\partial_L\mathcal{R}$. By the definition of $\partial_L\mathcal{R}$, we see that every rule in $\mathcal{H}_\mathcal{L}$ is admissible, and every rule not in

$\mathcal{H}_L$ is inadmissible. No proper subset C of $\mathcal{H}_L$ could be a complete class because each element of $\mathcal{H}_L \backslash C$ is not dominated by an element of C (or by anything else.) What remains to the first claim (about a minimal complete class) is to show that every rule not in $\mathcal{H}_L$ (i.e., is inadmissible) is dominated by a rule in $\mathcal{H}_L$. The proof will proceed in terms of the risk functions rather than the decision rules.

Let $f_0 \in \mathcal{R}\backslash\partial_L\mathcal{R}$, and let $\mathcal{F}_0$ be $\{f_0\}$ union with the set of all elements of $\mathcal{R}$ that dominate f_0. We show next that $\mathcal{F}_0$ is compact. If $\mathcal{F}_0 = \mathcal{R}$, then $\mathcal{F}_0$ is compact because $\mathcal{R}$ is compact by Lemma 1. If $\mathcal{R}\backslash\mathcal{F}_0 \neq \emptyset$, we will show that $\mathcal{F}_0$ is closed and apply Proposition 3. Let $h \in \mathcal{R}\backslash\mathcal{F}_0$. We need to show that h has a neighborhood that is disjoint from $\mathcal{F}_0$. Since $h \neq f_0$ and h does not dominate f_0, there exists θ_0 such that $h(\theta_0) > f_0(\theta_0)$. Let $N = \{g : g(\theta_0) > f_0(\theta_0)\}$, which is a neighborhood of h. It is clear that $N \bigcap \mathcal{F}_0 = \emptyset$, because for every element k of $\mathcal{F}_0$ $k(\theta_0) \leq f_0(\theta_0)$. Since $\mathcal{F}_0$ is compact, every net in $\mathcal{F}_0$ has a convergent subnet (Propositions 5 and 6.)

The bulk of the proof proceeds by transfinite induction. Well-order ($\prec$) the elements of Θ as $\{\theta_\gamma : \gamma \in \Gamma\}$. For $\gamma = 1$, let $\xi_1 = \inf_{g\in\mathcal{F}_0} g(\theta_1)$. Let $\{g_n\}_{n=1}^\infty$ be a sequence (hence a net) of elements of $\mathcal{F}_0$ such that $\lim_{n\to\infty} g_n(\theta_1) = \xi_1$. Proposition 6 says that there is $f_1 \in \mathcal{F}_0$ that is the limit of a convergent subnet. Let $\mathcal{F}_1 = \{g \in \mathcal{F}_0 : g(\theta_1) = \xi_1\}$, which we just showed was non-empty, and which is closed by Proposition 11 (in Sect. B.) We just proved the following induction hypothesis for $\gamma = 1$:

> *Induction hypothesis:* Let $\gamma \in \Gamma$. There is a non-empty closed set $\mathcal{F}_\gamma$ that is a subset of $\mathcal{F}_\eta$ for all $\eta < \gamma$ and a function $f_\gamma \in \mathcal{F}_\gamma$ such that $f_\gamma(\theta_\eta) = \inf_{g\in\mathcal{F}_\eta} g(\theta_\eta)$ for all $\eta \preceq \gamma$.

Let γ be a successor ordinal, and assume that *Induction hypothesis* is true for each $\eta \prec \gamma$. Let $\xi_\gamma = \inf_{g\in\mathcal{F}_{\gamma-1}} g(\theta_\gamma)$. Let $\{g_n\}_{n=1}^\infty$ be a sequence (net) of elements of $\mathcal{F}_{\gamma-1}$ such that $\lim_{n\to\infty} g_n(\theta_\gamma) = \xi_\gamma$. As in the $\gamma = 1$ case above, let $f_\gamma \in \mathcal{F}_{\gamma-1}$ be the limit of a convergent subnet. Let $\mathcal{F}_\gamma = \{g \in \mathcal{F}_{\gamma-1} : g(\theta_\gamma) = \xi_\gamma\}$, which is closed and which we just showed was non-empty. Then *Induction hypothesis* is true for γ.

Next, let γ be a limit ordinal. Define $\mathcal{F}_{\gamma-} = \bigcap_{\eta<\gamma} \mathcal{F}_\eta$. For each $\eta \prec \gamma$, let $h_\eta \in \mathcal{F}_\eta$. By Proposition 6, the net $\{h_\eta\}_{\eta\prec\gamma}$ has a convergent subnet, whose limit we will call $f_{\gamma-}$. By construction, for every $\eta \prec \gamma$, $h_\psi(\theta_\eta)$ is constant in ψ for $\psi \succeq \eta$, hence $f_{\gamma-}(\theta_\eta) = h_\eta(\theta_\eta)$ for every $\eta \prec \gamma$ and $f_{\gamma-} \in \mathcal{F}_{\gamma-}$ which is closed, non-empty, and a subset of every $\mathcal{F}_\eta$ for $\eta \prec \gamma$. Now, apply the successor ordinal argument to γ with $\gamma-$ standing in for $\gamma - 1$ wherever it appears. So, for every $\gamma \in \Gamma$, there is f_γ that satisfies *Induction hypothesis*. Let f be a cluster point of the net $\{f_\gamma\}_{\gamma\in\Gamma}$, which exists by Proposition 5.

Next, we show that f is in the lower boundary. Assume, to the contrary, that f is not in the lower boundary. Then, there exists g and θ_0 such that $g(\theta) \leq f(\theta)$ for all θ and $g(\theta_0) < f(\theta_0)$. Let $\gamma \in \Gamma$ be such that $\theta_\gamma = \theta_0$. Let $\mathcal{F}_* = \mathcal{F}_{\gamma-1}$ if γ is a successor and $\mathcal{F}_* = \mathcal{F}_{\gamma-}$ if γ is a limit ordinal. By construction,

$$f(\theta_0) = f_\gamma(\theta_0) = \inf_{g\in\mathcal{F}_*} g(\theta_0).$$

This contradicts $g(\theta_0) < f(\theta_0)$.

Finally, we show that all admissible rules are Bayes. Let δ be admissible with risk function k, which we just proved is in $\partial_L \mathcal{R}$. Now apply Proposition 12 (in Sect. B) with $G = \{g : g \geq f \text{ for some } f \in \mathcal{R}\}$ and $\mathcal{Z} = \Theta$. □

Theorem 1 together with Lemma 2 provide the following essentially complete class theorem.

Corollary 1 *The set $\mathcal{H}_1$ of decision rules whose randomizations are in $\Lambda_{\mathcal{H}_0}$ and whose risk functions are on $\partial_L \mathcal{R}_1$ is an essentially complete class. Each subset of $\mathcal{H}_1$ that contains one and only one decision rule with each of the risk functions in $\partial_L \mathcal{R}_1$ is a minimal essentially complete class.*

6 A minimax theorem

In this section, we state and prove a general minimax theorem. The proof refers to several results that appear in Sect. B of the appendix (with appropriate forward references) or in Sect. 4.

Theorem 2 (Minimax theorem) *There exist a minimax rule and a least-favorable prior. Every minimax rule is Bayes with respect to the least-favorable prior.*

Proof Let $\mathcal{R}_0$ be the subset of $\mathcal{R}$ consisting of the bounded elements. First, suppose that $\mathcal{R}_0 = \emptyset$. Define λ_0 in two stages. First, let λ_0 be an arbitrary finitely-additive expectation on the subset of $\mathcal{M}_\Theta$ that consists of the bounded functions. The maximum extension of λ_0 from Lemma 10 (in Sect. A.1) has $\lambda_0(f) = \infty$ for all $f \in \mathcal{R}$. Clearly, λ_0 is least-favorable. Every risk function is unbounded if and only if the minimax value of the decision problem is ∞, in which case every rule is minimax. Since $r_0(\lambda_0) = \infty$, every rule is Bayes with respect to λ_0.

For the rest of the proof, suppose that $\mathcal{R}_0 \neq \emptyset$. Without loss of generality, assume that the loss function L is non-negative. Let $f_* \in \mathcal{R}_0$, and let $s_* = \sup_\theta f_*(\theta)$. Let $\mathcal{F}$ be the set of all elements of $\mathcal{M}_\Theta$ that are bounded above by s_*. By Proposition 14 (in Sect. B,) $\mathcal{F}$ is a closed set, so $\mathcal{F} \cap \mathcal{R}$ is closed. Also, note that $\mathcal{F} \cap \mathcal{R} = \mathcal{F} \cap \mathcal{R}_0$. No decision rule outside of $\mathcal{F}$ could be a minimax rule because the supremum of its risk function would be larger than s_*. For each real $s \leq s_*$, let $A_s = \{f \in \mathcal{F} : \sup_\theta f(\theta) \leq s\}$. Let $s_0 = \inf\{s : A_s \cap \mathcal{R}_0 \neq \emptyset\}$. It is clear then that $s_0 = \inf_\delta \sup_\theta R(\theta, \delta)$. Let $\{f_n\}_{n=1}^\infty$ be a sequence of elements of $\mathcal{R}_0 \cap \mathcal{F}$ such that $\sup_\theta f_n(\theta) \leq s_0 + 1/n$. Since $\mathcal{R}_0 \cap \mathcal{F}$ is closed (and compact) Proposition 5 says that there is $f_0 \in \mathcal{R}_0 \cap \mathcal{F}$ that is a cluster point of $\{f_n\}_{n=1}^\infty$ and $\sup_\theta f_0(\theta) \leq s_0$ by Proposition 13 (in Sect. B.) So $\sup_\theta f_0(\theta) = s_0$ and f_0 is the risk function of a minimax rule.

Let

$$G = \left\{ f \in \widehat{\mathbb{R}}^\Theta : f \geq g, \quad \text{for some } g \in \mathcal{R} \right\}.$$

Let $\mathcal{Z} = \Theta$, and let $k(\theta) = s_0$ for all θ. Then the conditions of Proposition 15 (in Sect. B) hold. There exists a finitely-additive expectation λ_0 such that $\lambda_0(k) = s_0$ and $\lambda_0(g) \geq s_0$ for all $g \in \mathcal{R}$. In particular,

$$\lambda_0(f) \ge s_0, \quad \text{for each } f \text{ that is the risk function of a minimax rule.} \tag{8}$$

Because the risk function f of a minimax rule satisfies $\sup_\theta f(\theta) \le s_0$ we also have

$$\lambda(f) \le s_0, \text{ for all } \lambda. \tag{9}$$

In particular, $\lambda_0(f) \le s_0$ which combines with (8) to imply that $\lambda_0(f) = s_0$, and f is Bayes with respect to λ_0. This makes $r_0(\lambda_0) = s_0$. But (9) implies that $r_0(\lambda) \le s_0$ for all λ. This makes λ_0 least-favorable. □

7 Bayes rules

In this section, we present some results on the existence and admissibility of Bayes rules and one result about uniform dominance of rules that fail to be Bayes. The proofs and examples rely on results from the appendix.

Lemma 5 *Let $\lambda \in \mathcal{S}_\Theta$. Then, there is a Bayes rule with respect to λ with risk function on the lower boundary.*

Proof Let $c = \inf_{f \in \mathcal{R}} \lambda(f)$. For each positive integer m, there is $f_m \in \mathcal{R}$ such that $\lambda(f_m) \le c + 1/m$. Since $\mathcal{R}$ is compact, Proposition 5 says that there is a subnet of $\{f_m\}_{m=1}^\infty$ that converges to an element $f \in \mathcal{R}$. Proposition 16 (in Sect. B) says that λ is continuous, so $\lambda(f) = \lim_{n\to\infty} \lambda(f_n) = c$. If f is on the lower boundary, let $g = f$. If f is not on the lower boundary, Theorem 1 says that there is g on the lower boundary that dominates f, hence $\lambda(g) = c$ because it cannot be any lower. Every δ such that $R(\theta, \delta) = g$ is a Bayes rule. □

When the loss function is bounded, we can show that Bayes rules exist with respect to a larger class of priors. In this case, $\mathcal{R} = \mathcal{R}_1$ because all risk functions are integrals over $a \in \mathcal{A}$ of bounded functions of (θ, a).

Lemma 6 *Suppose that the loss function is bounded. Let λ be the minimum extension of a countably-additive discrete probability on Θ. Then, there is a Bayes rule with respect to λ with risk function on the lower boundary.*

Proof Let $c = r_0(\lambda) = \inf_{f \in \mathcal{R}} \lambda(f)$. For each positive integer m, there is $f_m \in \mathcal{R}$ such that $c \le \lambda(f_m) \le c + 1/m$. Since $\mathcal{R}$ is compact, $\{f_m\}_{m=1}^\infty$ has a cluster point $f \in \mathcal{R}$. Lemma 14 (in Sect. A.2) says that there is an ultrafilter $\mathcal{U}$ of subsets of $\mathbb{Z}^+$ with corresponding probability P such that $f(\theta) = \int_{\mathbb{Z}^+} f_m(\theta) P(dm)$. We then have

$$\begin{aligned}
\int_\Theta f(\theta)\lambda(d\theta) &= \int_\Theta \int_{\mathbb{Z}^+} f_m(\theta) P(dm)\lambda(d\theta) \\
&= \int_{\mathbb{Z}^+} \int_\Theta f_m(\theta)\lambda(d\theta) P(dm) \\
&= \int_{\mathbb{Z}^+} \lambda(f_m) P(dm) = c,
\end{aligned}$$

where the second equality follows from Proposition 18 (in Sect. B,) and the final equality follows from Lemma 14 again, this time applied to the net of constant functions $\{\lambda(f_m)\}_{m=1}^{\infty}$ which has only one cluster point, c. If f is on the lower boundary, let $g = f$. If f is not on the lower boundary, Theorem 1 says that there is g on the lower boundary that dominates f, hence $\lambda(g) = c$ because it cannot be any lower. Every δ such that $R(\theta, \delta) = g$ is a Bayes rule. □

Lemma 7 *If a Bayes rule with respect to* λ *exists, then there is an admissible Bayes rule with respect to* λ.

Proof Let δ_1 be a Bayes rule with respect to λ with risk function f_1. Then $r_0(\lambda) = r(\lambda, \delta_1)$. If $f_1 \in \partial_L \mathcal{R}$, then f_1 is admissible. If $f_1 \notin \partial_L \mathcal{R}$, Theorem 1 shows that there is $f_2 \in \partial_L \mathcal{R}$ that dominates f_1. So that $f_2 \leq f_1$,

$$r_0(\lambda) \leq r(\lambda, f_2) \leq r(\lambda, f_1) = r_0(\lambda),$$

and f_2 is also Bayes with respect to λ. □

The one theorem conspicuous by its absence is one that says that every prior has a Bayes rule. The following is a counterexample to the missing theorem.

Example 2 Let $\Theta = (0, 1)$, and let $\mathcal{A}$ be the set of all non-empty open sub-intervals of Θ. Let the loss function be $L(\theta, a) = I_a(\theta)/|a| + 0.1(1 - |a|)$, where $|a|$ is the length of the interval a. Consider the prior λ_0 which is the minimum extension of the probability corresponding to an ultrafilter containing sets of the form $(0, b)$ for $b \in (0, 1)$. Such an ultrafilter exists by Proposition 9 (in Sect. A.2.) Also, consider the sequence of non-randomized rules $\{\gamma_n\}_{n=2}^{\infty}$ with $\gamma_n = (1/n, 1)$. The Bayes risk of each such rule is $r(\lambda_0, \gamma_n) = 0.1/n$, and the infimum of the Bayes risks is 0, which is clearly the infimum of the Bayes risks over all decision rules. The remainder of this example is devoted to showing that no decision rule has Bayes risk equal to 0. In light of Corollary 1, we can restrict attention to rules whose randomizations are in $\Lambda_{\mathcal{H}_0}$. Proposition 17 (in Sect. B) says that every such decision rule δ_B has a risk function $R(\cdot, \delta_B)$ that is the limit of some convergent net $\{R(\cdot, \delta_\eta)\}_{\eta \in \mathcal{D}}$ of risk functions of simple randomized rules. Each element of the convergent net has the form

$$R(\theta, \delta_\eta) = \sum_{j=1}^{n_\eta} \frac{\alpha_{\eta,j}}{b_{\eta,j} - a_{\eta,j}} I_{(a_{\eta,j}, b_{\eta,j})}(\theta) + 0.1 \sum_{j=1}^{n_\eta} \alpha_{\eta,j}(1 - b_{\eta,j} + a_{\eta,j}), \tag{10}$$

where n_η is the number of components of the simple randomization δ_η, the jth component is $(a_{\eta,j}, b_{\eta,j})$, and the $\alpha_{\eta,j}$ are the coefficients in the convex combination. Let $f_\eta(\theta)$ denote the first sum on the right-hand side of (10), and let ℓ_η denote the second sum (together with the 0.1 factor.) Without loss of generality, we can assume that $\ell = \lim_\eta \ell_\eta$ exists and $f(\theta) = \lim_\eta f_\eta(\theta)$ exists for all θ. (If not, there exists a subnet of $\{\ell_\eta\}_{\eta \in \mathcal{D}}$ that converges and then the corresponding subnet of $\{f_\eta(\cdot)\}_{\eta \in \mathcal{D}}$ converges pointwise.) Then $R(\theta, \delta_B) = f(\theta) + \ell$. Since the ultrafilter that defines

λ_0 contains all sets of the form $(0, b)$ and others, Lemma 13 (in Sect. A.2) says that $\lambda_0(f) \geq \liminf_{\theta\downarrow 0} f(\theta)$. The Bayes risk of δ_B with respect to λ_0 then satisfies

$$r(\lambda_0, \delta_B) \geq \liminf_{\theta\downarrow 0} f(\theta) + \ell. \tag{11}$$

We will prove that the right-hand side of (11) is strictly positive.

If $\ell > 0$, then the right-hand side (11) is strictly positive. For the rest of the proof, assume that $\ell = 0$. Let $\theta \in (0, 1/2)$. Since $\ell = 0$, we know there exists η_θ such that $\eta \geq \eta_\theta$ implies $\ell_\eta < \theta/20$. It follows that $\sum_{j=1}^{n_\eta} \alpha_{\eta,j}(b_{\eta,j} - a_{\eta,j}) > 1 - \theta/2$, for $\eta \geq \eta_\theta$. Let $J = \{j : (b_{\eta,j} - a_{\eta,j}) > 1 - \theta\}$, which is clearly non-empty. Also, $\sum_{j\in J} \alpha_{\eta,j} \geq 1/2$, and $\theta \in (a_{\eta,j}, b_{\eta,j})$ for $j \in J$. Since $I_{(a_{\eta,j},b_{\eta,j})}(\theta)/(b_{\eta,j} - a_{\eta,j}) \geq 1$ for $j \in J$, $f_\eta(\theta) \geq 1/2$ for $\eta > \eta_\theta$. It follows that $f(\theta) \geq 1/2$ for $\theta \in (0, 1/2)$, and the right-hand side of (11) is strictly positive.

The failure of some priors to have Bayes rules may seem puzzling. For each prior λ, there are rules that have Bayes risk arbitrarily close to $r_0(\lambda)$, and the lower boundary is contained in every risk set. The problem is that not all finitely-additive expectations are continuous functions on $\mathcal{M}_\Theta$. The lack of continuity is due not to the finite-additivity but rather to the large number of nets that converge in the pointwise topology. By contrast, in the topology of uniform convergence, all finitely-additive expectations are continuous. We explore the relationship between priors and pointwise convergence in more detail in Section 5 of the Online Resource.

The following is a generalization of a theorem of Pearce (1984) which was restricted to finite action and parameter spaces. It is also an extension of Heath and Sudderth (1978, Theorem 2) to cases in which it matters in which order the data and parameter integrals are performed. For a different generalization of Pearce (1984) applied to ambiguity aversion, see Battigalli (2016, Lemma 1).

Lemma 8 *Assume that the there is a rule δ_0 with bounded risk function such that $r(\lambda, \delta_0) > r_0(\lambda)$ for every λ. That is, there is no λ such that δ_0 is Bayes with respect to λ. Then there is a rule δ_1 and $\epsilon > 0$ such that $R(\theta, \delta_0) > R(\theta, \delta_1) + \epsilon$ for all θ.*

Proof Consider a new decision problem with loss $L'(\theta, a) = L(\theta, a) - R(\theta, \delta_0)$, so that L' is bounded below. Let R' be the new risk function, so that

$$R'(\theta, \delta) = R(\theta, \delta) - R(\theta, \delta_0), \tag{12}$$

for all δ and θ. In this new problem, there is no λ such that δ_0 is a Bayes rule. Also $R'(\theta, \delta_0) = 0$ for all θ. Theorem 2 applies, and there is a minimax rule δ_1. and a least-favorable prior λ_1. Let $\epsilon = -\sup_\theta R'(\theta, \delta_1)$. Since δ_0 is not Bayes with respect λ_1, it is not minimax. Since $\sup_\theta R'(\theta, \delta_0) = 0$, it follows that

$$-\epsilon = \sup_\theta R'(\theta, \delta_1) < 0,$$

So, $R'(\theta, \delta_1) \leq -\epsilon < 0$ for all θ. It follows that $\epsilon > 0$ and

$$-\epsilon \geq R(\theta, \delta_1) - R(\theta, \delta_0),$$

for all θ. So $R(\theta, \delta_0) \geq R(\theta, \delta_1) + \epsilon$ for all θ. □

8 Discussion

We consider a general decision problem with general loss function. The only assumption that we make about the loss function is that it is bounded below. The loss is allowed to take the value ∞. We put no restrictions on the distributions of the data that might be observed and used in decision rules. We allow all finitely-additive and countably-additive randomized rules. We prove that the risk set contains its lower boundary, and we prove general complete class and minimax theorems. In particular, all decision problems have a minimal complete class consisting of Bayes rules and all decision problems have minimax rules and a least-favorable prior.

We have defined the various parts of the decision problem to match the classical countably-additive theory as closely as we could. In particular, we defined a risk function for each randomized rule before introducing prior distributions. The Bayes risk then becomes the expectation of the risk function with respect to the prior. For non-randomized rules, we defined the risk function as the expectation of the loss function with respect to the distribution of the data. We then defined the risk of a randomized rule to be the expectation of the risk functions of non-randomized rules with respect to a finitely-additive expectation over the set $\mathcal{H}_0$ of non-randomized rules. This type of randomization is what Wald and Wolfowitz (1951) called a special randomization in the countably-additive setting. Special randomizations differ from general randomizations where the loss function is redefined for randomized rules by integrating the loss function for a specific data value x with respect to an x-specific randomization over $\mathcal{A}$. We show that our risk set for special randomizations contains all of the risk functions that result from general randomizations in the countably-additive setting (Lemma 19 in Section 4 of the Online Resource.)

We have done no a posteriori Bayesian analysis. That is, we have computed neither posterior distributions of parameters nor formal Bayes rules, which minimize posterior risk. In fact, we have made no concessions to a Bayesian who might wish to start with a general finitely-additive joint distribution of data and parameter, as Heath and Sudderth (1972, 1978) do. Instead, we require that the data integral (given θ) be performed before the parameter integral (prior). A Bayesian with a general finitely-additive joint distribution of data and parameter might require that the integration be performed differently. Example 5 in Section 2.4 of the Online Resource illustrates how the order in which one does integrals can make a difference with finitely-additive distributions even when the the integrands are positive and/or bounded. This is in sharp contrast to the countably-additive theory in which Fubini and Tonelli theorems apply. In addition, decision problems involve a third integral over the action space, so there are six possible orders in which integrals could be computed. The one we have chosen is (as we mentioned above) as close as we could come to the order used in the classical countably-additive approach to decision theory, and it accommodates all of the risk functions that are available in the countably-additive theory.

Another fundamental difference between finitely-additive and countably-additive probabilities and expectations is their domains of definition. Countably-additive probabilities are traditionally defined on σ-fields of subsets of their underlying spaces and their associated expectation operators are defined on sets of functions that are measurable with respect to that same σ-field. A countably-additive probability can be extended in some cases beyond the σ-field on which it was originally defined. (Example 5 in Section 2.4 of the Online Resource contains an example of such an extension.) Such extensions, beyond the measure completion of a probability, are rarely studied in detail due to the fact that they are not unique and fail to admit regular conditional probabilities. (See Seidenfeld et al. 2001, Corollary 1.) Finitely-additive probabilities can be defined on arbitrary collections of subsets of their underlying spaces, including the power sets. The associated finitely-additive expectations can be defined on arbitrary sets of real-valued functions. Measurability of sets and/or functions is rarely an issue except when proving that some extension is unique.

Acknowledgements The authors wish to offer thanks to two excellent reviewers and the editors for their invaluable help and limitless patience in the processing of this article.

A Finitely-additive integrals and expectations

A.1 Unbounded and infinite-valued functions

In de Finetti (1974), de Finetti laid out the theory of coherent previsions for bounded random variables, which he related to finitely-additive probability. De Finetti's definition of coherent prevision is the following.

Definition 7 Let $\mathcal{F}$ be a set of bounded functions defined on a space $\mathcal{Z}$. For each $f \in \mathcal{F}$, let $P(f)$ be a real number. De Finetti called $\{P(f) : f \in \mathcal{F}\}$ *coherent previsions* for $\mathcal{F}$ if, for every finite integer n, every n-tuple $(f_1, \ldots, f_n) \in \mathcal{F}^n$, and every n-tuple $(\alpha_1, \ldots, \alpha_n)$ of real numbers,

$$\sup_z \sum_{j=1}^{n} \alpha_j [f_j(z) - P(f_j)] \geq 0. \tag{13}$$

If $\mathcal{F}$ consists solely of indicator functions of events, coherent previsions for $\mathcal{F}$ can be shown to form a finitely-additive probability. (See de Finetti 1974, Chapter 3 for a long-winded, but elementary treatment.) Also, a finitely-additive probability can be defined on an arbitrary collection of subsets of a general set $\mathcal{Z}$, including the power set. A number of theorems in Bhaskara Rao and Bhaskara Rao (1983, Chapter 3) show how to extend a partially defined finitely-additive probability to arbitrary larger domains. Theorems 3.2.9 and 3.2.10 on pages 69–70 of Bhaskara Rao and Bhaskara Rao (1983) are very general. For this reason, measurability conditions are often not included in theorems about finitely-additive probabilities. In this paper, we assume that each finitely-additive probability is defined on $2^{\mathcal{Z}}$. If the reader starts with a finitely-additive probability P defined on a smaller domain, our results will apply to every extension of P to $2^{\mathcal{Z}}$.

Extending a finitely-additive probability P defined on $2^{\mathcal{Z}}$ to a coherent prevision on the set of all bounded functions in $\mathcal{M}_{\mathcal{Z}}$ is straightforward. For each simple function $f = \sum_{j=1}^{n} \alpha_j I_{A_j}$ on $\mathcal{Z}$ (with all α_j finite), there is a unique value for the coherent previson $P(f) = \sum_{j=1}^{n} \alpha_j P(A_j)$. (See Bhaskara Rao and Bhaskara Rao 1983, Proposition 4.4.2 on page 97.) Because every bounded function is uniformly approximable both above and below by simple functions, for each bounded function f on $\mathcal{Z}$, there is a unique value for the finitely-additive expectation

$$P(f) = \sup_{\text{simple } g \leq f} P(g) = \inf_{\text{simple } g \geq f} P(g).$$

In this paper, we need an extension of the finitely-additive theory from bounded functions both to unbounded functions, as was done by Schervish et al. (2014), and to functions that assume the value ∞ as well. The extension to unbounded functions in Schervish et al. (2014) generalizes the concept of coherence to allow $P(f) = \infty$ without requiring the symbol ∞ to appear in (13). They then prove in Schervish et al. (2014, Definition 7, Lemmas 4 and 6) that previsions are coherent if and only if they are the values of a normalized monotone linear functional. The assumption that the loss function is bounded below allows us to avoid dealing with functions that both assume the value ∞ and are unbounded below. This has the added benefit of allowing us to avoid arithmetic that leads to $\infty - \infty$.

Because the set $\mathcal{M}_{\mathcal{Z}}$ of functions that are bounded below is not a linear space, we need to generalize the concept of normalized monotone linear functional.

Definition 8 Let $\mathcal{Z}$ be a space, and let $\mathcal{F}$ be a subset of $\mathcal{M}_{\mathcal{Z}}$ that contains all constant functions and has the following restricted linearity property:

$$\text{If } f, g \in \mathcal{F} \text{ and } h = \alpha f + \beta g \in \mathcal{M}_{\mathcal{Z}} \text{ with } \alpha, \beta \in \mathbb{R}, \text{ then } h \in \mathcal{F}. \tag{14}$$

We call such a set $\mathcal{F}$ a *restricted-linear space.* If $\mathcal{F}$ is a restricted-linear space, a function $L : \mathcal{F} \to \widehat{\mathbb{R}}$ that satisfies

$$L(\alpha f + \beta g) = \alpha L(f) + \beta L(g), \tag{15}$$

whenever $f, g, \alpha f + \beta g \in \mathcal{F}$ and the arithmetic on the right-hand side of (15) is well defined (i.e., does not involve $\infty - \infty$,) is called *a restricted-linear functional*. A restricted-linear functional is called *monotone* if $f \leq g$ implies $L(f) \leq L(g)$. A monotone restricted-linear functional is called a *finitely-additive Daniell integral*. If $L(1) = 1$, we say that L is *normalized*. A normalized finitely-additive Daniell integral is called a *finitely-additive expectation*.

Note that $\mathcal{M}_{\mathcal{Z}}$ is a restricted-linear space, as is the set of bounded functions. The unique coherent prevision on the bounded functions constructed above from a finitely-additive probability on $2^{\mathcal{Z}}$ is an example of a finitely-additive expectation.

The various conditions in Definition 8 prevent $\infty - \infty$ from appearing on either side of (15). Note that one cannot have α and β both negative in (14) unless both f and g are bounded. If at least one of the functions f, g is unbounded in (14), then at least one of the unbounded functions must have a positive coefficient in order for the linear combination to be well-defined and bounded below. Examples of each of the following situations arise, and they are the reason that we do not enforce (15) when the arithmetic on the right-hand side involves $\infty - \infty$:

- The difference between two functions that are unbounded above can be bounded below.
- The difference of two functions, each with infinite finitely-additive expectation, can have finite expectation.

Although there is no unique extension of a finitely-additive expectation from the set of bounded functions to $\mathcal{M}_{\mathcal{Z}}$, there are two special extensions that exist and prove useful. Lemmas 9 and 10 apply to unbounded and infinite-valued functions. Equation 1.2 of Heath and Sudderth (1978) states a result like Lemma 9 without proof and without being explicit about the fact that the extension might take the value ∞. The restriction to functions that are bounded below is important (even for the result of Heath and Sudderth 1978) in order to avoid $\infty - \infty$.

Lemma 9 *Let $\mathcal{J}$ be the set of all bounded real-valued functions defined on a set $\mathcal{Z}$, and let P be a finitely-additive expectation on $\mathcal{J}$. For each $f \in \mathcal{M}_{\mathcal{Z}}$, define*

$$Q(f) = \sup_{g \in \mathcal{J},\, g \le f} P(g).$$

Then $Q = P$ on $\mathcal{J}$, and Q is a finitely-additive expectation on $\mathcal{M}_{\mathcal{Z}}$.

Proof Because P is a finitely-additive expectation, $g \le f$ implies $P(g) \le P(f)$ for $f, g \in \mathcal{J}$. Hence $Q(f) = P(f)$ for $f \in \mathcal{J}$.

Next, we show that Q is restricted-linear. Let $f_1, f_2 \in \mathcal{M}_{\mathcal{Z}}$, $\alpha_1, \alpha_2 \in \mathbb{R}$, and $\alpha_1 f_1 + \alpha_2 f_2 \in \mathcal{M}_{\mathcal{Z}}$. We need to show that $Q(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 Q(f_1) + \alpha_2 Q(f_2)$ whenever the arithmetic on the right-hand side is well-defined. Since $Q(\alpha_j f_j) = \alpha_j Q(f_j)$ if $\alpha_j > 0$, there is no loss of generality in assuming that α_1, α_2 are each either 1 or -1. First, we show that $Q(f_1 + f_2) = Q(f_1) + Q(f_2)$ for $f_1, f_2 \in \mathcal{M}_{\mathcal{Z}}$. Notice that $Q(f) = \lim_{m\to\infty} P(f \wedge m)$ because every bounded $g \le f$ is bounded above by $f \wedge m$ where $m = \sup_z g(z)$. Then notice that, for all m,

$$(f_1 \wedge m/2) + (f_2 \wedge m/2) \le (f_1 + f_2) \wedge m \le (f_1 \wedge m) + (f_2 \wedge m).$$

The limits of the left-hand and right-hand expressions are both $Q(f_1) + Q(f_2)$ while the limit of the middle expression is $Q(f_1 + f_2)$. To complete the proof that Q is restricted-linear, we need to show that $Q(f_1 - f_2) = Q(f_1) - Q(f_2)$ if $f_1 - f_2 \in \mathcal{M}_{\mathcal{Z}}$ and $Q(f_2)$ is finite. What we just proved implies that

$$Q(f_1) = Q(f_1 - f_2 + f_2) = Q(f_1 - f_2) + Q(f_2). \tag{16}$$

Since $Q(f_2)$ is finite, we can subtract it from both sides of (16) to complete this part of the proof.

Next, we show that Q is monotone. Let $f, g \in \mathcal{M}_{\mathcal{Z}}$ with $f \leq g$. We need to show that $Q(f) \leq Q(g)$. Define

$$g'(z) = \begin{cases} g(z) - f(z) & \text{if } f(z) < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Then $g = f + g'$, $g' \geq 0$, and $g' \in \mathcal{M}_{\mathcal{Z}}$. Hence, $0 \leq Q(g')$ from the definition of Q. From the previous part of the proof,

$$Q(g) = Q(f + g') = Q(f) + Q(g') \geq Q(f).$$

□

At the other extreme from Lemma 9, we have the following alternative extension of a finitely-additive expectation from bounded functions to $\mathcal{M}_{\mathcal{Z}}$.

Lemma 10 *Assume the conditions of Lemma 9. Define L on $\mathcal{M}_{\mathcal{Z}}$ as follows: $L(f) = P(f)$ if $f \in \mathcal{J}$, and $L(f) = \infty$ if $f \in \mathcal{M}_{\mathcal{Z}} \setminus \mathcal{J}$. Then L is a finitely-additive expectation on $\mathcal{M}_{\mathcal{Z}}$ that extends P.*

Proof Clearly, L extends P. Also, L is monotone since every instance of $f \leq g$ either has both $f, g \in \mathcal{J}$ or $L(g) = \infty$. For (15), if either α or β is 0, the equation holds. When both α and β are nonzero, consider four cases:

(i) $\alpha f, \beta g \in \mathcal{J}$. In this case both sides of (15) are the same because P is linear on $\mathcal{J}$.
(ii) $\alpha f \in \mathcal{J}$ and $\beta g \notin \mathcal{J}$. In this case, $\beta > 0$ so that both sides of (15) are ∞.
(iii) $\alpha f \notin \mathcal{J}$ and $\beta g \in \mathcal{J}$. In this case, $\alpha > 0$ so that both sides of (15) are ∞.
(iv) $\alpha f, \beta g \notin \mathcal{J}$. In this case, at least one of α or β must be positive. If α and β are both positive, both sides of (15) are ∞. If one of them is negative, the right-hand side of (15) is $\infty - \infty$, and (15) has no force.

□

Definition 9 We refer to Q in Lemma 9 as the *minimum extension of P*. If a prevision Q is the minimum extension of its restriction to the bounded random variables, then we say that *Q is a minimum extension*. We refer to the L in Lemma 10 as the *maximum extension of P*. We also use the notation $\int_{\mathcal{Z}} f(z)P(dz)$ to stand for $Q(f)$ because $Q(f)$ is a finitely-additive Daniell integral and is determined uniquely from P and f.

The terms "minimum extension" and "maximum extension" are used for Q and L respectively because they assign the minimum and maximum of all possible finitely-additive expectations that extend P to $\mathcal{M}_{\mathcal{Z}}$.

It is also straightforward to see that the measure-theoretic definition of expectation with respect to a countably-additive probability P is the restriction of a minimum extension to the functions in $\mathcal{M}_{\mathcal{Z}}$ that are measurable with respect to the σ-field on which P is defined. Nothing in the countably-additive theory corresponds to the maximum extension. There are extensions that are neither minimum nor maximum, but these cannot generally be constructed in one step, like we did for the minimum and maximum extensions.

The following is a generalization of a well-know measure-theoretic result.

Lemma 11 *Let $\mathcal{Z}$ and $\mathcal{Y}$ be sets. Let P be a finitely-additive expectation on $\mathcal{M}_{\mathcal{Y}}$. Let $h : \mathcal{Y} \to \mathcal{Z}$ be a function. Then P' defined by $P'(f) = P[f(h)]$ is a finitely-additive expectation defined on $\mathcal{M}_{\mathcal{Z}}$ which we call* the finitely-additive expectation induced by *h.*

Proof Clearly $P' \in \widehat{\mathbb{R}}^{\mathcal{M}_{\mathcal{Z}}}$, and $P'(1) = P(1) = 1$. If $f \leq g$, then $f(h) \leq g(h)$ and $P'(f) \leq P'(g)$. We need to show that P' is extended-linear. Assume that $f, g, \alpha f + \beta g \in \mathcal{M}_{\mathcal{Z}}$ and $\alpha P'(f)$ and $\beta P'(g)$ are not both infinite of opposite signs. Then $\alpha P[f(h)]$ and $\beta P[g(h)]$ are not both infinite of opposite signs and $P[\alpha f(h) + \beta g(h)] = \alpha P[f(h)] + \beta P[g(h)]$, which implies $P'(\alpha f + \beta g) = \alpha P'(f) + \beta P'(g)$. □

A.2 Ultrafilter probabilities

Definition 10 Let S be a set and let $\mathcal{U}$ be a non-empty collection of subsets of S. We call $\mathcal{U}$ an *ultrafilter* on S if (i) $A \in \mathcal{U}$ and $A \subseteq B$ implies $B \in \mathcal{U}$, (ii) $A, B \in \mathcal{U}$ implies $A \cap B \in \mathcal{U}$, and (iii) for every $A \subseteq S$, $A \in \mathcal{U}$ if and only if $A^C \notin \mathcal{U}$. An ultrafilter on S is *principal* if there exists $s \in S$ such that $\mathcal{U}$ consists of all subset of S that contain s. Such an s is called the *atom* of $\mathcal{U}$. Other ultrafilters are called *non-principal*.

Proposition 9 gives general conditions under which ultrafilters exist. It requires a definition first.

Definition 11 A collection $\mathcal{F}$ of subsets of a set $\mathcal{Z}$ has *the finite-intersection property* if every finite subcollection has non-empty intersection.

Proposition 9 (Comfort and Negrepontis 1974: Special case of Theorem 2.18, p. 39) *Let $\mathcal{Z}$ be a set, and let $\mathcal{F}$ be a collection of subsets. If $\mathcal{F}$ has the finite intersection property, then there is an ultrafilter on $\mathcal{Z}$ that contains $\mathcal{F}$.*

We have two main uses for ultrafilters in this paper. One is to use them as examples of merely finitely-additive probabilities via Lemma 12 below. The other is to exploit the connection between limits of nets and integrals with respect to ultrafilter probabilities defined on directed sets via Lemma 14 below.

There is a correspondence between ultrafilters and 0–1-valued probabilities.

Lemma 12 *Let $\mathcal{Z}$ be a set, and let $\mathcal{F}$ be a field of subsets of $\mathcal{Z}$. A finitely-additive probability P defined on $\mathcal{F}$ takes only the values 0 and 1 if and only if (i) P can be extended to P' defined on $2^{\mathcal{Z}}$ and (ii) there is an ultrafilter $\mathcal{U}$ of subsets of $\mathcal{Z}$ such that $P'(E) = 1$ if and only if $E \in \mathcal{U}$.*

Proof For the "if" direction, the restriction of P' to $\mathcal{F}$ takes only the values 0 and 1. For the "only if" direction, $\mathcal{V} = \{E \in \mathcal{F} : P(E) = 1\}$ has the finite-intersection property, hence there is an ultrafilter $\mathcal{U}$ of subsets of $\mathcal{Z}$ such $\mathcal{V} \subseteq \mathcal{U}$. Define $P'(E) = 1$ if $E \in \mathcal{U}$ and $P'(E) = 0$ if $E \notin \mathcal{U}$. It is clear that P' is finitely-additive on $2^{\mathcal{Z}}$. □

Definition 12 Let $\mathcal{U}$ be an ultrafilter of subsets of some set $\mathcal{Z}$. We call the probability P defined by $P(E) = 1$ if $E \in \mathcal{U}$ and $P(E) = 0$ if $E \notin \mathcal{U}$ the *probability corresponding to* $\mathcal{U}$.

Lemma 13 *Let $\mathcal{D}$ be a set. Let $\mathcal{U}$ be an ultrafilter of subsets of $\mathcal{D}$. Let P be the minimum extension of the probability on $\mathcal{D}$ corresponding to $\mathcal{U}$. Then, for each $f \in \mathcal{M}_{\mathcal{D}}$,*

$$\int_{\mathcal{D}} f(\eta)P(d\eta) = \sup_{B\in\mathcal{U}} \inf_{\eta\in B} f(\eta) = \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f(\eta).$$

Proof Let $k_0 = \int_{\mathcal{D}} f(\eta)P(d\eta)$. Since (i) $P(B) = 1$ for all $B \in \mathcal{U}$, (ii) P is a monotone restricted-linear functional, and (iii) P is a minimum extension, $k_0 \le \sup_{\eta\in B} f(\eta)$, for all $B \in \mathcal{U}$. Similarly, $k_0 \ge \inf_{\eta\in B} f(\eta)$, for all $B \in \mathcal{U}$. It follows that

$$\sup_{B\in\mathcal{U}} \inf_{\eta\in B} f(\eta) \le k_0 \le \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f(\eta). \tag{17}$$

If the two endpoints, call them $a \le b$, of (17) are not equal, then for each $c \in (a, b)$ precisely one of $\{\eta : f(\eta) \le c\}$ or $\{\eta : f(\eta) > c\}$ is in $\mathcal{U}$. If it is the first of these, it contradicts $b = \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f(\eta)$. If it is the second one, it contradicts $a = \sup_{B\in\mathcal{U}} \inf_{\eta\in B} f(\eta)$. Hence $a = b$, and both inequalities in (17) are equality. □

The final result requires a definition first.

Definition 13 Let $x = \{x_\eta\}_{\eta\in\mathcal{D}}$ be a net in a topological space $\mathcal{T}$. We call a net $y = \{y_\gamma\}_{\gamma\in\mathcal{D}'}$ a *subnet* of x if there exists a function $h : \mathcal{D}' \to \mathcal{D}$ with the following properties: (i) $y_\gamma = x_{h(\gamma)}$ for all $\gamma \in \mathcal{D}'$, (ii) $\gamma_1 \le_{\mathcal{D}'} \gamma_2$ implies $h(\gamma_1) \le_{\mathcal{D}} h(\gamma_2)$ and (iii) for every $\eta \in \mathcal{D}$ there exists $\gamma \in \mathcal{D}'$ such that $h(\gamma) \ge_{\mathcal{D}} \eta$. A *cluster point* of x is a point $p \in \mathcal{T}$ such that, for every neighborhood N of p and every $\eta \in \mathcal{D}$, there exists $\eta' \in \mathcal{D}$ such that $\eta' \ge_{\mathcal{D}} \eta$ and $x_{\eta'} \in N$.

See Section 1 of the Online Resource for more detail about nets and subnets.

Lemma 14 *Let $\mathcal{Z}$ be a set. Let $\mathcal{D}$ be a directed set and let $f = \{f_\eta\}_{\eta\in\mathcal{D}}$ be a net in $\mathcal{M}_{\mathcal{Z}}$. An element g of $\widehat{\mathbb{R}}^{\mathcal{Z}}$ is a cluster point of f if and only if there is an ultrafilter $\mathcal{U}$ that contains all tails of $\mathcal{D}$ whose corresponding probability on $\mathcal{D}$ has minimum extension P such that $g(z) = \int_{\mathcal{D}} f_\eta(z)P(d\eta)$ for all $z \in \mathcal{Z}$.*

Proof For the "only if" direction, assume first that f converges to g. Let $\mathcal{U}$ be any ultrafilter that contains all tails of $\mathcal{D}$, and let P be the minimum extension of the corresponding probability on $\mathcal{D}$. Let $h(z) = \int_{\mathcal{D}} f_\eta(z)P(d\eta)$ for each $z \in \mathcal{Z}$. We need to show that, for each $z \in \mathcal{Z}$ and each neighborhood N of $g(z)$, $h(z) \in N$. Let $z \in \mathcal{Z}$. Lemma 13, applied to ultrafilters of subsets of $\mathcal{D}$, says that

$$\sup_{B\in\mathcal{U}} \inf_{\eta\in B} f_\eta(z) = h(z) = \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f_\eta(z). \tag{18}$$

If $g(z)$ is finite, let $\epsilon > 0$, and let N be the interval $(g(z) - \epsilon, g(z) + \epsilon)$. Because f converges to g, there exists $\eta \in \mathcal{D}$ such that $f_\beta(z) \in N$ for all $\beta \ge_{\mathcal{D}} \eta$. Let $B =$

$A_\eta = \{\beta \in \mathcal{D} : \eta \le_{\mathcal{D}} \beta\}$, which is in $\mathcal{U}$. It follows that the left and right sides of (18) are respectively at least $g(z) - \epsilon$ and at most $g(z) + \epsilon$. Hence $h(z) \in N$. If $g(z) = \infty$, let $N = (c, \infty]$. Then there exists $\eta \in \mathcal{D}$ such that $f_\beta(z) > c$ for all $\beta \ge_{\mathcal{D}} \eta$. Let $B = A_\eta$, which is in $\mathcal{U}$. It follows that the left side of (18) is at least c, so that $h(z) \in N$. Next, assume that g is merely a cluster point of f. By Proposition 6, there exists a subnet $f' = \{r_\tau\}_{\tau \in \mathcal{D}'}$ that converges to g. The previous argument shows that for every ultrafilter $\mathcal{U}'$ on $\mathcal{D}'$ that contains all tails of $\mathcal{D}'$, $g(z) = \int_{\mathcal{D}'} r_\tau(z) P'(d\tau)$, where P' is the probability corresponding to $\mathcal{U}'$. Let $h : \mathcal{D}' \to \mathcal{D}$ be the function in Definition 13 that embeds $\mathcal{D}'$ into $\mathcal{D}$. Then $r_\tau = f_{h(\tau)}$, so that $g(z) = \int_{\mathcal{D}'} f_{h(\tau)}(z) P'(d\tau)$. Let P be the probability on $\mathcal{D}$ induced by h from P' (Lemma 11.) Then

$$\int_{\mathcal{D}} f_\eta(z) P(d\eta) = \int_{\mathcal{D}'} f_{h(\tau)}(z) P'(d\tau).$$

For the "if" direction, let $\mathcal{U}$ be an ultrafilter that contains all tails of $\mathcal{D}$ such that $g(z) = \int_{\mathcal{D}} f_\eta(z) P(d\eta)$ for all $z \in \mathcal{Z}$, where P is the minimum extension of the probability on $\mathcal{D}$ corresponding to $\mathcal{U}$. We need to show that g is a cluster point of f. Specifically, we need to show that, for every neighborhood N of g and $\eta \in \mathcal{D}$, there exists $\beta \ge_{\mathcal{D}} \eta$ such that $f_\beta \in N$. It suffices to prove this claim for all neighborhoods of the form $N = \{h : h(z_j) \in N_j, \text{ for } j = 1, \ldots, n\}$, for arbitrary positive integer n, distinct $z_1, \ldots, z_n \in \mathcal{Z}$ and neighborhoods $N_1, \ldots, N_n$ of the form $N_j = (g(z_j) - \epsilon, g(z_j) + \epsilon)$ for $\epsilon > 0$ if $g(z_j)$ is finite and $N_j = (c, \infty]$ if $g(z_j) = \infty$. So, let N be of the form just described, and let $\eta \in \mathcal{D}$. We need to find $\beta \in \mathcal{D}$ such that $\eta \le_{\mathcal{D}} \beta$ and $f_\beta(z_j) \in N_j$ for $j = 1, \ldots, n$. We know that, for all $z \in \mathcal{Z}$,

$$\sup_{B \in \mathcal{U}} \inf_{\beta \in B} f_\beta(z) = g(z) = \inf_{B \in \mathcal{U}} \sup_{\beta \in B} f_\beta(z).$$

For each j such that $g(z_j)$ is finite, let $B_j \in \mathcal{U}$ be such that

$$\inf_{\eta \in B_j} f_\beta(z_j) > g(z_j) - \epsilon,$$
$$\sup_{\eta \in B_j} f_\beta(z_j) < g(z_j) + \epsilon.$$

For each j such that $g(z_j) = \infty$, let $B_j \in \mathcal{U}$ be such that

$$\inf_{\beta \in B_j} f_\beta(z_j) > c.$$

If we replace B_j by $B_j \cap A_\eta$, all of the last three inequalities above continue to hold. For each j, let $\beta_j \in B_j$, and let $\beta \ge_{\mathcal{D}} \beta_j$ for all j. Then $\eta \le_{\mathcal{D}} \beta$ and $f_\beta(z_j) \in N_j$ for all j. □

B Results whose proofs are in the online resource

This appendix contains results whose proofs are located in the Online Resource.

Proposition 10 (Lemma 10 in Section 2.3 of the Online Resource) *Let $\mathcal{Z}$ be a set. Then $\mathcal{P}_{\mathcal{Z}}$ is compact.*

Proposition 11 (Lemma 2 in Section 1.2 of the Online Resource) *Let $\mathcal{Z}$ be a set. For every $z \in \mathcal{Z}$ and every $c \in \widehat{I\!R}$, $H = \{f \in \widehat{I\!R}^{\mathcal{Z}} : f(z) = c\}$ is closed. Also, both $H_< = \{f \in \widehat{I\!R}^{\mathcal{Z}} : f(z) < c\}$ and $H_> = \{f \in \widehat{I\!R}^{\mathcal{Z}} : f(z) > c\}$ are open.*

Proposition 12 (Lemma 18 in Section 3 of the Online Resource) *Let G be a closed convex subset of $I\!R^{\mathcal{Z}}$ consisting of non-negative functions such that $g \in G$ and $f \geq g$ implies $f \in G$. Let $k \in \partial_L G$. Then there exists a finitely-additive expectation $\lambda_k \in \Lambda_{\mathcal{Z}}$ such that $\lambda_k(k) \leq \lambda_k(g)$ for all $g \in G$.*

Proposition 13 (Lemma 1 in Section 1.2 of the Online Resource) *Let $\{f_\eta\}_{\eta\in\mathcal{D}}$ be a convergent net (with limit f_0) in $\widehat{I\!R}^{\mathcal{Z}}$ for some set $\mathcal{Z}$.*

- *If $\sup_{z\in\mathcal{Z}} f_\eta(z) \leq c_\eta$ for all η and $\{c_\eta\}_{\eta\in\mathcal{D}}$ converges to c_0, then $\sup_{z\in\mathcal{Z}} f_0(z) \leq c_0$.*
- *If $\inf_{z\in\mathcal{Z}} f_\eta(z) \geq c_\eta$ for all η and $\{c_\eta\}_{\eta\in\mathcal{D}}$ converges to c_0, then $\inf_{z\in\mathcal{Z}} f_0(z) \geq c_0$.*

Proposition 14 (Corollary 1 in Section 1.2 of the Online Resource) *Let $c \in I\!R$. Then $\{f : \inf_{z\in\mathcal{Z}} f(z) \geq c\}$ and $\{f : \sup_{z\in\mathcal{Z}} f(z) \leq c\}$ are closed.*

Proposition 15 (Lemma 17 in Section 3 of the Online Resource) *Let G be a closed convex subset of $\widehat{I\!R}^{\mathcal{Z}}$ consisting of non-negative functions such that $g \in G$ and $f \geq g$ implies $f \in G$. Let k be a real-valued function, and define*

$$H_k = \{h \in \mathcal{M}_{\mathcal{Z}} : h(z) < k(z), \text{ for each } z\}.$$

Suppose that $H_k \cap G = \emptyset$. Then there exists $\lambda_k \in \Lambda_{\mathcal{Z}}$ such that $\lambda_k(k) \leq \lambda_k(g)$ for all $g \in G$.

Proposition 16 (Lemma 22 in Section 5 of the Online Resource) *Let $\mathcal{Z}$ be a set. A finitely-additive expectation λ on $\mathcal{M}_{\mathcal{Z}}$ is continuous in the pointwise topology if and only if λ is simple and is a minimum extension.*

Proposition 17 (Lemma 11 in Section 2.3 of the Online Resource) *Let $\mathcal{Z}$ be a set. Then $\mathcal{S}_{\mathcal{Z}}$ is dense in $\Lambda_{\mathcal{Z}}$.*

Proposition 18 (Lemma 12 in Section 2.4 of the Online Resource) *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets. Let P be a finitely-additive probability on $2^{\mathcal{Y}}$. Let Q be a countably-additive discrete probability on $2^{\mathcal{X}}$. Then $P[Q](f) = Q[P](f)$ for each bounded function f.*

References

Battigalli P, Cerreia-Vioglio S, Maccheroni F, Marinacci M (2016) A note on comparative ambiguity aversion and justifiability. Econometrica 84:1903–1916

Berger JO, Srinivasan C (1978) Generalized bayes estimators in multivariate problems. Ann Stat 6:783–801

Bhaskara Rao KPS, Bhaskara Rao M (1983) Theory of charges. Academic Press, London

Brown LD (1971) Admissible estimators, recurrent diffusions, and insoluble boundary value problems. Ann Math Stat 42:855–903
Comfort WW, Negrepontis S (1974) The theory of ultrafilters. Springer, New York
de Finetti B (1974) Theory of probability. Wiley, New York
Dunford N, Schwartz JT (1957) Linear operators. Part I: general theory. Wiley, Hoboken
Heath D, Sudderth W (1972) On a theorem of de finetti, oddsmaking, and game theory. Ann Math Stat 43:2072–2077
Heath D, Sudderth W (1978) On finitely additive priors, coherence, and extended admissibility. Ann Stat 6:333–345
Kelley JL (1955) General topology. Springer, New York
Lehmann EL, Casella G (1998) Theory of point estimation. Springer, New York
Pearce D (1984) Rationalizable strategic behavior and the problem of perfection. Econometrica 52:1029–1050
Royden HL (1968) Real analysis. Macmillan, London
Schervish MJ, Seidenfeld T, Kadane JB (2009) Proper scoring rules, dominated forecasts, and coherence. Decis Anal 6:202–221
Schervish MJ, Seidenfeld T, Kadane JB (2014) Infinite previsions and finitely additive expectations, 2014. Online supplement to "Dominating countably many forecasts". Ann Stat 42:728–756. https://doi.org/10.1214/14-AOS1203
Seidenfeld T, Schervish MJ, Kadane JB (2001) Improper regular conditional distributions. Ann Probab 29:1612–1624. Correction 34:423–426 (2006)
Wald A, Wolfowitz J (1951) Two methods of randomization in statistics and the theory of games. Ann Math 52:581–586
Willard S (1970) General topology. Addison-Wesley, Boston

Online Resource for "What Finite-Additivity Can Add to Decision Theory"

Mark J. Schervish,[1] Teddy Seidenfeld, Rafael B. Stern, and Joseph B. Kadane

Abstract This Online Resource includes background material for the paper Schervish et al. (2019). The background begins with topology and convergence in Section 1. At that point, we have enough tools to give an overview of the theory of finitely-additive expectations in Section 2. The final bit of background includes some results on separation of convex sets of unbounded functions in Section 3. The randomized rules constructed in Schervish et al. (2019) are seemingly less general than the randomized rules that typically arise in the countably-additive decision theory. Section 4 demonstrates that the risk set that we construct from our finitely-additive randomizations includes the risk functions for the more general randomizations when all probabilities are countably-additive. Section 5 presents an interesting property of pointwise convergence that helps to understand why some prior distributions (both finitely-additive and countably-additive) might not have Bayes rules.

1 Topology Background

Our results rely on the theory of product sets, nets, and ultrafilters.

1.1 Product Sets

Definition 1 Let $\{\mathcal{Z}_\alpha\}_{\alpha\in A}$ be a collection of sets, each of which has a topology. The *product topology* on the product set $\prod_{\alpha\in A}\mathcal{Z}_\alpha$ is the topology that has as a sub-base the collection of all sets of the form $\prod_{\alpha\in A}\mathcal{Y}_\alpha$ where $\mathcal{Y}_\alpha = \mathcal{Z}_\alpha$ for all but at most one α, and for that one α, $\mathcal{Y}_\alpha$ is an open subset of $\mathcal{Z}_\alpha$.

If we let $\mathcal{Z} = \bigcup_{\alpha\in A}\mathcal{Z}_\alpha$, then the product set $\prod_{\alpha\in A}\mathcal{Z}_\alpha$ can be thought of as the set of all functions $f : A \to \mathcal{Z}$ such that $f(\alpha) \in \mathcal{Z}_\alpha$ for all α. If all $\mathcal{Z}_\alpha$ are the same set $\mathcal{Z}$, then the product set is often written $\mathcal{Z}^A$.

Definition 2 Let $\{\mathcal{Z}_\alpha\}_{\alpha\in A}$ be a collection of sets. For each $\beta \in A$, the function $f_\beta : \prod_{\alpha\in A}\mathcal{Z}_\alpha \to \mathcal{Z}_\beta$ defined by $f_\beta(g) = g(\beta)$ is called an *evaluation functional* or a *coordinate-projection function.*

It is not difficult to show that the product topology has two other equivalent characterizations: (i) the smallest topology such that all of the evaluation functionals are continuous, and (ii) the topology of pointwise convergence of the functions in the product set.

[1] Corresponding author affiliation: Carnegie Mellon University, Pittsburgh, PA 15213 USA, email: mark@cmu.edu

Every open set in a topological space is the union of arbitrarily many basic open sets. Each basic open set is the intersection of finitely many sub-basic open sets. Hence, the following result is straightforward.

Proposition 1 *For every open set N in a product space $\prod_{\alpha\in A}\mathcal{Z}_\alpha$, there exist a finite integer n, points $\alpha_1,\ldots,\alpha_n \in A$, and open sets $N_j \subseteq \mathcal{Z}_{\alpha_j}$ for $j = 1,\ldots,n$ such that N contains*

$$\{f : f(\alpha_j) \in N_j, \text{ for } j = 1,\ldots,n\}. \tag{1}$$

In particular, every neighborhood of a function g in the product space must contain a set of the form (1) such that $g(\alpha_j) \in N_j$ for $j = 1,\ldots,n$.

1.2 Nets

Our main use for nets is to identify the closure of a set.

Definition 3 A *partial order* on a set $\mathcal{D}$ is a binary relation $\leq_{\mathcal{D}}$ with the following properties: (i) for all $\eta \in \mathcal{D}$, $\eta \leq_{\mathcal{D}} \eta$ (reflexive), and (ii) if $\eta \leq_{\mathcal{D}} \beta$ and $\beta \leq_{\mathcal{D}} \gamma$, then $\eta \leq_{\mathcal{D}} \gamma$ (transitive). A *directed set* is a set $\mathcal{D}$ with a partial order $\leq_{\mathcal{D}}$ that has the following additional property: (iii) for all $\eta, \beta \in \mathcal{D}$, there exists $\gamma \in \mathcal{D}$ such that $\eta \leq_{\mathcal{D}} \gamma$ and $\beta \leq_{\mathcal{D}} \gamma$.

Definition 4 Let $\mathcal{D}$ be a directed set with partial order $\leq_{\mathcal{D}}$. A function $r : \mathcal{D} \to \mathcal{T}$ is called a *net on $\mathcal{D}$ in $\mathcal{T}$*. Such a net is denoted either $(\mathcal{D}, r)$ or $x = \{x_\eta\}_{\eta\in\mathcal{D}}$, where $x_\eta = r(\eta)$ for $\eta \in \mathcal{D}$. If $\{x_\eta\}_{\eta\in\mathcal{D}}$ is a net, then for each $\eta \in \mathcal{D}$, the set

$$A_\eta = \{x_\beta : \eta \leq_{\mathcal{D}} \beta\}$$

is called a *tail of x*. If $\mathcal{T}$ is a topological space, we say that the net x *converges to* $x_* \in \mathcal{T}$ if every neighborhood of x_* contains a tail of x.

Note that the set $\mathbb{Z}^+$ of positive integers is a directed set, and each sequence is a net on $\mathbb{Z}^+$. Every result that holds for all nets holds *a fortiori* for all sequences. There are nets that are not sequences.

Example 1 Let $\mathcal{D}$ be the collection of all finite subsets of an uncountable set $\mathcal{X}$. Create the partial order $\leq_{\mathcal{D}}$ on $\mathcal{D}$ defined by $\eta_1 \leq_{\mathcal{D}} \eta_2$ if $\eta_1 \subseteq \eta_2$. Notice that $\eta_j \leq_{\mathcal{D}} (\eta_1 \cup \eta_2)$ for $j = 1, 2$, making $\mathcal{D}$ a directed set. Let $\mathcal{T}$ be the collection of all indicator functions of subsets of $\mathcal{X}$, and let $x_\eta = I_\eta$ for each $\eta \in \mathcal{D}$. It is straightforward to show that $\{x_\eta\}_{\eta\in\mathcal{D}}$ converges to $I_{\mathcal{X}}$. It is also easy to see that no sequence of indicators of finite sets can converge to the indicator of an uncountable set.

Definition 5 Let $x = \{x_\eta\}_{\eta\in\mathcal{D}}$ be a net in a topological space $\mathcal{T}$. We call a net $y = \{y_\gamma\}_{\gamma\in\mathcal{D}'}$ a *subnet* of x if there exists a function $h : \mathcal{D}' \to \mathcal{D}$ with the following properties: (i) $y_\gamma = x_{h(\gamma)}$ for all $\gamma \in \mathcal{D}'$, (ii) $\gamma_1 \leq_{\mathcal{D}'} \gamma_2$ implies $h(\gamma_1) \leq_{\mathcal{D}} h(\gamma_2)$ and (iii) for every $\eta \in \mathcal{D}$ there exists $\gamma \in \mathcal{D}'$ such that $h(\gamma) \geq_{\mathcal{D}} \eta$. A *cluster point* of x is a point $p \in \mathcal{T}$ such that, for every neighborhood N of p and every $\eta \in \mathcal{D}$, there exists $\eta' \in \mathcal{D}$ such that $\eta' \geq_{\mathcal{D}} \eta$ and $x_{\eta'} \in N$.

It is trivial to see that a net is a subnet of itself. If we think of a sequence as a net, then it is also trivial that a subsequence is a subnet. As Willard (1970, Example 11.4(b)) points out, a subnet of a sequence need not be a subsequence because the directed set $\mathcal{D}'$ for the subnet might be merely partially ordered, unlike the integers. An example of such a subnet can be constructed from the proof of Kelley (1955, Lemma 5, p.70).

The following result is straightforward.

Proposition 2 *Let $\mathcal{Z}$ be a set, and let $\{f_\eta\}_{\eta\in\mathcal{D}}$ and $\{g_\eta\}_{\eta\in\mathcal{D}}$ be convergent nets in $\widehat{I\!R}^{\mathcal{Z}}$. Let f and g be the respective limits. If $f_\eta(z) \le g_\eta(z)$ for all $z \in \mathcal{Z}$ and $\eta \in \mathcal{D}$, then $f(z) \le g(z)$ for all $z \in \mathcal{Z}$.*

The supremum and infimum are functionals defined on spaces of $\widehat{I\!R}$-valued functions. They are not continuous in general, but they do have the following property.

Lemma 1 *Let $\{f_\eta\}_{\eta\in\mathcal{D}}$ be a convergent net (with limit f_0) in $\widehat{I\!R}^{\mathcal{Z}}$ for some set $\mathcal{Z}$.*

- *If $\sup_{z\in\mathcal{Z}} f_\eta(z) \le c_\eta$ for all η and $\{c_\eta\}_{\eta\in\mathcal{D}}$ converges to c_0, then $\sup_{z\in\mathcal{Z}} f_0(z) \le c_0$.*
- *If $\inf_{z\in\mathcal{Z}} f_\eta(z) \ge c_\eta$ for all η and $\{c_\eta\}_{\eta\in\mathcal{D}}$ converges to c_0, then $\inf_{z\in\mathcal{Z}} f_0(z) \ge c_0$.*

Proof Consider the claim about the supremum first. The claim is vacuous if $c_0 = \infty$. If $c_0 < \infty$, suppose, to the contrary, that $\sup_{z\in\mathcal{Z}} f_0(z) > c_0$. Then there exists $\epsilon > 0$ and $z_0 \in \mathcal{Z}$ such that $f_0(z_0) > c_0 + \epsilon$. Because f_η converges to f_0, there exists η_0 such that $\eta \ge_{\mathcal{D}} \eta_0$ implies that $f_\eta(z_0) > c_0 + 2\epsilon/3$. Since c_η converges to c_0, there exists η_1 such that $\eta \ge_{\mathcal{D}} \eta_1$ implies that $c_\eta < c_0+\epsilon/3$, hence $f_\eta(z_0) < c_0+\epsilon/3$. There exists η_2 such that $\eta_j \le_{\mathcal{D}} \eta_2$ for $j = 0, 1$. Hence, $\eta \ge_{\mathcal{D}} \eta_2$ implies both $f_\eta(z_0) > c_0+2\epsilon/3$ and $f_\eta(z_0) < c_0+\epsilon/3$, a contradiction. A similar argument works for the claim about the infimum.

Lemma 1 has a useful corollary.

Corollary 1 *Let $c \in I\!R$. Then $\{f : \inf_{z\in\mathcal{Z}} f(z) \ge c\}$ and $\{f : \sup_{z\in\mathcal{Z}} f(z) \le c\}$ are closed.*

The following result is used in the proof of Theorem 1 of Schervish et al. (2019).

Lemma 2 *Let $\mathcal{Z}$ be a set. For every $z \in \mathcal{Z}$ and every $c \in \widehat{I\!R}$, $H = \{f \in \widehat{I\!R}^{\mathcal{Z}} : f(z) = c\}$ is closed. Also, both $H_< = \{f \in \widehat{I\!R}^{\mathcal{Z}} : f(z) < c\}$ and $H_> = \{f \in \widehat{I\!R}^{\mathcal{Z}} : f(z) > c\}$ are open.*

Proof For finite c, each of $H_>$ and $H_<$ are sub-basic open sets, hence their union H^C is open. For $c = \infty$, $H_> = \emptyset$ and every cluster point f of every net in H has $f(z) = \infty$, hence H is closed.

1.3 Ultrafilters

Definition 6 Let S be a set and let $\mathcal{U}$ be a non-empty collection of subsets of S. We call $\mathcal{U}$ an *ultrafilter* on S if (i) $A \in \mathcal{U}$ and $A \subseteq B$ implies $B \in \mathcal{U}$, (ii) $A, B \in \mathcal{U}$ implies $A \cap B \in \mathcal{U}$, and (iii) for every $A \subseteq S$, $A \in \mathcal{U}$ if and only if $A^C \notin \mathcal{U}$. An ultrafilter on S is *principal* if there exists $s \in S$ such that $\mathcal{U}$ consists of all subset of S that contain s. Such an s is called the *atom* of $\mathcal{U}$. Other ultrafilters are called *non-principal*.

Proposition 3 gives general conditions under which ultrafilters exist. It requires a definition first.

Definition 7 A collection $\mathcal{F}$ of subsets of a set $\mathcal{Z}$ has *the finite-intersection property* if every finite subcollection has non-empty intersection.

Proposition 3 (Comfort and Negrepontis 1974: Special case of Theorem 2.18, p. 39) *Let $\mathcal{Z}$ be a set, and let $\mathcal{F}$ be a collection of subsets. If $\mathcal{F}$ has the finite intersection property, then there is an ultrafilter on $\mathcal{Z}$ that contains $\mathcal{F}$.*

Example 2 Let $\mathcal{D}$ be a directed set. Then the collection of all tails of $\mathcal{D}$ has the finite-intersection property. (Every finite subcollection has non-empty intersection.) By Proposition 3, there is an ultrafilter $\mathcal{U}_\mathcal{D}$ that contains all tails of $\mathcal{D}$.

We have two main uses for ultrafilters in this paper. One is to use them as examples of merely finitely-additive probabilities via Lemma 7 in Section 2.2. The other is to exploit the connection between limits of nets and integrals with respect to ultrafilter probabilites defined on directed sets via Lemma 9 in Section 2.2.

2 Finitely-Additive Integrals and Expectations

2.1 Unbounded and Infinite-Valued Functions

In de Finetti (1974), de Finetti laid out the theory of coherent previsions for bounded random variables, which he related to finitely-additive probability. De Finetti's definition of coherent prevision is the following.

Definition 8 Let $\mathcal{F}$ be a set of bounded functions defined on a space $\mathcal{Z}$. For each $f \in \mathcal{F}$, let $P(f)$ be a real number. De Finetti called $\{P(f) : f \in \mathcal{F}\}$ *coherent previsions* for $\mathcal{F}$ if, for every finite integer n, every n-tuple $(f_1, \ldots, f_n) \in \mathcal{F}^n$, and every n-tuple $(\alpha_1, \ldots, \alpha_n)$ of real numbers,

$$\sup_z \sum_{j=1}^n \alpha_j [f_j(z) - P(f_j)] \geq 0. \tag{2}$$

If $\mathcal{F}$ consists solely of indicator functions of events, coherent previsions for $\mathcal{F}$ can be shown to form a finitely-additive probability. (See de Finetti 1974, Chapter 3 for a long-winded, but elementary treatiment.) Also, a finitely-additive probability can be defined on an arbitrary collection of subsets of a general set $\mathcal{Z}$, including the power set. A number of theorems in Bhaskara Rao and Bhaskara Rao (1983, Chapeter 3) show how to extend a partially defined finitely-additive probability to arbitrary larger domains. Theorems 3.2.9 and 3.2.10 on pages 69–70 of Bhaskara Rao and Bhaskara Rao (1983) are very general. For this reason, measurability conditions are often not included in theorems about finitely-additive probabilities. In this paper, we assume that each finitely-additive probability is defined on $2^{\mathcal{Z}}$. If the reader starts with a finitely-additive probability P defined on a smaller domain, our results will apply to every extension of P to $2^{\mathcal{Z}}$.

Extending a finitely-additive probability P defined on $2^{\mathcal{Z}}$ to a coherent prevision on the set of all bounded functions in $\mathcal{M}_{\mathcal{Z}}$ is straightforward. For each simple function $f = \sum_{j=1}^{n} \alpha_j I_{A_j}$ on $\mathcal{Z}$ (with all α_j finite), there is a unique value for the coherent previson $P(f) = \sum_{j=1}^{n} \alpha_j P(A_j)$. (See Bhaskara Rao and Bhaskara Rao 1983, Proposition 4.4.2 on page 97.) Because every bounded function is uniformly approximable both above and below by simple functions, for each bounded function f on $\mathcal{Z}$, there is a unique value for the finitely-additive expectation

$$P(f) = \sup_{\text{simple } g \leq f} P(g) = \inf_{\text{simple } g \geq f} P(g).$$

In this paper, we need an extension of the finitely-additive theory from bounded functions both to unbounded functions, as was done by Schervish et al. (2014), and to functions that assume the value ∞ as well. The extension to unbounded functions in Schervish et al. (2014) generalizes the concept of coherence to allow $P(f) = \infty$ without requiring the symbol ∞ to appear in (2). They then prove in Schervish et al. (2014, Definition 7, Lemmas 4 and 6) that previsions are coherent if and only if they are the values of a normalized monotone linear functional. The assumption that the loss function is bounded below alows us to avoid dealing with functions that both assume the value ∞ and are unbounded below. This has the added benefit of allowing us to avoid arithmetic that leads to $\infty - \infty$.

Because the set $\mathcal{M}_{\mathcal{Z}}$ of functions that are bounded below is not a linear space, we need to generalize the concept of normalized monotone linear functional.

Definition 9 Let $\mathcal{Z}$ be a space, and let $\mathcal{F}$ be a subset of $\mathcal{M}_{\mathcal{Z}}$ that contains all constant functions and has the following restricted linearity property:

$$\text{If } f, g \in \mathcal{F} \text{ and } h = \alpha f + \beta g \in \mathcal{M}_{\mathcal{Z}} \text{ with } \alpha, \beta \in \mathbb{R}, \text{ then } h \in \mathcal{F}. \quad (3)$$

We call such a set $\mathcal{F}$ a *restricted-linear space.* If $\mathcal{F}$ is a restricted-linear space, a function $L : \mathcal{F} \to \widehat{\mathbb{R}}$ that satisfies

$$L(\alpha f + \beta g) = \alpha L(f) + \beta L(g), \quad (4)$$

whenever $f, g, \alpha f + \beta g \in \mathcal{F}$ and the arithmetic on the right-hand side of (4) is well defined (i.e., does not involve $\infty - \infty$,) is called *a restricted-linear functional.* A restricted-linear functional is called *monotone* if $f \leq g$ implies $L(f) \leq L(g)$. A monotone restricted-linear functional is called a *finitely-additive Daniell integral.* If $L(1) = 1$, we say that L is *normalized.* A normalized finitely-additive Daniell integral is called a *finitely-additive expectation.*

Note that $\mathcal{M}_{\mathcal{Z}}$ is a restricted-linear space, as is the set of bounded functions. The unique coherent prevision on the bounded functions constructed above from a finitely-additive probability on $2^{\mathcal{Z}}$ is an example of a finitely-additive expectation.

The various conditions in Definition 9 prevent $\infty - \infty$ from appearing on either side of (4). Note that one cannot have α and β both negative in (3) unless both f and g are bounded. If at least one of the functions f, g is unbounded in (3), then at least one of the unbounded functions must have a positive coefficient in order for the linear combination to be well-defined and bounded below. Examples of each of the following situations arise, and they are the reason that we do not enforce (4) when the arithmetic on the right-hand side involves $\infty - \infty$:

- The difference between two functions that are unbounded above can be bounded below.
- The difference of two functions, each with infinite finitely-additive expectation, can have finite expectation.

Although there is no unique extension of a finitely-additive expectation from the set of bounded functions to $\mathcal{M}_{\mathcal{Z}}$, there are two special extensions that exist and prove useful. Lemmas 3 and 4 apply to unbounded and infinite-valued functions. Equation 1.2 of Heath and Sudderth states a result like Lemma 3 without proof and without being explicit about the fact that the extension might take the value ∞. The restriction to functions that are bounded below is important (even for the result of Heath and Sudderth) in order to avoid $\infty - \infty$.

Lemma 3 *Let $\mathcal{J}$ be the set of all bounded real-valued functions defined on a set $\mathcal{Z}$, and let P be a finitely-additive expectation on $\mathcal{J}$. For each $f \in \mathcal{M}_{\mathcal{Z}}$, define*

$$Q(f) = \sup_{g \in \mathcal{J},\ g \leq f} P(g).$$

Then $Q = P$ on $\mathcal{J}$, and Q is a finitely-additive expectation on $\mathcal{M}_{\mathcal{Z}}$.

Proof Because P is a finitely-additive expectation, $g \leq f$ implies $P(g) \leq P(f)$ for $f, g \in \mathcal{J}$. Hence $Q(f) = P(f)$ for $f \in \mathcal{J}$.

Next, we show that Q is restricted-linear. Let $f_1, f_2 \in \mathcal{M}_{\mathcal{Z}}$, $\alpha_1, \alpha_2 \in \mathbb{R}$, and $\alpha_1 f_1 + \alpha_2 f_2 \in \mathcal{M}_{\mathcal{Z}}$. We need to show that $Q(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 Q(f_1) + \alpha_2 Q(f_2)$ whenever the arithmetic on the right-hand side is well-defined. Since $Q(\alpha_j f_j) = \alpha_j Q(f_j)$ if $\alpha_j > 0$, there is no loss of generality in assuming that α_1, α_2 are each either 1 or -1. First, we show that $Q(f_1 + f_2) = Q(f_1) + Q(f_2)$

for $f_1, f_2 \in \mathcal{M}_{\mathcal{Z}}$. Notice that $Q(f) = \lim_{m\to\infty} P(f \wedge m)$ because every bounded $g \le f$ is bounded above by $f \wedge m$ where $m = \sup_z g(z)$. Then notice that, for all m,

$$(f_1 \wedge m/2) + (f_2 \wedge m/2) \le (f_1 + f_2) \wedge m \le (f_1 \wedge m) + (f_2 \wedge m).$$

The limits of the left-hand and right-hand expressions are both $Q(f_1)+Q(f_2)$ while the limit of the middle expression is $Q(f_1 + f_2)$. To complete the proof that Q is restricted-linear, we need to show that $Q(f_1 - f_2) = Q(f_1) - Q(f_2)$ if $f_1 - f_2 \in \mathcal{M}_{\mathcal{Z}}$ and $Q(f_2)$ is finite. What we just proved implies that

$$Q(f_1) = Q(f_1 - f_2 + f_2) = Q(f_1 - f_2) + Q(f_2). \tag{5}$$

Since $Q(f_2)$ is finite, we can subtract it from both sides of (5) to complete this part of the proof.

Next, we show that Q is monotone. Let $f, g \in \mathcal{M}_{\mathcal{Z}}$ with $f \le g$. We need to show that $Q(f) \le Q(g)$. Define

$$g'(z) = \begin{cases} g(z) - f(z) & \text{if } f(z) < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Then $g = f + g'$, $g' \ge 0$, and $g' \in \mathcal{M}_{\mathcal{Z}}$. Hence, $0 \le Q(g')$ from the definition of Q. From the previous part of the proof,

$$Q(g) = Q(f + g') = Q(f) + Q(g') \ge Q(f).$$

At the other extreme from Lemma 3, we have the following alternative extension of a finitely-additive expectation from bounded functions to $\mathcal{M}_{\mathcal{Z}}$.

Lemma 4 *Assume the conditions of Lemma 3. Define L on $\mathcal{M}_{\mathcal{Z}}$ as follows: $L(f) = P(f)$ if $f \in \mathcal{J}$, and $L(f) = \infty$ if $f \in \mathcal{M}_{\mathcal{Z}} \setminus \mathcal{J}$. Then L is a finitely-additive expectation on $\mathcal{M}_{\mathcal{Z}}$ that extends P.*

Proof Clearly, L extends P. Also, L is monotone since every instance of $f \le g$ either has both $f, g \in \mathcal{J}$ or $L(g) = \infty$. For (4), if either α or β is 0, the equation holds. When both α and β are nonzero, consider four cases:
(i) $\alpha f, \beta g \in \mathcal{J}$. In this case both sides of (4) are the same because P is linear on $\mathcal{J}$.
(ii) $\alpha f \in \mathcal{J}$ and $\beta g \notin \mathcal{J}$. In this case, $\beta > 0$ so that both sides of (4) are ∞.
(iii) $\alpha f \notin \mathcal{J}$ and $\beta g \in \mathcal{J}$. In this case, $\alpha > 0$ so that both sides of (4) are ∞.
(iv) $\alpha f, \beta g \notin \mathcal{J}$. In this case, at least one of α or β must be positive. If α and β are both positive, both sides of (4) are ∞. If one of them is negative, the right-hand side of (4) is $\infty - \infty$, and (4) has no force.

Definition 10 We refer to Q in Lemma 3 as the *minimum extension of P*. If a prevision Q is the minimum extension of its restriction to the bounded random variables, then we say that Q *is a minimum extension.* We refer to the L in Lemma 4 as the *maximum extension of P*. We also use the notation $\int_{\mathcal{Z}} f(z)P(dz)$ to stand for $Q(f)$ because $Q(f)$ is a finitely-additive Daniell integral and is determined uniquely from P and f.

The terms "minimum extension" and "maximum extension" are used for Q and L respectively because they assign the minimum and maximum of all possible finitely-additive expectations that extend P to $\mathcal{M}_{\mathcal{Z}}$.

It is also straightforward to see that the measure-theoretic definition of expectation with respect to a countably-additive probability P is the restriction of a minimum extension to the functions in $\mathcal{M}_{\mathcal{Z}}$ that are measurable with respect to the σ-field on which P is defined. Nothing in the countably-additive theory corresponds to the maximum extension. There are extensions that are neither minimum nor maximum, but these cannot generally be constructed in one step, like we did for the minimum and maximum extensions. The next example includes such an extension.

Example 3 Let $\mathcal{F}_0$ be the set of all bounded functions on a set $\mathcal{Z}$, and let f_0 be an unbounded function (possibly infinite-valued.) Let $T_{f_0} = \{z : f_0(z) = \infty\}$. Let P_0 be a finitely-additive expectation on $\mathcal{F}_0$. Define

$$\begin{aligned}
\mathcal{F} &= \{\alpha f_0 + g : \alpha \geq 0, g \in \mathcal{F}_0\},\\
c &= \sup_{h \in \mathcal{F}_0,\, h \leq f_0} P_0(f_0),\\
q &= P_0(T_{f_0}).
\end{aligned}$$

If $T_{f_0} \neq \emptyset$, the representation of elements of $\mathcal{F}$ as $f = \alpha f_0 + g$ is not unique because arbitrary changes to g on T_{f_0} do not affect f. This degree of non-uniqueness does not affect the example. However, if f_0 is bounded on $T_{f_0}^C$, there is a more serious non-uniqueness in $f = \alpha f_0 + g$. For example, let $h = f_0 I_{T_{f_0}^c}$. Then $\alpha f_0 + g = (\alpha+1) f_0 + g - h$. In this case, which we will call the "bounded case," replace f_0 by $\infty I_{T_{f_0}}$ in the definition of $\mathcal{F}$ above. In this way the set $\mathcal{F}$ does not change. Then, the representation of elements of $\mathcal{F}$ as $\alpha \infty I_{T_{f_0}} + g$ with $\alpha \in \{0, 1\}$ and $g \in \mathcal{F}_0$ is unique except for the values that g takes on the set T_{f_0}. Throughout this example, when we refer to generic elements f_1, f_2 of $\mathcal{F}$, we will denote $f_j = \alpha_j f_0 + g_j$ with $\alpha_j \geq 0$ (or $\alpha_j \in \{0, 1\}$ in the bounded case) and $g_j \in \mathcal{F}_0$ for $j = 1, 2$.

To see that $\mathcal{F}$ is restricted-linear, let $f_1, f_2 \in \mathcal{F}$. The proof depends on whether or not f_0 takes the value ∞. If there is $z \in \mathcal{Z}$ such that $f_0(z) = \infty$, then $\alpha f_1 + \beta f_2 \in \mathcal{F}$ if and only if both $\alpha\alpha_1$ and $\beta\alpha_2$ are non-negative. Similarly, $\alpha f_1 + \beta f_2 \in \mathcal{M}_{\mathcal{Z}}$ if and only if both $\alpha\alpha_1$ and $\beta\alpha_2$ are non-negative. If f_0 is real-valued, then $\alpha f_1 + \beta f_2 \in \mathcal{F}$ if and only if $\alpha\alpha_1 + \beta\alpha_2 \geq 0$. Similarly, $\alpha f_1 + \beta f_2 \in \mathcal{M}_{\mathcal{Z}}$ if and only if $\alpha\alpha_1 + \beta\alpha_2 \geq 0$.

Next, extend P_0 to the domain $\mathcal{F}$. Let $f = \alpha f_0 + g$ with $\alpha \geq 0$ (or $\alpha \in \{0, 1\}$ for the bounded case) and $g \in \mathcal{F}_0$. If $q > 0$ or $c = \infty$, define

$$P(f) = \begin{cases} \infty & \text{if } \alpha > 0, \\ P_0(g) & \text{if } \alpha = 0. \end{cases}$$

If $q = 0$ and $c < \infty$, let $d \geq c$ be arbitrary ($d = \infty$ is allowed,) and define $P(f) = \alpha d + P_0(g)$. This extension is well-defined because (i) if $\alpha = 0$, g is unique, (ii) if $\alpha > 0$ and $q = 0$, the values of g on T_{f_0} are irrelevant to $P(f)$, and (iii) if $\alpha > 0$ and $q > 0$, all of g is irrelevant to $P(f)$. Note that

$$P(f_0) = \begin{cases} \infty \text{ if } q > 0 \text{ or } c = \infty, \\ \alpha d \text{ if } q = 0 \text{ and } c < \infty. \end{cases}$$

If we choose $d = c$, P will agree with the minimum extension on $\mathcal{F}$. If we choose $d = \infty$, P will agree with the maximum extension. If we choose $d \in (c, \infty)$, P will agree with an extension that is between the two extremes.

For those cases in which $P(f_0) = \infty$, it is not difficult to see that P is a finitely-additive expectation on $\mathcal{F}$. The key step is to notice that $f_1 \leq f_2$ implies that $\alpha_2 \geq \alpha_1$. If at least one of α_1, α_2 is non-zero, then $P(f_2) = \infty$ and $P(f_1) \leq P(f_2)$. If $\alpha_1 = \alpha_2 = 0$, then $P(f_1) \leq P(f_2)$ because $P(f_j) = P_0(f_j)$ for $j = 1, 2$ and P_0 is is a finitely-additive expectation.

For the rest of the example, assume that $P(f_0)$ is finite. In particular, $q = 0$, d is finite, and $P(f)$ is finite for all $f \in \mathcal{F}$. First, we show that P is restricted-linear. If $f_1, f_2, \alpha f_1 + \beta f_2 \in \mathcal{F}$, then $\alpha P(f_1) + \beta P(f_2)$ is not $\infty - \infty$, and both sides of (4) are

$$(\alpha\alpha_1 + \beta\alpha_2)d + \alpha P_0(g_1) + \beta P_0(g_2).$$

Next, we show that P is monotone. For $j = 1, 2$, we have $P(f_j) = \alpha_j d + P_0(g_j)$, so

$$P(f_2) - P(f_1) = d(\alpha_2 - \alpha_1) + P_0(g_2 - g_1). \tag{6}$$

Let $f_1 \leq f_2$, so that (as above) $\alpha_2 \geq \alpha_1$ and for all $z \notin T_{f_0}$,

$$g_2(z) - g_1(z) \geq (\alpha_1 - \alpha_2) f_0(z).$$

If $\alpha_2 = \alpha_1$, then, $P_0(g_2 - g_1) \geq 0$ because P_0 is monotone. If $\alpha_2 > \alpha_1$, we have

$$\frac{g_1(z) - g_2(z)}{\alpha_2 - \alpha_1} \leq f_0(z),$$

for all $z \notin T_{f_0}$. Since $P_0(T_{f_0}) = 0$,

$$(\alpha_2 - \alpha_1)c \geq P_0(g_1 - g_2).$$

Combining this with (6) and $d \geq c$ yields $P(f_2) - P(f_1) \geq 0$.

The following is a useful result concerning minimum extensions.

Lemma 5 *Let f be bounded below. Let P be a minimum extension such that $P(f) < \infty$. Then $P(f) = \lim_{m\to\infty} P[fI_{\{f\leq m\}}]$.*

Proof Since $P[fI_{\{f\leq m\}}]$ is non-decreasing in m, it converges to some number $c_1 \leq P(f)$. Hence $mP(f > m)$ converges to $c_2 = P(f) - c_1 \geq 0$. We need to prove that $c_2 = 0$. Assume, to the contrary, that $c_2 > 0$. Then

$$\lim_{n\to\infty} \frac{2^n P(f > 2^n)}{2^{n-1} P(f > 2^{n-1})} = 1.$$

Also,

$$P(f) \geq \sum_{n=1}^{\infty} 2^{n-1} P(2^{n-1} < f \leq 2^n). \tag{7}$$

For each $n \geq 1$, let $d_n = 2P(f > 2^n)/P(f > 2^{n-1})$. We have assumed that $\lim_{n\to\infty} d_n = 1$. So $P(2^{n-1} < f \leq 2^n) = P(f > 2^n)(1 + \epsilon_n)$ where $\lim_{n\to\infty} \epsilon_n = 0$. It follows from (7) that

$$P(f) \geq \sum_{n=1}^{\infty} 2^{n-1} P(f > 2^{n-1})(1 + \epsilon_n) = \infty,$$

which contradicts $P(f) < \infty$.

The following is a generalization of a well-know measure-theoretic result.

Lemma 6 *Let $\mathcal{Z}$ and $\mathcal{Y}$ be sets. Let P be a finitely-additive expectation on $\mathcal{M}_{\mathcal{Y}}$. Let $h : \mathcal{Y} \to \mathcal{Z}$ be a function. Then P' defined by $P'(f) = P[f(h)]$ is a finitely-additive expectation defined on $\mathcal{M}_{\mathcal{Z}}$ which we call* the finitely-additive expectation induced by h.

Proof Clearly $P' \in \widehat{\mathbb{R}}^{\mathcal{M}_{\mathcal{Z}}}$, and $P'(1) = P(1) = 1$. If $f \leq g$, then $f(h) \leq g(h)$ and $P'(f) \leq P'(g)$. We need to show that P' is extended-linear. Assume that $f, g, \alpha f + \beta g \in \mathcal{M}_{\mathcal{Z}}$ and $\alpha P'(f)$ and $\beta P'(g)$ are not both infinite of opposite signs. Then $\alpha P[f(h)]$ and $\beta P[g(h)]$ are not both infinite of opposite signs and $P[\alpha f(h) + \beta g(h)] = \alpha P[f(h)] + \beta P[g(h)]$, which implies $P'(\alpha f + \beta g) = \alpha P'(f) + \beta P'(g)$.

2.2 Ultrafilter Probabilities

There is a correspondence between ultrafilters and 0-1-valued probabilities.

Lemma 7 *Let $\mathcal{Z}$ be a set, and let $\mathcal{F}$ be a field of subsets of $\mathcal{Z}$. A finitely-additive probability P defined on $\mathcal{F}$ takes only the values 0 and 1 if and only if (i) P can be extended to P' defined on $2^{\mathcal{Z}}$ and (ii) there is an ultrafilter $\mathcal{U}$ of subsets of $\mathcal{Z}$ such that $P'(E) = 1$ if and only if $E \in \mathcal{U}$.*

Proof For the "if" direction, the restriction of P' to $\mathcal{F}$ takes only the values 0 and 1. For the "only if" direction, $\mathcal{V} = \{E \in \mathcal{F} : P(E) = 1\}$ has the finite-intersection property, hence there is an ultrafilter $\mathcal{U}$ of subsets of $\mathcal{Z}$ such $\mathcal{V} \subseteq \mathcal{U}$. Define $P'(E) = 1$ if $E \in \mathcal{U}$ and $P'(E) = 0$ if $E \notin \mathcal{U}$. It is clear that P' is finitely-additive on $2^{\mathcal{Z}}$.

Definition 11 Let $\mathcal{U}$ be an ultrafilter of subsets of some set $\mathcal{Z}$. We call the probability P defined by $P(E) = 1$ if $E \in \mathcal{U}$ and $P(E) = 0$ if $E \notin \mathcal{U}$ the *probability corresponding to $\mathcal{U}$*.

Lemma 8 *Let $\mathcal{D}$ be a set. Let $\mathcal{U}$ be an ultrafilter of subsets of $\mathcal{D}$. Let P be the minimum extension of the probability on $\mathcal{D}$ corresponding to $\mathcal{U}$. Then, for each $f \in \mathcal{M}_{\mathcal{D}}$,*

$$\int_{\mathcal{D}} f(\eta)P(d\eta) = \sup_{B\in\mathcal{U}} \inf_{\eta\in B} f(\eta) = \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f(\eta).$$

Proof Let $k_0 = \int_{\mathcal{D}} f(\eta)P(d\eta)$. Since (i) $P(B) = 1$ for all $B \in \mathcal{U}$, (ii) P is a monotone restricted-linear functional, and (iii) P is a minimum extension, $k_0 \leq \sup_{\eta\in B} f(\eta)$, for all $B \in \mathcal{U}$. Similarly, $k_0 \geq \inf_{\eta\in B} f(\eta)$, for all $B \in \mathcal{U}$. It follows that

$$\sup_{B\in\mathcal{U}} \inf_{\eta\in B} f(\eta) \leq k_0 \leq \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f(\eta). \tag{8}$$

If the two endpoints, call them $a \leq b$, of (8) are not equal, then for each $c \in (a, b)$ precisely one of $\{\eta : f(\eta) \leq c\}$ or $\{\eta : f(\eta) > c\}$ is in $\mathcal{U}$. If it is the first of these, it contradicts $b = \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f(\eta)$. If it is the second one, it contradicts $a = \sup_{B\in\mathcal{U}} \inf_{\eta\in B} f(\eta)$. Hence $a = b$, and both inequalities in (8) are equality.

Lemma 9 *Let $\mathcal{Z}$ be a set. Let $\mathcal{D}$ be a directed set and let $f = \{f_\eta\}_{\eta\in\mathcal{D}}$ be a net in $\mathcal{M}_{\mathcal{Z}}$. An element g of $\widehat{I\!R}^{\mathcal{Z}}$ is a cluster point of f if and only if there is an ultrafilter $\mathcal{U}$ that contains all tails of $\mathcal{D}$ whose corresponding probability on $\mathcal{D}$ has minimum extension P such that $g(z) = \int_{\mathcal{D}} f_\eta(z)P(d\eta)$ for all $z \in \mathcal{Z}$.*

Proof For the "only if" direction, assume first that f converges to g. Let $\mathcal{U}$ be any ultrafilter that contains all tails of $\mathcal{D}$, and let P be the minimum extension of the corresponding probability on $\mathcal{D}$. Let $h(z) = \int_{\mathcal{D}} f_\eta(z)P(d\eta)$ for each $z \in \mathcal{Z}$. We need to show that, for each $z \in \mathcal{Z}$ and each neighborhood N of $g(z)$, $h(z) \in N$. Let $z \in \mathcal{Z}$. Lemma 8, applied to ultrafilters of subsets of $\mathcal{D}$, says that

$$\sup_{B\in\mathcal{U}} \inf_{\eta\in B} f_\eta(z) = h(z) = \inf_{B\in\mathcal{U}} \sup_{\eta\in B} f_\eta(z). \tag{9}$$

If $g(z)$ is finite, let $\epsilon > 0$, and let N be the interval $(g(z)-\epsilon, g(z)+\epsilon)$. Because f converges to g, there exists $\eta \in \mathcal{D}$ such that $f_\beta(z) \in N$ for all $\beta \geq_{\mathcal{D}} \eta$. Let $B = A_\eta = \{\beta \in \mathcal{D} : \eta \leq_{\mathcal{D}} \beta\}$, which is in $\mathcal{U}$. It follows that the left and right sides of (9) are respectively at least $g(z)-\epsilon$ and at most $g(z)+\epsilon$. Hence $h(z) \in N$. If $g(z) = \infty$, let $N = (c, \infty]$. Then there exists $\eta \in \mathcal{D}$ such that $f_\beta(z) > c$ for all $\beta \geq_{\mathcal{D}} \eta$. Let $B = A_\eta$, which is in $\mathcal{U}$. It follows that the left side of (9) is at least c, so that $h(z) \in N$. Next, assume that g is merely a cluster point of f. By Proposition 6 in Schervish et al. (2019), there exists a subnet $f' = \{r_\tau\}_{\tau\in\mathcal{D}'}$ that converges to g. The previous argument shows that for every ultrafilter $\mathcal{U}'$ on $\mathcal{D}'$ that contains all tails of $\mathcal{D}'$, $g(z) = \int_{\mathcal{D}'} r_\tau(z)P'(d\tau)$, where P' is the probability corresponding to $\mathcal{U}'$. Let $h : \mathcal{D}' \to \mathcal{D}$ be the function in Definition 5 that embeds $\mathcal{D}'$ into $\mathcal{D}$. Then $r_\tau = f_{h(\tau)}$, so that $g(z) = \int_{\mathcal{D}'} f_{h(\tau)}(z)P'(d\tau)$. Let P be the probability on $\mathcal{D}$ induced by h from P' (Lemma 6.) Then

$$\int_{\mathcal{D}} f_\eta(z)P(d\eta) = \int_{\mathcal{D}'} f_{h(\tau)}(z)P'(d\tau).$$

For the “if” direction, let $\mathcal{U}$ be an ultrafilter that contains all tails of $\mathcal{D}$ such that $g(z) = \int_{\mathcal{D}} f_\eta(z)P(d\eta)$ for all $z \in \mathcal{Z}$, where P is the minimum extension of the probability on $\mathcal{D}$ corresponding to $\mathcal{U}$. We need to show that g is a cluster point of f. Specifically, we need to show that, for every neighborhood N of g and $\eta \in \mathcal{D}$, there exists $\beta \geq_{\mathcal{D}} \eta$ such that $f_\beta \in N$. It suffices to prove this claim for all neighborhoods of the form $N = \{h : h(z_j) \in N_j, \text{ for } j = 1, \ldots, n\}$, for arbitrary positive integer n, distinct $z_1, \ldots, z_n \in \mathcal{Z}$ and neighborhoods $N_1, \ldots, N_n$ of the form $N_j = (g(z_j) - \epsilon, g(z_j) + \epsilon)$ for $\epsilon > 0$ if $g(z_j)$ is finite and $N_j = (c, \infty]$ if $g(z_j) = \infty$. So, let N be of the form just described, and let $\eta \in \mathcal{D}$. We need to find $\beta \in \mathcal{D}$ such that $\eta \leq_{\mathcal{D}} \beta$ and $f_\beta(z_j) \in N_j$ for $j = 1, \ldots, n$. We know that, for all $z \in \mathcal{Z}$,

$$\sup_{B \in \mathcal{U}} \inf_{\beta \in B} f_\beta(z) = g(z) = \inf_{B \in \mathcal{U}} \sup_{\beta \in B} f_\beta(z).$$

For each j such that $g(z_j)$ is finite, let $B_j \in \mathcal{U}$ be such that

$$\inf_{\eta \in B_j} f_\beta(z_j) > g(z_j) - \epsilon,$$
$$\sup_{\eta \in B_j} f_\beta(z_j) < g(z_j) + \epsilon.$$

For each j such that $g(z_j) = \infty$, let $B_j \in \mathcal{U}$ be such that

$$\inf_{\beta \in B_j} f_\beta(z_j) > c.$$

If we replace B_j by $B_j \cap A_\eta$, all of the last three inequalities above continue to hold. For each j, let $\beta_j \in B_j$, and let $\beta \geq_{\mathcal{D}} \beta_j$ for all j. Then $\eta \leq_{\mathcal{D}} \beta$ and $f_\beta(z_j) \in N_j$ for all j.

2.3 Topology of Finitely-Additive Expectations

Let $\mathcal{Z}$ be a set with $\mathcal{M}_{\mathcal{Z}}$ the set of $\widehat{\mathbb{R}}$-valued functions defined on $\mathcal{Z}$ that are bounded below. The set of finitely-additive expectations $\mathcal{P}_{\mathcal{Z}}$ is a set of $\widehat{\mathbb{R}}$-valued functions defined on $\mathcal{M}_{\mathcal{Z}}$. As such, $\mathcal{P}_{\mathcal{Z}}$ has a product topology, which is also the topology of pointwise convergence. That is, a net $\{Q_\eta\}_{\eta \in \mathcal{D}}$ in $\mathcal{P}_{\mathcal{Z}}$ converges to Q if and only if $\{Q_\eta(f)\}_{\eta \in \mathcal{D}}$ converges to $Q(f)$ for all $f \in \mathcal{M}_{\mathcal{Z}}$. The key topological feature of $\mathcal{P}_{\mathcal{Z}}$ is that it is compact.

Lemma 10 *Let $\mathcal{Z}$ be a set. Then $\mathcal{P}_{\mathcal{Z}}$ is compact.*

Proof First, we show that $\mathcal{P}_{\mathcal{Z}}$ is a closed subset of $\widehat{\mathbb{R}}^{\mathcal{M}_{\mathcal{Z}}}$. To do this, we show that the limit of every convergent net in $\mathcal{P}_{\mathcal{Z}}$ is an element of $\mathcal{P}_{\mathcal{Z}}$. Let $\{P_\eta\}_{\eta \in \mathcal{D}}$ be a net of elements of $\mathcal{P}_{\mathcal{Z}}$ such that $P_\eta(f)$ converges to $P(f)$ for each $f \in \mathcal{M}_{\mathcal{Z}}$. We will show that $P \in \mathcal{P}_{\mathcal{Z}}$. Since $P_\eta(1) = 1$ for all η, $P(1) = 1$. If $f \leq g$, then $P_\eta(f) \leq P_\eta(g)$ for all $\eta \in \mathcal{D}$, so $\lim_\eta P_\eta(f) \leq \lim_\eta P_\eta(g)$ by Proposition 2. Let $f, g, \alpha f + \beta g \in \mathcal{M}_{\mathcal{Z}}$ be such that $\alpha P(f) + \beta P(g)$ is not $\infty - \infty$. Then

$$P_\eta(\alpha f + \beta g) = \alpha P_\eta(f) + \beta P_\eta(g),$$

for each η such that the right-hand side is not $\infty - \infty$. Since $\alpha P_\eta(f)$ and $\beta P_\eta(g)$ converge to $\alpha P(f)$ and $\beta P(g)$ respectively, the two limits cannot be infinite of opposite signs. Since $P_\eta(\alpha f + \beta g)$ converges to $P(\alpha f + \beta g)$, we have that

$$P(\alpha f + \beta g) = \alpha P(f) + \beta P(g).$$

Hence, P is a finitely-additive expectation on $\mathcal{M}_{\mathcal{Z}}$.

The rest of the proof is to show that $\mathcal{P}_{\mathcal{Z}}$ is compact. Let $\mathcal{Q}$ be the set of restrictions of the elements of $\mathcal{P}_{\mathcal{Z}}$ to elements of $\mathcal{M}_{\mathcal{Z}}$ that are non-negative, i.e., to the subset $\mathcal{G} = [0, \infty]^{\mathcal{Z}} \subset \mathcal{M}_{\mathcal{Z}}$. Note that the elements of $\mathcal{Q}$ are not restricted-linear functionals because $\mathcal{G}$ is not a restricted-linear space.

Next, we prove that for each element Q of $\mathcal{Q}$, there is a unique $P \in \mathcal{P}_{\mathcal{Z}}$ such that Q is the restriction of P to $\mathcal{G}$. Let $P_j \in \mathcal{P}_{\mathcal{Z}}$ be such that Q is the restriction of P_j to $\mathcal{G}$ for $j = 1, 2$. Let $f \in \mathcal{M}_{\mathcal{Z}}$, and let $g = f - \inf_z f(z)$ which is an element of $[0, \infty]^{\mathcal{Z}}$. So, $P_j(g) = Q(g)$ for $j = 1, 2$. But

$$P_j(f) = P_j(f - \inf_z f(z)) + \inf_z f(z) = Q(f - \inf_z f(z)) + \inf_z f(z), \tag{10}$$

for $j = 1, 2$, hence $P_1 = P_2$.

Next, we prove that $\mathcal{Q}$ is compact. Each $Q \in \mathcal{Q}$ is an element of $[0, \infty]^{\mathcal{G}}$, which is compact. It now suffices to show that $\mathcal{Q}$ is closed. Let $\{Q_\eta\}_{\eta \in \mathcal{D}}$ be a convergent net in $\mathcal{Q}$ with limit Q. We need to show that $Q \in \mathcal{Q}$. For each $\eta \in \mathcal{D}$, there is a unique $P_\eta \in \mathcal{P}_{\mathcal{Z}}$ such that Q_η is the restriction of P_η to $\mathcal{G}$. We know that $P_\eta(f) = Q_\eta(f)$ converges to $Q(f)$ for all $f \in \mathcal{G}$. Let $f \in \mathcal{M}_{\mathcal{Z}}$. Then $P_\eta(f - \inf_z f(z)) = P_\eta(f) - \inf_z f(z)$ converges to $Q(f - \inf_z f(z))$. Hence, for each $f \in \mathcal{M}_{\mathcal{Z}}$, $P_\eta(f)$ converges to $Q(f - \inf_z f(z)) + \inf_z f(z)$, which is the right-hand side of (10). Since $\mathcal{P}_{\mathcal{Z}}$ is closed, the right-hand side of (10) defines the limit P of the net $\{P_\eta\}_{\eta \in \mathcal{D}}$, which makes P an element of $\mathcal{P}_{\mathcal{Z}}$. Clearly, Q is the restriction of P to $\mathcal{G}$, hence $\mathcal{Q}$ is closed. As a closed subset of a compact set, $\mathcal{Q}$ is compact.

Finally, let $\{P_\eta\}_{\eta \in \mathcal{D}}$ be a net in $\mathcal{P}_{\mathcal{Z}}$. We need to show that it has a cluster point in $\mathcal{P}_{\mathcal{Z}}$. For each η, let Q_η be the restriction of P_η to $\mathcal{G}$. Then $\{Q_\eta\}_{\eta \in \mathcal{D}}$ has a cluster point Q, which is in $\mathcal{Q}$ because $\mathcal{Q}$ is compact. Let $\{R_\gamma\}_{\gamma \in \mathcal{D}'}$ be a subnet of $\{Q_\eta\}_{\eta \in \mathcal{D}}$ that converges to Q. Let P be the unique element of $\mathcal{P}_{\mathcal{Z}}$ that extends Q to $\mathcal{M}_{\mathcal{Z}}$. For each $\gamma \in \mathcal{D}'$, let T_γ be the unique element of $\mathcal{P}_{\mathcal{Z}}$ that extends R_γ. Then $\{T_\gamma\}_{\gamma \in \mathcal{D}'}$ is a subnet of $\{P_\eta\}_{\eta \in \mathcal{D}}$ that converges to P by the argument in the previous paragraph.

The set of minimum extensions $\Lambda_{\mathcal{Z}}$ is not compact, as the next example illustrates.

Example 4 Let $\mathcal{Z} = \mathbb{Z}^+$. Let $\{P_n\}_{n=1}^{\infty}$ be the sequence of minimum extensions of the following countably-additive simple probabilities for $n = 1, 2, \ldots$:

$$P_n(z) = \begin{cases} 2^{-z} & \text{for } z = 1, \ldots, n, \\ 2^{-n} & \text{for } z = 2^n, \\ 0 & \text{otherwise.} \end{cases}$$

For each bounded f, $P_n(f)$ converges to the countably-additive expectation of f under the geometric distribution with parameter $1/2$. However, for the function $g(z) = z$, $P_n(g)$ converges to 3, which is 1 plus the countably-additive expectation of g. Every cluster point P of the sequence $\{P_n\}_{n=1}^{\infty}$ will have $P(g) = 3$, so none of the cluster points will be minimum extensions.

Note that the sequence of finitely-additive expectations in Example 4 is contained in $\mathcal{S}_{\mathcal{Z}}$, the simple expectations. Even though the closure of $\mathcal{S}_{\mathcal{Z}}$ contains finitely-additive expectations outside of $\Lambda_{\mathcal{Z}}$, the following result says that $\Lambda_{\mathcal{Z}}$ is a subset of the closure of $\mathcal{S}_{\mathcal{Z}}$.

Lemma 11 *Let $\mathcal{Z}$ be a set. Then $\mathcal{S}_{\mathcal{Z}}$ is dense in $\Lambda_{\mathcal{Z}}$.*

Proof Let $\lambda_0 \in \Lambda_{\mathcal{Z}}$. Let N' be a neighborhood of λ_0. Then N contains a basic open set of the form

$$N = \{\lambda \in \Lambda_{\mathcal{Z}} : \lambda(f_j) \in N_j, \text{ for } j = 1, \ldots, n\}, \tag{11}$$

where each $f_j \in \widehat{\mathbb{R}}^{\mathcal{Z}}$ is bounded below and each N_j is a neighborhood of $\lambda_0(f_j)$ in $\widehat{\mathbb{R}}$. Define

$$\mathcal{J}_1 = \{j : \lambda_0(f_j) < \infty\},$$
$$\mathcal{J}_2 = \{j : \lambda_0(f_j) = \infty\}.$$

Without loss of generality, we can assume that there exist $\epsilon_j > 0$ for each $j \in \mathcal{J}_1$ and c_j for each $j \in \mathcal{J}_2$ such that $N_j = (P(f_j) - \epsilon_j, P(f_j) + \epsilon_j)$ for each $j \in \mathcal{J}_1$, and $N_j = (c_j, \infty]$ for each $j \in \mathcal{J}_2$. Let $\epsilon = \min_{j \in \mathcal{J}_1} \epsilon_j$ and $c = \max_{j \in \mathcal{J}_2} c_j$. For each $j \in \mathcal{J}_2$, there exists m_j, such that $\lambda_0(f_j \wedge m_j) > c + \epsilon$. Let $a = \min_j \inf_z f_j(z)$. Let $g_j = f_j - a$ for $j \in \mathcal{J}_1$ and $g_j = (f_j \wedge m_j) - a$ for $j \in \mathcal{J}_2$. It follows that each g_j is non-negative and has $\lambda_0(g_j) < \infty$. We will find a simple probability measure λ_N such that

$$|\lambda_0(g_j) - \lambda_N(g_j)| < \epsilon \text{ for all } j \in \mathcal{J}_1 \text{ and } \lambda_N(g_j) > c - a \text{ for all } j \in \mathcal{J}_2.$$

Since $f_j \geq g_j + a$, for $j \in \mathcal{J}_2$, the above clearly implies that $\lambda_N \in N$. Let $g = \sum_{j=1}^n g_j$. Since $\lambda_0(g_j) < \infty$ for all j, we have $\lambda_0(g) < \infty$. Let m be large enough so that $\lambda_0(g) < \lambda_0(gI_{\{g \leq m\}}) + \epsilon/3$, which exists by Lemma 5. Note that

$$\begin{aligned} \sum_{j=1}^n \lambda_0(g_j I_{\{g \leq m\}}) &= \lambda_0(gI_{\{g \leq m\}}) \\ &> \lambda_0(g) - \frac{\epsilon}{3}, \\ &= \sum_{j=1}^n \lambda_0(g_j) - \frac{\epsilon}{3}. \end{aligned}$$

Hence

$$\frac{\epsilon}{3} > \sum_{j \in \mathcal{J}_1} \left[\lambda_0(g_j) - \lambda_0(g_j I_{\{g \leq m\}})\right].$$

Since each $\lambda_0(g_j) \geq \lambda_0(g_j I_{\{g \leq m\}})$, we have $\lambda_0(g_j) < \lambda_0(g_j I_{\{g \leq m\}}) + \epsilon/3$ for each j. Let $A = \{z : g(z) \leq m\}$. Let $\ell = \lceil 3m/\epsilon \rceil$. For each $j \in \mathcal{J}_1$ and $k = 1, \ldots, \ell$, let $A_{j,k} = A \cap \{z : (k-1)\epsilon \leq g_j(z) < k\epsilon\}$. Then, for each $j \in \mathcal{J}_1$, $A_{j,1}, \ldots, A_{j,\ell}$ partitions A. Let $B_1, \ldots, B_r$ be the elements of the common refinement of these partitions that satisfy $\lambda_0(B_k) > 0$ for each k. If $z_k \in B_k$ for all k, then

$$\left| \sum_{k=1}^{r} \lambda_0(B_k) g_j(z_k) - \lambda_0(g_j I_A) \right| < \frac{2\epsilon}{3},$$

for all j. Let $z_0 \in A$ and $B_0 = A^C$. For $E \subseteq \mathcal{Z}$, define

$$\lambda_N(E) = \sum_{k=0}^{r} \lambda_0(B_k) I_E(z_k).$$

Then

$$\lambda_N(g_j) = \sum_{k=1}^{r} \lambda_0(B_k) g_j(z_k) + \lambda_0(A^C) g_j(z_0).$$

Since $g_j(z_0) \geq 0$, for each $j \in \mathcal{J}_2$, we have

$$\lambda_N(g_j) \geq \lambda_0(g_j I_A) - \frac{2\epsilon}{3} > \lambda_0(g_j) - \epsilon > c + \epsilon - a \geq c - a.$$

Because $z_0 \in A$, we have $0 \leq g_j(z_0) \leq g(z_0) \leq m$. Hence, for $j \in \mathcal{J}_1$,

$$g_j(z_0) \lambda_0(A^C) \leq \lambda_0(g I_{A^C}) \leq \frac{\epsilon}{3}.$$

It follows that, for $j \in \mathcal{J}_1$,

$$\lambda_0(g_j) - \epsilon < \lambda_0(g_j I_A) - \frac{2\epsilon}{3} \leq \lambda_N(g_j) \leq \lambda_0(g_j I_A) + \frac{2\epsilon}{3} + g_j(z_0)\lambda_0(A^C) \leq \lambda_0(g_j) + \frac{2\epsilon}{3}.$$

So $|\lambda_N(g_j) - \lambda_0(g_j)| < \epsilon$ for all $j \in \mathcal{J}_1$ and $\lambda_N(g_j) > c - a$ for all $j \in \mathcal{J}_2$ and $\lambda_N \in N$.

2.4 Order Matters

In this section, we deal only with bounded functions, so all finitely-additive expectations are integrals. Since risk functions and Bayes risks are defined in terms of integrals of functions of multiple variables with respect to probability measures over multiple spaces, we need to understand how finitely-additive multiple integration behaves. Let $\mathcal{X}$ and $\mathcal{Y}$ be sets. Let P and Q be finitely-additive probabilities defined on $2^{\mathcal{Y}}$ and $2^{\mathcal{X}}$ respectively. For each $A \in 2^{\mathcal{X} \times \mathcal{Y}}$, define the x-section and y-section for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ as follows

$$A_x = \{y \in \mathcal{Y} : (x, y) \in A\},$$
$$A^y = \{x \in \mathcal{X} : (x, y) \in A\}.$$

For each $x \in \mathcal{X}$, $A_x \in 2^{\mathcal{Y}}$, which has a probability $P(A_x)$. Hence $P(A_x)$ is a bounded function of x which we temporarily denote $f_{A,1}(x)$. The integral of $f_{A,1}$ with respect to Q is denoted

$$Q[P](A) = \int_{\mathcal{X}} f_{A,1}(x) Q(dx) = \int_{\mathcal{X}} \left[\int_{\mathcal{Y}} I_A(x,y) P(dy) \right] Q(dx).$$

Similarly, if $f_{A,2}$ is the bounded function on $\mathcal{Y}$ defined by $f_{A,2}(y) = Q(A^y)$, then $P[Q](A)$ denotes $\int_{\mathcal{Y}} f_{A,2}(y) P(dy)$. Each of $P[Q]$ and $Q[P]$ has a unique extension to bounded functions on $\mathcal{X} \times \mathcal{Y}$:

$$\begin{aligned} P[Q](f) &= \int_{\mathcal{Y}} \left[\int_{\mathcal{X}} f(x,y) Q(dx) \right] P(dy), \\ Q[P](f) &= \int_{\mathcal{X}} \left[\int_{\mathcal{Y}} f(x,y) P(dy) \right] Q(dx). \end{aligned}$$

If P and Q were countably-additive probabilities defined on σ-fields $\Sigma_{\mathcal{Y}}$ and $\Sigma_{\mathcal{X}}$ respectively, then the theorems of Fubini and Tonelli imply that $P[Q](f) = Q[P](f)$ for all bounded f that are measurable with respect to the product σ-field $\Sigma_{\mathcal{X}} \otimes \Sigma_{\mathcal{Y}}$ (as well as for some unbounded f.) The same is not generally true for finitely-additive probabilities.

The following is as close as we come to a Fubini/Tonelli theorem involving finitely-additive integrals.

Lemma 12 *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets. Let P be a finitely-additive probability on $2^{\mathcal{Y}}$. Let Q be a countably-additive discrete probability on $2^{\mathcal{X}}$. Then $P[Q](f) = Q[P](f)$ for each bounded function f.*

Proof If a finitely-additive prevision is defined on all indicators, it extends uniquely to all bounded functions, so we prove the result for indicators only. Let $A \subseteq \mathcal{X} \times \mathcal{Y}$. For each $x \in \mathcal{X}$, let A_x and A^y be the sections defined above.

First, consider the case in which the support of Q is a finite subset $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$. Then

$$\begin{aligned} P[Q](A) = \int_{\mathcal{Y}} Q(A^y) P(dy) &= \int_{\mathcal{Y}} \sum_{j=1}^{n} Q(\{x_j\}) I_{A^y}(x_j) P(dy) \\ &= \sum_{j=1}^{n} Q(\{x_j\}) \int_{\mathcal{Y}} I_{A^y}(x_j) P(dy) \\ &= \sum_{j=1}^{n} Q(\{x_j\}) P(A_{x_j}) = \int_{\mathcal{X}} P(A_x) Q(dx) = Q[P](A). \end{aligned}$$

Finally, assume that there is a countably infinite subset $\{z_1, z_2, \ldots\}$ of $\mathcal{X}$ such that $\sum_{j=1}^{\infty} Q(\{z_j\}) = 1$. Let $\epsilon > 0$, and let $C_\epsilon = \{x_1, \ldots, x_n\}$ be a large enough finite subset of C_A so that $q_\epsilon = Q(C_\epsilon) > 1 - \epsilon/2$. Let $x_{n+1} \notin C_\epsilon$. Define $Q_\epsilon(B)$ for $B \subseteq \mathcal{X}$ by

$$Q_\epsilon(B) = Q(B \cap C_\epsilon) + (1 - q_\epsilon) I_{C_\epsilon^C}(x_{n+1}),$$

That is, Q_ϵ is like Q on subsets of C_ϵ but has all of $Q(C_\epsilon^C)$ assigned to the one point x_{n+1}. Then $Q_\epsilon[P] = P[Q_\epsilon]$. The proof will be complete when we show the following two facts hold for all $A \subseteq \mathcal{X} \times \mathcal{Y}$:

$$|P[Q](A) - P[Q_\epsilon](A)| \leq \frac{\epsilon}{2}, \tag{12}$$

$$|Q[P](A) - Q_\epsilon[P](A)| \leq \frac{\epsilon}{2}. \tag{13}$$

First, notice that $x_{n+1} \in A^y$ if and only if $y \in A_{x_{n+1}}$. Then

$$\begin{aligned}
P[Q_\epsilon](A) &= \int_{\mathcal{Y}} \left[\int_{\mathcal{X}} I_A(x,y) Q_\epsilon(dx) \right] P(dy) \\
&= \int_{\mathcal{Y}} Q_\epsilon(A^y) P(dy) \\
&= \int_{\mathcal{Y}} Q(A^y \cap C_\epsilon) P(dy) + (1 - q_\epsilon) \int_{\mathcal{Y}} I_{A^y}(x_{n+1}) P(dy) \\
&= \int_{\mathcal{Y}} Q(A^y) P(dy) - \int_{\mathcal{Y}} Q(A^y \cap C_\epsilon^C) P(dy) + (1 - q_\epsilon) P(A_{x_{n+1}}) \\
&= P[Q](A) - \int_{\mathcal{Y}} Q(A^y \cap C_\epsilon^C) P(dy) + (1 - q_\epsilon) P(A_{x_{n+1}}),
\end{aligned}$$

which is between $P[Q](A) - \frac{\epsilon}{2}$ and $P[Q](A) + \frac{\epsilon}{2}$, so, (12) is true. Next, notice that

$$\begin{aligned}
Q_\epsilon[P](A) &= \int_{\mathcal{X}} \left[\int_{\mathcal{Y}} I_A(x,y) P(dy) \right] Q(dx) \\
&= \int_{\mathcal{X}} P(A_x) Q_\epsilon(dx) \\
&= \sum_{j=1}^{n} P(A_{x_j}) Q(\{x_j\}) + P(A_{x_{n+1}})(1 - q_\epsilon) \\
&= Q[P](A) - \sum_{j=n+1}^{\infty} [P(A_{x_j}) - P(A_{x_{n+1}})] Q(\{x_j\}),
\end{aligned}$$

which is between $Q[P](A) - \frac{\epsilon}{2}$ and $Q[P](A) + \frac{\epsilon}{2}$, so, (13) is true, and the proof is complete.

Here is an example to show why Lemma 12 has the condition that Q be a discrete countably-additive probability rather than a general countably-additive probability. A few cautions are in order first. If a countably-additive probabilitiy Q assigns probability 0 to every singleton, it is often the case that every extension to the power set is merely finitely-additive. Hence, the Q in Example 5 will violate the assumption that we use for all finitely-additive probabilities in this paper, namely that it be defined for every event. The point of the example is that, even if we relax that assumption and allow Q to be defined on a proper σ-field of $2^{\mathcal{X}}$, we still get different results by integrating in the two different orders.

Example 5 Let $\mathcal{X} = \{0,1\}^\infty$, the set of countable binary sequences, and let Σ_0 be the product σ-field of subsets of $\mathcal{X}$. Let Q_0 be the probability on Σ_0 that is the distribution of a sequence $F = \{f_n\}_{n=1}^\infty$ of independent Bernoulli random variables each with $\Pr(f_n = 1) = 1/2$. Let $\mathcal{Y}$ be the positive integers, and let P be the finitely-additive probability corresponding to a non-principal ultrafilter $\mathcal{U}$ of subsets of the integers. Let $E \subseteq \mathcal{X} \times \mathcal{Y}$ be defined as follows: For each $y \in \mathcal{Y}$, let $E^y = \{x : x_y = 1\}$, i.e., the set of sequences in which the yth coordinate is 1. Each E^y is a measurable set. Set $E = \bigcup_{y=1}^\infty E^y$, which is also measurable, and each E^y is the y-section of E. For each $x \in \mathcal{X}$, the x-section of E is $E_x = \{y : x_y = 1\}$, i.e. the subscripts of those terms in the sequence x that equal 1. We will also need to use the set $A = \{x : E_x \in \mathcal{U}\}$, which is not measurable. To see this, suppose that A is measurable. Then A is in the tail σ-field of the sequence F. By the Kolmogorov 0-1 law, $Q_0(A) \in \{0,1\}$. Similarly, A^C is in the tail σ-field, so $Q_0(A^C) \in \{0,1\}$. But $A^C = \{x : E_x^C \in \mathcal{U}\}$ where $E_x^C = \{y : x_y = 0\}$. By the inherent symmetry in the distribution of F, we must have $Q_0(A) = Q_0(A^C)$. But we cannot have $Q_0(A) = Q_0(A^C) = 0$, and we cannot have $Q_0(A) = Q_0(A^C) = 1$, so A is not measurable. It follows that the inner and outer measures of A are not equal, so there is a number $c \neq 1/2$ that is between the inner and outer measures. By symmetry the interval between the inner and outer measures is symmetric around $1/2$, so we can take $c > 1/2$. Extend Q_0 to the measure Q on the σ-field Σ generated by $\Sigma_0 \cup \{A\}$ using a construction similar to that of Doob (1953, p. 624). That is, each $B \in \Sigma$ has the form $B = (A \cap B_1) \cup (A^C \cap B_2)$ with $B_1, B_2 \in \Sigma_0$. Set $Q(B) = cQ_0(B_1) + (1-c)Q_0(B_2)$. Then $Q(A) = c$.

The remainder of the example is devoted to showing that $P[Q](E) \neq Q[P](E)$. First, note that $Q(E^y) = Q(f_y = 1) = 1/2$ for all $y \in \mathcal{Y}$, hence

$$P[Q](E) = \int_{\mathcal{Y}} Q(E^y)P(dy) = \frac{1}{2}.$$

Next, note that $P(E_x) = I_A(x)$, so

$$Q[P](E) = \int_{\mathcal{X}} P(E_x)Q(dx) = \int_{\mathcal{X}} I_A(x)Q(dx) = Q(A) = c > \frac{1}{2}.$$

3 Separation of Convex Sets

The following definition extends some common terms to deal with $\widehat{\mathbb{R}}$-valued functions.

Definition 12 Let $\mathcal{Z}$ be a set. Let $F \subseteq \widehat{\mathbb{R}}^{\mathcal{Z}}$ be convex. For each $p \in F$ and $f \in \widehat{\mathbb{R}}^{\mathcal{Z}}$, define

$$A_{p,f,F} = \{a \in (0,\infty) : p + f/a \in F\}.$$

We say that *p is an internal point of F* if for every bounded f, $A_{p,f,F} \neq \emptyset$. If 0 is an internal point of F, define

$$t_F(f) = \inf A_{0,f,F}.$$

We call t_F the *support function of* F.

Example 6 Let $g \in \mathbb{R}^{\mathcal{Z}}$ and $F = \{f \in \mathbb{R}^{\mathcal{Z}} : f(z) < g(z), \text{ for all } z\}$. Then F is convex and $g - \epsilon$ is an internal point for each $\epsilon > 0$. In fact, for every $\epsilon > 0$ and every bounded $h \geq \epsilon$, $g - h$ is an internal point of F. Functions that get arbitrarily close to g are not internal.

In discussing separation of convex sets A and B, it is common to use the notation $A - B$ to stand for $\{a - b : a \in A, b \in B\}$. We use this notation freely througout the following results and proofs.

Lemma 13 *Let g and k be real-valued functions. Let F be a convex set with g as an internal point. Let F' be another convex set that contains k. Suppose that $f' - f$ is well defined for every $f' \in F'$ and every $f \in F$. Then $k - g$ is internal to $F' - F$.*

Proof Let h be a bounded function. We need to show that there exists $a > 0$ such that $k - g + h/a \in F' - F$. Since g is internal to F and $-h$ is bounded, there exists $a > 0$ such that $g + (-h)/a \in F$. Then $k - [g + (-h)/a] = k - g + h/a \in F' - F$. (Note that none of the arithmetic in this proof involved infinite values.)

Lemma 14 *Let F be a convex set with 0 as an internal point. Then*

(i) *if $f \in F$, $t_F(f) \in [0, 1]$,*
(ii) *if $f \notin F$, $t_F(f) \in [1, \infty]$,*
(iii) *$t_F(f_1 + f_2) \leq t_F(f_1) + t_F(f_2)$ for all f_1, f_2,*
(iv) *$t_F(\alpha f) = \alpha t_F(f)$ for all f and all $\alpha > 0$, and*
(v) *$\mathcal{F}_F = \{f : t_F(f) < \infty \text{ and } t_F(-f) < \infty\}$ is a linear space that contains all bounded functions.*

Proof Part (i): Let $f \in F$. Since $f/1 = f$, we see that $1 \in A_{0,f,F}$ and $t_F(f) \leq 1$.

Part (ii): Let $f \notin F$. If $A_{0,f,F} = \emptyset$, then $t_F(f) = \infty > 1$. For the rest of the proof of part (ii), suppose that $A_{0,f,F} \neq \emptyset$. Note that $f = a(f/a) + (1 - a)0$ for all $a \in (0, 1]$. So $f/a \notin F$ for $a \in (0, 1]$ and $t_F(f) \geq 1$.

Part (iii): The inequality holds if either $t_F(f_1) = \infty$ or $t_F(f_2) = \infty$. For the rest of the proof of part (iii), assume that $\alpha_j = t_F(f_j) < \infty$ for both $j = 1, 2$. Let $\epsilon > 0$. Then $f_j/(\alpha_j + \epsilon) \in F$ for $j = 1, 2$. Since F is convex

$$\frac{\alpha_1 + \epsilon}{\alpha_1 + \alpha_2 + 2\epsilon} \frac{f_1}{\alpha_1 + \epsilon} + \frac{\alpha_2 + \epsilon}{\alpha_1 + \alpha_2 + 2\epsilon} \frac{f_2}{\alpha_2 + \epsilon} = \frac{f_1 + f_2}{\alpha_1 + \alpha_2 + 2\epsilon},$$

is in F. Hence $t_F(f_1 + f_2) \leq \alpha_1 + \alpha_2 + 2\epsilon$ for all ϵ, which implies $t_F(f_1 + f_2) \leq \alpha_1 + \alpha_2$.

Part (iv): Let $\alpha > 0$. For every $a > 0$ and every f, $\alpha f/(\alpha a) \in F$ if and only if $f/a \in F$, so that $A_{0,\alpha f,F} = \alpha A_{0,f,F}$ for all f. Hence $t_F(\alpha f) = \alpha t_F(f)$.

Part (v): Let $f_1, f_2 \in \mathcal{F}_F$, $\alpha_1, \alpha_2 \in \mathbb{R}$. Let

$$f'_j = \begin{cases} f_j \text{ if } \alpha_j \geq 0, \\ -f_j \text{ if } \alpha_j < 0. \end{cases}$$

Then $\alpha_1 f_1+\alpha_2 f_2 = |\alpha_1|f_1'+|\alpha_2|f_2'$. By part (iv), $|\alpha_j|f_j' \in \mathcal{F}_F$ for $j=1,2$, and by part (iii), $|\alpha_1|f_1'+|\alpha_2|f_2' \in \mathcal{F}_F$. So $\alpha_1 f_1+\alpha_2 f_2 \in \mathcal{F}_F$. That $\mathcal{F}_F$ contains all bounded functions is immediate from the fact that 0 is an internal point.

Lemma 15 *Let F be a convex set with 0 as an internal point. Suppose that, for every $f \in F$, every $g \geq f$ is in F. If $f_1 \leq f_2$, then $t_F(f_1) \geq t_F(f_2)$.*

Proof Let $f_1 \leq f_2$. We prove that $t_F(f_1) \geq t_F(f_2)$ in cases. Case 1: Let $f_2 \notin F$. Then $A^C_{0,f_2,F} \neq \emptyset$. Let $a \in A^C_{0,f_2,F}$ be such that $f_2/a \notin F$. If $f_1/a \in F$, $f_1/a \leq f_2/a$ would contradict $f_2/a \notin F$. So $f_1/a \notin F$, and $A^C_{0,f_2,F} \subseteq A^C_{0,f_1,F}$. Then $A_{0,f_1,F} \subseteq A_{0,f_2,F}$, which means $\inf A_{0,f_1,F} \geq \inf A_{0,f_2,F}$ and $t_F(f_1) \geq t_F(f_2)$. Case 2: Let $f_2 \in F$ with $f_1 \notin F$. Then $t_F(f_1) \geq 1 \geq t_F(f_2)$. Case 3: Let $f_2 \in F$ with $f_1 \in F$. Then $A_{0,f_1,F} \neq \emptyset$. For each $a \in A_{0,f_1,F}$ we have $f_1/a \leq f_2/a \in F$, which implies $f_2/a \in F$. So $A_{0,f_1,F} \subseteq A_{0,f_2,F}$ and $\inf A_{0,f_1,F} \geq \inf A_{0,f_2,F}$, so that $t_F(f_1) \geq t_F(f_2)$.

Lemma 16 *Let G be a closed convex subset of $\widehat{\mathbb{R}}^{\mathcal{Z}}$ consisting of non-negative functions. Let G' be a superset of G such that for each $h \in G' \setminus G$ there exists $g \in G$ with $h \geq g$. Then $\partial_L G' = \partial_L G$.*

Proof Clearly, $G \subseteq G'$ so $G \subseteq \overline{G'}$, and $\partial_L G \subseteq \overline{G'}$. Note that every $h \in G' \setminus G$ is dominated by an element of G. Let $h \in \overline{G'} \setminus G$. Then there is a net $f = \{f_\eta\}_{\eta \in \mathcal{D}}$ in G' such that $\lim_\eta f_\eta = h$. Clearly, every tail of f must contain elements of $G' \setminus G$, otherwise f would have a cluster point (hence its limit) in G. Let $g = \{g_\gamma\}_{\gamma \in \Gamma}$ be a subnet consisting of elements of $G' \setminus G$. Then g converges to h. For each $\gamma \in \Gamma$, let $k_\gamma \in G$ be such that k_γ dominates g_γ. Since G is compact, $k = \{k_\gamma\}_{\gamma \in \Gamma}$ has a convergent subnet whose limit we call ℓ. The corresponding subnet of g converges to h, and $\ell \leq h$ by Proposition 2. Since $\ell \in G$ and $h \notin G$, it must be that $\ell \neq h$, so ℓ dominates h. Hence, every element of $\overline{G'} \setminus G$ is dominated by an element of G. Hence, no element of $\overline{G'} \setminus G$ can dominate an element of $\partial_L G$, which makes $\partial_L G = \partial_L G'$.

Lemma 17 *Let G be a closed convex subset of $\widehat{\mathbb{R}}^{\mathcal{Z}}$ consisting of non-negative functions such that $g \in G$ and $f \geq g$ implies $f \in G$. Let k be a real-valued function, and define*

$$H_k = \{h \in \mathcal{M}_{\mathcal{Z}} : h(z) < k(z), \text{ for each } z\}.$$

Suppose that $H_k \cap G = \emptyset$. Then there exists $\lambda_k \in \Lambda_{\mathcal{Z}}$ such that $\lambda_k(k) \leq \lambda_k(g)$ for all $g \in G$.

Proof Note that H_k is convex. Every element of $F = G - H_k - 1$ is a well-defined function. Note that $-1 \notin F$ because $H_k \cap G = \emptyset$. However, for each $c > -1$, the constant function $c = k - (k - c - 1) - 1$ is in F. In particular, $0 \in F$.

The first step in the proof is to show that F satisfies the conditions of Lemma 15. Let $g_1 - h_1 - 1$ and $g_2 - h_2 - 1$ be in F. Then $h' = \alpha h_1 + (1-\alpha) h_2 \in H_k$ and $g' = \alpha g_1 + (1-\alpha) g_2 \in G$. It follows that F is convex. Next, since

$k-1$ is internal to H_k, $k-(k-1)=1$ is internal to $G-H_k$ by Lemma 13. Hence 0 is internal to F. Let $f \in F$ so that $f = g - h - 1$ for some $h \in H_k$ and some $g \in G$. If $f' \geq f$, we need to show $f' \in F$. Let $\ell \geq 0$ be such that $f' = f + \ell$. (If $f(z) = f'(z) = \infty$ for some z, $\ell(z)$ can be any non-negative number.) Then, $g + \ell \in G$ and $f' = g + \ell - h - 1 \in F$. This completes the proof that F satisfies the conditions of Lemma 15.

It is easy to show that

$$t_F(c) = \begin{cases} 0 \text{ if } c > 0, \\ -c \text{ if } c \leq 0. \end{cases} \tag{14}$$

The next step in the proof is to perform a construction similar to the construction of a separating hyperplane between H_k and G. Specifically, we attempt to find a finitely-additive expectation λ on a linear space $\mathcal{L} \subseteq \widehat{\mathbb{R}}^{\mathcal{Z}}$ such that

$$-\lambda(f) \leq t_F(f), \text{ for all } f \in \mathcal{L}. \tag{15}$$

The importance of satisfying (15) is that, if $\mathcal{L}$ contains $F \cup \{-1\}$, then $\lambda(f) \geq -1$ for all $f \in F$, which implies that $\lambda(h) \leq \lambda(g)$ for all $h \in H_k$ and all $g \in G$.

We start with a finitely-additive expectation λ_0 on the space $\mathcal{L}_0$ of constant functions and which satisfies (15). We then extend it to a finitely-additive expectation λ_1 on the set of all bounded functions in $\widehat{\mathbb{R}}^{\mathcal{Z}}$, while satifying (15). We then follow the proof of the Hahn-Banach theorem. We cannot apply the Hahn-Banach theorem directly because not all linear functionals are finitely-additive expectations. Let $\mathcal{L}_0$ be the set of all constant real-valued functions in $\widehat{\mathbb{R}}^{\mathcal{Z}}$. Define $\lambda_0(f) = c$ if f is the constant function $f(z) = c$ for all z. Then λ_0 is trivially a finitely-additive prevision on $\mathcal{L}_0$. Also, $-\lambda_0(c) \leq t_F(c)$ by (14). Next, partially order all extensions that satisfy (15) of λ_0 to finitely-additive expectations on linear spaces of bounded functions. That is, we denote an extension λ of λ_0 that satisfies (15) on a linear space $\mathcal{L}$ of bounded functions that contains $\mathcal{L}_0$ by $(\lambda, \mathcal{L})$, and we define the partial order $(\lambda_a, \mathcal{L}_a) \preceq (\lambda_b, \mathcal{L}_b)$ to mean that $\mathcal{L}_a \subseteq \mathcal{L}_b$. Since every chain in this partial order has the union of its domains as an upper bound, Zorn's lemma says that there is a maximal element, $(\lambda_1, \mathcal{L}_1)$.

The next step is to show that $\mathcal{L}_1$ contains all bounded functions. Assume to the contrary that there is a bounded $h \notin \mathcal{L}_1$. Then there is a bounded h that satisfies $h \geq 0$, $t_F(h) < \infty$, and $t_F(-h) < \infty$. (Every bounded function is the difference between two bounded non-negative functions.) Let $\mathcal{F}_1$ be the linear span of $\mathcal{L}_1 \cup \{h\}$ so that

$$\mathcal{F}_1 = \{f + \alpha h : \alpha \in \mathbb{R}, f \in \mathcal{L}_1\},$$

and for each $g \in \mathcal{F}_1$ there is a unique $\alpha_g \in \mathbb{R}$ and $f_g \in \mathcal{L}_1$ such that $g = f_g + \alpha_g h$. Define

$$\lambda_1^c(g) = \lambda_1(f_g) + \alpha_g c,$$

for each real c. We need to show that there exists $c > 0$ such that $-\lambda_1^c(g) \leq t_F(g)$ for all $g \in \mathcal{F}_1$. Let $f, \ell \in \mathcal{L}_1$. Then

$$\begin{aligned}
\lambda_1(f) - \lambda_1(\ell) &= \lambda_1(f - \ell) \\
&\geq -t_F(f - \ell) \\
&= -t_F(f + h - h - \ell) \\
&\geq -t_F(f + h) - t_F(-\ell - h), \\
\lambda_1(f) + t_F(f + h) &\geq \lambda_1(\ell) - t_F(-h - \ell).
\end{aligned}$$

(Note that all quantities involved in the above equations are finite.) It follows that

$$\sup_{\ell \in \mathcal{L}_1} [\lambda_1(\ell) - t_F(-h - \ell)] \leq \inf_{f \in \mathcal{L}_1} [\lambda_1(f) + t_F(f + h)]. \tag{16}$$

Because $h \geq 0$, $\lambda_1(\ell) = -\lambda_1(-\ell) \leq t_F(-\ell) \leq t_F(-h - \ell)$ for all $f \in \mathcal{L}_1$. So the left-hand side of (16) is at most 0. Let $c \geq 0$ be such that $-c$ is in the closed interval between the two sides of (16). If $\alpha_g > 0$, let $f = f_g/\alpha_g$ on the right-hand side of (16), so that

$$\begin{aligned}
-c &\leq \lambda_1(f_g/\alpha_g) + t_F(f_g/\alpha_g + h), \\
-\alpha_g c &\leq \lambda_1(f_g) + t_F(f_g + \alpha_g h), \\
-\lambda_1(f_g) - \alpha_g c &\leq t_F(f_g + \alpha_g h), \\
-\lambda_1^c(g) &\leq t_F(g).
\end{aligned}$$

If $\alpha_g < 0$, let $\ell = f_g/\alpha_g$ on the left-hand side of (16), so that

$$\begin{aligned}
\lambda_1(f_g/\alpha_g) - t_F(-h - f_g/\alpha_g) &\leq -c, \\
-\lambda_1(f_g) - t_F(f_g + \alpha_g h) &\leq \alpha_g c, \\
-\lambda_1(f_g) - \alpha_g c &\leq t_F(f_g + \alpha_g h), \\
-\lambda_1^c(g) &\leq t_F(g).
\end{aligned}$$

Hence, we have extended λ_1 to a finitely-additive expectation on a larger domain while satisfying (15), which contradicts $\mathcal{L}_1$ being maximal. It follows that $\mathcal{L}_1$ contains all bounded functions.

Next, we extend λ_1 to a domain that includes all functions that are bounded below. Let λ_k be the minimum extension of λ_1, namely

$$\lambda_k(f) = \sup_{\text{bounded } g \leq f} \lambda_1(g).$$

It follows that, for every f that is bounded below, and every bounded $g \leq f$

$$\begin{aligned}
\lambda_k(f) &\geq \lambda_1(g), \\
-\lambda_k(f) &\leq -\lambda_1(g) \\
&\leq t_F(g) \\
&\leq t_F(f),
\end{aligned}$$

where the third inequality follows from what we proved for bounded g, and the fourth inequality follows from Lemma 15.

Next, we prove that λ_k provides the desired separation between G and H_k. We know that $\lambda_k(f) \geq -1$ for each $f \in F$ that is bounded below. For each $g \in G$ and bounded $h \in H_k$, $f = g - h - 1 \in F$ and f is bounded below, hence $\lambda_k(f) \geq -1$. If h is bounded, so is $h + 1$ and

$$\lambda_k(g) \geq \lambda_k(h+1) - 1 = \lambda_k(h), \tag{17}$$

for all $g \in G$ and all bounded $h \in H_k$. For each $\epsilon > 0$ and $m > 0$, $(k-\epsilon) \wedge m \in H_k$, is bounded, and is no greater than $k - \epsilon$. So

$$\lambda_k(k-\epsilon) = \lim_{m\to\infty} \lambda_k((k-\epsilon) \wedge m) \leq \sup_{h \in H_k} \lambda_k(h) \leq \lambda_k(g),$$

for all $g \in G$. So

$$\lambda_k(k) \leq \epsilon + \sup_{h\in H_k} \lambda_k(h) \leq \epsilon + \lambda_k(g),$$

for all $g \in G$, and $\lambda_k(g) \geq \lambda_k(k) - \epsilon$ for every $g \in G$ and $\epsilon > 0$. It follows that $\lambda_k(g) \geq \lambda_k(k)$ for all $g \in G$.

Lemma 18 *Let G be a closed convex subset of $\mathbb{R}^{\mathcal{Z}}$ consisting of non-negative functions such that $g \in G$ and $f \geq g$ implies $f \in G$. Let $k \in \partial_L G$. Then there exists a finitely-additive expectation $\lambda_k \in \Lambda_{\mathcal{Z}}$ such that $\lambda_k(k) \leq \lambda_k(g)$ for all $g \in G$.*

Proof If k is real-valued, Lemma 17 implies the conclusion. If $k(z) = \infty$ for all z, then G is the singleton $\{k\}$, and every λ can be λ_k. If k takes the value ∞ and at least one finite value, let $\mathcal{Z}_k = \mathcal{Z} \setminus \{z : k(z) = \infty\}$. Let G_k be the restrictions of all elements of G to the sub-domain $\mathcal{Z}_k$. Let k' be the restriction of k to $\mathcal{Z}_k$. Then G_k is a convex set of non-negative functions such that $g \in G_k$ and $f \geq g$ implies $f \in G_k$. Also $k' \in G_k$ is real-valued.

First, we show that $k' \in \partial_L G_k$. Assume to the contrary that there is $g \in \overline{G_k}$ such that $g(z) \leq k'(z)$ for all $z \in \mathcal{Z}_k$ and $g(z_0) < k'(z_0)$ for some $z_0 \in \mathcal{Z}_k$. Let $g' = \{g_\eta\}_{\eta\in\mathcal{D}}$ be a net in G_k that converges to g. Each g_η is the restriction to $\mathcal{Z}_k$ of $h_\eta \in G$. Let h be a cluster point of $h' = \{h_\eta\}_{\eta\in\mathcal{D}}$. Let h'' be a subnet that converges to h, and let g'' be the corresponding subnet of g', which still converges to g. It follows that g is the restriction of h to $\mathcal{Z}_k$. Since $k(z) = \infty$ for all $z \in \mathcal{Z} \setminus \mathcal{Z}_k$, $h(z) \leq k(z)$ for all z and $h(z_0) < k(z_0)$, contradicting $k \in \partial_L G$.

Apply Lemma 17 with k replaced by k' and G replaced by G_k to get a finitely-additive expectation $\lambda' \in \Lambda_{\mathcal{Z}_k}$ such that $\lambda'(k') \leq \lambda'(g)$ for all $g \in G_k$. For each $g \in \widehat{\mathbb{R}}^{\mathcal{Z}}$, define $\lambda_k(g)$ to be λ' of the restriction of g to $\mathcal{Z}_k$. Then, $\lambda_k(k) \leq \lambda_k(g)$ for all $g \in G$, and $\lambda_k \in \Lambda_{\mathcal{Z}}$.

4 Countably-Additive General Randomizations

In this section, we prove that the risk functions of the form

$$R(\theta, \delta) = P_\theta \left[\int_{\mathcal{A}} L(\theta, a) \delta(\cdot)(da) \right], \tag{18}$$

with all probabilities countably additive, are included in the set $\mathcal{R}$.

Lemma 19 *Assume that each P_θ is countably-additive and a minimum extension. Also, assume that every randomization included in each randomized rule δ is countably-additive and a minimum extension. Then every function of the form (18) is in $\mathcal{R}$.*

Proof Let H be the set of all functions of the form (18) with all probabilities countably-additive. We will show that every neighborhood V of every element $h \in H$ contains an element of $\mathcal{R}$. In particular, we will find a simple randomization $\delta_* \in \mathcal{S}_{\mathcal{H}_0}$ such that $R(\cdot, \delta_*) \in V$.

For a general element of $\widehat{\mathbb{R}}^{\Theta}$, a neighborhood in the product topology is the union of arbitrarily many basic open sets, hence we can restrict attention to basic open sets. A basic open set is the intersection of finitely many one-dimensional open sets, each of which has one of the following two forms:

- $\{f : |f(\theta) - c| < \epsilon\}$ for some $\epsilon > 0$, some θ, and some real c, or
- $\{f : f(\theta) > c\}$ for some θ and some real c.

Let $h = R(\cdot, \delta) \in H$ so that δ becomes fixed for the remainder of the proof. The neighborhoods of an element $h \in H$ that we need to show intersect $\mathcal{R}$ have the form

$$V = \{f : f(\theta_j) \in N_j, \text{ for } j = 1, \dots, m\}, \tag{19}$$

for some m and some $\theta_1, \dots, \theta_m$ with

$$N_j = \begin{cases} (h(\theta_j) - \epsilon_j, h(\theta_j) + \epsilon_j) \text{ for some } \epsilon_j > 0 & \text{if } h(\theta_j) < \infty, \\ (c_j, \infty] \text{ for some finite } c_j & \text{if } h(\theta_j) = \infty. \end{cases}$$

Let V have the form (19). We need to find an element of $\mathcal{R}$ that is in V. The rule $\delta_* \in \mathcal{R}$ that we find will have the following form. There will be finitely many elements of $\mathcal{A}$, $a_0, \dots, a_M$ and a partition of $\mathcal{X}$, $D_1, \dots, D_R$ such that $\delta_* = \sum_{v \in \mathcal{V}} r_v \delta_v$, where $\mathcal{V}$ is a finite set and for each $v \in \mathcal{V}$, (i) δ_v is constant on each D_t, $t = 1, \dots, R$ and (ii) δ_v takes values only from the set $\{a_0, a_1, \dots, a_M\}$. There are $(M+1)^R$ possible δ_v. The remainder of the proof is devoted to finding an appropriate $a_0, \dots, a_M$ along with an appropriate partition $D_1, \dots, D_R$ and the corresponding weights $\{r_v\}_{v \in \mathcal{V}}$.

For each $j = 1, \dots, m$, let Q_j be the (countably-additive) expectation on $\mathcal{M}_{\mathcal{A}}$ defined by $Q_j(f) = \int_{\mathcal{X}} \int_{\mathcal{A}} f(a) \delta(x)(da) P_{\theta_j}(dx)$ for $f \in \mathcal{M}_{\mathcal{A}}$. This makes

$h(\theta_j) = Q_j[L(\theta_j, \cdot)]$. For $j = 1, \ldots, m$, let

$$\begin{aligned}
B_j &= \{a : L(\theta_j, a) = \infty\}, \\
\mathcal{J}_3 &= \{j : Q_j(B_j) > 0\}\}, \\
\mathcal{J}_2 &= \{j \notin \mathcal{J}_3 : h(\theta_j) = \infty\}, \text{ and} \\
\mathcal{J}_1 &= \{j : h(\theta_j) < \infty\}.
\end{aligned}$$

Note that $h(\theta_j) = \infty$ if and only if $j \in \mathcal{J}_2 \cup \mathcal{J}_3$. Let $\epsilon = \min\{\epsilon_j : j \in \mathcal{J}_1\}$, let $c = \max\{c_j : j \in \mathcal{J}_2 \cup \mathcal{J}_3\}$. By construction, $Q_j(B_j) = 0$ for $j \in \mathcal{J}_1$, and $Q_j(B_j) > 0$ for $j \in \mathcal{J}_3$. For each $j \in \mathcal{J}_3$, let $b_j \in B_j$ be such that $L(\theta_j, b_j) = \infty$. Let w be the number of distinct b_j for $j \in \mathcal{J}_3$.

Define

$$\begin{aligned}
E_\infty &= \bigcup_{j=1}^{m} B_j, \\
E_n &= E_\infty \cup \{a : L(\theta_j, a) < n, \text{ for } j = 1, \ldots, m\}.
\end{aligned}$$

Then $\bigcup_{n=1}^\infty E_n = \mathcal{A}$. For each $j \in \mathcal{J}_1$, there exists n_j such that

$$\left| h(\theta_j) - Q_j \left[I_{E_{n_j}} L(\theta_j, \cdot) \right] \right| < \frac{\epsilon}{4}. \tag{20}$$

For each $j \in \mathcal{J}_2$, there exists n_j such that

$$Q_j \left[I_{E_{n_j}} L(\theta_j, \cdot) \right] > c + \epsilon$$

Let $n = \max_{j \in \mathcal{J}_1 \cup \mathcal{J}_2} n_j$.

Next, we select M and the $M + 1$ elements of $\mathcal{A}$, $a_0, \ldots, a_M$. For each $j \in \mathcal{J}_1 \cup \mathcal{J}_2$, partition $\bigcup_{j \in \mathcal{J}_1 \cup \mathcal{J}_2} \{a : L(\theta_j, a) \leq n\}$ into sets of the form $\{a : L(\theta_j, a) \in [(k-1)\epsilon/4, k\epsilon/4)\}$, for $1 \leq k \leq \lfloor 1 + 4n/\epsilon \rfloor$. Then form the common refinement of these partitions. Let the non-empty sets in that common refinement be $A_1, \ldots, A_K$. For each $\ell \geq 1$ and each $j \in \mathcal{J}_1 \cup \mathcal{J}_2$, $L(\theta_j, a)$ (as a function of a) varies by less than $\epsilon/4$ on A_ℓ. Let $a_\ell \in A_\ell$ for each $\ell = 1, \ldots, K$. Let $a_0 = a_1$. If $w > 0$, let $a_{K+1}, \ldots, a_{K+w}$ be the distinct values of b_j for $j \in \mathcal{J}_3$ in some order. Then $M = K + w$. Let

$$\begin{aligned}
A_0 &= \left(\bigcup_{j \in \mathcal{J}_1 \cup \mathcal{J}_2} \{a : L(\theta_j, a) > n\} \right) \setminus E_\infty, \\
A_{K+1} &= E_\infty.
\end{aligned}$$

It follows that $\{A_0, \ldots, A_{K+1}\}$ is a partiton of $\mathcal{A}$. For $j \in \mathcal{J}_1$,

$$\left| Q_j \left[I_{E_{n_j}} L(\theta_j, \cdot) \right] - \sum_{j=1}^{M} L(\theta_j, a_j) Q_j(A_j) \right| < \frac{\epsilon}{4}, \tag{21}$$

and

$$\frac{\epsilon}{4} \geq Q_j \left[I_{E^C_{n_j}} L(\theta_j, \cdot) \right] \geq L(\theta_j, a_0) Q_j(A_0).$$

Next, we form the partition $D_1, \ldots, D_R$. Let $\gamma = 1/\lceil 4n/\epsilon \rceil$, and for each $\ell = 0, \ldots, K+1$, partition $\Lambda_{\mathcal{A}}$ into sets of the form $\{P : P(A_\ell) = 0\}$ and $\{P : P(A_\ell) \in ([k-1]\gamma, k\gamma]\}$ for $1 \leq k \leq \lfloor 1 + 1/\gamma \rfloor$. Then form the common refinement of these partitions for $\ell = 0, \ldots, K+1$. Let the non-empty sets in that refinement be $C_1, \ldots, C_R$. Then, as a function of P for fixed ℓ and t, $P(A_\ell)$ varies less than $\epsilon/(4n)$ as P varies over C_t. For each x, $\delta(x)(\cdot)$ lies in one and only one C_t. For $t = 1, \ldots, R$, let $D_t = \{x : \delta(x) \in C_t\}$.

Next, we construct the set of non-randomized rules $\{\delta_v\}_{v \in \mathcal{V}}$. Let $\mathcal{V}$ be the set of functions from the set $\{1, \ldots, R\}$ to $\{0, \ldots, M\}$, which has $(M+1)^R$ elements. For each $v \in \mathcal{V}$, define δ_v to be the function defined by

$$\delta_v(x) = \sum_{t=1}^{R} a_{v(t)} I_{D_t}(x).$$

That is, δ_v is the non-randomized rule such that, for each t, $\delta_v(x) = a_{v(t)}$ for all $x \in D_t$.

Next, we construct the weights for each δ_v in the convex combination that defines δ_*. For each $t = 1, \ldots, R$, let $\mu_t \in C_t$. Let $s_{t,\ell} = \mu_t(A_\ell)$ for $\ell = 0, \ldots, K$. If $w > 0$, let $s_{t,\ell} = \mu_t(A_{K+1})/w$ for $\ell = K+1, \ldots, M$. Clearly, $\sum_{\ell=0}^{M} s_{t,\ell} = 1$ for each $t = 1, \ldots, R$. The weights are $r_v = \prod_{t=1}^{R} s_{t,v(t)}$. It is straightforward to show that for each $\ell = 0, \ldots, M$ and each $t = 1, \ldots, R$,

$$\sum_{\{v : v(t) = \ell\}} r_v = \sum_{\{v : v(t) = \ell\}} \prod_{t=1}^{R} \mu_t(A_\ell) = s_{t,\ell}, \tag{22}$$

by induction on R. (The case $R = 1$ is trivial, and for $R > 1$, $\{v : v(t) = \ell\}$ looks just like the $R-1$ case with each weight having an extra factor of $s_{t,\ell}$.) It then follows that $\sum_{v \in \mathcal{V}} r_v = 1$.

Finally, we define $\delta_* = \sum_{v \in \mathcal{V}} r_v \delta_v \in V$ and show that $R(\cdot, \delta_*) \in V$. The risk function of δ_* is

$$\begin{aligned}
R(\theta, \delta_*) &= \sum_{t=1}^{R} P_\theta(D_t) \left[\sum_{v \in \mathcal{V}} r_v L(\theta, a_{v(t)}) \right] \\
&= \sum_{t=1}^{R} P_\theta(D_t) \sum_{\ell=0}^{M} \left[\sum_{v : v(t) = \ell} r_v L(\theta, a_\ell) \right] \\
&= \sum_{t=1}^{R} P_\theta(D_t) \sum_{\ell=0}^{M} s_{t,\ell} L(\theta, a_\ell),
\end{aligned}$$

by (22). For $j \in \mathcal{J}_1 \cup \mathcal{J}_2$ and each $x \in D_t$,

$$\left| \int_{\{a : L(\theta_j, a) \leq n\}} L(\theta_j, a) \delta(x)(da) - \sum_{\ell=0}^{M-w} s_{t,\ell} L(\theta_j, a_\ell) \right| \leq \frac{\epsilon}{4}. \tag{23}$$

Integrate both sides of (23) with respect to P_{θ_j} to see that

$$\left| Q_j\left[I_{\{L(\theta_j,\cdot)\le n\}}L(\theta_j,\cdot)\right] - \sum_{t=1}^{R} P_{\theta_j}(D_t)\sum_{\ell=0}^{M-w} s_{t,\ell}L(\theta_j,a_\ell)\right| < \frac{\epsilon}{4}. \tag{24}$$

For $j\in\mathcal{J}_1$, combine (24) with (20) and the fact that $P_{\theta_j}(D_t)=0$ for each $j\in\mathcal{J}_1$ and each t such that $D_t\subseteq A_{K+1}$ to get

$$|h(\theta_j)-R(\theta_j,\delta_*)| < \frac{\epsilon}{2}.$$

For $j\in\mathcal{J}_2$, we get $R(\theta_j,\delta_*)>c$. For $j\in\mathcal{J}_3$, $R(\theta_j,\delta_*)\ge\epsilon/(8w)L(\theta_j,b_j)=\infty$. So, $\delta_*\in V$.

5 A Sore Point About Pointwise Convergence

The pointwise topology allows convergence of nets to occur where one might not expect it.

Lemma 20 *Let $\mathcal{Z}$ be a set. Let f and g be elements of $\widehat{\mathbb{R}}^{\mathcal{Z}}$. There exists a net $\{f_\eta\}_{\eta\in\mathcal{D}}$ that converges pointwise to g such that for each η, $f_\eta(z)=f(z)$ for all but finitely many z.*

Proof Let $\mathcal{D}$ be the directed set of all finite subsets of $\mathcal{Z}$, as in Example 1. Define the net

$$f_\eta(z)=\begin{cases} g(z) & \text{if } z\in\eta,\\ f(z) & \text{if } z\notin\eta.\end{cases}$$

Clearly, $f_\eta(z)=f(z)$ for all but finitely many z, namely those in the set η. To see that $f_\eta\to g$, let N be a neighborhood of g. Then N contains a set of the form

$$N'=\{h: h(z_j)\in N_j,\ \text{for } j=1,\dots,n\},$$

where n is a finite integer, $z_1,\dots,z_n\in\mathcal{Z}$, $N_1,\dots,N_n$ are open sets in $\widehat{\mathbb{R}}$, and $g\in N'$. The proof will be complete if we can show that there exists η_N such that $\eta_N\le_{\mathcal{D}}\eta$ implies $f_\eta\in N'$. Because $g\in N'$, $g(z_j)\in N_j$ for $j=1,\dots,n$. Let $\eta_N=\{z_1,\dots,z_n\}$. Then $\eta_N\le\eta$ implies $f_\eta(z_j)=g(z_j)$ for $j=1,\dots,n$, and $f_\eta\in N'$.

For $c\in\widehat{\mathbb{R}}$, let

$$H_{\lambda,c}=\{f:\lambda(f)=c\}.$$

When trying to find a Bayes rule with respect to a finitely-additive expectation λ on $\mathcal{M}_\Theta$, we seek functions that lie in $H_{\lambda,c}\cap\mathcal{R}$, for $c=\inf_{f\in\mathcal{R}}\lambda(f)$. A consequence of Lemma 20 is the following result that makes it difficult to find Bayes rules in the pointwise topology.

Lemma 21 *Let λ be a finitely-additive expectation on $\widehat{\mathbb{R}}^{\Theta}$ such that $\lambda(f) = \lambda(g)$ whenever f and g differ at only finitely many values. (That is, $\lambda(A) = 0$ for each finite subset of Θ.) For each pair (c, d) of elements of $\widehat{\mathbb{R}}$ and each $g \in H_{\lambda,d}$, there exists a net contained in $H_{\lambda,c}$ that converges to g.*

Proof Let $f \in H_{\lambda,c}$ and $g \in H_{\lambda,d}$, and apply Lemma 20.

Lemma 21 says that, if λ assigns 0 probability to each finite set, then each $H_{\lambda,c}$ is dense in $\widehat{\mathbb{R}}^{\Theta}$. No matter where you look in the risk set, you will find risk functions "in the neighborhood" that have every possible Bayes risk. Lemma 21 applies to both finitely-additive and countably-additive previsions.

There is one set of prior distributions for which Bayes rules can be found.

Lemma 22 *Let $\mathcal{Z}$ be a set. A finitely-additive expectation λ on $\mathcal{M}_{\mathcal{Z}}$ is continuous in the pointwise topology if and only if λ is simple and is a minimum extension.*

Proof For the "if" direction, note that all evaluation functionals are continuous in the pointwise topology, so every convex combination is continuous. A simple prevision that is a minimum extension is a convex combination of evaluation functionals. For the "only if" direction, we show that all other finitely-additive expectations are discontinuous. First, let λ be a simple finitely-additive expectation that is not a minimum extension. Then there exists a function f such that $\lim_{m\to\infty} \lambda(f \wedge m) < \lambda(f)$ while $\lim_{m\to\infty}(f \wedge m) = f$ pointwise. Next, assume that the restriction of λ to indicators of events takes infinitely many different values. Then there is a countable partition $\mathcal{Z} = \bigcup_{n=1}^{\infty} \mathcal{Z}_n$ such that $\lambda(\mathcal{Z}_n) > 0$ for all n. Let $f_n(z) = I_{\mathcal{Z}_n}(z)/\lambda(\mathcal{Z}_n)$ for each n. Then f_n converges pointwise to 0, but $\lambda(f_n) = 1$ for all n. Next, assume that the restriction of λ to indicator functions takes only finitely many different values. Then there exist finitely many distinct ultrafilters $\mathcal{U}_1, \ldots, \mathcal{U}_n$ such that $\lambda = \sum_{j=1}^{n} \alpha_j P_j$ where each P_j is the probability associated with $\mathcal{U}_j$, each $\alpha_j > 0$ and $\sum_{j=1}^{n} \alpha_j = 1$. Then there is a partion $\mathcal{Z} = \bigcup_{j=1}^{n} \mathcal{Z}_j$ where $\mathcal{Z}_j \in \mathcal{U}_j$ for each j. If each $\mathcal{U}_j$ is principal, then λ is simple. Finally, assume that there is j such that $\mathcal{U}_j$ is non-principal. Then P_j is not countably-additive. Since P_j takes only the values 0 and 1, there must be a partition of $\mathcal{Z}_j = \bigcup_{k=1}^{\infty} A_k$ such that $P_j(A_k) = 0$ for all k. Define $g_m(z) = I_{\bigcup_{i=m}^{\infty} A_i}(z)$. Then g_m converges pointwise to 0, but $\lambda(g_m) = \alpha_j$ for all m.

As a purely mathematical aside, the proof of Lemma 22 could be modified to prove the following.

Proposition 4 *Let $\mathcal{Z}$ be a set. The set of continuous linear functionals on $\mathbb{R}^{\mathcal{Z}}$ with the pointwise topology is the linear span of the evaluation functionals.*

References

K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges.* Academic Press, London, 1983.

W. W. Comfort and S. Negrepontis. *The Theory of Ultrafilters*. Springer-Verlag, New York, NY, 1974.

B. de Finetti. *Theory of Probability*. Wiley, New York, NY, 1974.

J. L. Doob. *Stochastic Processes*. Wiley, New York, NY, 1953.

D. Heath and W. Sudderth. On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.*, 6:333–345, 1978.

J. L. Kelley. *General Topology*. Springer-Verlag, New York, NY, 1955.

M. J. Schervish, T. Seidenfeld, and J. B. Kadane. Infinite previsions and finitely additive expectations, 2014. Online supplement to "Dominating countably many forecasts. *Ann. Statist.* **42**, 728–756 2014, DOI=10.1214/14-AOS1203".

M. J. Schervish, T. Seidenfeld, and J. B. Kadane. What finite additivity can add to decision theory, 2019. *Statist. Meth. and Appl.* to appear.

S. Willard. *General Topology*. Addison-Wesley, Reading, MA, 1970.