The Annals of Statistics
1982, Vol. 10, No. 3, 741–761

TOWARDS A FREQUENTIST THEORY OF UPPER AND LOWER PROBABILITY

BY PETER WALLEY[1] AND TERRENCE L. FINE[2]

Cornell University and Bell Laboratories; and Cornell University

We present elements of a frequentist theory of statistics for concepts of upper and lower (interval-valued) probability (IVP), defined on finite event algebras. We consider IID models for unlinked repetitions of experiments described by IVP and suggest several generalizations of standard notions of independence, asymptotic certainty and estimability. Instability of relative frequencies is favoured under our IID models. Moreover, generalizations of Bernoulli's Theorem give some justification for the estimation of an underlying IVP mechanism from fluctuations of relative frequencies. Our results indicate that an objectivist, frequency- or propensity-oriented, view of probability does not necessitate an additive probability concept, and that IVP models can represent a type of indeterminacy not captured by additive probability.

1. Introduction.

1.1 *Objectives and background.* We present here a frequentist account of upper and lower (interval-valued) probabilities (IVP). Our results parallel, sometimes with noteworthy differences, the elements of the familiar frequentist account of the usual additive numerical probability (NP) concept, and they provide the rudiments of a frequentist theory of statistics for IVP. Although we concentrate on frequentist notions in this paper, our philosophical position does not restrict us to just frequentist views of probability. We accept some subjective and epistemic views of probability as well (see Fine, 1981).

By IVP we refer to a pair of functions taking their values in the unit interval, called the lower ($\underline{P}$) and upper ($\bar{P}$) probabilities, that are defined on an event algebra $\mathscr{A}$ and satisfy the axioms given in Section 2. The lower probability $\underline{P}$ is superadditive, $\bar{P}$ is subadditive, and $\bar{P}$ is always at least as large as $\underline{P}$. Section 2 also contains elementary consequences of the axioms as well as definitions of important subclasses of lower and upper probabilities, notably upper and lower envelopes, to which most of our results refer.

Of particular note is the fact that the IVP concepts we consider have instances that are not compatible with the usual NP concept satisfying Kolmogorov's axioms; if

$$(\exists A \in \mathscr{A})\bar{P}(A) > \underline{P}(A),$$

then neither $\bar{P}$ nor $\underline{P}$ are probability measures. Given the possible nonadditivity of upper and lower probabilities, it may be surprising that we propose to link them to relative frequencies. After all, relative frequencies have hitherto provided a basic interpretation of the additive NP concept and the most commonly invoked motivation for the axioms of probability as presented in introductory courses. We discuss, in Section 4, a limiting frequentist interpretation of $\underline{P}$ and $\bar{P}$ as the lim inf and lim sup of relative frequencies in hypothetical unlinked repetitions of an experiment, that generalizes the usual limiting frequentist interpretation of additive probability.

In fact, the mathematical framework developed here provides a basis for frequentist interpretations of other concepts of probability, including comparative probability (CP)

Received September 1980; revised September 1981.

[1] Now at Dept. of Statistics, University of Warwick, Coventry, CV4 7AL, England.

[2] Prepared with partial support from NSF Grant SOC 78-12278.

AMS 1970 *subject classification.* Primary 60A05.

Key words and phrases. Non-additive probability, interval-valued probability, upper and lower probability, independence, unstable relative frequencies, propensities, Bernoulli's Theorem, upper and lower envelopes.

and modal probability (Walley and Fine, 1979). There is a natural connection between CP and upper and lower probabilities, which gives some support to our characterizations of independence and asymptotic certainty or favourability.

While upper and lower probabilities have been previously treated in the literature, these treatments have largely been devoted either to mathematical aspects of structure or to those applications and interpretations associated with the subjectivist and epistemic schools of probability (e.g., Good, 1962, Levi, 1980, and Shafer, 1976). Fine (1974, 1977) and Walley and Fine (1979) indicated objective bases for CP and IVP, although these bases (quantum mechanics, relative frequency lead times, imprecise set-valued observations) were largely suggestive. The present goal is the development of an objective, frequentist interpretation for IVP that parallels the relative frequency development of the usual NP concept and that can provide the statistical basis whereby we can infer IVP models of random experiments from observations made on unlinked repetitions.

1.2 *Sketch of the argument.* We now sketch the link between relative frequencies and IVP that is developed in technical detail in the following sections. IVP will be regarded as a representation of propensities of outcomes. The frequency-propensity connection will be given by laws of large numbers expressed in terms of IVP. There are several propensity accounts of probability that disagree with each other. Perhaps Giere (1973) is closest to our present view.

Suppose that propensities of events in unlinked repetitions $\varepsilon_1, \cdots, \varepsilon_n$ are represented through the lower probability $\underline{P}$. To provide a connection between frequency and propensity which extends the classical laws of large numbers, we examine our ability to infer or estimate the marginal $\underline{P}$ from relative frequency data. Clearly we cannot base an estimator of $\underline{P}(A)$ upon just the terminal relative frequency $r_n(A)$ of event A in $\varepsilon_1, \cdots, \varepsilon_n$; for relative frequencies observed at any time n form a probability measure on the event algebra $\mathscr{A}$. Rather we consider an estimator

$$\underline{r}_n(A) = \min\{r_j(A): k(n) \le j \le n\}, \quad \text{where } k(n) \to \infty,$$

that makes use of the additional information concerning the past evolution of the sequence of relative frequencies. Let $\underline{P}^\infty$ describe the IID repetitions $\varepsilon_1, \cdots$. In Section 5 we show that the estimation process succeeds in the sense that, letting

$$G_{n,\delta} = [\,|\underline{r}_n(A) - \underline{P}(A)| < \delta], \quad \lim_{n\to\infty} \underline{P}^\infty(G^c_{n,\delta})/\underline{P}^\infty(G_{n,\delta}) = 0, \quad \forall\delta > 0;$$

we call this notion of convergence "asymptotic favourability" of $(G_{n,\delta})$. Hence our "confidence" that our estimator $\underline{r}_n(A)$ is within δ of $\underline{P}(A)$ grows with increasing sample size. This result parallels the Bernoulli law of large numbers. However, we must warn the reader that the inference situation for IVP has some novel puzzling elements that do not confront us in NP.

We also consider the problem of hypothesis testing, i.e., of selecting one element from a finite family of possible lower probabilities governing ε on the basis of the outcomes of $\varepsilon_1, \cdots, \varepsilon_n$. We provide conditions in Section 5 under which this problem can be solved with increasing "confidence" as the sample size n increases.

Finally, we supplement the preceding results by studying the stability of relative frequencies. It might be expected from our ability to estimate $\underline{P}(A)$ that convergence of the sequence of relative frequencies might be accorded little support, and this is in fact the case. We show in Section 4 that apparent divergence of relative frequencies is asymptotically favored over their apparent convergence for nonadditive models. Hence our theory of IVP expects our nonadditive propensities to support only partially stable relative frequencies, the stationarity of our model notwithstanding.

1.3 *Our philosophical position.* The following comments on our philosophical position are not required for an understanding of the technical developments of this paper but will clarify our motivation in carrying out this study. We have both ontological (nature of reality) and epistemological (nature of knowledge) commitments to IVP. Our ontological commitment leads us to view IVP models as representations of properties (propensities) of classes of physical/empirical phenomena. Our epistemological commitment leads us to IVP descriptions of our knowledge concerning some "random experiment." Ontological probability relates to the familiar term "chance" while epistemological probability relates to "uncertainty."

A discussion of our view of probability is available in Walley and Fine (1979) and Fine (1982). Central to our philosophical position is an acceptance of a concept of indeterminacy that is distinguished from the concepts of chance and uncertainty. We admit both physical and epistemic concepts of indeterminacy and have employed IVP models in both domains. Our position on epistemic indeterminacy is one with which non-Bayesian statisticians should feel at ease. The unknown but non-random parameter is to us a case of an ontologically determinate but epistemologically indeterminate quantity. We do not follow the Bayesians in supposing that our knowledge concerning the value of an unknown parameter can always be cast in the form of a NP description. However, unlike the non-Bayesian statisticians, we expect to be able to differentiate degrees of epistemic indeterminism which can then be represented by IVP or by other probability structures.

The thornier issue is that of an ontological indeterminism that is distinct from chance. We believe that there are empirical phenomena that exhibit the sort of indeterminism for which our theory is appropriate. For example, the utterances of an individual speaker have statistical regularities for some features but seemingly not for all. Speech does not seem to be a stochastic process, a view compatible with the Chomskyean view of the creativity of speech. One can also speculate that the confusing probabilistic interpretation of quantum mechanics, confusing because it is far from clear what specific frequentist concept is being invoked, should be revised to take account of indeterminacy as well as the more regular concept of chance (see Fine, 1974). Finally, we might also suggest that the familiar view, that in the long run observed relative frequencies will converge to a fixed number that is then regarded as either a measurement of the probability of the event in question or the probability itself, is often implicitly recognized to be a fiction. Critical reflection on even such familiar physical processes as die tossing reveals that in the long run we can expect to encounter erratic time variations that admit only rough, qualitative explanations (e.g. abrasions of the die). This qualitative information can perhaps be accommodated by an IVP model but seemingly not be an additive probability model. See Section 6 for further discussion of this point.

2. Basic properties of upper and lower probability.

We start by defining a hierarchy of regularity properties for upper and lower probabilities. Although this paper is mainly concerned with upper and lower *envelopes*, the other properties will be referred to and have some general importance.

Let $\mathscr{A}$ be a finite algebra. Since any finite algebra is isomorphic to the algebra of all subsets of some Ω, we will assume $\mathscr{A} = 2^{\Omega}$. Throughout, $\underline{P}: \mathscr{A} \to [0, 1]$ will be a non-negative set function satisfying $\underline{P}(\phi) = 0$ and $\underline{P}(\Omega) = 1$, and its conjugate function $\bar{P}: \mathscr{A} \to [0, 1]$ will be defined by $\bar{P}(A) = 1 - \underline{P}(A^c)$.

$\underline{P}$ is called a *lower probability*, and $\bar{P}$ its conjugate *upper probability*, when they satisfy

$$(\forall A \cap B = \phi)\, \underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) \quad \text{(super-additivity), and}$$

$$\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B) \quad \text{(sub-additivity).}$$

$\bar{P}$ is called a *lower envelope* when

$$(\forall A \in \mathscr{A})\, \underline{P}(A) = \inf\{\pi(A): \pi \in \mathscr{M}\}$$

for some non-empty class $\mathscr{M}$ of probability measures on $\mathscr{A}$. Then $\bar{P}(A) = \sup\{\pi(\mathscr{A}): \pi \in \mathscr{M}\}$ is the conjugate *upper envelope*. $\mathscr{M}$ will usually be compact under the natural topology, so that the inf and sup are achieved.

If P and Q are any set functions on $\mathscr{A}$, P is said to *dominate* Q, written $P \geq Q$, when

$(\forall A \in \mathscr{A}) P(A) \geq Q(A)$. Note that $\underline{P} \geq \underline{Q} \Rightarrow \bar{Q} \geq \bar{P}$. Denote the class of all probability measures dominating $\underline{P}$ by $\mathscr{M}(\underline{P})$, so that $\pi \in \mathscr{M}(\underline{P})$ iff $\bar{P} \geq \pi \geq \underline{P}$. $\mathscr{M}(\underline{P})$ may be empty when $\underline{P}$ is a lower probability.

$\underline{P}$ is called *2-monotone* when

$$(\forall A, B \in \mathscr{A}) \underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B) + \underline{P}(A \cap B).$$

It can be shown that $\underline{P}$ is 2-monotone if and only if

$$(*) \qquad (\forall A \cap B = \phi)(\exists \pi \in \mathscr{M}(\underline{P})) \pi(A) = \underline{P}(A), \pi(B) = \bar{P}(B).$$

The *probability assignment* m for a set function $\underline{P}$ is defined by

$$(\forall A \in \mathscr{A}) m(A) = \textstyle\sum_{B \subset A} (-1)^{\|A-B\|} \underline{P}(B),$$

where $\|C\|$ denotes the cardinality of C. By the Mobius Inversion Theorem (Shafer, 1976, Lemma 2.3), $\underline{P}$ can be recovered from m by $(\forall A \in \mathscr{A}) \underline{P}(A) = \sum_{B \subset A} m(B)$. $\underline{P}$ is called a *belief function* when m is non-negative.

LEMMA 2.1. *$\underline{P}$ is an additive probability measure $\Rightarrow$ $\underline{P}$ is a belief function $\Rightarrow$ $\underline{P}$ is 2-monotone $\Rightarrow$ $\underline{P}$ is a lower envelope $\Rightarrow$ $\underline{P}$ is a lower probability.*

PROOF. The first implication holds because a probability measure has probability assignment concentrated on atoms and equal there to its probability mass function. If $\underline{P}$ is a belief function with probability assignment m,

$$\underline{P}(A \cup B) = \textstyle\sum_{C \subset A \cup B} m(C) \geq \sum_{C \subset A} m(C) + \sum_{C \subset B} m(C) - \sum_{C \subset A \cap B} m(C)$$

$$= \underline{P}(A) + \underline{P}(B) - \underline{P}(A \cap B).$$

The third implication is a consequence of $(*)$ above, and the fourth follows from sub- and super-additivity of sup and inf. □

Examples on 4 or fewer atoms can easily be constructed to show that none of the converse implications holds in Lemma 2.1.

We shall use without comment the following easily verified properties of upper and lower probabilities.

LEMMA 2.2. *If $\underline{P}$ is a lower probability and $\bar{P}$ its conjugate upper probability, then for all A, B, $\in \mathscr{A}$*

a) $\bar{P}(A) \geq \underline{P}(A)$
b) $A \supset B \Rightarrow \underline{P}(A) \geq \underline{P}(B)$ *and* $\bar{P}(A) \geq \bar{P}(B)$
c) $A \cap B = \phi \Rightarrow \bar{P}(A \cup B) \geq \bar{P}(A) + \underline{P}(B) \geq \underline{P}(A \cup B)$
d) $\underline{P}(A) + \underline{P}(B) \leq \bar{P}(A \cup B) + \underline{P}(A \cap B) \leq 1 + \underline{P}(A \cap B)$.

LEMMA 2.3. (Dempster, 1967). *If $\underline{P}$ is a lower probability on $\mathscr{A}$ then $\mathscr{M}(\underline{P})$ is a closed convex polyhedron (possibly empty) in the simplex of all probability measures on $\mathscr{A}$.*

It appears that a frequentist theory of upper and lower probability can be developed most naturally in terms of upper and lower *envelopes* (see Theorem 4.2), and envelopes have the major role in this work. The extra regularity of 2-monotonicity is needed for the Huber-Strassen Theorem (Theorem 5.2), and also seems to be needed to establish such basic frequentist properties involving *pairs* of events as the following: if $A \cap B = \phi$, $\bar{P}(A) > \bar{P}(B)$ and $\underline{P}(A) > \underline{P}(B)$ then $\underline{P}^n[r_n(A) > r_n(B)] > \underline{P}^n[r_n(A) < r_n(B)]$, where r_n denotes relative frequency and $\underline{P}^n$ describes n IID repetitions of $\underline{P}$ (to be defined in Section 3.1).

The further regularity of belief functions will be used in an alternative definition of the independent product, given in an Appendix.

It is noteworthy that upper and lower envelopes are central also in a personalist theory of upper and lower probability. It can be shown that $\underline{P}$: $\mathscr{A} \to \mathbb{R}$ is a lower envelope if and only if it satisfies the following "coherence" condition. Write $G(A, \omega) = I_A(\omega) - \underline{P}(A)$. Then $\underline{P}$ is said to be coherent if ($\nexists$ positive integers m, n; $A_0, A_1, \cdots, A_n \in \mathscr{A}$)

$$(\forall \omega \in \Omega) mG(A_0, \omega) > \sum_{i=1}^{n} G(A_i, \omega).$$

The personalist interpretation of this condition is that $\underline{P}(A)$ represents the maximum price you are willing to pay in order to receive 1 unit if A occurs, and $G(A, \omega)$ is your gain from a "marginally acceptable" bet on A. The condition then denies the existence of acceptable bets on A_i whose overall outcome is uniformly worse than the outcome of an unacceptable bet of m units on A_0; it implies that no combination of acceptable bets can lead to sure loss, by taking $A_0 = \phi$ (clearly, $\underline{P}(\phi) = 0$). The personalist and limiting frequency interpretations seem to lead to similar mathematical structures (upper and lower envelopes), and similar definitions of upper and lower expectations and conditional probabilities. See Section 4.3; details are in Williams (1976) and Walley (1981).

Further general results about lower probabilities and related set functions may be found in Dempster (1967, 1968), Good (1962), Huber (1973), Huber and Strassen (1973), Levi (1980), Shafer (1976), Smith (1961), Suppes (1974), and Wolfenson (1979).

3. Independence and asymptotics.

Our results involve a particular generalization to non-additive set functions of the usual definition of stochastic independence. (An alternative generalization is treated in an Appendix). As is obvious from our notation, our discussion will be restricted to product spaces with identical marginal spaces (Ω, $\mathscr{A}$), but the definitions and results of this Section extend in an obvious way to arbitrary finite marginals (Ω_i, $\mathscr{A}_i$).

3.1 *Independence.*

NOTATION. $\mathscr{A} = 2^{\Omega}$ is a finite algebra. We write Ω^n for the n-fold Cartesian product of Ω, and $\mathscr{A}^n = 2^{\Omega^n}$. Let $A^i = \{\omega \in \Omega^n: \omega_i \in A\}$ denote the cylinder set with base A at the ith co-ordinate. We write $X_{j=1}^{n} A_j = \cap_{j=1}^{n} A_j^j$ for "rectangles" in $\mathscr{A}^n$.

DEFINITION. Suppose $(1 \le i \le n)\underline{P}_i: \mathscr{A} \to [0, 1]$ are lower envelopes. Their *independent product* is written as $\prod_{i=1}^{n} \underline{P}_i$ or (for simplicity) $\underline{P}^n$ and defined by

$$(\forall A \in \mathscr{A}^n)\underline{P}^n(A) = \min\{(\textstyle\prod_{j=1}^{n} \pi_j)(A): \pi_j \in \mathscr{M}(\underline{P}_j), 1 \le j \le n\},$$

where $\prod_{j=1}^{n} \pi_j$ is the product probability measure, defined by

$$(\textstyle\prod_{j=1}^{n} \pi_j)(X_{j=1}^{n} A_j) = \prod_{j=1}^{n} \pi_j(A_j) \quad \text{and extended to } \mathscr{A}^n \text{ by additivity.}$$

Note that $\underline{P}^n(A^i) = \underline{P}_i(A)$, so it is proper to refer to $\underline{P}_i$ as the ith marginal of $\underline{P}^n$. $\underline{P}^n$ is a probability measure (the usual product measure) iff all its marginals $\underline{P}_i$ are additive. Most of our results concern IID products $\underline{P}^n$, for which the marginals $\underline{P}_i$ are all equal to some $\underline{P}$.

$\underline{P}^n(A)$ may be interpreted as the greatest lower bound to the probability of $A \in \mathscr{A}^n$ in independent trials governed by probability measures π_i, where π_i is chosen in an unknown way from $\mathscr{M}(\underline{P}_i)$. We emphasize however that this is not the interpretation that most interests us here. In particular, our use of $\underline{P}^n$ to describe repetitions of an experiment does not commit us to a belief in "underlying" probability measures $\pi_i \in \mathscr{M}(\underline{P}_i)$ that are operative on the individual trials. Indeed, we shall outline in Section 4 a different justification for our definition of the IID product $\underline{P}^n$, in which $\underline{P}$ is interpreted as the lower envelope of a class $\mathscr{M}$ of limit points of relative frequencies in an (ideal) infinite sequence of repetitions. Such an interpretation does not require measures in $\mathscr{M}$ to be actually operative in individual experiments. Our use of standard probability theory to establish most of our mathematical results, possible because of our definition of $\underline{P}^n$ in terms of $\mathscr{M}(\underline{P}_i)$, should not mislead the reader into thinking that we are dealing with such standard statistical problems as estimation of characteristics of measures $\pi_1, \pi_2, \cdots$ governing independent trials. Of course, our results can always be given this interpretation, but it may not be the most interesting one.

Both $\underline{P}^n$ and the alternative product $\underline{Q}^n$ defined in an Appendix have several desirable qualitative properties of "independence" that can be expressed in terms of the comparative probability relations they induce. The limiting frequency interpretation of Section 4.3 supports $\underline{P}^n$ over $\underline{Q}^n$, but our present results do not rule out the possibility of other characterizations of independence for "ontologically indeterminate" experiments.

Denote the upper envelope conjugate to $\underline{P}^n$ by $\bar{P}^n$. The next results shows that $\underline{P}^n$ and $\bar{P}^n$ factorize on rectangles.

LEMMA 3.1. *Let $\alpha, \beta \subset \{1, \cdots, n\}$, $\alpha \cap \beta = \phi$. Write $\mathscr{A}^n_\alpha$ for the sub-algebra of $\mathscr{A}^n$ generated by $\{A^i: A \in \mathscr{A}, i \in \alpha\}$. Then*

$$(\forall A \in \mathscr{A}^n_\alpha, B \in \mathscr{A}^n_\beta)\underline{P}^n(A \cap B) = \underline{P}^n(A)\underline{P}^n(B) \quad and \quad \bar{P}^n(A \cap B) = \bar{P}^n(A)\bar{P}^n(B).$$

PROOF. Routine verification.

COROLLARY 3.1. $(\forall A_j \in \mathscr{A})\underline{P}^n(X^n_{j=1}A_j) = \prod^n_{j=1} \underline{P}_j(A_j)$, *and similarly for $\bar{P}^n$.*

3.2 *Asymptotics.* As the remaining results refer to asymptotic properties in an unlimited sequence of trials, it is convenient to define a single independent product envelope on the Cartesian product of infinite sequences of outcomes.

NOTATION. As before, $\mathscr{A} = 2^\Omega$ is a finite algebra. Let Ω^∞ denote the countably infinite Cartesian product of Ω. Let $\mathscr{A}^n$ now denote the algebra of subsets of Ω^∞ generated by the outcomes of the first n trials (isomorphic to $\mathscr{A}^n$ as previously defined). Then $(\mathscr{A}^n)$ is an increasing sequence of finite algebras. Write $\mathscr{A}^\infty = \cup_{n\geq 1}\mathscr{A}^n$ for the infinite algebra of finite-dimensional cylinder sets.

In the rest of the paper all events considered are in $\mathscr{A}^\infty$, and their propensities will be represented by a single lower envelope $\underline{P}^\infty$, generally an IID product, defined on $\mathscr{A}^\infty$. In order to discuss versions of the laws of large numbers for envelopes we need to define some notion of convergence to certainty for a sequence of events $A_n \in \mathscr{A}^\infty$. We now define two such notions.

DEFINITION. Suppose that $\underline{P}^\infty: \mathscr{A}^\infty \to [0, 1]$ is a lower envelope, i.e., $\underline{P}^\infty(A) = \inf\{\pi(A): \pi \in \mathscr{M}(\underline{P}^\infty)\}$ where $\mathscr{M}(\underline{P}^\infty)$ is the class of all probability measures on $\mathscr{A}^\infty$ that dominate $\underline{P}^\infty$, and $\bar{P}^\infty$ is the conjugate upper envelope. Suppose also that $(\forall n \geq 1)A_n \in \mathscr{A}^\infty$. If $\bar{P}^\infty(A^c_n)/\bar{P}^\infty(A_n) \to 0$ as $n \to \infty$, we say that the sequence of events (A_n) is *asymptotically certain* (a.c.) under $\underline{P}^\infty$. If $\underline{P}^\infty(A^c_n)/\underline{P}^\infty(A_n) \to 0$, we say that (A_n) is *asymptotically favoured* (a.f.) under $\underline{P}^\infty$.

REMARK (i). Any consistent sequence of set functions $\underline{P}^n$ defined on $\mathscr{A}^n$ induces $\underline{P}^\infty$ on $\mathscr{A}^\infty$ in the obvious way. If $\underline{P}^n$ are lower envelopes, so is $\underline{P}^\infty$.

PROOF. It follows from the Bolzano-Weierstrass Theorem and a diagonal argument that any sequence $\pi_n \in \mathscr{M}(\underline{P}^n)$ (π_n defined on $\mathscr{A}^n$) has a subsequence converging to some $\pi \in \mathscr{M}(\underline{P}^\infty)$. Since $\underline{P}^n$ are envelopes, for any $A \in \mathscr{A}^n$ we can take $\pi_j(A) = \underline{P}^n(A) = \underline{P}^\infty(A)$ for $j \geq n$, so that the limit π of any convergent subsequence satisfies $\pi(A) = \underline{P}^\infty(A)$. Thus $\underline{P}^\infty$ is the lower envelope of $\mathscr{M}(\underline{P}^\infty)$. □

Conversely, the restrictions $\underline{P}^n$ of a lower envelope $\underline{P}^\infty$ to $\mathscr{A}^n$ are lower envelopes. Thus, the above definition could be written equivalently in terms of a consistent sequence of lower envelopes $\underline{P}^n$ defined on an increasing sequence of algebras $\mathscr{A}^n$.

It is clear that we are still essentially concerned only with finite algebras, and need not worry about the extra conditions (e.g., continuity) that need to be imposed on envelopes defined on infinite spaces.

REMARK (ii). (A_n) a.c. under $\underline{P}^\infty$ iff $\underline{P}^\infty(A_n) \to 1$ (by sub-additivity of $\bar{P}^\infty$). Hence, (A_n) a.c. under $\underline{P}^\infty$ implies (A_n) a.f. under $\underline{P}^\infty$: asymptotic certainty is stronger than asymptotic favourability.

REMARK (iii). When $\underline{P}^\infty$ is additive, (A_n) a.c. iff (A_n) a.f. iff $\underline{P}^\infty(A_n) \to 1$, so the two concepts agree with the usual notion of asymptotic certainty.

REMARK (iv). (A_n) a.f. under $\underline{P}^\infty$ implies $\bar{P}^\infty(A_n) \to 1$, but it is possible that $\underline{P}^\infty(A_n) \to 0$ (in which case we are becoming "increasingly ignorant" about the possible occurrence of A_n). Thus, asymptotic favourability is a rather weak property. The results of Sections 4 and 5 using this property must therefore be interpreted with some caution. One rationale for using a.f. is that it (and a.c.) have a simple interpretation through measurement scales in comparative probability. (A_n) a.f. under $\underline{P}^\infty$ implies, for example, that A_n is eventually more probable than A_n^c under comparative probability relations induced in a natural way by $\underline{P}^\infty$. (A_n) a.f. implies that $\underline{P}^\infty$ gives unboundedly greater support to A_n than to A_n^c as $n \to \infty$.

Note that if (A_n) a.c. under $\underline{P}^\infty$ then $\pi(A_n) \to 1$ for any $\pi \in \mathscr{M}(\underline{P}^\infty)$. We might therefore expect asymptotic certainty to have most of the properties familiar in the special case of additive probability. In particular we have the following,

LEMMA 3.2. *Suppose $\underline{P}^\infty$ is a lower envelope on $\mathscr{A}^\infty$, $(\forall n \geq 1, 1 \leq j \leq J) A_{jn} \in \mathscr{A}^\infty$, and $(\forall 1 \leq j \leq J)(A_{jn})$* a.c. *as $n \to \infty$ under $\underline{P}^\infty$. Then $(\cap_{j=1}^J A_{jn})$* a.c. *as $n \to \infty$ under $\underline{P}^\infty$.*

PROOF. $1 - \underline{P}^\infty(\cap_{j=1}^J A_{jn}) = \bar{P}^\infty(\cup_{j=1}^J A_{jn}^c) \leq \sum_{j=1}^J \bar{P}^\infty(A_{jn}^c) = \sum_{j=1}^J [1 - \underline{P}^\infty(A_{jn})] \to 0$ by hypothesis. □

We will see (Theorems 5.3 and 5.4) that the corresponding property fails in general for asymptotic favourability, which need not be preserved under finite intersections.

4. Unstable relative frequencies. We now take $\underline{P}^\infty$ (defined on $\mathscr{A}^\infty$) to be the infinite IID product of the lower envelope $\underline{P}$ (defined on the finite algebra $\mathscr{A} = 2^\Omega$). $\underline{P}^\infty$ is then the lower envelope of the class of all infinite products of $\pi_i \in \mathscr{M}(\underline{P})$. In this Section we examine the behaviour of the sequence (r_n) of relative frequency measures under $\underline{P}^\infty$, and show that such IID repetitions of a non-additive envelope $\underline{P}$ favour apparent divergence of relative frequencies over their apparent convergence.

4.1 *Apparent convergence of relative frequencies.*

DEFINITION. For $A \in \mathscr{A}$, define the *relative frequency* $r_n(A)\colon \Omega^\infty \to [0, 1]$ by $r_n(A)(\omega) = \|\{i: 1 \leq i \leq n, \omega_i \in A\}\|/n$, where $\|S\|$ denotes the cardinality of S. Write r_n for the random relative frequency measure on $\mathscr{A}$. Write

$$C_n(A; k, \varepsilon) = \cap_{j=k}^n [|r_j(A) - r_n(A)| < \varepsilon]$$

for the event that $r_1(A), \cdots, r_n(A)$ *apparently converges* (k, ε). (See Fine, 1973, page 89, for discussion). Use $C_n(k, \varepsilon) = \cap_{A \in \mathscr{A}} C_n(A; k, \varepsilon)$ to denote the event that $r_1, \cdots, r_n$ apparently converges (k, ε), and $D_n(A; k, \varepsilon) = [C_n(A; k, \varepsilon)]^c$, $D_n(k, \varepsilon) = [C_n(k, \varepsilon)]^c$ to denote *apparent divergence*.

THEOREM 4.1. *Suppose $\varepsilon > 0$, $k\colon N \to N$, and $k(n) \to \infty$ as $n \to \infty$.*

(a) $(\forall \eta > 0)(\cap_{A \in \mathscr{A}} \cap_{j=k(n)}^n [\bar{P}(A) + \eta > r_j(A) > \underline{P}(A) - \eta])$ a.c. *as $n \to \infty$ under $\underline{P}^\infty$.*

(b) *If $\varepsilon > \bar{P}(A) - \underline{P}(A)$ then $(C_n(A; k(n), \varepsilon))$* a.c. *under $\underline{P}^\infty$. Hence, if $\varepsilon > \max\{\bar{P}(A) - \underline{P}(A): A \in \mathscr{A}\}$ then $(C_n(k(n), \varepsilon))$* a.c. *under $\underline{P}^\infty$. For example, if $\underline{P}$ is additive then $(\forall \varepsilon > 0)(C_n(k(n), \varepsilon))$* a.c.

(c) $(\forall \underline{P})\underline{P}^\infty(D_n(k(n), \varepsilon)) \to 0$ *as* $n \to \infty$.

(d) *If* $0 < \underline{P}(A) < \bar{P}(A) < 1$, $\varepsilon < \bar{P}(A) - \underline{P}(A)$, *and* lim sup $k(n)/n$ *is sufficiently small, then* $(D_n(A; k(n), \varepsilon))$ a.f. *under* $\underline{P}^\infty$. *Hence, if* $\underline{P}$ *is non-additive,* $\varepsilon < \max\{\bar{P}(A) - \underline{P}(A): A \in \mathscr{A}, 0 < \underline{P}(A) < \bar{P}(A) < 1\}$ *and* $k(n)/n \to 0$ *as* $n \to \infty$, *then* $(D_n(k(n), \varepsilon))$ a.f. *under* $\underline{P}^\infty$.

PROOF. (a) Let $\pi \in \mathscr{M}(\underline{P})$ with $\pi(A) = \underline{P}(A)$. For $\eta > 0$,

$$\underline{P}^\infty\left(\cap_{j=k(n)}^n [r_j(A) > \underline{P}(A) - \eta]\right) = \pi^n\left(\cap_{j=k(n)}^n [r_j(A) > \pi(A) - \eta]\right) \to 1 \quad \text{as} \quad k(n) \to \infty,$$

by the strong law of large numbers. Apply Lemma 3.2 to the intersection over $A \in \mathscr{A}$, and note that

$$[\bar{P}(A) + \eta > r_j(A) > \underline{P}(A) - \eta] = [r_j(A) > \underline{P}(A) - \eta] \cap [r_j(A^c) > \underline{P}(A^c) - \eta].$$

(b)
$$\begin{aligned} &\cap_{j=k(n)}^n [\bar{P}(A) + \eta > r_j(A) > \underline{P}(A) - \eta] \\ &\subset \cap_{j=k(n)}^n [|r_j(A) - r_n(A)| < \bar{P}(A) - \underline{P}(A) + 2\eta] \\ &\subset C_n(A; k(n), \varepsilon) \quad \text{provided} \quad 2\eta < \varepsilon - |\bar{P}(A) - \underline{P}(A)|. \end{aligned}$$

Thus, (a) implies $(C_n(A; k(n), \varepsilon))$ a.c. under $\underline{P}^\infty$ when $\varepsilon > \bar{P}(A) - \underline{P}(A)$. If $\varepsilon > \max\{\bar{P}(A) - \underline{P}(A) : A \in \mathscr{A}\}$, $(C_n(k(n), \varepsilon)) = (\cap_{A \in \mathscr{A}} C_n(A; k(n), \varepsilon))$ a.c. by Lemma 3.2.

(c) $\underline{P}^\infty(D_n(k(n), \varepsilon)) = \underline{P}^n(D_n(k(n), \varepsilon)) \le \min\{\pi^n(D_n(k(n), \varepsilon)) : \pi \in \mathscr{M}(\underline{P})\} \to 0$ by (b).

(d) We use the following consequence of Chernoff's Theorem (Chernoff, 1952).

LEMMA 4.1. *Write* $b(n, p, \lambda) = \sum_{\lambda n \le j \le n} \binom{n}{j} p^j q^{n-j}$ *for the probability that a binomial* (n, p) *random variable is at least* λn, *where* $q = 1 - p$. *Suppose* $0 < p < \lambda < 1$. *Then*

$$n^{-1} \log b(n, p, \lambda) \to -\alpha(p, \lambda) \quad \text{as} \quad n \to \infty,$$

where $\alpha(p, \lambda) = \lambda \log\left(\frac{\lambda}{p}\right) + (1 - \lambda) \log\left(\frac{1 - \lambda}{q}\right) > 0.$

To prove (d) of the Theorem, suppose that $0 < \underline{P}(A) < \bar{P}(A) < 1$, $\varepsilon < \bar{P}(A) - \underline{P}(A)$, and $\delta >$ lim sup $k(n)/n$. Eventually $k(n) < \delta n$, so it suffices to prove the result for $k(n) = [\delta n]$. To simplify notation, assume δn is integral (the general argument is similar, with δn replaced by $[\delta n]$). Write $r_n = r_n(A)$, $C_n = C_n(A; \delta n, \varepsilon)$, $D_n = C_n^c$, and let $r'_{\delta n}$ denote the relative frequency of A on trials $\delta n + 1, \cdots, 2\delta n$, so that $r_{\delta n}$ and $r'_{\delta n}$ are IID under $\underline{P}^\infty$. Then

$$\begin{aligned} D_n &\supset [|r_{\delta n} - r_{2\delta n}| \ge 2\varepsilon] = [\tfrac{1}{2}|r_{\delta n} - r'_{\delta n}| \ge 2\epsilon] \\ &\supset [r_{\delta n} \ge \tfrac{1}{2} + 2\varepsilon] \cap [r'_{\delta n} \le \tfrac{1}{2} - 2\varepsilon]. \end{aligned}$$

Hence $\underline{P}^\infty(D_n) = \underline{P}^n(D_n) \ge b(\delta n, \underline{P}(A), \tfrac{1}{2} + 2\varepsilon) b(\delta n, \underline{P}(A^c), \tfrac{1}{2} + 2\varepsilon)$ and lim inf $n^{-1} \log \underline{P}^\infty(D_n) \ge -\delta\alpha_1$ for some $\alpha_1 > 0$ (not depending on δ) by Lemma 4.1.

Similarly, if $0 < \gamma < 1 - \delta$,

$$C_n \subset [|r_{(1-\gamma)n} - r_n| < \varepsilon] = [|r_{(1-\gamma)n} - r'_{\gamma n}| < \varepsilon/\gamma] \subset [r_{(1-\gamma)n} \le \rho + \eta] \cup [r'_{\gamma n} \ge \rho - \eta]$$

where $\rho = [\bar{P}(A) + \underline{P}(A)]/2$, $\eta = \varepsilon/2\gamma$. Hence $\underline{P}^\infty(C_n) \le b((1 - \gamma)n, \underline{P}(A^c), 1 - \rho - \eta) + b(\gamma n, \underline{P}(A), \rho - \eta)$ and lim sup $n^{-1} \log \underline{P}^\infty(C_n) \le -\min\{(1 - \gamma)\alpha_2, \gamma\beta_2\}$ for some $\alpha_2, \beta_2 > 0$, provided $1 > \gamma > \varepsilon/[\bar{P}(A) - \underline{P}(A)]$ so that $\eta < \rho - \underline{P}(A) = [\bar{P}(A) - \underline{P}(A)]/2$. For $\delta < \min\{1 - \gamma, (1 - \gamma)\alpha_2/\alpha_1, \gamma\beta_2/\alpha_1\}$, this gives

$$\underline{P}^\infty(C_n)/\underline{P}^\infty(D_n) \to 0 \quad \text{as} \quad n \to \infty. \qquad \square$$

Thus, if $k(n) \to \infty$ and $k(n)/n \to 0$, apparent convergence $(k(n), \varepsilon)$ of $r_1(A), \cdots, r_n(A)$ is asymptotically certain when $\bar{P}(A) - \underline{P}(A) < \varepsilon$, but apparent divergence $(k(n), \varepsilon)$ is asymptotically favoured when $\bar{P}(A) - \underline{P}(A) > \varepsilon$.

REMARKS. (i) Theorem 4.1 can be generalized to upper and lower expectations as follows. Define the *upper* and *lower expectations* of an $\mathscr{A}$-measurable random variable X: $\Omega \to \mathbb{R}$ under $\underline{P}$ to be

$$\bar{E}X = \max\{E_\pi X : \pi \in \mathscr{M}(\underline{P})\} \quad \text{and} \quad \underline{E}X = \min\{E_\pi X : \pi \in \mathscr{M}(\underline{P})\},$$

where $E_\pi X = \int X\, d\pi$ is expectation under π. If $\underline{P}$ is a lower envelope then clearly $\underline{E}I_A = \underline{P}(A)$. Indeed, by the Hahn-Banach Theorem, there is a super-linear functional $\underline{E}$ defined for all real X satisfying $\underline{E}X \geq \inf X$, $(\forall\lambda \geq 0)\ \underline{E}(\lambda X) = \lambda \underline{E}X$ and $(\forall A \in \mathscr{A})\underline{E}I_A = \underline{P}(A)$ iff $\underline{P}$ is a lower envelope, and in that case $\underline{E}X = \min\{E_\pi X : \pi \in \mathscr{M}\}$ for some class of probability measures with lower envelope $\underline{P}$.

For 2-monotone $\underline{P}$ the minimum in $\underline{E}X$ is achieved by a probability measure which induces on $\mathbb{R}$ the distribution function $\bar{F}(x) = \bar{P}(\{\omega : X(\omega) \leq x\})$ for X, but more generally there may be no such measure in $\mathscr{M}(\underline{P})$.

Write $X_i = X(\omega_i)(i \geq 1)$. It can be easily seen that

$$(\forall\varepsilon > 0)\underline{P}^\infty([\underline{E}X - \varepsilon < n^{-1} \textstyle\sum_{i=1}^n X_i < \bar{E}X + \varepsilon]) \to 1 \quad \text{and}$$

$$\underline{P}^\infty([\underline{E}X + \varepsilon \leq n^{-1} \textstyle\sum_{i=1}^n X_i]) \to 0.$$

The first statement generalizes the usual weak law of large numbers and may be strengthened as in Theorem 4.1(a).

(ii) If $\underline{P}^\infty$ is any *stationary* lower envelope on $\mathscr{A}^\infty$, i.e. $(\forall A \in \mathscr{A}^\infty)\ \underline{P}^\infty(TA) = \underline{P}^\infty(A)$ where $TA = \{\omega : \omega' \in A,\ \omega_i = \omega'_{i-1},\ i \geq 2\}$, then there is a stationary probability measure $\pi \in \mathscr{M}(\underline{P}^\infty)$ defined on $\mathscr{A}^\infty$. It follows that $\underline{P}^\infty(D_n(k(n), \varepsilon)) \to 0$ as $k(n) \to \infty$. Thus Theorem 4.1(c) holds more generally.

(iii) Define, for $A \in \mathscr{A}^\infty$, $\underline{R}^\infty(A) = \min\{\pi^\infty(A) : \pi \in \mathscr{M}(\underline{P})\}$. Then $(\forall A \in \mathscr{A}^\infty)\underline{R}^\infty(A) \geq \underline{P}^\infty(A)$. Since $\underline{R}^\infty(C_n(k(n), \varepsilon)) = \min\{\pi^n(C_n(k(n), \varepsilon)) : \pi \in \mathscr{M}(\underline{P})\} \to 1$, Theorem 4.1(d) distinguishes $\underline{P}^\infty$ from $\underline{R}^\infty$. The latter models the standard approximate specification of additive probability in which the non-additivity of $\underline{P}$ results merely from our ignorance about the "true underlying" probability measure ("epistemological indeterminacy"), rather than from any imprecision inherent in the repeated experiment ("ontological indeterminacy"). When such imprecision exists there is no reason to expect convergence of relative frequencies. Moreover, according to Theorem 4.1(d), apparent *divergence* is favoured asymptotically.

4.2 *Minima of relative frequencies.* Despite the weakness of asymptotic favourability, Theorem 4.1(d) suggests that there may be more information contained in outcomes of a repeated experiment about an underlying marginal envelope $\underline{P}$ than is contained in the terminal relative frequencies r_n alone. We might therefore attempt to estimate $\underline{P}$ from the fluctuations of (r_n) over long intervals. We will consider the class of estimators for $\underline{P}$ of the form

$$(\forall A \in \mathscr{A})\ \underline{r}_n(A) = \min\{r_j(A) : k(n) \leq j \leq n\},$$

where k is some function such that $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$ (e.g. $k(n) = [\sqrt{n}]$). Then

$$D_n(A; k(n), \varepsilon) \subset [\underline{r}_n(A) + \underline{r}_n(A^c) < 1 - \varepsilon] \subset D_n(A; k(n), \varepsilon/2)$$

so that Theorem 4.1 may be interpreted in terms of $\underline{r}_n$; for example, (d) asserts that non-additivity of $\underline{r}_n$ is asymptotically favoured under $\underline{P}^\infty$ when $\underline{P}$ is itself non-additive.

Just as the finite relative frequency measures r_n support the usual additive probability theory (all r_n are additive probability measures, and all such measures can be attained as limits of sequences r_n), so do the functions $\underline{r}_n$ support an approach based on envelopes.

Moreover all sequences of outcomes, including those whose relative frequencies diverge, give rise to envelopes.

THEOREM 4.2. *$\underline{r}_n$ is a lower envelope on $\mathscr{A}$, $\underline{P}$ defined by $\underline{P}(A) = \liminf_{n\to\infty} \underline{r}_n(A) = \liminf_{n\to\infty} r_n(A)$ is the lower envelope of the class $\mathscr{M}_\omega$ of all limits of pointwise convergent subsequences of (r_n), and the class of set functions $\underline{P}_\omega$ generated in this way by infinite sequences ω of sample points is just the class of all lower envelopes on $\mathscr{A}$.*

PROOF. $\underline{r}_n$ is the lower envelope of the measures $r_{k(n)}, \cdots, r_n$. Since $\inf\{\underline{r}_n(A): n \geq m\} = \inf\{r_j(A): j \geq \min_{n\geq m} k(n)\}$, and $k(n) \to \infty$, $\liminf \underline{r}_n = \liminf r_n$. If $\underline{P} = \liminf r_n$, clearly $\underline{P}(A) \leq \pi(A)$ for all $\pi \in \mathscr{M}_\omega$ (such π are probability measures since $\mathscr{A}$ is finite). For any $A \in \mathscr{A}$ there is a subsequence (n_k) such that $r_{n_k}(A) \to \underline{P}(A)$, and by the Bolzano-Weierstrass Theorem (r_{n_k}) has a subsequence converging to a probability measure $\pi \in \mathscr{M}_\omega$. Clearly $\pi(A) = \underline{P}(A)$. Thus $\underline{P}$ is the lower envelope of $\mathscr{M}_\omega$.

To see that any lower envelope $\underline{P}$ on $\mathscr{A}$ can be generated in this way by some ω, use Lemma 2.3 to write $\underline{P}$ as the minimum of a finite set $\pi_1, \cdots, \pi_N$ of extreme measures. Find measures $\pi_{j0}, \pi_{j1}, \cdots$ such that $2^m\pi_{jm}(A)$ are integers and $\pi_{jm} \to \pi_j$. Let b_{jm} be a sequence of 2^m sample points generating relative frequencies π_{jm}. The sequence ω then consists of successive segments of length 2^{n^2}, in which the nth segment is made up of 2^{n^2-m} blocks b_{jm}, where $j(n) = (n-1)\bmod N + 1$, $m(n) = (n-j)/N$. Write $\ell_n = \sum_{j=1}^{n} 2^{j^2}$. Because $\ell_{n-1}/2^{n^2} \to 0$, the nth segment dominates previous segments, and $r_{\ell_n}(A) - \pi_{j(n)}(A) \to 0$. Hence $\liminf r_n \leq \underline{P}$. Because $m(n) < n$, any block b_{jm} of the nth segment is dominated by previous segments, so that all relative frequencies in the nth block can be approximated by a mixture of $\pi_{j(n)}$ and $\pi_{j(n-1)}$. Hence $\liminf r_n \geq \underline{P}$, which establishes the result. It can be seen also that if $k(n) = O(n^\lambda)$ with $\lambda < 1$ then $\underline{r}_n \to \underline{P}$ for the constructed sequence. Hence all lower envelopes can be attained as limits of $\underline{r}_n$. □

Popper (1959, Section 63–66) calls the limit points of (r_n) "middle frequencies," and admits "chance-like" sequences ω with more than one middle frequency. In his Appendix iv, Popper effectively demonstrates the existence of infinite sequences ω in which $\liminf r_n$ and $\limsup r_n$ are insensitive to certain types of subsequence-selection.

Results on the estimability of $\underline{P}$ through $\underline{r}_n$ are given in Section 5.3.

4.3 *A limiting frequentist interpretation of $\underline{P}$ and $\bar{P}$.* Theorem 4.2 suggests that, as a generalization of the limiting frequency interpretation of additive probability, we characterize the probabilities of events in $\mathscr{A}$ by some closed set $\mathscr{M}$ of probability measures on $\mathscr{A}$, to be interpreted as (ideal) limit points of relative frequencies in an infinite sequence of repetitions. We would expect $\mathscr{M}$ generated by an "IID" sequence of outcomes to be convex, but we will not assume convexity here. For a particular infinite sequence $\omega \in \Omega^\infty$, let $\mathscr{M}$ be the class of all limit points of (r_n). The following result shows that $\mathscr{M}$ characterizes the limiting behaviour of ω in the sense that it determines the upper and lower limits of continuous functionals of r_n.

LEMMA 4.2. *Let f be any real functional defined on all probability measures on $\mathscr{A}$ and continuous (under the natural topology) at all $\pi \in \mathscr{M}$. Then* $\liminf f(r_n) = \min\{f(\pi): \pi \in \mathscr{M}\}$.

PROOF. If $r_{n_k} \to \pi$ then $f(r_{n_k}) \to f(\pi)$, so $\liminf f(r_n) \leq \min\{f(\pi): \pi \in \mathscr{M}\}$. If $f(r_{n_k}) \to \liminf f(r_n)$, (r_{n_k}) has a subsequence converging to some $\pi \in \mathscr{M}$, and $f(\pi) = \liminf f(r_n)$ by continuity of f at π. □

As in the standard theory, we can use the limiting frequentist interpretation to suggest definitions of expectation, conditional probability, and independence. It is natural to identify *upper* and *lower expectations* $\bar{E}X$ and $\underline{E}X$, where $X: \Omega \to \mathbb{R}$, with the lim sup and lim inf of the average value of X in repetitions, i.e., $\underline{E}X = \liminf n^{-1}\sum_{j=1}^{n} X(\omega_j) = \liminf \int X dr_n$. Since $f(\pi) = \int X d\pi$ is continuous, Lemma 4.2 implies

$$\underline{E}X = \min\left\{\int X\, d\pi : \pi \in \mathscr{M}\right\}, \quad \bar{E}X = \max\left\{\int X\, d\pi : \pi \in \mathscr{M}\right\}.$$

Similarly, we might identify the *upper* and *lower probabilities* of A *conditional* on B with the upper and lower limits of the relative frequency of A in those trials on which B occurs, $\underline{P}(A|B) = \liminf r_n(A|B)$, where $r_n(A|B) = r_n(A \cap B)/r_n(B)$. Since $f(\pi) = \pi(A|B)$ is continuous at π for which $\pi(B) > 0$, Lemma 4.2 gives $\underline{P}(A|B) = \min\{\pi(A|B): \pi \in \mathscr{M}\}$ and $\bar{P}(A|B) = \max\{\pi(A|B): \pi \in \mathscr{M}\} = 1 - \underline{P}(A^c|B)$, both defined whenever $\underline{P}(B) = \min\{\pi(B): \pi \in \mathscr{M}\} > 0$. We note that $\liminf r_n(A|B)$ need not be determined by $\mathscr{M}$ when $\bar{P}(B) > \underline{P}(B) = 0$. When $\Omega = \{a, b, c\}$, for example, sequences ω can be constructed for which $\mathscr{M}$ is the convex hull of (1, 0, 0) and (0, 1, 0) but $\liminf r_n(\{a\}|\{a, c\})$ can take any value in [0, 1]. The personalist interpretation of upper and lower probabilities based on "coherence" leads to the same definitions of expectations and conditional probabilities, except that $\underline{P}(A|B)$ can be meaningfully defined whenever $\bar{P}(B) > 0$, by $\underline{P}(A|B) = \inf\{\pi(A|B): \pi \in \mathscr{M}, \pi(B) > 0\}$; see Walley (1981).

It might be argued that $\underline{P}(A|B)$ defined as $\liminf r_n(A|B)$, though it correctly describes limiting behaviour, may be a misleading representation of the information about the possible occurrence of A provided by the knowledge that B has occurred. The reason is that if $\pi \in \mathscr{M}$ and $\pi(B)$ is small, observation of B provides "evidence against" π in some sense, so that measures in $\mathscr{M}$ are no longer on an equal footing. As a simple example, suppose $A \subset B$ and $\mathscr{M}$ is the convex hull of π_1 and π_2, where $\varepsilon < ½$, $\pi_1(A) = \pi_2(A) = \varepsilon(1 - \varepsilon)$, $\pi_1(B) = 1 - \varepsilon$, $\pi_2(B) = \varepsilon$. Then $\underline{P}(A|B) = \underline{P}(A^c|B) = \varepsilon$, $\bar{P}(A|B) = \bar{P}(A^c|B) = 1 - \varepsilon$, yet B seems to support π_1 over π_2, and hence A^c over A.

The rule of conditioning for belief functions proposed by Dempster (1967),

$$\bar{Q}(A|B) = \bar{P}(A \cap B)/\bar{P}(B)$$

and $\underline{Q}(A|B) = 1 - \bar{Q}(A^c|B)$, each defined when $\bar{P}(B) > 0$, gives in this example (for which $\underline{P}$ is indeed a belief function) $\underline{Q}(A|B) = \bar{Q}(A|B) = \varepsilon$, which takes no account at all of π_2!

Dempster's rule gives in general more precise conditional probabilities than the above frequentist definition when $\mathscr{M} = \mathscr{M}(\underline{P})$. This is because, if $\underline{P}$ is defined as above, $\bar{P}(A|B) \geq \bar{P}(A \cap B)/\bar{P}(B)$. Hence, if $\underline{P}$ is a belief function and $\underline{P}(B) > 0$,

$$\bar{P}(A|B) \geq \bar{Q}(A|B) \geq \underline{Q}(A|B) \geq \underline{P}(A|B).$$

Finally, consider *independence* of experiments. Let ω be an infinite sequence of outcomes of a joint experiment, $\omega_j \in \Omega^2$. For additive probability, independence of the two marginal experiments can be characterized by the condition

$$(\forall A, B \in \mathscr{A}) \lim r_n(A \times B) = \lim r_n(A \times \Omega) \lim r_n(\Omega \times B).$$

If probability is identified with limiting relative frequency, the marginal probabilities π_1, π_2 then determine the additive joint probabilities π^2 by $\pi^2(A \times B) = \pi_1(A)\pi_2(B)$.

In general, in order to describe the marginals as "independent" we would require that the sequence ω satisfies

(a) $(\forall A, B \in \mathscr{A}) \liminf r_n(A \times B) = \liminf r_n(A \times \Omega) \liminf r_n(\Omega \times B)$,
(b) $(\forall A, B \in \mathscr{A}) \liminf r_n(\Omega \times B) > 0 \Rightarrow \liminf r_n(A \times \Omega | \Omega \times B) = \liminf r_n(A \times \Omega)$, and similarly for $r_n(\Omega \times B | A \times \Omega)$.

(a) asserts that lower limits of relative frequencies factorise over rectangles. (b) asserts that lower limits of relative frequencies of events in one marginal experiment are unaffected by restricting to a subsequence of repetitions in which an "independent" event B occurs, provided the relative frequency of B is eventually bounded away from zero.

Now let $\mathscr{M}^2$ be the set of limit points of (r_n), $\underline{P}^2(A) = \liminf r_n(A) = \min\{\pi(A): \pi \in \mathscr{M}^2\}$ for $A \subset \Omega^2$, $\underline{P}_1$ and $\underline{P}_2$ the marginals of $\underline{P}^2$, and $\underline{P}^2(A|B) = \liminf r_n(A|B) = \min\{\pi(A|B): \pi \in \mathscr{M}^2\}$ when $\underline{P}^2(B) > 0$. The $\underline{P}^2$ corresponding to "independent" sequences ω must then satisfy

(a′) $(\forall A, B \in \mathscr{A})\, \underline{P}^2(A \times B) = \underline{P}_1(A)\underline{P}_2(B)$

(b′) $(\forall A, B \in \mathscr{A})\, \underline{P}_2(B) > 0 \Rightarrow \underline{P}^2(A \times \Omega | \Omega \times B) = \underline{P}_1(A)$ and $\underline{P}_1(A) > 0 \Rightarrow \underline{P}^2(\Omega \times B | A \times \Omega) = \underline{P}_2(B)$.

Properties (a′) and (b′) may be required of any definition of the "independent product" ($\underline{P}^2$) of $\underline{P}_1$ and $\underline{P}_2$. The independent product defined in Section 3.1 satisfies (a′), by Corollary 3.1. Note that (b′) refers implicitly to $\mathscr{M}^2$, since conditional lower probabilities are not determined by the unconditional $\underline{P}^2$. If we take $\mathscr{M}^2 = \{\pi_1\pi_2: \pi_1 \in \mathscr{M}(\underline{P}_1), \pi_2 \in \mathscr{M}(\underline{P}_2)\}$, the product of Section 3.1 will satisfy (b′) also.

The IID products $\underline{R}^2(A) = \min\{\pi^2(A): \pi \in \mathscr{M}(\underline{P})\}$ satisfy condition (b′), with $\mathscr{M}^2 = \{\pi^2: \pi \in \mathscr{M}(\underline{P})\}$, but fail to satisfy (a′). The belief function products $\underline{Q}^2$, defined in an Appendix, satisfy (a′) but are incompatible with (b′). (Note that (a′) and (b′) are essentially the same condition for additive probability, but are quite distinct in general.) The limiting frequency interpretation therefore supports the definition of independence through $\underline{P}^2$ of Section 3.1, rather than $\underline{Q}^2$ or $\underline{R}^2$. It is not clear to us whether the product defined in Section 3.1 can be characterized through simple conditions like (a′) and (b′).

With $\underline{P}^2$ the independent product of Section 3.1, the convex hull of $\mathscr{M}^2 = \{\pi_1\pi_2: \pi_1 \in \mathscr{M}(\underline{P}_1), \pi_2 \in \mathscr{M}(\underline{P}_2)\}$ is generally strictly contained in $\mathscr{M}(\underline{P}^2)$. As can be seen from examples with $\|\Omega\| = 2$, $\pi(A \times \Omega | \Omega \times B)$ may be minimised over $\pi \in \mathscr{M}(\underline{P}^2)$ by some π not in $\mathscr{M}^2$, so that (b′) fails when conditional lower probabilities are defined through $\mathscr{M}(\underline{P}^2)$ rather than $\mathscr{M}^2$. The situation is simpler if we regard (a) and (b) as conditions on $\mathscr{M}^2$. Let $\mathscr{M}_1$ and $\mathscr{M}_2$ be the classes of projections onto the marginal spaces of measures in $\mathscr{M}^2$, and let $\text{ext}(\mathscr{M})$ denote the set of extreme points of $\mathscr{M}$. Call $\mathscr{M}^2$ an *independent class of measures* when

$$\text{ext}(\mathscr{M}^2) = \{\pi_1\pi_2: \pi_1 \in \text{ext}(\mathscr{M}_1), \pi_2 \in \text{ext}(\mathscr{M}_2)\}.$$

It is easily checked that infinite sequences corresponding to independent classes $\mathscr{M}^2$ satisfy (a) and (b). Moreover, independent classes of measures correspond to the independent products defined in Section 3, in the sense that $\underline{P}^2$ is such an independent product iff it is the lower envelope of an independent class of measures for which $\mathscr{M}_1 = \mathscr{M}(\underline{P}_1)$, $\mathscr{M}_2 = \mathscr{M}(\underline{P}_2)$. (As noted above, $\mathscr{M}(\underline{P}^2)$ is generally not itself an independent class). Thus, if "independent" marginal experiments are characterized by $\mathscr{M}_1$ and $\mathscr{M}_2$, interpreted as hypothetical limit points of relative frequencies in repetitions, the joint experiment is characterized by $\mathscr{M}^2$ whose extreme points are all products $\pi_1\pi_2$ with $\pi_i \in \text{ext}(\mathscr{M}_i)$. If convexity is required, $\mathscr{M}^2$ is then completely determined by $\mathscr{M}_1$ and $\mathscr{M}_2$.

5. Estimability.

Again consider an IID sequence of repetitions of $(\mathscr{A}, \underline{P})$, modeled by the IID product $(\mathscr{A}^\infty, \underline{P}^\infty)$. In this Section we define three notions of estimability of the unknown marginal lower envelope $\underline{P}$ from an observed sequence of outcomes, and show that in a weak sense the marginal is estimable through the minimum estimators $\underline{r}_n$.

5.1 *Concepts of estimability.* We define estimability of the marginal envelope with respect to a class ξ of lower envelopes on $\mathscr{A}$, interpreted as a class of possible marginals governing IID repetitions of the experiment. This generalizes the usual statistical specification in which ξ consists only of additive measures. We will say that the marginal is "estimable from ξ" when we can (in some sense) discriminate amongst the envelopes in ξ. "Estimability from ξ" is thus a property of the class ξ. For example, the marginal is estimable, in all three senses of the next definition, from the class of all additive probability measures on $\mathscr{A}$.

DEFINITION. The marginal envelope is said to be *partially estimable from* ξ when

$$(\forall n \geq 1)\, (\forall A \in \mathscr{A})\, \exists\, \hat{\underline{P}}_n(A): \Omega^\infty \to [0, 1],\ \hat{\underline{P}}_n(A)\mathscr{A}^n - \text{measurable},$$

such that $(\forall \underline{P} \in \xi)(\forall \varepsilon > 0)([|\hat{\underline{P}}_n(A) - \underline{P}(A)| < \varepsilon])$ a.f. under $\underline{P}^\infty$. (For example, Corollary 5.2 will establish partial estimability from a wide class of lower envelopes with $\hat{\underline{P}}_n = \underline{r}_n$, the minimum estimators of Section 4.2). The marginal is said to be *completely estimable from* ξ when $(\forall n \geq 1)(\forall A \in \mathscr{A}) \exists\ \hat{\underline{P}}_n(A): \Omega^\infty \to [0, 1]$, $\hat{\underline{P}}_n(A)\ \mathscr{A}^n$ – measurable, such that

$$(\forall \underline{P} \in \xi)(\forall \varepsilon > 0)(\cap_{A \in \mathscr{A}} [|\hat{\underline{P}}_n(A) - \underline{P}(A)| < \varepsilon]) \text{ a.f. under } \underline{P}^\infty.$$

Say that the marginal is *strongly estimable from* ξ when either definition is satisfied with "a.f." replaced by "a.c." (It follows from Lemma 3.2 that the two properties involving a.c. are equivalent).

We have therefore three concepts of estimability: strong estimability, which implies complete estimability, which implies partial estimability (from a fixed ξ). When ξ consists only of additive probability measures the three concepts coincide with the kind of estimability established by the weak law of large numbers (since then a.c. coincides with a.f.). We are interested in weaker restrictions on ξ sufficient to guarantee estimability.

Consider the case of finite $\xi = \{\underline{P}_1, \cdots, \underline{P}_m\}$, which corresponds to simple hypothesis testing. Write $\underline{P}_j^\infty$ for the infinite IID product of $\underline{P}_j$. In this case strong (complete) estimability from ξ is easily seen to be equivalent to the following property

$$(\forall n \geq 1)(\exists \text{ disjoint } E_1^n, \cdots, E_m^n \in \mathscr{A}^n)$$
$$(\forall 1 \leq j \leq m)(E_j^n)\text{a.c.[a.f.]} \quad \text{as} \quad n \to \infty \quad \text{under} \quad \underline{P}_j^\infty.$$

Similar properties correspond to discrimination amongst finitely many compound hypotheses.

5.2 *Strong estimability.* We show next that strong estimability is not possible in general, even in the case of finite ξ.

THEOREM 5.1. *For the marginal envelope to be strongly estimable from ξ it is necessary that $\{\mathscr{M}(\underline{P}): \underline{P} \in \xi\}$ are disjoint. If ξ is finite, this condition is also sufficient.*

PROOF. For necessity, suppose $\pi \in \mathscr{M}(\underline{P}_1) \cap \mathscr{M}(\underline{P}_2)$. If the marginal is strongly estimable there are $E_1^n \cap E_2^n = \phi$ with $\underline{P}_1^\infty(E_1^n) \to 1, \underline{P}_2^\infty(E_2^n) \to 1$. But $\pi^\infty \geq \underline{P}_1^\infty$ and $\pi^\infty \geq \underline{P}_2^\infty$ by definition of the IID product, contradicting disjointness of E_1^n and E_2^n.

For sufficiency, let $\xi = \{\underline{P}_1, \cdots, \underline{P}_m\}$. By Lemma 2.3, $\xi_j = \mathscr{M}(\underline{P}_j)$ are closed in the compact space of all probability measures on $\mathscr{A}$. Write $\xi_j^k = \mathscr{M}(\underline{P}_j - k^{-1})$. Then $\xi_j^k \supset \xi_j^{k+1}$, ξ_j^k are compact, and $\cap_{k \geq 1} \xi_j^k = \xi_j$ so that $\cap_{k \geq 1} (\xi_i^k \cap \xi_j^k) = \phi$. The Finite Intersection Property implies that $\xi_i^k \cap \xi_j^k = \phi$ for some k. Similarly, $\xi_1^k, \cdots, \xi_m^k$ are disjoint for some k. Write $\delta = k^{-1}$. Define

$$E_j^n = \cap_{A \in \mathscr{A}} [r_n(A) \geq \underline{P}_j(A) - \delta] \in \mathscr{A}^n.$$

Then $E_1^n, \cdots, E_m^n$ are disjoint. But

$$\underline{P}_j^\infty([r_n(A) \geq \underline{P}_j(A) - \delta]) = \pi^\infty([r_n(A) \geq \pi(A) - \delta]) \to 1,$$

where $\pi(A) = \underline{P}_j(A)$. Applying Lemma 3.2, (E_j^n) a.c. under $\underline{P}_j^\infty$, which establishes strong estimability. By Lemma 4.1, the upper error probabilities $e_{jn} = \bar{P}_j^\infty(E_j^n)^c$ are in fact bounded by $\limsup n^{-1} \log e_{jn} \leq -\alpha_j$, where $\alpha_j > 0$ depends on $\underline{P}_j$ and δ. □

Thus, it is possible to discriminate with asymptotic certainty amongst a finite set of envelopes if and only if their sets of dominating probability measures are disjoint. If this condition is satisfied, the marginal is strongly estimable through the estimator $\hat{\underline{P}}_n$ that takes the value $\underline{P}_j$ whenever the terminal relative frequency r_n "almost dominates" $\underline{P}_j$, i.e. $(\forall A \in \mathscr{A}) r_n(A) \geq \underline{P}_j(A) - \delta$, with $\hat{\underline{P}}_n$ otherwise defined arbitrarily. Alternatively, using Theorem 4.1(a), define $\hat{\underline{P}}_n = \underline{P}_j$ whenever the minimum estimator $\underline{r}_n$ "almost dominates" $\underline{P}_j$.

The main result of Huber and Strassen (1973), which characterizes "most powerful" tests for discriminating between two lower envelopes that are 2-monotone, fits naturally into the above framework.

THEOREM 5.2. (Huber-Strassen). *Let $\underline{P}_0$ and $\underline{P}_1$ be 2-monotone functions on $\mathscr{A}$, with $\mathscr{M}(\underline{P}_0) \cap \mathscr{M}(\underline{P}_1) = \phi$, and $\underline{P}_0^\infty$, $\underline{P}_1^\infty$ their infinite* IID *products. Let $A_\alpha^n \in \mathscr{A}^n$ be a critical region for the minimax test of $\underline{P}_0$ against $\underline{P}_1$ based on n* IID *repetitions, chosen to maximize the minimum power $\underline{P}_1^\infty(A)$ subject to level $\bar{P}_0^\infty(A) \leq \alpha$. Then there are "least-favorable" distributions $\pi_0 \in \mathscr{M}(\underline{P}_0)$, $\pi_1 \in \mathscr{M}(\underline{P}_1)$, not depending on n or α, such that A_α^n may be taken to be an optimal level-α critical region for testing π_0^n against π_1^n, with $\underline{P}_1^\infty(A_\alpha^n) = \pi_1^n(A_\alpha^n)$, $\bar{P}_0^\infty(A_\alpha^n) = \pi_0^n(A_\alpha^n)$.*

Although Huber and Strassen state their result in terms of $\underline{R}^\infty = \min\{\pi^\infty: \pi \in \mathscr{M}(\underline{P})\}$, both Theorems 5.1 and 5.2 apply to each of the products $\underline{P}^\infty$, $\underline{R}^\infty$, and $\underline{Q}^\infty$ defined in the Appendix. (It is easily seen, using the Theorem in the Appendix, that $\pi_1^\infty(A_\alpha^n) = \underline{Q}_1^\infty(A_\alpha^n) = \underline{P}_1^\infty(A_\alpha^n)$). The above version of Theorem 5.2 involving $\underline{P}^\infty$, allowing models in which the underlying probability measure varies between trials, will often be appropriate when robustness is a concern. (An important special case, for which Huber gives the least-favorable π_i, is ε-contamination in which the contaminating distribution may depend on the trial.)

5.3 *Partial and complete estimability.* It is well known that the extension of the class of deterministic models to the class of additive probability models loses *strict falsifiability*—typically, all experimental outcomes are compatible with many probability models. As we have seen, the further extension to include non-additive models loses *strong estimability.* For example, the marginal is strongly estimable from the class of all additive probability measures, but strong estimability fails as soon as this class is extended to include any non-additive envelope. "Typical" sample sequences under $\underline{P}$ are those with $\liminf r_n \geq \underline{P}$, and "typical" sets for non-additive envelopes with common dominating measures will overlap. Again, non-additivity favors divergence of relative frequencies (Theorem 4.1(d)), but *not* with asymptotic certainty (Theorem 4.1(c)). We seem forced therefore to consider the weaker notions of asymptotic favorability and complete or partial estimability.

First, we show that the marginal envelope is partialy estimable from a wide class of envelopes, through the minimum estimators

$$\underline{r}_n(A) = \min\{r_j(A): k(n) \leq j \leq n\}$$

discussed in Section 4.2. We shall assume that $k(n)/n \to 0$ and $k(n)/\log n \to \infty$ as $n \to \infty$.

LEMMA 5.1. *Suppose $\bar{P}(A) < 1$ and $\varepsilon > 0$. Write*

$$g_n(\pi_1, \cdots, \pi_n) = (\textstyle\prod_{j=1}^n \pi_j)\,[\underline{r}_n(A) \leq \underline{P}(A) - \varepsilon \,|\, \underline{r}_n(A) < \underline{P}(A) + \varepsilon].$$

Then

$$\sup\{g_n(\pi_1, \cdots, \pi_n): \pi_1, \cdots, \pi_n \in \mathscr{M}(\underline{P})\} \to 0 \quad \text{as} \quad n \to \infty.$$

PROOF. We use the results of Samuels (1965). Write $p = \underline{P}(A)$, $\sigma = p - \varepsilon$, $\tau = p - \varepsilon/2$, $\pi^{(n)} = \prod_{j=1}^n \pi_j$. We may suppose $\underline{P}(A) \geq \varepsilon$, so $0 \leq \sigma < \tau < p < 1$. Let m maximize $\pi^{(m)}[r_m(A) \leq \sigma]$ over $k(n) \leq m \leq n$.

$$\frac{\pi^{(n)}[\underline{r}_n(A) \leq p - \varepsilon]}{\pi^{(n)}[\underline{r}_n(A) < p + \varepsilon]} \leq \frac{n \cdot \pi^{(m)}[r_m(A) \leq \sigma]}{\pi^{(m)}[r_m(A) \leq \tau]} \leq \frac{n^2 f_{[m\sigma]}}{f_{[m\tau]}},$$

where $f_k = \pi^{(m)}[mr_m(A) = k]$ and $[x]$ denotes the integer part of x, since f_k is increasing in

k for $k < \sum_{i=1}^{m} \pi_i(A)$ (Samuels, Theorem 1). But

$$\frac{f_{[m\tau]}}{f_{[m\sigma]}} = \prod_{k=[m\sigma]+1}^{[m\tau]} \frac{f_k}{f_{k-1}} \geq \left(\frac{f_{[m\tau]}}{f_{[m\tau]-1}}\right)^{[m\tau]-[m\sigma]}$$

since f_k/f_{k-1} is decreasing in k (Samuels, equation 6). But f_k/f_{k-1} is increasing in each $\pi_i(A)$ (Samuels, equation 7); hence

$$\frac{f_k}{f_{k-1}} \geq \left(\frac{m-k+1}{k}\right)\left(\frac{p}{1-p}\right), \quad \text{its value when all} \quad \pi_i(A) = p.$$

Thus,

$$\frac{f_{[m\tau]}}{f_{[m\sigma]}} \geq \left[\left(\frac{m-[m\tau]+1}{[m\tau]}\right)\frac{p}{1-p}\right]^s,$$

$$\text{where} \quad s = [m\tau] - [m\sigma] > m(\tau - \sigma) - 1,$$

$$> \left[\left(\frac{m}{m\tau} - 1\right)\left(\frac{p}{1-p}\right)\right]^s$$

$$= \gamma^s,$$

where

$$\gamma = \left(\frac{p}{1-p}\right)\Big/\left(\frac{\tau}{1-\tau}\right) > 1.$$

Thus,

$$\frac{\pi^{(n)}[\underline{r}_n(A) \leq p - \varepsilon]}{\pi^{(n)}[\underline{r}_n(A) < p + \varepsilon]} < n^2\gamma^{-m(\tau-\sigma)+1} \leq n^2\gamma^{-k(n)(\tau-\sigma)+1} = \delta_n,$$

and $\delta_n \to 0$ provided $k(n)/\log n \to \infty$.

LEMMA 5.2. *If* $0 < \lambda < \bar{P}(A) < 1$ *then* $k(n)^{-1}\log \underline{P}^\infty[\underline{r}_n(A) < \lambda] \to -\alpha(\bar{P}(A), \lambda)$, *where* $\alpha(p, \lambda) = \lambda \log(\lambda/p) + (1-\lambda)\log((1-\lambda)/(1-p))$.

PROOF. $\underline{P}^\infty[\underline{r}_n(A) < \lambda]$ is clearly achieved by π^∞, where $\pi(A) = \bar{P}(A)$.

$$\pi^\infty[r_{k(n)}(A) < \lambda] \leq \pi^\infty[\underline{r}_n(A) < \lambda] \leq \textstyle\sum_{j=k(n)}^{n} \pi^\infty[r_j(A) < \lambda].$$

Apply Lemma 4.1. □

COROLLARY 5.1. *If* $\varepsilon > \bar{P}(A) - \underline{P}(A)$,

$$\underline{P}^\infty[|\underline{r}_n(A) - \underline{P}(A)| < \varepsilon] \to 1.$$

If $0 < \varepsilon < \bar{P}(A) - \underline{P}(A)$ *and* $\bar{P}(A) < 1$,

$$k(n)^{-1}\log \underline{P}^\infty[|\underline{r}_n(A) - \underline{P}(A)| < \varepsilon] \to -\alpha(\bar{P}(A), \underline{P}(A) + \varepsilon).$$

PROOF. The first statement is a consequence of Theorem 4.1(a). For the second,

$$P^\infty[|\underline{r}_n(A) - \underline{P}(A)| < \varepsilon]/\underline{P}^\infty[\underline{r}_n(A) < \underline{P}(A) + \varepsilon] \to 1$$

by Lemma 5.1 (let $\pi_1, \cdots, \pi_n$ attain the minimum in the numerator). But $k(n)^{-1}$ $\log \underline{P}^\infty[\underline{r}_n(A) < \underline{P}(A) + \varepsilon] \to -\alpha(\bar{P}(A), \underline{P}(A) + \varepsilon)$ by Lemma 5.2. □

THEOREM 5.3. *Suppose* $\bar{P}(A) < 1$ *or* $\underline{P}(A) = 1$. *Then* $(\forall\varepsilon > 0)([|\underline{r}_n(A) - \underline{P}(A)| < \varepsilon])$ *is asymptotically favoured under* $\underline{P}^\infty$.

PROOF. If $\underline{P}(A) = \bar{P}(A)$, this follows from Corollary 5.1. Suppose then that $0 < \varepsilon <$

$\bar{P}(A) - \underline{P}(A)$, $\bar{P}(A) < 1$. Let $\ell(n) = [k(n)^{2/3}n^{1/3}]$, $m(n) = [k(n)^{1/3}n^{2/3}]$, $\omega(n) = n - m(n)$, so that $k(n)/\ell(n)$, $\ell(n)/m(n)$, $m(n)/n$ all $\to 0$. Let r_j' denote relative frequency on trials $m(n) + 1, \cdots m(n) + j$. For n sufficiently large,

$$G_n = [\,|\underline{r}_n(A) - \underline{P}(A)| < \varepsilon] \supset D_n \cap E_n \cap F_n,$$

where

$$D_n = [\min_{k(n)\le j\le m(n)} r_j(A) \ge \underline{P}(A)]$$

$$E_n = [\min_{\ell(n)\le j\le \omega(n)} r_j'(A) > \underline{P}(A) - \varepsilon]$$

and

$$F_n = [r'_{\omega(n)}(A) \le \underline{P}(A) + \varepsilon/2].$$

Let

$$\pi_j(A) = \begin{cases} \bar{\pi}(A) = \bar{P}(A) & \text{for } 1 \le j \le m(n) \\ \underline{\pi}(A) = \underline{P}(A) & \text{for } m(n) < j \le n. \end{cases}$$

$$\underline{P}^\infty(G_n^c) \le (\textstyle\prod_{j=1}^n \pi_j)(D_n^c \cup E_n^c \cup F_n^c) \le \bar{\pi}^{m(n)}(D_n^c) + \underline{\pi}^{\omega(n)}(E_n^c \cup F_n^c).$$

By Lemmas 4.1 and 5.2,

$$k(n)^{-1}\log \bar{\pi}^{m(n)}(D_n^c) \to -\alpha(\bar{P}(A), \underline{P}(A)),$$

$$\ell(n)^{-1}\log \underline{\pi}^{\omega(n)}(E_n^c) \to -\alpha(\underline{P}(A), \underline{P}(A) - \varepsilon),$$

$$\omega(n)^{-1}\log \underline{\pi}^{\omega(n)}(F_n^c) \to -\alpha(\underline{P}(A), \underline{P}(A) + \varepsilon/2).$$

Thus

$$\limsup k(n)^{-1}\log \underline{P}^\infty(G_n^c) \le -\alpha(\bar{P}(A), \underline{P}(A)) < -\alpha(\bar{P}(A), \underline{P}(A) + \varepsilon)$$
$$= \lim k(n)^{-1}\log \underline{P}^\infty(G_n).$$

So $\underline{P}^\infty(G_n^c)/\underline{P}^\infty(G_n) \to 0$.

Definition. A set function $\underline{P}$ on $\mathscr{A}$ is called *almost positive* when $(\forall A \in \mathscr{A})\underline{P}(A) = 0 \Rightarrow \bar{P}(A) = 0$.

Corollary 5.2. *The marginal lower envelope is partially estimable from any class of almost positive lower envelopes, through the minimum estimators*

$$\underline{r}_n(A) = \min\{r_j(A): k(n) \le j \le n\},$$

provided $k(n)/n \to 0$ *and* $k(n)/\log n \to \infty$.

Proof. Almost positivity implies that $(\forall A \in \mathscr{A})\bar{P}(A) < 1$ or $\underline{P}(A) = 1$. Partial estimability then follows from Theorem 5.3. □

(In fact, $\underline{P}$ is partially estimable through $\underline{r}_n$ if and only if ξ contains no lower envelopes which are not almost positive).

Corollary 5.3. *The marginal is completely estimable from any finite class* ξ *of almost positive lower envelopes.*

Proof. Let $\xi = \{\underline{P}_1, \cdots, \underline{P}_u\}$, where $\underline{P}_j$ is trivial (i.e., $(\exists\omega \in \Omega)\underline{P}_j(\{\omega\}) = 1$) for $m < j \le u$, non-trivial for $1 \le j \le m$. For some $i \ge 1$, it is possible to construct $A \in \mathscr{A}^i$ such that the values $\underline{P}_1^i(A), \cdots, \underline{P}_m^i(A)$ are all distinct, and $(\forall 1 \le j \le m) 0 < \underline{P}_j^i(A) < 1$ (details omitted). Let $\underline{r}_\ell^{(i)}: \mathscr{A}^i \to [0, 1]$ denote the minimum estimator for ℓ repetitions of the i-fold joint experiment. For $1 \le j \le m$ write $E_j^n = [\,|\underline{r}_\ell^{(i)}(A) - \underline{P}_j^i(A)| < \delta] \in \mathscr{A}^n$, where $\ell = [n/i]$. If $\underline{P}_j(\{\omega\}) = 1$, let $E_j^n = \cap_{k=1}^n \{\omega : \omega_k = \omega\}$, so that $\underline{P}_j^\infty(E_j^n) = 1$. For δ sufficiently small, $\{E_1^n, \cdots, E_u^n\}$ are disjoint.

Theorem 5.3 cannot be applied directly to $\underline{P}_j^i(A)$, because $(\underline{P}_j^i)^\infty \ne \underline{P}_j^\infty$. But Theorem 5.3 can be generalized to apply to any $\underline{P}^\infty$ of the form $\underline{P}^\infty(A) = \min\{(\prod_{j=1}^\infty \pi_j)(A): \pi_j \in \mathscr{M}\}$, for any closed $\mathscr{M}$ with lower envelope $\underline{P}$ (since $\exists\underline{\pi}, \bar{\pi} \in \mathscr{M}$ with $\underline{\pi}(A) = \underline{P}(A)$, $\bar{\pi}(A) = \bar{P}(A)$, so that the bounds in the proof of Theorem 5.3 must be achieved by measures in $\mathscr{M}$). Applying this generalization to the event $A \in \mathscr{A}^{\ell}$ and the class $\mathscr{M}^{i}_{j} = \{(\prod_{k=1}^{i} \pi_k) : \pi_k \in \mathscr{M}(\underline{P}_j)\}$, which has lower envelope $\underline{P}^{\ell}_j$,

$$\frac{\underline{P}_j^{\infty}(E_j^n)^c}{\underline{P}_j^{\infty}(E_j^n)} = \frac{\min\{(\prod_{s=1}^{\infty} \mu_s)(E_j^n)^c : \mu_s \in \mathscr{M}_j^i\}}{\min\{(\prod_{s=1}^{\infty} \mu_s)(E_j^n) : \mu_s \in \mathscr{M}_j^i\}} \to 0$$

as $n \to \infty$ (so that $\ell \to \infty$). Thus, for $1 \le j \le m$, (E_j^n) a.f. under $\underline{P}_j^{\infty}$. □

Is complete estimation possible in general? For the minimum estimator $\underline{r}_n$ we have, by Theorem 4.1(a),

$$(\forall \varepsilon > 0)(\cap_{A \in \mathscr{A}}[\underline{r}_n(A) > \underline{P}(A) - \varepsilon]) \text{ a.c. under } \underline{P}^{\infty}.$$

Moreover, we can show that if $\underline{P}$ is almost positive then

$$(\forall \varepsilon > 0)(\cap_{A \in \mathscr{A}}[\underline{r}_n(A) < \underline{P}(A) + \varepsilon]) \text{ a.f. under } \underline{P}^{\infty}.$$

But, as the next result shows, it does *not* follow that $\underline{P}$ is completely estimable through $\underline{r}_n$ – Lemma 3.2 fails for asymptotic favourability.

THEOREM 5.4. *Suppose $\underline{P}$ is a lower envelope, with $0 < \underline{P}(A) < \bar{P}(A) < 1$ and $\alpha(\underline{P}(A), \bar{P}(A)) \neq \alpha(\bar{P}(A), \underline{P}(A))$, α as in Lemma 5.2. Write*

$$G_n(A) = [\,|\underline{r}_n(A) - \underline{P}(A)| < \varepsilon],$$

$$G_n = G_n(A) \cap G_n(A^c).$$

Then for ε sufficiently small (G_n^c) is asymptotically favoured under $\underline{P}^{\infty}$.

PROOF. Suppose $\alpha(\underline{P}(A), \bar{P}(A)) < \alpha(\bar{P}(A), \underline{P}(A)) = \alpha(\underline{P}(A^c), \bar{P}(A^c))$ (otherwise, substitute A^c for A). Let ε be so small that

$$\alpha_1 = \alpha(\underline{P}(A), \bar{P}(A) + \varepsilon) < \alpha(\bar{P}(A), \underline{P}(A) + \varepsilon) = \alpha_2.$$

Since $G_n \subset G_n(A)$, Corollary 5.1 gives

$$\limsup k(n)^{-1} \log \underline{P}^{\infty}(G_n) \le -\alpha_2.$$

But $G_n^c \supset G_n(A^c)^c \supset [r_{k(n)}(A) \ge \bar{P}(A) + \varepsilon]$, so by Lemma 4.1

$$\liminf k(n)^{-1} \log \underline{P}^{\infty}(G_n^c) \ge -\alpha_1 > -\alpha_2.$$

Hence (G_n^c) a.f. under $\underline{P}^{\infty}$. □

Thus, the marginal is not in general completely estimable through the minimum estimator. We conjecture that it is not completely estimable from any wide class of lower envelopes, e.g. all positive lower envelopes, through *any* function $\hat{\underline{P}}_n$.

6. Discussion. The two results preceding Theorem 5.4, together with the partial estimability property (Corollary 5.2), give some support to the class of minimum estimators $\underline{r}_n$ as estimators for a lower envelope $\underline{P}$ governing unlinked repetitions of an experiment. But even if we ignore the admittedly weak property of asymptotic favourability, we might still be led to something like the minimum estimators. By Theorem 4.1(a), we can assert with asymptotic certainty that $\underline{r}_n \ge P - \varepsilon$ in IID repetitions. We might then invoke a criterion of *maximal falsifiability* to support the most precise model $\hat{\underline{P}}$ consistent with $\underline{r}_n \ge \hat{\underline{P}} - \varepsilon(n)$, so that $\hat{\underline{P}}_n = \underline{r}_n$ approximately. Estimation of $\underline{P}$ by $\underline{r}_n$ is conservative with respect to the standard additive probability concept, in that the model chosen is close to additive ($\underline{r}_n$ close to r_n) except when relative frequencies appear to be diverging. We make no claim, however, that the minimum estimators are "optimal" in any sense; alternative estimators with the partial estimability property can be found, and more study is needed to distinguish the "good" estimators.

Although we have concentrated in this paper on the extreme case of divergent relative frequencies, it could be argued that upper and lower probability models are appropriate also for some time series whose relative frequencies appear to converge more slowly than expected under an additive IID model, so that different long blocks of outcomes have quite different relative frequencies. An estimator such as

$$\hat{\underline{P}}_n(A) = \min\{k^{-1} \textstyle\sum_{i=j-k+1}^{j} I_A(\omega_i) : j = k, k+1, \cdots, n\},$$

for suitable $k(n)$, would reflect such variation in frequency of occurrence, and could give non-additive estimates even when relative frequencies converged.

The assumption of stability of relative frequencies is of course central to the application of additive probability theory. It is often claimed to be an "empirical fact" that relative frequencies apparently converge for many phenomena of interest (Fine, 1970, argues that apparent convergence is a consequence of our data-processing procedures rather than of any "laws of nature"). This is taken to support an additive model which predicts or explains apparent convergence as in Theorem 4.1(b). We see that non-additive models with sufficiently small interval widths $\bar{P}(A) - \underline{P}(A)$ can equally well explain apparent convergence. Moreover, apparent convergence is evidence against some non-additive models with large interval widths (those lower envelopes dominated by the terminal relative frequency measure) only in the weaker a.f. sense. If nonetheless one accepted apparent convergence (for sufficiently small $\varepsilon(n)$) as evidence for an underlying additive IID measure, on grounds of maximal falsifiability or the greater familiarity of additive probability, then $\underline{r}_n$ and its conjugate $\bar{r}_n$ would be interpreted as lower and upper bounds for this measure. Assuming an underlying additive measure, the appropriate model for IID repetitions is $\underline{R}^\infty$, and the maximal interval width max $\{\bar{r}_n(A) - \underline{r}_n(A) : A \in \mathscr{A}\}$ will tend to zero with asymptotic certainty under $\underline{R}^\infty$. The imprecision of $\underline{r}_n$ in this case reflects merely incomplete knowledge about the true underlying measure due to the limited number of observations ("epistemological indeterminacy").

In the case of apparent divergence of relative frequencies we would argue that non-additive models $\underline{P}$, representing "ontological indeterminacy", should be considered. Several interpretations are possible. First, the observed outcomes might be viewed as realizations of independent additive measures $\pi_j \in \mathscr{M}$ which varied between trials. If this variation is arbitrary or more detailed models were either not of interest or not estimable from data, the IID product model $\underline{P}^n$ would be appropriate. The estimator $\underline{r}_n$ would then estimate the lower envelope of the underlying class $\mathscr{M}$ of probability measures that generate the observed outcomes.

A second interpretation takes $\mathscr{M}$ to be the set of limit points, and its lower envelope $\underline{P}$ to be the lim inf, of relative frequencies in hypothetical unlinked repetitions. As in Section 4.3, this interpretation again supports the independent product $\underline{P}^n$ as a characterization of "unlinkedness" of experiments. Note that $\underline{P}^\infty$[lim inf $r_n \geq \underline{P}$] = 1, but $\underline{P}^\infty$[lim inf $r_n = \underline{P}$] = 0 if $\underline{P}$ is nonadditive, so that the model is not self-supporting in the same way as additive probability models.

Thirdly, $\underline{r}_n$ might be regarded as estimating a distribution of propensities over events, to be represented in general by a non-additive envelope $\underline{P}$. That is, the physical tendencies of an experiment to give rise to various possible outcomes are represented directly by $\underline{P}$ without intermediate reference to its dominating measures. There is no obvious reason for modern propensity interpretations of probability to insist that propensites have an additive representation, although this restriction seems to have been taken for granted. Propensities are theoretical, dispositional terms referring to physical properties of experimental arrangements which are related to (but not defined through) relative frequencies in repetitions. A propensity account of probability should explicate this relation. On our account, propensities $\underline{P}$ imply a disposition to produce lim inf $r_n(A) \geq \underline{P}(A)$ in unlinked repetitions. Thus there is a connection between propensity and frequency, without propensity distributions necessarily sharing the additivity of relative frequencies. Note however that $\underline{P}$ is no longer interpreted as a bound on underlying measures, so that the definition of independence through $\underline{P}^n$ rather than, say, $\underline{Q}^n$ becomes an issue.

Finally, a more pragmatic justification for the non-additive models studied here is that they can extend the class of non-deterministic phenomena that we can usefully model. Most propensity accounts of additive probability (e.g. Giere, 1973) regard propensities as characterizing some "ultimate" randomness that is inherent in nature. It is very difficult to argue convincingly that particular phenomena are ultimately random, though there is strong evidence in the case of quantum phenomena. Nevertheless, additive probability models are successfully used to model data generated by a wide variety of incompletely understood phenomena. The use of such models does not rule out the possibility of more refined (perhaps deterministic) models for the same phenomena. Upper and lower probability models might have a similar descriptive role, without any implications of "ultimate indeterminacy," in cases where no useful additive probability model is available. For instance, repeated trials which are physically unlinked and indistinguishable may nevertheless produce diverging relative frequencies. Many geophysical, economic and sociological time series, for which little is known about any dependence between successive observations, display similar instability. Such behaviour can be accommodated within standard probability theory only through a non-stationary model, possibly a complex model with no basis in our understanding of the phenomenon and no predictive value. A non-additive IID model may be much simpler, and may be sufficiently precise to give useful predictions, e.g. of the type in Theorem 4.1(a). (The notion of complexity involved here can be given a precise, though possibly too narrow, explication in terms of the Kolmogorov-Chaitin-Solomonoff conditional complexity of empirical time series).

These issues arise, at least on a conceptual level, even for such a paradigm of the standard theory as die tossing. Toss any real die long enough and it abrades, corners round, and its propensities at least would have to be held to be time varying. But how precisely can this time variation be described? The variation is evidently not deterministic, but is it then a stochastic process? If it is a stochastic process, are the relevant propensities precisely calculable, at least in principle? If one is unwilling to postulate the detailed regularities implied by a stochastic process governing fluctuations in outcome probabilities, one might accept that the propensities of the die to produce outcomes are not as well determined as is envisaged in propensity accounts of additive probability. One might then represent the intrinsically indeterminate propensities of the die by a non-additive structure incorporating the right degree of numerical imprecision (which need not be large) to represent the degree of indeterminacy. The usual Bernoulli model for die tossing is of course adequate for most practical purposes, but one can imagine applications for which it would be an over-idealization.

Our purpose in modeling is to extract from data simple regularities that are generalizable, predictively useful, and related to our theoretical knowledge of the phenomenon modeled. Other aspects of the data are regarded as merely "accidental". Just as "randomness" (chance) is introduced in additive probability models to account for poorly understood ("accidental") variation in outcomes, so "indeterminacy" might be introduced in upper and lower probability models to account for poorly understood variations in chance behaviour. The choice between deterministic, additive probability and non-additive probability models must depend on our background knowledge concerning the regularities they extract from data; in particular, on theoretical understanding of the mechanisms involved, generalizability of time variation to related series, and confidence in reproducibility of past regularities in future observations, as well as precision and complexity. These are difficult issues, which need to be carefully examined in the context of particular applications.

Acknowledgment. We wish to thank Dr. Anurag Kumar and one of the referees for their many helpful suggestions.

APPENDIX—AN ALTERNATIVE MODEL FOR INDEPENDENCE

The notation is that of Section 3.1. Suppose the marginal lower envelopes $\underline{P}_j\colon \mathscr{A} \to [0, 1]$, for $1 \le j \le n$, are belief functions, with probability assignments m_j. Define the *belief* *function product* $\underline{Q}^n$: $\mathscr{A}^n \to [0, 1]$ of $\underline{P}_1, \cdots, \underline{P}_n$ to be the belief function on $\mathscr{A}^n$ whose probability assignment m is concentrated on rectangles, and given by

$$(\forall A_j \in \mathscr{A}) m(X_{j=1}^n A_j) = \prod_{j=1}^n m_j(A_j).$$

Then $\underline{Q}^n$ has marginals $\underline{P}_j$ and is additive iff all its marginals are additive.

In a similar way to the independent product $\underline{P}^n$ of Section 3.1, the belief function product $\underline{Q}^n(A)$ may be interpreted as the greatest lower bound to the probability of A in trials whose outcomes ω_j are chosen in an unknown way from subsets $A_j \in \mathscr{A}$ which occur independently according to measures m_j. Thus, $\underline{Q}^n$ models observations of independent trials with imprecise outcomes, as discussed by Walley and Fine (1979). $\underline{Q}^n$ may also be interpreted through the related multivalued mappings of Dampster (1967, 1968).

In general, $\underline{Q}^n$ differs from the independent product $\underline{P}^n$ with same marginals $\underline{P}_j$. Write $\bar{Q}^n$ for the upper envelope conjugate to $\underline{Q}^n$. As in Lemma 3.1 and Corollary 3.1, both $\underline{Q}^n$ and $\bar{Q}^n$ factorize on rectangles so that, whenever all the $\underline{P}_j$ are belief functions,

$$\underline{Q}^n(X_{j=1}^n A_j) = \underline{P}^n(X_{j=1}^n A_j) = \prod_{j=1}^n \underline{P}_j(A_j) \quad \text{and} \quad \bar{Q}^n(X_{j=1}^n A_j) = \bar{P}^n(X_{j=1}^n A_j) = \prod_{j=1}^n \bar{P}_j(A_j).$$

Thus $\underline{P}^n$ and $\underline{Q}^n$ agree on rectangles. The next result (whose proof we omit) shows that $\underline{P}^n$ always dominates $\underline{Q}^n$, and characterizes the sets on which they agree.

THEOREM. *Suppose all the marginals* $\underline{P}_j$ $(1 \leq j \leq n)$ *are belief functions. Then*

$$(\forall A \in \mathscr{A}^n) \underline{P}^n(A) \geq \underline{Q}^n(A),$$

with equality iff *A can be written as*

$$A = [\cup_{i=1}^N X_{j=1}^n C_{ij}] \cup A_0$$

for some $A_0 \in \mathscr{A}^n$ *and* $C_{ij} \in \mathscr{A}$ *such that for each* $j (1 \leq j \leq n)$ *there is* $\pi_j \in \mathscr{M}(\underline{P}_j)$ *with* $(1 \leq i \leq N)$ $\pi_j(C_{ij}) = \underline{P}_j(C_{ij})$, *and* $(\prod_{j=1}^n \pi_j)(A_0) = 0$.

$\underline{P}^n$ and $\underline{Q}^n$ therefore agree on rectangles, on complements of rectangles, and on sets such as $[r_n(A) > \rho]$ and $[r_n(A) > r_n(B)]$. In general, $\underline{P}^n$ is *more precise* than $\underline{Q}^n$, in that it gives rise to narrower intervals $[\underline{P}^n(A), \bar{P}^n(A)]$.

If we define $\underline{Q}^\infty$ on $\mathscr{A}^\infty$ through $\underline{Q}^\infty(A) = \underline{Q}^n(A)$ when $A \in \mathscr{A}^n$, we have $(\forall A \in \mathscr{A}^\infty)$ $\underline{Q}^\infty(A) \leq \underline{P}^\infty(A) \leq \underline{R}^\infty(A)$. Many of the results of Sections 4 and 5, notably Theorems 4.1, 5.1, and 5.2, hold for $\underline{Q}^\infty$ as well as $\underline{P}^\infty$.

Clearly, (A_n) a.c. under $\underline{Q}^\infty$ implies (A_n) a.c. under $\underline{P}^\infty$. As an example of the failure of the converse, suppose $A \in \mathscr{A}$, identical marginals with $\underline{P}(A) < \rho < \bar{P}(A)$, and $A_n = [\,|r_n(A) - \rho| \geq 1/n]$. It is clear that $\underline{P}^\infty(A_n) \to 1$, and it can be shown that $\underline{Q}^\infty(A_n) \to 0$, so that (A_n) a.c. under $\underline{P}^\infty$ but not under $\underline{Q}^\infty$. [The proof relies on showing that, if $B_n = [r_n(A) - \rho \geq 1/n]$ and $C_n = [\rho - r_n(A) \geq 1/n]$, $D_n = X_{j=1}^n D_j \subset A_n$ implies $D_n \subset B_n$ or $D_n \subset C_n$. Hence

$$\underline{Q}^\infty(A_n) = \underline{Q}^n(B_n) + \underline{Q}^n(C_n) = \underline{P}^n(B_n) + \underline{P}^n(C_n) \to 0.]$$

If one uses Dempster's rule of conditioning for $\underline{Q}^n$ and the limiting frequentist rule of conditioning for $\underline{P}^n$, as in Section 4.3 one has

$$[\underline{Q}^n(A|B), \bar{Q}^n(A|B)] \subset [\underline{P}^n(A|B), \bar{P}^n(A|B)]$$

when A and B are rectangles, in contrast to the above result that

$$[\underline{Q}^n(A), \bar{Q}^n(A)] \supset [\underline{P}^n(A), \bar{P}^n(A)]$$

in general. Belief function products are less precise than independent products formed from the same marginals, but belief functions conditioned by Dempster's rule are more precise than the same functions conditioned by the limiting frequentist rule.

Note that $\underline{Q}^2$ satisfies the "independence" condition (a′) of Section 4.3, and also satisfies (b′) under Dempster's rule of conditioning (though not under the limiting frequentist rule, which gives $\underline{Q}^2(A \times \Omega \,|\, \Omega \times B) \leq \underline{P}_1(A)$ with strict inequality possible, as below). But the Dempster conditional probabilities $\underline{Q}(A \,|\, B)$ cannot be interpreted as lim inf $r_n(A \,|\, B)$ for any infinite sequence of outcomes, and the belief function product is incompatible with the limiting frequentist interpretation of Section 4.3, as the following example shows.

Let $\mathscr{A} = \{\phi, A, A^c, \Omega\}$, identical marginals $\underline{P}(A) = \underline{P}(A^c) = 1/4$. Then simple computations give $Q^2(A \times A \cup A^c \times A^c) = 1/8$. Suppose $(\forall D \in \mathscr{A}^2)Q^2(D) = \liminf r_n(D) = \min\{\pi(D): \pi \in \mathscr{M}^2\}$, where $\mathscr{M}^2$ is the set of limit points of (r_n). Let $\pi \in \mathscr{M}^2$, $\pi(A \times A \cup A^c \times A^c) = 1/8$. Suppose $\pi(A \times A) \le 1/16$ (otherwise, replace A by A^c) and $\pi(A^c \times A) \ge 7/16$ (otherwise, consider $\pi(\Omega \times A \mid A \times \Omega)$). Then $\pi(A \times \Omega \mid \Omega \times A) \le 1/(1+7) = 1/8 < 1/4 = Q^2(A \times \Omega)$. Thus, for some $B, C \in \mathscr{A}$ with $\underline{P}(B) > 0$,

$$\liminf r_n(C \times \Omega \mid \Omega \times B) = \min\{\pi(C \times \Omega \mid \Omega \times B) : \pi \in \mathscr{M}^2\}$$
$$< \liminf r_n(C \times \Omega) = Q^2(C \times \Omega),$$

which violates the "independence" condition (b) of Section 4.3. This shows that the belief function product Q^2 cannot arise as $\liminf r_n$ for any infinite sequence of points in Ω^2 satisfying $(\forall B, C \in \mathscr{A})$

$$\liminf r_n(\Omega \times B) > 0 \Rightarrow \liminf r_n(C \times \Omega \mid \Omega \times B) = \liminf r_n(C \times \Omega).$$

REFERENCES

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38** 325–339.

DEMPSTER, A. P. (1968). A generalization of Bayesian inference. *J. Roy. Statist. Soc., Ser. B* **30** 205–247.

FINE, T. L. (1970). On the apparent convergence of relative frequency and its implications. *IEEE Trans. Inform. Theory* **IT-16** 251–257.

FINE, T. L. (1973). *Theories of Probability.* Academic, New York.

FINE, T. L. (1974). Towards a revised probabilistic basis for quantum mechanics. *Synthese* **29** 187–201.

FINE, T. L. (1977). An argument for comparative probability. In R. E. Butts and J. Hintikka (eds.), *Basic Problems in Methodology and Linguistics*, 105–119. Reidel, Dordrecht-Holland.

FINE, T. L. (1982). The foundations of probability. In N. L. Johnson and S. Kotz (eds.), *The Encyclopedia of Statistical Sciences*, Vol. E–G. Wiley, New York.

GIERE, R. N. (1973). Objective single-case probabilities and the foundations of statistics. In P. Suppes et al. (eds.), *Logic, Methodology and the Philosophy of Science IV*, 467–483, North-Holland, Amsterdam.

GOOD, I. J. (1962). Subjective probability as the measure of a non-measurable set. In E. Nagel, Suppes and Tarski (eds.), *Logic, Methodology and Philosophy of Science*, 319–329, Stanford University Press.

HUBER, P. J. (1973). The use of Choquet capacities in statistics. *ISI Bulletin 45*, Book 4, 181–188.

HUBER, P. J., and STRASSEN, V. (1973). Minimax tests and the Neyman-Pearson Lemma for capacities. *Ann. Statist.* **1** 251–263.

LEVI, I. (1980). *The Enterprise of Knowledge.* MIT Press, Cambridge, Mass.

POPPER, K. R. (1959). *The Logic of Scientific Discovery.* Hutchinson, London.

SAMUELS, S. M. (1965). On the number of successes in independent trials. *Ann. Math. Statist.* **36** 1272–1278.

SHAFER, G. (1976). *A Mathematical Theory of Evidence.* Princeton University Press, Princeton.

SMITH, C. A. B. (1961). Consistency in statistical inference and decision. *J. Roy. Statist. Soc., Ser. B* **23** 1–25.

SUPPES, P. (1974). The measurement of belief. *J. Roy. Statist. Soc., Ser. B* **36** 160–191.

WALLEY, P. (1981). Coherent lower (and upper) probabilities. *Statistics Research Report,* University of Warwick, Coventry.

WALLEY, P., and FINE, T. L. (1979). Varieties of modal (classificatory) and comparative probability. *Synthese* **41** 321–374.

WILLIAMS, P. M. (1976). Indeterminate probabilities. In M. Przelecki, K. Szaniawski and R. Wojcicki (eds.), *Formal Methods in the Methodology of Empirical Sciences.* Reidel, Dordrecht-Holland, 229–246.

WOLFENSON, M. (1979). *Inference and decision making based on interval-valued probability.* Ph.D. thesis, Cornell University, Ithaca.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL, ENGLAND

SCHOOL OF ELECTRICAL ENGINEERING
CORNELL UNIVERSITY
PHILLIPS HALL
ITHACA, NEW YORK 14853