Available online at www.sciencedirect.com
SCIENCE DIRECT

Journal of Economic Theory 113 (2003) 32–50

JOURNAL OF Economic Theory

http://www.elsevier.com/locate/jet

# IID: independently and indistinguishably distributed

Larry G. Epstein[a,*] and Martin Schneider[b]

[a] *Department of Economics, University of Rochester, Rochester NY, 14627, USA*
[b] *Department of Economics, UCLA, CA, 90095-1477, USA*

Received 1 May 2002; final version received 21 November 2002

**Abstract**

The inability of the Bayesian model to accommodate Ellsberg-type behavior is well-known. This paper focuses on another limitation of the Bayesian model, specific to a dynamic setting, namely the inability to permit a distinction between experiments that are identical and those that are only indistinguishable. It is shown that such a distinction is afforded by recursive multiple-priors utility. Two related technical contributions are the proof of a strong LLN for recursive multiple-priors utility and the extension to sets of priors of the notion of regularity of a probability measure.


*JEL classification:* D81; D9

*Keywords:* Ambiguity; Multiple-priors; Regular measures; Law of large numbers; Independent beliefs; Independent experiments; Dynamic consistency; Ellsberg Paradox; Robust control; Recursive utility

## 1. Introduction

### 1.1. Objectives

In [6], we provide an axiomatic model of intertemporal utility that accommodates aversion to ambiguity and exhibits dynamic consistency. Because utility is recursive and the model builds on the atemporal multiple-priors model due to Gilboa and Schmeidler [10], we call it *recursive multiple-priors utility*. An attractive feature of

*Corresponding author. Fax: 585-256-2309.
*E-mail addresses:* lepn@troi.cc.rochester.edu (L.G. Epstein), martin@econ.ucla.edu (M. Schneider).


doi:10.1016/S0022-0531(03)00121-2

recursive multiple-priors utility is that, as shown in the companion paper [5], it permits a model of learning under ambiguity.

This paper studies recursive multiple-priors when there is no learning because the random variables generating observations through time are viewed by the agent as independent. Besides being of interest for modeling situations where the agent has learned all she can, the study of independence and the complete absence of learning are of interest also because they constitute benchmarks that are important for proper understanding of the associated model of learning. (Imagine, for example, Bayesian theory in the absence of a compelling notion of probabilistic independence.) In particular, some issues that are discussed in the more general context of [5] permit a sharper treatment here in the absence of learning.

We make four principal contributions. First, we explicate the notion of independence that emerges from recursive multiple-priors utility and compare it with the notion appearing in [10,21]. More particularly, we focus on the counterpart of the IID assumption for our model. Second, we adapt from Walley [20] the distinction between experiments that are viewed as indistinguishable as opposed to identical, and we show how such a distinction can be accommodated by recursive multiple-priors utility. Third, inspired by Marinacci [12], we prove a strong LLN for recursive multiple-priors utility satisfying the IID property.[1] Besides the formal interest in a generalization of the Bayesian LLN, this is of interest because it supports our claims, described shortly, about the way in which our model improves upon the reference Bayesian model. The usual LLN is typically formulated for countably additive probability measures, or equivalently in our setting, for regular probability measures. Our final contribution is to extend the notion of regularity of a probability measure to IID sets of priors. Because this material may be of interest only to the more technically inclined readers and because it is not essential for understanding the other contributions, the treatment of regularity is for the most part confined to an appendix.

### 1.2. The IID bayesian model

Motivation for the study of non-Bayesian models stems from limitations of the Bayesian model. One that is well-known is the inability to accommodate aversion to ambiguity such as demonstrated by the Ellsberg Paradox. Here we borrow from Walley [20, pp. 457–471] and highlight other problematic features of the Bayesian model that are specific to a dynamic setting.

For concreteness, consider a sequence of coins to be tossed sequentially. Suppose that the agent is told that the coins are unrelated. Further she is told the same about each coin, though possibly very little about any of them. The natural state space is $S^\infty$, where $S = \{H, T\}$, and the Bayesian agent forms a prior $P$ on $S^\infty$. Given the symmetry of information about the coins and their unrelatedness, an IID prior is called for. Then the usual strong LLN applies and states that the agent assigns probability 1 to the event that the empirical frequency of the outcome $H$ converges

[1] The relation to Marinacci's LLN result is described below.

to the probability of $H$ on a single toss. In particular, she is certain at the outset that empirical frequencies converge. This expectation is intuitive if she is confident that the coins are identical so that it is as though the same coin is being tossed repeatedly. However, what is the basis for such confidence when the agent is told nothing (or very little) about the coins beyond their being unrelated? In such cases, she would presumably admit the possibility that the coins are *not* identical and thus that empirical frequencies may fail to converge. On the other hand, given the symmetry of information about the coins, she has no reason to distinguish between them.

In summary, the IID Bayesian model imposes that the coins, or more generally, the underlying experiments, are independently and *identically* distributed, while frequently it might be appropriate to adopt only the weaker assumption that the experiments are independently and *indistinguishably* distributed. Such a distinction is not possible within the Bayesian framework, but we show that it is possible within the framework of recursive multiple-priors.[2] In particular, the corresponding LLN reflects the agent's relaxed view about empirical frequencies whereby she is certain that they will lie in an interval, whose size reflects her ambiguity about the degree to which coins differ. (Such a connection between subjective beliefs and empirical frequencies is put forth also by Marinacci [12] in interpreting his results.)

Yet another way to describe the difference between the Bayesian IID model and ours is that, while no learning takes place in either, the reasons differ. In the former, the agent understands the data generating mechanism to the point that she is confident that the coins are identical and thus that there is nothing left to learn. In contrast, the non-Bayesian does not believe that the coins are identical. However, she does not understand how they differ well enough to even formulate a theory about these differences. Thus lack of understanding is the reason that she does not learn.

## 2. Recursive multiple-priors

We work with a finite period state space $S$, identical for all times, so that the (full) state space is $S^\infty = \prod_{t=1}^\infty S_t$, $S_t = S$ all $t \geqslant 1$. At $t$, the agent has observed the realized history $s_1^t = (s_1, \ldots, s_t)$; denote by $\{\mathscr{F}_t\}$ the corresponding filtration. Unless otherwise specified, measures on $S^\infty$ are understood to be defined on $\mathscr{F}_\infty = \sigma(\bigcup_1^\infty \mathscr{F}_t)$ and those on any $S_t$ are understood to be defined on the power set of $S$. Both finitely additive and countably additive measures on $S^\infty$ will be relevant. The corresponding sets are denoted $ba(S^\infty)$ and $ca(S^\infty)$. The corresponding subsets of probability measures are denoted $ba_+^1(S^\infty)$ and $ca_+^1(S^\infty)$. We denote by $\Delta(S)$ the probability simplex for the finite set $S$.

The agent ranks consumption processes $c = (c_t)$ that are adapted to the filtration $\{\mathscr{F}_t\}$. At any time $t = 0, 1, \ldots$, and given the history $s_1^t$, her ordering is represented

[2] See Section 3.2 for a behavioral definition of the distinction.

by the conditional utility function $V_t$, defined recursively by

$$V_t(c; s_1^t) = \min_{Q \in \mathscr{P}_t(s_1^t)} E_Q[u(c_t) + \beta V_{t+1}(c)], \tag{2.1}$$

where $\beta$ and $u$ satisfy the usual properties and where $\mathscr{P}_t(s_1^t)$ is a primitive set of 1-step-ahead measures conditional on the history $s_1^t$. These embody beliefs about the next step (about $s_{t+1}$) given the history of observations $s_1^t$. Such beliefs reflect ambiguity when $\mathscr{P}_t(s_1^t)$ is a nonsingleton. (See [6] for the model's axiomatic underpinnings and for a more detailed discussion.)

The hypothesis that successive realizations of $s_t$ are viewed as independent is naturally expressed by specializing 1-step-ahead conditionals to be independent of history, while indistinguishability is expressed by assuming also time stationarity. Thus suppose that

$$\mathscr{P}_t = \mathscr{L}, \tag{2.2}$$

where $\mathscr{L} \subset \Delta(S)$ is a convex and closed set of probability measures. Assume also mutual absolute continuity within $\mathscr{L}$.

When $\mathscr{L}$ is the singleton $\{\ell\}$, then the 1-step-ahead conditional $\ell$ determines a unique countably additive measure $P$ on $S^\infty$ whose 1-step-ahead conditionals equal $\ell$, namely, $P$ is the countably additive product measure $\otimes_1^\infty \ell$. The Kolmogorov extension theorem ensures this uniqueness but only within $ca(S^\infty)$.

For the general case of nonsingleton $\mathscr{L}$, the latter generates the following set of priors $\mathscr{P}$ on $S^\infty$:[3]

$$\mathscr{P} = cl(\{P \in ca_+^1(S^\infty)\colon\ P_t \in \mathscr{L} \text{ all } t\}), \tag{2.3}$$

where $P_t$ denotes the 1-step-ahead $\mathscr{F}_t$-conditional of $P$ and $cl(\cdot)$ denotes closure with respect to the weak topology on $ba(S^\infty)$, namely the topology induced by bounded measurable functions. We refer to $\mathscr{P}$ as modeling IID beliefs, where IID means *independently and indistinguishably* distributed.

We offer a number of observations to reassure the reader that this is a natural specification of multiple priors $\mathscr{P}$ on the full state space given that 1-step-ahead beliefs are given by $\mathscr{L}$; see Appendix A for a more detailed rationale for (2.3). First, $\mathscr{P}$ is consistent with $\mathscr{L}$ in that it satisfies (2.2). Second, the subset of $\mathscr{P}$ equal to $\{P \in ca_+^1(S^\infty)\colon\ P_t \in \mathscr{L} \text{ all } t\}$ would seem to capture the 'countable additivity' appropriate for nonsingleton sets of priors. However, the multiple-priors model, as in both [10] and [6], imposes the technical condition that the set of priors be weakly closed in $ba(S^\infty)$. Thus we define $\mathscr{P}$ to be the weak closure of the noted set. Note that so defined it is convex, which is another technical condition imposed by the theory.

A critical property of $\mathscr{P}$ is rectangularity, a notion defined in [6]. For the present IID specification, rectangularity takes the form: $P \in \mathscr{P}$ *if and only if* there exist $\ell \in \mathscr{L}$

[3] More precisely, the definition requires that $P_t(\omega) \in \mathscr{L}$ with probability 1 according to every measure generated by $\mathscr{L}$. Because all measures in $\mathscr{L}$ are equivalent, any two induced measures are equivalent on $\bigcup_1^\infty \mathscr{F}_t$ and thus 'every measure' is equivalent to 'some measure.' To simplify notation, '*a.e.*' qualifications are not stated explicitly.

and a collection $\{Q^s \in \mathscr{P}: s \in S\}$ satisfying

$$P(A_1 \times A_{-1}) = \sum_{s \in A_1} \ell(s) Q^s(A_{-1}), \quad \text{for all } A_1 \subset S \quad \text{and} \quad A_{-1} \subset \prod_{t \geqslant 2} S. \tag{2.4}$$

Any $P$ that is a product measure $\otimes_1^\infty \ell_t^*$ with every $\ell_t^* \in \mathscr{L}$ can be written in this form by taking $\ell = \ell_1^*$ and $Q^s = \otimes_2^\infty \ell_t^*$ for each $s$. The essential content of rectangularity is that the measure $P$ defined in (2.4) is in $\mathscr{P}$ for arbitrary choices of $\ell \in \mathscr{L}$ and measures $Q^s$ in $\mathscr{P}$ that may vary with $s$.

Rectangularity has an immediate and revealing implication for the minimum probability of events. If we define

$$\mathscr{P}(A) \equiv \min_{P \in \mathscr{P}} P(A), \quad A \in \mathscr{F}_\infty, \tag{2.5}$$

then we have the recursive relation

$$\mathscr{P}(A) = \min_{\ell \in \mathscr{L}} \left[ \sum_{s \in S} \ell(s) \mathscr{P}(A_s) \right], \tag{2.6}$$

where $A_s = \{(s_t)_2^\infty : (s, (s_t)_2^\infty) \in A\}$. This recursivity plays a critical role below in establishing the LLN.

A final noteworthy implication of rectangularity is the following explicit formula for utility equivalent to (2.1):

$$V_t(c; s_1^t) = \min_{Q \in \mathscr{P}} E_Q \left[ \sum_{s \geqslant t} \beta^{s-t} u(c_s) \mid s_1^t \right]. \tag{2.7}$$

In particular, conditional utility conforms to the multiple-priors model [10], with the set of priors for time $t$ determined by updating the set $\mathscr{P}$ prior-by-prior via Bayes' Rule.

## 3. Independence and identical/indistinguishable

This section is designed to clarify both our notion of independence and the distinction between identical and indistinguishable mentioned in the introduction.

### 3.1. Independence

Our definition of independence is most easily clarified by comparing it with that adopted by Gilboa and Schmeidler [10].[4] Since the time invariance of $\mathscr{L}$ imposed to model indistinguishability is not germane to this comparison of notions of independence, we continue to assume it. With this qualification, the notion of independence used by Gilboa and Schmeidler is captured in the present setting via

[4] Essentially, the same definition is adopted by Walley and Fine [21] to prove a LLN. However, their analysis, unlike that of Gilboa and Schmeidler, is not tied explicitly to the multiple-priors model of utility. For other discussions of independence see [9,11].

the set of priors $\mathcal{P}^{\mathrm{GS}}$, where

$$\mathcal{P}^{\mathrm{GS}} = \overline{co}\left\{ \bigotimes_{t=1}^{\infty} \ell_t \colon \ell_t \in \mathcal{L} \right\};$$

$\overline{co}(\cdot)$ denotes the weakly closed convex hull. It is readily seen that $\mathcal{P}^{\mathrm{GS}}$ is a strict subset of $\mathcal{P}$.[5] We proceed to examine differences between the two sets. Convexity and closure are both technical regularity conditions; the former is imposed to ensure uniqueness of the representing set of priors as in [10, Theorem 1], and the latter to justify writing *min* rather than *inf* when defining utility as in (2.7), for example. Thus the essential comparison is between

$$\hat{\mathcal{P}}^{\mathrm{GS}} = \left\{ \bigotimes_{t=1}^{\infty} \ell_t \colon \ell_t \in \mathcal{L} \right\} \quad \text{and} \quad \hat{\mathcal{P}} = \{ P \in ca_+^1(S^\infty) \colon P_t \in \mathcal{L} \text{ all } t \}.$$

One difference is that while $\hat{\mathcal{P}}^{\mathrm{GS}}$ consists entirely of product measures, that is not the case for $\hat{\mathcal{P}}$. For example, in the coin tossing setting with $S = \{H, T\}$, suppose that $\mathcal{L}$ is the convex hull of the measures $(\frac{1}{3}, \frac{2}{3})$ and $(\frac{2}{3}, \frac{1}{3})$, that is, beliefs about the outcome $H$ on any single toss are modeled by the probability interval $[\frac{1}{3}, \frac{2}{3}]$. Then it is apparent from (2.4) that $\hat{\mathcal{P}}$ contains $P$ such that

$$P(H_1) = \tfrac{2}{3}, \quad P(H_1, H_2) = \tfrac{2}{3} \cdot \tfrac{2}{3} \quad \text{and} \quad P(T_1, H_2) = \tfrac{1}{3} \cdot \tfrac{1}{3},$$

so that $P(H_2 \mid H_1) \neq P(H_2 \mid T_1)$.

At a formal (functional form) level, this feature of our definition may strike the reader as odd. However, sets of priors are properly interpreted only as part of the representation of preference and properly evaluated only via their behavioral implications. For our model, these are clear from (2.1) and (2.2) —the conditional preference order at any time-event pair does not depend on history, reflecting perceived independence between experiments. On the other hand, the significance of the specification $\hat{\mathcal{P}}^{\mathrm{GS}}$ for dynamic behavior is unclear. That is because it violates rectangularity and hence, by Epstein and Schneider [6], is not compatible with dynamically consistent preferences. Thus, preferences alone do not determine behavior without a specification of how intrapersonal conflicts are resolved and this is not addressed in [10].

The upshot is that our model and the associated notion of independence take time ‘seriously,’ while Gilboa and Schmeidler are concerned with a formally atemporal, one-shot choice setting. In their model, the experiments could be viewed as occurring simultaneously, where no economic decisions are made between realizations, in which case $\hat{\mathcal{P}}^{\mathrm{GS}}$ arguably captures an intuitive notion of cross-sectional independence. On the other hand, our approach is not applicable to a cross-sectional setting. That is because recursive multiple-priors utility presumes a given sequential ordering

[5] A curiosity is that if used to define utility as in (2.7), then both $\mathcal{P}$ and $\mathcal{P}^{\mathrm{GS}}$ lead to the same utility for consumption processes $c$ such that $c_\tau$ depends only on the current state $s_\tau$. That utility equals $\sum_{\tau \geqslant t} \beta^{\tau - t} \min_{\ell \in \mathcal{L}} E_\ell u(c_\tau)$.

of experiments and the latter is available only in a temporal setting, where the order in which experiments are conducted is typically given.

### 3.2. Identical vs. indistinguishable

The importance of this distinction rests on its being behaviorally meaningful. This raises the question: "what behavior would reveal that the agent views the experiments as indistinguishable but not necessarily identical?"

To illustrate our response, turn again to the coins example from the introduction. Consider the choice at time 0 between bets on the first two outcomes where prizes are awarded at time 2 and are denominated in utils.[6] If evidence about the coins is symmetric, then indifference between betting on $HT$ and $TH$ seems intuitive. Consider now the choice between either of the above bets and $\frac{1}{2}HT + \frac{1}{2}TH$, the act paying $\frac{1}{2}$ if $\{HT, TH\}$ and 0 otherwise. If the coins are identical, then there is nothing to be gained by mixing and one would expect the rankings

$$HT \sim TH \quad \text{and} \quad \tfrac{1}{2}HT + \tfrac{1}{2}TH \sim HT.$$

On the other hand, uncertainty that the coins are identical plausibly generates an incentive to smooth utility across $HT$ and $TH$, leading to a strict preference for the mixture. Thus we take the rankings

$$HT \sim TH \quad \text{and} \quad \tfrac{1}{2}HT + \tfrac{1}{2}TH \succ HT \tag{3.1}$$

as reflecting that the coins are viewed as indistinguishable but not identical.

For our IID model consisting of (2.2), (2.3) and (2.7), where we assume that $\min_{\ell \in \mathscr{L}} \ell(H)$ and $\min_{\ell \in \mathscr{L}} \ell(T)$ are both positive, the rankings in (3.1) are valid if and only if $\mathscr{L}$ is not a singleton (the proof follows below). Thus the Bayesian IID model (singleton $\mathscr{L}$) is incapable of modeling 'indistinguishable but not identical.' The way in which this is accommodated within the multiple-priors framework is roughly that the use of the same set $\mathscr{L}$ at every time delivers indistinguishability, while the nonsingleton nature of $\mathscr{L}$ admits the possibility that coins differ.

The preceding claim can be proven as follows: Abbreviate $\min_{\ell \in \mathscr{L}} \ell(H)$ by $\mathscr{L}(H)$ and similarly for $T$. Then by (2.6) $V_0(HT) = \beta^2 \min_{P \in \mathscr{P}} P(HT) = \beta^2 \mathscr{L}(H)\mathscr{L}(T) = V_0(TH)$. From (2.4),

$$\begin{aligned} V_0(\tfrac{1}{2}HT + \tfrac{1}{2}TH) &= \beta^2 \min_{P \in \mathscr{P}} [\tfrac{1}{2}P(HT) + \tfrac{1}{2}P(TH)] \\ &= \beta^2 \min_{\substack{\ell \in \mathscr{L} \\ Q^H, Q^T \in \mathscr{P}}} [\tfrac{1}{2}\ell(H)Q^H(T) + \tfrac{1}{2}\ell(T)Q^T(H)] \\ &= \beta^2 \min_{\ell \in \mathscr{L}} [\tfrac{1}{2}\ell(H)\mathscr{L}(T) + \tfrac{1}{2}\ell(T)\mathscr{L}(H)] \\ &= \beta^2 \begin{cases} \tfrac{1}{2}\mathscr{L}(H)\mathscr{L}(T) + \tfrac{1}{2}(1 - \mathscr{L}(H))\mathscr{L}(H) & \text{if } \mathscr{L}(T) \geqslant \mathscr{L}(H), \\ \tfrac{1}{2}(1 - \mathscr{L}(T))\mathscr{L}(T) + \tfrac{1}{2}\mathscr{L}(T)\mathscr{L}(H) & \text{if } \mathscr{L}(T) \leqslant \mathscr{L}(H). \end{cases} \end{aligned}$$

[6] Formally, the bets correspond to Anscombe–Aumann acts.

Thus $V_0(\frac{1}{2}HT + \frac{1}{2}TH) = \frac{1}{2}V_0(HT) + \frac{1}{2}V_0(TH) = \beta^2 \mathscr{L}(H)\mathscr{L}(T) \Rightarrow \mathscr{L}(T) + \mathscr{L}(H) = 1 \Rightarrow \min_{\ell \in \mathscr{L}} \ell(H) = \max_{\ell \in \mathscr{L}} \ell(H)$, in other words, $\mathscr{L}$ is a singleton.

The preceding is readily generalized beyond the specific example to the setting of the abstract state space $S$. For any $t \geqslant 2$, let $h$ denote the bet on the outcome sequence $\bar{s}_1^t = (\bar{s}_1, \ldots, \bar{s}_t)$ that delivers payoff $h_t$ time $t$, where

$$h_t(s_1, \ldots, s_t) = \begin{cases} 1 & \text{if } s_1^t = \bar{s}_1^t, \\ 0 & \text{otherwise,} \end{cases}$$

and where the payoff is denominated in utils. By associating a zero payoff to all other time periods we may identify $h$ with a random payoff stream and thus compute its utility via $V_0$ defined in (2.7). For any permutation $\pi$ on $\{1, \ldots, t\}$, $\pi h$ denotes the 'permuted act'

$$\pi h(s_1, \ldots, s_t) = h(\pi^{-1} s_1^t) \quad \text{for all } (s_1, \ldots, s_t) \text{ in } S^t,$$

which is a bet on the permuted sequence $\pi \bar{s}_1^t = (s_{\pi(1)}, \ldots, s_{\pi(t)})$. Then the generalization of (3.1) is: for all $t$, $\pi$ and $h$ as above,

$$V_0(h) = V_0(\pi h) \quad \text{and} \quad V_0(\tfrac{1}{2}h + \tfrac{1}{2}\pi h) \geqslant V_0(h),$$

with strict inequality on the right occurring for every $t$ and $\pi$ for some choice of $h$. This is delivered by our model if and only if $\mathscr{L}$ is not a singleton. (The proof is a straightforward extension of the above argument.)

## 4. Strong LLN

Let $X : S \to \mathbf{R}^1$ and $X_t \equiv X(s_t)$ for each $t$. For the Bayesian special case of our model, where $\mathscr{L} = \{\ell^*\}$ and $\mathscr{P} = \{P^*\}$ are singletons, with $P^*$ being IID, the strong LLN states that

$$P^* \left\{ \lim \sum_{t=1}^{k} X_t / k = E_{\ell^*} X \right\} = 1. \tag{4.1}$$

As discussed in Section 1.2, certainty that sample averages converge is not intuitive in many situations. Here we establish a version of the LLN for the IID recursive multiple-priors model that reflects a less confident agent, in particular, one who is not certain that the experiments underlying the $X_t$'s are identical.

To proceed, restrict $\mathscr{L}$ further by assuming that its lower envelope $v$ defined by

$$v(A) = \min_{\ell \in \mathscr{L}} \ell(A), \quad A \subset S \tag{4.2}$$

is supermodular.[7] It is well-known that this characterizes sets $\mathscr{L}$ that conform to the Choquet model [19] in the sense that

$$\min_{\ell \in \mathscr{L}} \int f \, d\ell = \int f \, dv \equiv E_v f, \tag{4.3}$$

for all $f : S \to \mathbf{R}^1$, where the integral $\int f \, dv$ is in the sense of Choquet.

[7] $v$ is said to be supermodular or convex if $v(A \cup B) + v(A \cap B) \geqslant v(A) + v(B)$, for all $A, B \subset S$.

Though restrictive, assumption (4.3) is common and admits many specifications for $\mathscr{L}$. Moreover, it is totally unrestrictive in the binary case. Thus when $S = \{R, B\}$, and wlog if $f(R) \geqslant f(B)$, then for any $\mathscr{L}$,

$$\min_{\ell \in \mathscr{L}} [\ell(R)f(R) + \ell(B)f(B)] = \left(\min_{\ell \in \mathscr{L}} \ell(R)\right)f(R) + \left(1 - \min_{\ell \in \mathscr{L}} \ell(R)\right)f(B) = E_v f,$$

for the capacity defined in (4.2).

**Theorem 4.1.** *For the finite IID model, where $\mathscr{L}$ conforms to the Choquet model as in* (4.3), *we have*

$$\min_{P \in \mathscr{P}} P\left\{E_v X \leqslant \liminf \sum_{t=1}^{k} X_t/k \leqslant \limsup \sum_{t=1}^{k} X_t/k \leqslant -E_v(-X)\right\} = 1, \quad (4.4)$$

$$\min_{P \in \mathscr{P}} P\left\{E_v X < \liminf \sum_{t=1}^{k} X_t/k\right\} = 0 = \min_{P \in \mathscr{P}} P\left\{\limsup \sum_{t=1}^{k} X_t/k < -E_v(-X)\right\}. \quad (4.5)$$

In the Bayesian case, $v$ is additive and $E_v X = -E_v(-X)$, which delivers the standard strong LLN.

Condition (4.4) is the central result and constitutes the counterpart of $P^*$-a.s. convergence in the Bayesian case. The agent is certain that sample averages lie in the interval

$$[E_v X, \ -E_v(-X)] = \left[\min_{\ell \in \mathscr{L}} \int X \, d\ell, \ \max_{\ell \in \mathscr{L}} \int X \, d\ell\right],$$

whose size is determined by the set of likelihoods $\mathscr{L}$. In the coins example, where $X = 1$ or 0 according as the outcome is $H$ or $T$, the endpoints of the interval are the minimum and maximum probabilities for $H$.

The second result (4.5) describes a sense in which the interval provides tight bounds. Further results follow immediately from the superadditivity property: for all disjoint $B$ and $C$ in $\mathscr{F}_\infty$,

$$\min_{P \in \mathscr{P}} P(B \cup C) \geqslant \min_{P \in \mathscr{P}} P(B) + \min_{P \in \mathscr{P}} P(C).$$

For example, an immediate consequence of the preceding and (4.4) is that

$$\min_{P \in \mathscr{P}} P\left\{\liminf \sum_{t=1}^{k} X_t/k < E_v X\right\} = 0 = \min_{P \in \mathscr{P}} P\left\{-E_v(-X) < \limsup \sum_{t=1}^{k} X_t/k\right\}.$$

Marinacci [12] proves a (strong) LLN for nonadditive probabilities that is very general in many respects, but it does not cover our result.[8] One reason is that he assumes that his set $\mathscr{P}$ conforms to the Choquet model, while we adopt the strictly weaker assumption that 1-step-ahead beliefs in the form of $\mathscr{L}$ conform to Choquet as in (4.3). In fact, our set of priors $\mathscr{P}$, because it is rectangular, conforms to the Choquet model only in the degenerate case where $\mathscr{L}$ and $\mathscr{P}$ are singletons, which means that Marinacci's analysis has little to say about recursive multiple-priors.[9] Though we assume structure, in the form of rectangularity, that Marinacci does not, this added structure has two obvious advantages. First, it ensures that our analysis is tied to a coherent model of dynamic choice, namely recursive multiple-priors utility. Second, rectangularity facilitates a simpler and more transparent proof. In particular, our proof exploits heavily the recursivity ((2.6), for example) delivered by rectangularity and reveals the analytical power that such recursivity affords. For these reasons we view the proof as an important part of the message and thus we include it in the main body of the paper (see the next section).

Finally, turn from the formal result to the behavioral interpretation of our LLN. Though one is tempted to interpret the LLN as above in terms of willingness to bet on or against the events appearing in (4.4) and (4.5), such an interpretation is not justified by the decision-theoretic foundations in [6] and outlined in Section 2. That is because the indicator function for $\{E_v X < \liminf \sum_{t=1}^{k} X_t/k\}$, for example, is not an adapted consumption process and thus does not lie in the domain of recursive multiple-priors utility. In fact, this difference is decidedly nontrivial because the events appearing in the statement of the LLN are tail events while the realized consumption profile depends only on the true event in $\bigcup_1^\infty \mathscr{F}_t$.[10] Similar remarks apply *even in the Bayesian case*—the fact that the Savage prior $P^*$ assigns probability 1 to the event that sample averages converge, as in (4.1), has no implications for the ranking of consumption processes. This is a restatement of the point emphasized by de Finetti that because tail events are unobservable, limit laws formulated in terms of such events are of questionable importance for applications (see [17] for extensive discussion of de Finetti's view).

One response in the Bayesian case has been to formulate limit laws that may be expressed exclusively in terms of events in $\bigcup_1^\infty \mathscr{F}_t$ and that are equivalent to usual statements if countable additivity is assumed (see [3,17], for example). We suspect that such reformulations are possible also in our multiple-priors framework, but further examination is beyond the scope of this paper. Since the assumption of countable additivity is so widespread and permits much simpler formulations of the LLN, we feel that the corresponding analysis for multiple-priors is of interest in spite of the gap in supporting behavioral interpretations.

[8] As explained in [12, p. 148], Marinacci's LLN generalizes Walley and Fine [21]. The following comparison with Marinacci's LLN applies also to [21].

[9] A special case of this claim is proven in [18, Theorem 3.1].

[10] The tail $\sigma$-field $\mathscr{F}^{\text{tail}} = \bigcap_{j=1}^\infty \sigma(\bigcup_{t=j}^\infty \mathscr{S}_t)$, where $\mathscr{S}_t$ is the (power set) $\sigma$-algebra on $S_t$, consists of all events that can never be known from finitely many observations.

## 5. Proof of LLN

The structure of the proof is as follows: First, show that (4.4) is equivalent to the conjunction of

$$\min_{P \in \mathscr{P}} P\left\{ E_v X \leqslant \liminf \sum_{t=1}^{k} X_t/k \right\} = 1 \tag{5.1}$$

and

$$\min_{P \in \mathscr{P}} P\left\{ \limsup \sum_{t=1}^{k} X_t/k \leqslant -E_v(-X) \right\} = 1. \tag{5.2}$$

The key then is to show that in each case there is a minimizing measure $P^*$, different across cases, that is IID; rectangularity of $\mathscr{P}$ renders this step straightforward. Finally, apply the Bayesian LLN to $P^*$ . Some details that rely (implicitly) on regularity of $\mathscr{P}$ complete the proof.

**Proof.** Condition (4.4) is equivalent to (5.1) and (5.2): It is clear that (4.4) implies (5.1) and (5.2). To prove the converse, write $A = \{E_v X \leqslant \liminf \sum_{t=1}^{k} X_t/k\}$ and $B = \{\limsup \sum_{t=1}^{k} X_t/k \leqslant -E_v(-X)\}$. Then $\min_{P \in \mathscr{P}} P(A \cap B) = \min_{P \in \mathscr{P}} \{1 - P(A^c \cup B^c)\} = 1 - \max_{P \in \mathscr{P}} [P(A^c \cup B^c)] \geqslant 1 - \max_{P \in \mathscr{P}} [P(A^c) + P(B^c)] \geqslant 1 - \max_{P \in \mathscr{P}} P(A^c) - \max_{P \in \mathscr{P}} P(B^c) = 1$.

Prove (5.1); the argument for (5.2) is similar. The assumption (4.3) for $\mathscr{L}$ that it corresponds to a Choquet integral implies that there exists $\ell^* \in \mathscr{L}$ satisfying

$$\sum_{s \in S} \ell(s) X(s) \geqslant E_v X = \sum_{s \in S} \ell^*(s) X(s)$$

and

$$\sum_{s \in S} \ell(s) f(s) \geqslant E_v f = \sum_{s \in S} \ell^*(s) f(s) \tag{5.3}$$

for all $\ell$ and for any $f$ comonotone with $X$.[11] Define $P^*$ to be the (countably additive) IID product measure induced by $\ell^*$. The strong law for $P^*$ implies that

$$1 = P^*\left\{ E_v X = \lim \sum_{t=1}^{k} X_t/k \right\} \leqslant P^*\left\{ E_v X \leqslant \liminf \sum_{t=1}^{k} X_t/k \right\} \leqslant 1.$$

Thus for (5.1), it suffices to establish that

$$P\left\{ E_v X \leqslant \liminf \sum_{t=1}^{k} X_t/k \right\} \geqslant P^*\left\{ E_v X \leqslant \liminf \sum_{t=1}^{k} X_t/k \right\} \tag{5.4}$$

for all $P$ in $\mathscr{P}$.

[11] $X$ and $f$ are comonotone if $(X(s') - X(s))(f(s') - f(s)) \geqslant 0$ for all $s'$ and $s$ in $S$.

Intuition for (5.4) is as follows: Denote by $\Psi_{X,\ell}$ the cdf on $\mathbf{R}^1$ induced by the random variable $X$ and the measure $\ell$. Then (5.3) implies

$$\Psi_{X,\ell} \geqslant_{\text{FSD}} \Psi_{X,\ell^*}.$$

($\geqslant_{\text{FSD}}$ denotes the first-order stochastic dominance relation for cdf's on the real line.) Thus, at each $t$, the 1-step-ahead distribution under $P$ dominates that under $P^*$ in the FSD sense. One expects that this should lead to first-order dominance also for distributions induced on $\mathbf{R}^\infty$. In other words, we claim that

$$P\{(X_1, \dots, X_t, \dots) \in A\} \geqslant P^*\{(X_1, \dots, X_t, \dots) \in A\}$$

for any measurable $A \subset \mathbf{R}^\infty$ that is *increasing* in the sense that

$$x \in A \text{ and } x' \geqslant x \Rightarrow x' \in A.$$

Intuition is sharper in the binary case. Let $S = \{H, T\}$, $X(H) = 1$ and $X(T) = 0$. Then $\ell(H) \geqslant \ell^*(H)$ for all $\ell$ implies that every $P$ attaches higher probability than $P^*$ to the next step being $H$. It 'should' follow that any event $\{a \leqslant \liminf \sum_{t=1}^k X_t/k\}$ has weakly higher probability under $P$; similarly for other increasing sets.

To proceed with the formal argument, write $\tilde{X}(s_1, s_2, \dots) = (X(s_t))_1^\infty$ and, for any given $A$, define

$$\tilde{X}^{-1}(A) = \{(s_t) : \tilde{X}(s_1, s_2, \dots) \in A\}.$$

Denote by $P\tilde{X}^{-1}$ the induced measure on $\mathbf{R}^\infty$ and adopt the notation

$$\mathscr{P}(B) \equiv \min_{P \in \mathscr{P}} P(B), \quad \text{all } B \in \mathscr{F}_\infty.$$

Our objective is to prove, for all $P$ in $\mathscr{P}$,

$$P(\tilde{X}^{-1}(A)) \geqslant P^*(\tilde{X}^{-1}(A)) \text{ for all increasing } A \subset \mathbf{R}^\infty. \tag{5.5}$$

Then, because $P^*$ lies in $\mathscr{P}$, it would follow that

$$\mathscr{P}(\tilde{X}^{-1}(A)) = P^*(\tilde{X}^{-1}(A)) \text{ for all increasing } A \subset \mathbf{R}^\infty.$$

We proceed in three steps.

*Step* 1: Prove (5.5) for increasing sets $A$ such that

$$\tilde{X}^{-1}(A) \in \mathscr{F}_t \text{ for some finite } t. \tag{5.6}$$

For any such $A$, rectangularity and IID deliver (recall (2.4))

$$P(\tilde{X}^{-1}(A)) = \sum_{s_1 \in S} \ell(s_1) Q^{s_1}(\{(s_t)_2^\infty : (X(s_1), (X(s_t))_2^\infty) \in A\}).$$

In computing the minimum as $P$ varies over $\mathscr{P}$, note that each $Q^{s_1}$ can be varied independently over $\mathscr{P}$. Thus one obtains the following recursion analogous to (2.6):

$$\mathscr{P}(\tilde{X}^{-1}(A)) = \min_{\ell \in \mathscr{L}} \left[ \sum_{s_1 \in S} \ell(s_1) \mathscr{P}(\tilde{X}^{-1}(A_{s_1})) \right], \tag{5.7}$$

where

$$A_{s_1} = \{(X(s_t))_2^\infty : (X(s_1), (X(s_t))_2^\infty) \in A,\ (s_t))_2^\infty \in S^\infty\}.$$

Note that $X(s_1') \geqslant X(s_1) \Rightarrow A_{s_1'} \supset A_{s_1} \Rightarrow \mathscr{P}(\tilde{X}^{-1}(A_{s_1'})) \geqslant \mathscr{P}(\tilde{X}^{-1}(A_{s_1}))$. The implied comonotonicity implies further, by (5.3), that $\ell^*$ is a minimizer in (5.7), and hence that

$$\mathscr{P}(\tilde{X}^{-1}(A)) = \sum_{s_1 \in S} \ell^*(s_1) \mathscr{P}(\tilde{X}^{-1}(A_{s_1})). \tag{5.8}$$

Argue by induction on $t$. If $t = 1$, then $A$ has the form $A = \{X(s_1) \in I\}$ for some increasing interval $I$ in the real line, $A_{s_1} = \emptyset$ or $\mathbf{R}^\infty$ according as $X(s_1) \notin I$ or $\in I$; hence

$$\mathscr{P}(\tilde{X}^{-1}(A)) = \sum_{s_1 \in X_1^{-1}(I)} \ell^*(s_1) = P^*(\tilde{X}^{-1}(A)),$$

which is the appropriate version of (5.5). If $t > 1$, then $A_{s_1}$ is increasing and $\tilde{X}^{-1}(A_{s_1}) \in \mathscr{F}_{t-1}$. Thus the induction hypothesis combined with (5.8) implies that

$$\mathscr{P}(\tilde{X}^{-1}(A)) = \sum_{s_1 \in S} \ell^*(s_1) P^*(\tilde{X}^{-1}(A_{s_1})) = P^*(\tilde{X}^{-1}(A)).$$

*Step* 2: Denote by $\pi_t$ the projection operator from $\mathbf{R}^\infty$ to $\mathbf{R}^t$ and let

$$A_t = \pi_t A \times \mathbf{R} \times \mathbf{R} \times \dots\ .$$

Then $A_t \searrow A$ if $A$ is closed:[12] Let $x \in \cap A_t$ so that for every $t$, $x = (x_1, \dots, x_t, y^t) \in A$ for some $y^t$ in $\prod_{i>t} \mathbf{R}$. By the nature of the product topology, $(x_1, \dots, x_t, y^t) \to x$, which must lie within the closed set $A$.

*Step* 3: Prove (5.5) for countably additive measures $P$. First, if $A$ is closed, then $A_t$ is increasing and $\tilde{X}^{-1}(A_t)$ lies in $\mathscr{F}_t$. By Step 1, $P(A_t) \geqslant P^*(A_t)$. This implies that $P(A_t) \searrow P(A)$ and similarly for $P^*$, which proves (5.5) for $A$ closed.

To generalize (5.5) to arbitrary increasing sets, observe that

$$P\tilde{X}^{-1}(A) = \sup\{P\tilde{X}^{-1}(K) \colon K \text{ closed},\ K \subset A\}, \tag{5.9}$$

because $P\tilde{X}^{-1}$ is a regular measure. Define $K^{\text{inc}} = \{x \in \mathbf{R}^\infty : \exists y \in K, x \geqslant y\}$. Then $K^{\text{inc}}$ is closed and increasing. Moreover, if $A$ is increasing, then $K \subset A \Rightarrow K^{\text{inc}} \subset A$ and

$$P\tilde{X}^{-1}(A) = \sup\{P\tilde{X}^{-1}(K) \colon K \text{ closed and increasing},\ K \subset A\}.$$

Since a similar relation holds for $P^*$, (5.5) follows.

[12] Note that tail events are not closed and thus cannot be approximated in this way. Indeed, if $A$ is a tail event in $\mathbf{R}^\infty$, then $\pi_t A = \mathbf{R}^t$ and $A_t = \mathbf{R}^\infty$ for all $t$.

*Step* 4: Countably additive measures are dense in $\mathscr{P}$; see (2.3). Thus (5.5) is established.

Because (5.5) applies to all increasing sets $A$, it yields also that

$$\min_{P \in \mathscr{P}} P\left\{ E_v X < \liminf \sum_{t=1}^{k} X_t/k \right\} = P^*\left\{ E_{\ell^*} X < \liminf \sum_{t=1}^{k} X_t/k \right\} = 0,$$

where the last equality follows from the LLN for $P^*$. This proves the first equality in (4.5).

The remainder of the proof is evident. □

**Acknowledgments**

Epstein gratefully acknowledges the financial support of the NSF (Grant SES-9972442). We are grateful also to Tan Wang and especially Massimo Marinacci for comments.

## Appendix A. Regularity

This appendix is an attempt to clarify and provide foundations for (2.3). The issue is that, while in the finite horizon context 1-step-ahead conditionals uniquely determine a rectangular $\mathscr{P}$ (apply backward induction to (2.4)), that is not true if the horizon is infinite. For example, rectangularity restricts measures only on $\bigcup_1^\infty \mathscr{F}_t$, leaving scope for arbitrary assignments on the tail $\sigma$-field. Even in the Bayesian special case, we know from the Kolmogorov Theorem that uniqueness can be ensured only given regularity, which in our setting is equivalent to countable additivity. Thus one perspective on (2.3) to be described below is that it embodies a form of regularity that is appropriate for nonsingleton sets of priors. (See [15] for the definition of regularity of a measure and a statement of the Kolmogorov Theorem.)

More precisely, the objective of this appendix is to define a notion of regularity for sets of priors that extends the usual definition for probability measures and for which the following theorem can be proven:

**Theorem A.1.** *The set* $\mathscr{P}$ *defined in* (2.3) *is the unique regular and rectangular set of priors conforming with* $\mathscr{L}$, *that is*, *satisfying* (2.4).

In the present setting with state space $S^\infty$, regularity of a probability measure is equivalent to countable additivity. Thus one might view the relevant task as extending the notion of countable additivity from single to multiple-priors. Some seemingly natural ways of doing so have been studied in the literature. For example, Marinacci et al. [13] adopt an axiom of monotone continuity very much like the one used in [1] to deliver countable additivity of the Savage prior, and they show that it characterizes a form of compactness for $\mathscr{P}$; a related analysis and compactness property appear in [7, pp. 43–45]. However, these compactness properties imply that all measures in $\mathcal{P}$ are absolutely continuous with respect to some $P^*$ in $\mathcal{P}$ and thus that there is asymptotic merging to a single measure as in [2]. Such asymptotic vanishing of ambiguity is intuitive in some but possibly not in all situations; for example, it may not be intuitive in the IID environment of this paper where nothing is learned because experiments are not viewed as identical. Thus we do not want to impose it a priori on our theoretical framework. In particular, note that for the set $\mathcal{P}$ defined in (2.3), there is no single $P^*$ for which $P \ll P^*$ for all $P$ in $\mathcal{P}$.

We proceed in several steps towards Theorem A.1. First, we list some properties of $\mathcal{P}$ that follow from definition (2.3), restated here for convenience:

$$\mathcal{P} = cl(\{P \in ca_+^1(S^\infty):\ P_t \in \mathcal{L} \text{ all } t\}).$$

Below refer to the weak topology on $ba(S^\infty)$ induced by bounded measurable functions as the weak topology. By the weak convergence (wc) topology on $ca(S^\infty)$ we mean the topology on $S^\infty$ induced by continuous functions; since $S$ is finite, it has a natural topology and $S^\infty$ is endowed with the product topology, which renders it compact metric.

**Lemma A.2.** (i) *$\mathcal{P}$ is nonempty, rectangular, weakly closed and convex.*

(ii) *$\mathcal{P} \cap ca(S^\infty)$ is nonempty, wc-closed in $ca(S^\infty)$ and weakly dense in $\mathcal{P}$.*

(iii) *$\mathcal{P} \cap ca(S^\infty)$ is weakly closed relative to $ca(S^\infty)$, that is,*

$$P^n \to P \in ca(S^\infty),\ P^n \in \mathcal{P} \cap ca(S^\infty)] \Rightarrow P \in \mathcal{P}.$$

(iv) *For any $P$ in $\mathcal{P}$, there exists $P^*$ in $\mathcal{P} \cap ca(S^\infty)$ such that $P_t = P_t^*$ for all $t$.*

**Proof.** (i) Verify that $\{P \in ca_+^1(S^\infty):\ P_t \in \mathcal{L} \text{ all } t\}$ is a rectangular set and observe that rectangularity is preserved by taking the weak closure.

(ii) Denseness is by construction. Nonemptiness follows from Kolmogorov. Alternatively, it is implied by denseness of $\mathcal{P} \cap ca(S^\infty)$ and nonemptiness of $\mathcal{P}$. For wc-closed, let $P^n \in \mathcal{P} \cap ca(S^\infty)$ and $P^n \overset{wc}{\to} P \in ca(S^\infty)$. Then the 1-step-ahead conditionals $P_t^n$ lie in $\mathcal{L}$, which is wc-closed ($S$ is finite), and $P_t^n \overset{wc}{\to} P_t$. Hence $P_t \in \mathcal{L}$ and $P \in \mathcal{P}$.

(iii) Straightforward.

(iv) Let $Q$ be any countably additive measure in $\mathcal{P}$; it exists by (ii). Define a sequence of measures on $S^\infty$ via

$$P^n(\cdot) = \int Q(\cdot \mid \mathcal{F}_n)(\omega)\, dP(\omega).$$

Roughly, $P^n$ is constructed so as to agree with $P$ on $\mathcal{F}_n$. The construction ensures also that $P^n$ is countably additive (because $S$ is finite and $Q$ is countably additive) and, by rectangularity, that it lies in $\mathcal{P}$. Thus $\{P^n\}$ is a sequence in $\mathcal{P} \cap ca(S^\infty)$, which is wc-compact. Let $P^{n_k} \overset{wc}{\to} P^*$. Given any $t$, eventually every $P^{n_k}$ agrees with $P$

on $\mathscr{F}_t$. Therefore, $P^*$ agrees with $P$ on any $\mathscr{F}_t$ as asserted. (We can say more: The Kolmogorov Theorem ensures that $P^*$ is unique, that is, it equals the limit also for any other subsequence of $\{P^n\}$. In other words, $P^n \overset{\text{wc}}{\to} P^*$.) □

**Corollary A.3.** (a) *The set $\mathscr{P}$ defined by* (2.3) *is the smallest rectangular weakly closed and convex set of priors $\mathscr{P}'$ conforming with $\mathscr{L}$ and such that $\mathscr{P}' \cap ca(S^\infty)$ is nonempty and wc-closed in $ca(S^\infty)$.*

(b) *For any rectangular weakly closed and convex set of priors $\mathscr{P}'$ that conforms with $\mathscr{L}$, $\mathscr{P}' \cap ca(S^\infty) \subset \mathscr{P} \cap ca(S^\infty)$. Moreover, the latter two sets coincide if $\mathscr{P}' \cap ca(S^\infty)$ is nonempty and wc-closed in $ca(S^\infty)$. In the latter case, if also $\mathscr{P}' \cap ca(S^\infty)$ is weakly dense in $\mathscr{P}'$, then $\mathscr{P}' = \mathscr{P}$.*

**Proof.** (a) Argue as in the proof of part (iv) of the lemma that $\mathscr{P}'$ contains $\{P \in ca^1_+(S^\infty)\colon P_t \in \mathscr{L} \text{ all } t\}$. Because $\mathscr{P}'$ is also weakly closed, it contains $\mathscr{P}$.

(b) $Q \in \mathscr{P}' \cap ca(S^\infty) \Rightarrow Q_t \in \mathscr{L} \Rightarrow Q \in \mathscr{P} \cap ca(S^\infty)$. The second assertion follows from (a). The asserted equality follows from $\mathscr{P}' \cap ca(S^\infty) = \mathscr{P} \cap ca(S^\infty) = \{P \in ca^1_+(S^\infty)\colon P_t \in \mathscr{L} \text{ all } t\}$. □

The corollary provides a characterization of sorts for (2.3), though admittedly not one expressed exclusively in terms of the given $\mathscr{P}$.

Now we apply the lemma to study 'regularity'. The two properties of $\mathscr{P}$ established next reduce to the standard definition of regularity of the probability measure $P$ when $\mathscr{P} = \{P\}$. Further, for general $P$, and viewed as properties of the lower envelope $v(\cdot) = \min_{\mathscr{P}} P(\cdot)$, they are the standard notions of regularity for capacities (see [8, p. 1356] and the references therein).

**Lemma A.4.** (i) *For any compact $K \subset S^\infty$,*

$$\min_{\mathscr{P}} P(K) = \inf_G \left\{ \min_{\mathscr{P}} P(G)\colon G \text{ open},\ G \supset K \right\}. \tag{A.1}$$

(ii) *For any measurable $A \subset S^\infty$,*

$$\min_{\mathscr{P}} P(A) = \sup_K \left\{ \min_{\mathscr{P}} P(K)\colon K \text{ compact},\ K \subset A \right\}. \tag{A.2}$$

**Proof.** (i) Given $\varepsilon$, the denseness portion of Lemma A.2(ii) implies that there exists $P^* \in \mathscr{P} \cap ca(S^\infty)$ such that

$$P^*(K) < \min_{\mathscr{P}} P(K) + \varepsilon.$$

But $P^*$ is regular and hence $\inf_G \{\min_{\mathscr{P}} P(G)\colon G \text{ open},\ G \supset K\} \leqslant \inf_G \{P^*(G)\colon \text{ open},\ G \supset K\} < P^*(K) + \varepsilon < \min_{\mathscr{P}} P(K) + 2\varepsilon$.

(ii) Henceforth, $K$ and $K'$ denote compact sets even where not stated explicitly.

Suppose contrary to the assertion that for every $K \subset A$, there exists $P^K$ in $\mathscr{P}$ such that

$$P^K(K) < \min_{Q \in \mathscr{P}} Q(A) - \varepsilon \equiv v(A) - \varepsilon. \tag{A.3}$$

The collection of compact subsets of $A$ forms a directed set with respect to the partial order of set inclusion. Thus $\{P^K\colon K \subset A\}$ is a net in $\mathscr{P}$. Because the latter is weakly compact, assume w.l.o.g. that $P^K$ converges weakly to $P \in \mathscr{P}$.

For every $K$ and $K' \subset K$,

$$P^K(K') \leqslant P^K(K) < v(A) - \varepsilon.$$

But $P^K(K') \underset{K}{\rightarrow} P(K')$. Conclude that

$$P(K') < v(A) - \varepsilon \leqslant P(A) - \varepsilon, \quad \text{for all } K' \subset A. \tag{A.4}$$

If we knew that $P$ were countably additive, this would contradict regularity of $P$ and complete the proof. Thus it remains only to show that there exists a measure $P^{ca} \in \mathscr{P} \cap ca(S^\infty)$ satisfying an appropriate version of (A.4).

By the Yosida–Hewitt decomposition [16, Theorem 10.2.1], we can write $P$ as

$$P = P^{ca} + P^{ch},$$

where $P^{ca}$ is countably additive and $P^{ch}$ is a pure charge. Because the latter is necessarily nonnegative,

$$P^{ca}(K') < v(A) - \varepsilon, \quad \text{for all } K' \subset A.$$

The final step is to show that $P^{ca} \in \mathscr{P}$: By the nature of a pure charge [16, Theorem 10.1.2], $P^{ch}$ vanishes on $\bigcup_1^\infty \mathscr{F}_t$. Thus $P = P^{ca}$ there. Let $P^*$ be the measure in $\mathscr{P} \cap ca(S^\infty)$ provided by Lemma A.2(iv). Then also $P^* = P^{ca}$ on $\bigcup_1^\infty \mathscr{F}_t$. But both measures are countably additive and hence equality holds on all of $\mathscr{F}_\infty$. Conclude that $P^{ca} \in \mathscr{P}$ as desired. □

Our objective is to define regularity for sets of priors in such a way that $\mathscr{P}$ defined in (2.3) is the unique regular and rectangular set of priors conforming with $\mathscr{L}$. Think for the moment of using (A.1) and (A.2) as the definition of regularity and suppose that $\mathscr{P}'$ is another regular and rectangular set of priors. The question then is whether $\mathscr{P}' = \mathscr{P}$.

One can reason as follows: Since both sets conform with the same $\mathscr{L}$, they agree on $\bigcup_1^\infty \mathscr{F}_t$. Since basic open sets in the product topology are cylinders, (A.1) implies that their lower envelopes agree also on compact sets in $S^\infty$. Agreement of the lower envelopes on all events is subsequently implied by (A.2). Thus indeed,

$$v(A) \equiv \min_{Q \in \mathscr{P}} Q(A) = \min_{Q \in \mathscr{P}'} Q(A) \equiv v'(A)$$

for all measurable $A$. However, identity of the lower envelopes does not imply equality of the two sets of priors. It is apparent that (A.1) and (A.2) are too weak to constitute the sought-after definition of regularity because they (implicitly) deal with the utilities of binary acts only.

An appropriate strengthening of these conditions is borrowed from [8], which describes a general notion of regularity and to which the reader is referred for further details and discussion. Call $h : S^\infty \to R^1$

- *simple* if $\{h((s_t)) : (s_t) \in S^\infty\}$ is finite;
- *usc* if $\{(s_t) \in S^\infty : h((s_t)) \geqslant x\}$ is closed for every real number $x$;
- *lsc* if $\{(s_t) \in S^\infty : h((s_t)) > x\}$ is open for every real number $x$.

The set of bounded and $\mathscr{F}_\infty$-measurable acts $h$ is $\mathscr{B}$. The subset of simple usc (lsc) acts is $\mathscr{B}^{\mathrm{u}}$ ($\mathscr{B}^{\ell}$).

Say that the set of priors $\mathscr{P}'$ is *regular* if both of the following conditions are satisfied for $V' : \mathscr{B} \to R^1$ defined by $V'(h) = \min_{Q \in \mathscr{P}'} \int h\, dQ$:

*Outer regularity*: $V'(k) = \inf\{V'(g) : g \geqslant k, \; g \in \mathscr{B}^{\ell}\}$ for all $k \in \mathscr{B}^{\mathrm{u}}$.

*Inner regularity*: $V'(h) = \sup\{V'(k) : h \geqslant k, \; k \in \mathscr{B}^{\mathrm{u}}\}$ for all $h \in \mathscr{B}$.

Because the indicator function of a closed (open) subset of $S^\infty$ is simple and usc (lsc), it follows that regularity implies the counterparts of (A.1) and (A.2). For a singleton $\mathscr{P}' = \{P\}$, the converse is also true and thus $\mathscr{P}'$ is regular if and only if $P$ is regular in the usual sense [8, Theorem 4.1].

**Proof of Theorem A.1.** The proof that $\mathscr{P}$ is regular is similar to the proof of Lemma A.4.

Suppose that $\mathscr{P}' \neq \mathscr{P}$ is another regular and rectangular set. W.l.o.g. let $P^* \in \mathscr{P} \backslash \mathscr{P}'$. When $ba(S^\infty)$ is endowed with the weak topology, its dual space is isomorphic to $\mathscr{B}$ [14, p. 223]. Thus a separation theorem [4, Theorem V.2.10] implies that $P^*$ can be separated from the convex and weakly closed set $\mathscr{P}'$ by some $h \in \mathscr{B}$, that is,

$$\min_{Q \in \mathscr{P}} \int hdQ \leqslant \int hdP^* < \min_{Q \in \mathscr{P}'} \int hdQ. \tag{A.5}$$

Denote by $V'$ and $V$ the functionals corresponding to $\mathscr{P}'$ and $\mathscr{P}$ as defined prior to the definition of regularity. Rectangularity (2.4) implies that $\mathscr{P}'$ and $\mathscr{P}$ agree on $\bigcup_1^\infty \mathscr{F}_t$, that is, they induce the same set of measures on $\bigcup_1^\infty \mathscr{F}_t$. Hence,

$$V'(h) = V(h) \text{ for all } \mathscr{F}_t\text{-measurable } h \text{ and all } t.$$

Then outer regularity implies also equality for all acts $k \in \mathscr{B}^{\mathrm{u}}$. (Use the fact that basic open sets are cylinders and also [8, Lemma A.1] to argue that

$$V'(k) = \inf\{V'(g) : g \geqslant k,\; g \in \mathscr{B}^{\ell} \text{ and } g \text{ is } \mathscr{F}_t\text{-measurable for some } t\};$$

and similarly for $V$.) Finally, inner regularity implies that $V'$ and $V$ agree on $\mathscr{B}$, contradicting (A.5). $\square$

## References


[1] K. Arrow, Essays in the Theory of Risk-Bearing, North-Holland, Amsterdam, 1970.
[2] D. Blackwell, L. Dubins, Merging of opinions with increasing information, Ann. Statist. 38 (1962) 882–886.
[3] L.E. Dubins, On Lebesgue-like extensions of finitely additive measures, Ann. Probab. 2 (1974) 456–463.
[4] N. Dunford, J.T. Schwartz, Linear Operators Part I: General Theory, Wiley, New York, 1958.
[5] L.G. Epstein, M. Schneider, Learning under ambiguity, May 2002.
[6] L.G. Epstein, M. Schneider, Recursive multiple-priors, J. Econom. Theory, forthcoming.
[7] L.G. Epstein, T. Wang, Uncertainty, risk-neutral measures and security price booms and crashes, J. Econom. Theory 67 (1995) 40–82.
[8] L.G. Epstein, T. Wang, "Beliefs about beliefs" without probabilities, Econometrica 64 (1996) 1343–1373.
[9] P. Ghirardato, On independence for non-additive measures with a Fubini Theorem, J. Econom. Theory 73 (1997) 261–291.
[10] I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique priors, J. Math. Econom. 18 (1989) 141–153.
[11] E. Hendon, H.J. Jacobson. B. Sloth, T. Tranaes, The product of capacities and belief functions, Math. Soc. Sci. 32 (1996) 95–108.
[12] M. Marinacci, Limit laws for non-additive probabilities and their frequentist interpretation, J. Econom. Theory 84 (1999) 145–195.
[13] M. Marinacci, F. Maccheroni, A. Chateauneuf, J.M. Tallon, Monotone continuous multiple priors, 2002.
[14] R.E. Megginson, An Introduction to Banach Space Theory, Springer, Berlin, 1998.
[15] K.R. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, New York, 1967.
[16] K.P.S.B. Rao, M.B. Rao, Theory of Charges, Academic Press, New York, 1983.
[17] E. Regazzini, Observability and probabilistic limit theorems, in: Proceedings of the International Conference on The Notion of Event in Probabilistic Epistemology, Trieste, 27–29 Maggio 1996.
[18] R. Sarin, P. Wakker, Dynamic choice and nonexpected utility, J. Risk Uncertainty 17 (1998) 87–119.
[19] D. Schmeidler, Subjective probability and expected utility without additivity, Econometrica 57 (1989) 571–587.
[20] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall, London, 1991.
[21] P. Walley, T.L. Fine, Towards a frequentist theory of upper and lower probability, Ann. Statist. 10 (1982) 741–761.