Z. Wahrscheinlichkeitstheorie verw. Gebiete
69, 1–17 (1985)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Stationary Lower Probabilities and Unstable Averages

Anurag Kumar, and Terrence L. Fine

Bell Telephone Laboratories, Holmdel. N.J. 07733, USA;
School of Electrical Engineering, Cornell University, Ithaca, N.Y. 14853, USA

Summary. This paper explores the possibilities for probability-like models of stationary nondeterministic phenomena that possess divergent but bounded time averages. A random sequence described by a stationary probability measure must have almost surely convergent time averages whenever it has almost surely bounded time averages. Hence, no measure can provide the mathematical model we desire. In turning to lower probability-based models we first explore the relationships between divergence, stationarity, and monotone continuity and those between monotone continuity and unicity of extensions. We then construct several examples of stationary lower probabilities for sequences of uniformly bounded random variables such that divergence of time averages occurs with lower probability one. We conclude with some remarks on the problem of estimating lower probability models on the basis of cylinder set observations.

1. Introduction

If $\{X_i\}$ is a strictly stationary random process then its time averages $\left\{\overline{X}_n = \frac{1}{n}\sum_1^n X_i\right\}$ converge almost surely whenever the process has finite EX or is such that the lim inf and lim sup of its time averages are almost surely finite (Kalikow, 1984). If the random process is wide sense stationary then its time averages converge in mean square so long as EX^2 is finite. Hence, when modeling a physical or social process that we believe to be empirically stationary (i.e., we can discern no lawlike patterns of evolution and have no a priori theoretical arguments to suggest generating mechanisms that are time-varying), we must either assume stability of long-term time averages or else postulate infinite moments or unbounded time averages. Yet data may suggest that the time averages are both unstable and bounded. For example, flicker noise is a nondeterministic phenomenon that has resisted fully acceptable probabilistic modeling in that the data on such processes suggest both unstable time averages (e.g., positive Allan variance (Kroupa, 1983)) and physical considerations suggest that moments are finite and that the time averages do not grow without limit. Furthermore, from a philosophical point of view it would be surprising if a commitment to stationarity automatically implied a commitment to statistical stability, at least in the presence of physically reasonable assumptions of finiteness of moments or time averages. Hence, we are motivated to explore a lower probability-based methodology in the hope that this generalization of probability might enable us to reconcile the presently conflicting claims of physical/empirical stationarity, finite moments, and data suggesting unstable long term averages.

In Sect. 2 we present the axioms for upper and lower probability, provide examples of lower probabilities, and identify several familes of interest. Since the models we seek have to be defined on infinite product spaces, we explore those mathematical issues that arise when constructing probability models on infinite spaces. Section 3 discusses monotone continuity and its implications for extensions of lower probabilities from algebras of events to σ algebras. In Sect. 4 we consider the property of stationarity (shift invariance) and its implications, particularly when combined with monotone continuity, for support of sequences having divergent time averages. In Sect. 5 we construct examples of lower probabilities that are stationary and give a lower probability of one to the divergence event. Section 6 addresses our ability to estimate lower probability models on the basis of cyclinder set observations. We conclude in Sect. 7 and observe that we must rely on the ill-understood class of undominated lower probabilities if we hope to achieve a stationary lower probability that is monotonely continuous over the class of cylinder sets and supports divergence. This paper is based on Kumar (1982). Research into the uses and properties of undominated lower probabilities has been reported in Grize (1984) and Papamarcou (1983).

Notation. Most of the special notation is introduced as it arises in the discussion. Some repeatedly used, standard and non-standard, notation is described here. In the following A denotes an arbitrary set and $\mathscr{A}$ a class of sets.

(i) $\emptyset$ always denotes the empty set.

(ii) 2^A denotes the power set of A.

(iii) $\tilde{}$ is used in the infix notation for set difference.

(iv) I_A is the indicator function of A.

(v) If P and Q are real valued set functions on $\mathscr{A}$, then $P =_{\mathscr{A}} Q$ denotes that P and Q agree on $\mathscr{A}$, and similarly we write $P \geqq_{\mathscr{A}} Q$, $P >_{\mathscr{A}} Q$, etc.

(iv) $(\mathbb{R}, \mathscr{B})$ denotes the measurable space consisting of the set of real numbers ($\mathbb{R}$) and the σ-algebra of Borel subsets ($\mathscr{B}$) of $\mathbb{R}$.

2. The Basic Upper and Lower Probability Structure

An upper/lower probability structure consists of a pair of set functions, on an algebra of events, which possess properties that are simple generalizations of the properties of a probability measure. This structure has previously been studied in the literature in the context of coherent preference between gambles (Smith, 1961; Williams, 1976; Walley, 1981), upper and lower bounds to degrees of belief on an algebra, given exact degrees of belief on a sub-algebra (Good, 1962), inexact measurement of belief (Suppes, 1974), a theory of evidence based on belief functions (Shafer, 1976) and unstable relative frequencies in i.i.d. trials (Walley and Fine, 1982).

2.1. Definition. Let $(\Omega, \mathscr{A})$ be a measurable space. A lower/upper probability on $(\Omega, \mathscr{A})$ is a pair of functions $(\underline{P}, \bar{P})$ on $\mathscr{A}$ satisfying.

$$\underline{P}: \mathscr{A} \to [0,1]$$
$$\bar{P}: \mathscr{A} \to [0,1]$$

(i) $\underline{P}(\Omega)=1$,
(ii) $(\forall A \in \mathscr{A})\, \underline{P}(A) \geqq 0$,
(iii) $(\forall A, B \in \mathscr{A})\, A \cap B = \emptyset$ only if $\underline{P}(A)+\underline{P}(B) \leqq \underline{P}(A \cup B)$ (super-additivity) and $\bar{P}(A)+\bar{P}(B) \geqq \bar{P}(A \cup B)$ (sub-additivity),
(iv) $(\forall A \in \mathscr{A})\, \underline{P}(A)+\bar{P}(A^c)=1$. □

Remarks. All the axioms can be expressed in terms of $\underline{P}$ and $\bar{P}$ defined through (iv) in Definition 2.1. We shall find it convenient to work only with $\underline{P}$ and shall henceforth denote by $(\Omega, \mathscr{A}, \underline{P})$ a lower probability space, the upper probability being implicit in this notation.

2.2. Lemma. *Let $(\Omega, \mathscr{A}, \underline{P})$ be a lower probability space*
(i) $(\forall A \in \mathscr{A})\, \underline{P}(A) \leqq \bar{P}(A)$. *(Hence the occasionally used term "Interval valued probability".)*
(ii) $(\forall A, B \in \mathscr{A})\, A \subset B$ *only if* $(\underline{P}(A) \leqq \underline{P}(B)$ *and* $\bar{P}(A) \leqq \bar{P}(B))$. *Thus* $\underline{P}$ *and* $\bar{P}$ *are monotone set functions.*

Proof. Simple computations from Definition 2.1. □

2.3. Examples. Let $(\Omega, \mathscr{A})$ be a measurable space.
(i) Every probability measure μ on $(\Omega, \mathscr{A})$ is degenerately a lower probability $(\forall A \in \mathscr{A}, \underline{P}(A)=\mu(A)=\bar{P}(A))$.
(ii) If $\underline{P}$ and $\underline{Q}$ are lower probabilities then $(\forall \lambda, 0<\lambda<1)$ so is $\lambda \underline{P}+(1-\lambda)\underline{Q}$, i.e., the class of lower probabilities on $(\Omega, \mathscr{A})$ is convex.
(iii) If $\underline{P}_1, \ldots, \underline{P}_n$ are lower probabilities then so is $\underline{R}$ defined by $(\forall A \in \mathscr{A})\, \underline{R}(A) = \inf_{1 \leqq i \leqq n} \underline{P}_i(A)$. In particular combine (i) and (ii) to see that the infimum of any class of probability measures is a lower probability.

2.4. Definition. Let $(\Omega, \mathscr{A}, \underline{P})$ be a lower probability space.
(i) A probability measure μ on $(\Omega, \mathscr{A})$ *dominates* $\underline{P}$ on $\mathscr{A}$ (denoted $\mu \geqq_{\mathscr{A}} \underline{P}$) if $(\forall A \in \mathscr{A})\, \mu(A) \geqq \underline{P}(A)$.
(ii) The class of dominating measures of a lower probability $\underline{P}$ is denoted by $\mathscr{M}_{\underline{P}}$ and is defined as

$$\mathscr{M}_{\underline{P}} = \{\mu: \mu \text{ is a probability measure on } (\Omega, \mathscr{A})\ \mu \geqq_{\mathscr{A}} \underline{P}\}. \quad \square$$

Remarks. (i) $\mathscr{M}_{\underline{P}}$ may, in general, be empty.
(ii) There exists $\underline{P}$ for which $\mathscr{M}_{\underline{P}}$ is nonempty (trivially, every probability measure μ has nonempty $\mathscr{M}_{\mu}$).

2.5. Definition. Let $(\Omega, \mathscr{A}, \underline{P})$ be a lower probability space
(i) $\underline{P}$ is said to be *dominated* if $\mathscr{M}_{\underline{P}} \neq \emptyset$.
(ii) A dominated $\underline{P}$ is said to be a *lower envelope* if

$$(\forall A \in \mathscr{A})\, \underline{P}(A) = \inf\{\mu(A): \mu \in \mathscr{M}_{\underline{P}}\}. \quad \square$$

Note. Not every dominated lower probability is a lower envelope (for an example see Huber, 1976, p. 84).

3. Lower Probabilities on Infinite Spaces

In Sect. 4 and 5 we shall be studying models for unstable relative frequency, models that will necessarily have to be defined on infinite (product) spaces. In this section we present some of the mathematical details associated with upper and lower probability structures on general infinite spaces. Our discussion will focus on lower envelopes and dominated lower probabilities, as these are the models we use in the sequel. The issues that concern us are the continuity of lower probability and the unicity of extensions of lower probability.

Let Ω be an arbitrary, infinite space and let $\mathscr{A}$ be an infinite algebra of subsets of Ω. $\mathscr{F}(\mathscr{A})$ denotes the smallest σ-algebra containing $\mathscr{A}$. μ will always denote a countably additive probability measure on $\mathscr{F}(\mathscr{A})$.

To begin with, we examine continuity of lower probability.

3.1. Lemma. *Let $\underline{P}$ be the lower envelope of a class of countably additive measures on $(\Omega, \mathscr{F}(\mathscr{A}))$*

(i) *(continuity from above of $\underline{P}$). If $\{A_k, k \geqq 1\}$ is a decreasing sequence of sets in $\mathscr{F}(\mathscr{A})$,* s.t. *$A_k \downarrow A$, then $\underline{P}(A) = \lim_{k\to\infty} \underline{P}(A_k)$,*

(ii) *(countable super-additivity of $\underline{P}$). If $\{A_i, i \geqq 1\}$ is a pairwise disjoint sequence of sets in $\mathscr{F}(\mathscr{A})$,* s.t. *$\bigcup_{i=1}^{\infty} A_i = A$, then $\underline{P}(A) \geqq \sum_{i=1}^{\infty} \underline{P}(A_i)$.*

Proof. (i) $(\forall k \geqq 1)(\forall \mu \geqq \underline{P})\ \underline{P}(A_k) \leqq \mu(A_k)$

$$\begin{aligned} &\therefore (\forall \mu \geqq \underline{P}) \limsup_{k\to\infty} \underline{P}(A_k) \leqq \lim_{k\to\infty} \mu(A_k) = \mu(A). \\ &\therefore (\forall \mu \geqq \underline{P}) \limsup_{k\to\infty} \underline{P}(A_k) \leqq \inf_{\mu \geqq \underline{P}} \mu(A) = \underline{P}(A). \end{aligned} \tag{a}$$

Also $(\forall k \geqq 1)\ \underline{P}(A) \leqq \underline{P}(A_k)$ (by monotonicity of $\underline{P}$).

$$\therefore \underline{P}(A) \leqq \liminf_{k\to\infty} \underline{P}(A_k). \tag{b}$$

Combining (a) and (b) $\{\underline{P}(A_k)\}$ converges and

$$\lim_{k\to\infty} \underline{P}(A_k) = \underline{P}(A).$$

(ii) $\underline{P}(A) = \inf_{\mu \geqq \underline{P}} \mu(A) = \inf_{\mu \geqq \underline{P}} \sum_{i=1}^{\infty} \mu(A_i) \geqq \sum_{i=1}^{\infty} \inf_{\mu \geqq \underline{P}} \mu(A_i) = \sum_{i=1}^{\infty} \underline{P}(A_i).$ $\square$

The following example demonstrates that, in general, lower envelopes are not continuous from below.

3.2. Example. Define a lower envelope on $(\mathbb{R}, \mathscr{B})$ as follows.

Let $(\forall n \geqq 1)\, \mu_n$ denote the probability measure satisfying $\mu_n(\{n\})=1$.

Let $(\forall A \in \mathscr{B})\ \underline{P}(A) \triangleq \inf_{n \geqq 1} \mu_n(A)$. Clearly $\underline{P}$ is a lower envelope of countably additive measures.

Let $(\forall i \geqq 1)\, A_i = (-\infty, i]$.

Clearly $A_i \uparrow \mathbb{R}$.

But $(\forall i \geqq 1)\, \underline{P}(A_i)=0$ and $\underline{P}(\mathbb{R})=1$. Hence $\underline{P}$ is not continuous from below at $\mathbb{R}$. □

We now turn to the issue of the extension of lower probability.

If $\underline{P}$ is a lower probability on $(\Omega, \mathscr{F}(\mathscr{A}))$ and μ is a countably additive probability measure on $(\Omega, \mathscr{A})$ such that $\mu \geqq_{\mathscr{A}} \underline{P}$, then, in general, the unique countably additive extension of μ to $\mathscr{F}(\mathscr{A})$ need not dominate $\underline{P}$ on $\mathscr{F}(\mathscr{A})$. The following result yields a condition on $\underline{P}$ under which the extension of μ does dominate $\underline{P}$. Note that the result applies to more general set functions than lower probabilities.

3.3. Proposition. *Let $\underline{P}$ be a monotone set function on $(\Omega, \mathscr{F}(\mathscr{A}))$. If $\underline{P}$ is continuous from below on $\mathscr{A}$ (i.e. $(\forall \{A_i, A_i \in \mathscr{A}, i \geqq 1\})\, A_i \uparrow A (\in \mathscr{F}(\mathscr{A})) \Rightarrow \underline{P}(A_i) \uparrow \underline{P}(A)$), then $\mu \geqq_{\mathscr{A}} \underline{P} \Leftrightarrow \mu \geqq_{\mathscr{F}(\mathscr{A})} \underline{P}$.*

Proof. It suffices to show $\mu \geqq_{\mathscr{A}} \underline{P} \Rightarrow \mu \geqq_{\mathscr{F}(\mathscr{A})} \underline{P}$.

Let $\mathscr{A}_\sigma$ (resp. $\mathscr{A}_\delta$) denote the class of sets obtained by taking unions (resp. intersection) of countable families of sets in $\mathscr{A}$.

Let μ be a countably additive probability measure on $(\Omega, \mathscr{F}(\mathscr{A}))$ s.t. $\mu \geqq_{\mathscr{A}} \underline{P}$. We first show $\mu \geqq_{\mathscr{A}_\sigma} \underline{P}$. If $B \in \mathscr{A}_\sigma$, $\exists \{B_i \in \mathscr{A}, i \geqq 1\}\, B_i \uparrow B$, then since $\underline{P}$ is continuous from below on $\mathscr{A}$ we get

$$\mu(B) = \lim_{i \to \infty} \mu(B_i) \geqq \lim_{i \to \infty} \underline{P}(B_i) = \underline{P}(B).$$

Let $A \in \mathscr{F}(\mathscr{A})$. Then (see e.g. Royden, 1968, p. 256) $\exists(\{A_i \in \mathscr{A}_\sigma, i \geqq 1\}, (\forall i \geqq 1)\, A_i \supset A_{i+1} \supset A)$ s.t. $\mu(A_i) \downarrow \mu(A)$. Hence, using the monotonicity of $\underline{P}$ and the fact that $\mu \geqq_{\mathscr{A}_\sigma} \underline{P}$,

$$\mu(A) = \lim_{i \to \infty} \mu(A_i) \geqq \lim_{i \to \infty} \underline{P}(A_i) \geqq \underline{P}(A).$$

Hence $\mu \geqq_{\mathscr{F}(\mathscr{A})} \underline{P}$. □

In Sect. 6 we shall encounter examples which show that, unlike countably additive probability measures, lower envelopes of countably additive probability measures may not have unique extensions. Lemma 3.4 demonstrates a simple, natural extension to $(\Omega, \mathscr{F}(\mathscr{A}))$ of a lower envelope of countably additive measures on $(\Omega, \mathscr{A})$.

Let $\underline{P}_0$ be a lower envelope of countably additive measures on $(\Omega, \mathscr{A})$.

Let

$$\mathcal{M} = \{\mu: \mu \text{ is a probability measure on } (\Omega, \mathcal{F}(\mathcal{A})) \text{ and } \mu \geqq_{\mathcal{A}} \underline{P}_0\}$$

and let $\underline{P}$ be the lower envelope of $\mathcal{M}$ on $(\Omega, \mathcal{F}(\mathcal{A}))$.

3.4. Proposition. (i) $\underline{P} =_{\mathcal{A}} \underline{P}_0$ *i.e.* $\underline{P}$ *is an extension of* $\underline{P}_0$ *to* $\mathcal{F}(\mathcal{A})$.

(ii) *If* $\underline{Q}$ *is a lower envelope on* $(\Omega, \mathcal{F}(\mathcal{A}))$ s.t.$\underline{Q} =_{\mathcal{A}} \underline{P}_0$ *then* $\underline{Q} \geqq_{\mathcal{F}(\mathcal{A})} \underline{P}$, *i.e.* $\underline{P}$ *is the minimal extension of* $\underline{P}_0$ *to a lower envelope on* $(\Omega, \mathcal{F}(\mathcal{A}))$.

Proof. Follows easily from the fact that every countably additive probability measure on $(\Omega, \mathcal{A})$ has a unique extension to a countably additive probability measure on $(\Omega, \mathcal{F}(\mathcal{A}))$. □

We call $\underline{P}$ the *canonical extension* of $\underline{P}_0$ to $(\Omega, \mathcal{F}(\mathcal{A}))$.

4. Stationary Lower Probabilities

In this section we develop some results about stationary set functions (in particular stationary lower probabilities) on the countably infinite product of a finite measurable space. Henceforth, let Ω be an arbitrary *finite* space and $\mathcal{A}$ an algebra of subsets of Ω. Let Ω_∞ be the countably infinite product of Ω, $\mathcal{C}_\infty$ the algebra of cylinder sets (or finite dimensional sets) and $\mathcal{A}^\infty$ the σ-algebra generated by $\mathcal{C}_\infty$. To discuss stationarity we need to introduce the following shift operator.

$(\forall k \geqq 1)\, T^k: \Omega^\infty \to \Omega^\infty$ denotes the *left-shift-by-k* operator, i.e. if $\underset{\sim}{\omega} = \{\omega_1, \omega_2, \ldots\}$ then $T^k \underset{\sim}{\omega} = \{\omega_{k+1}, \omega_{k+2}, \ldots\}$. We also denote, for all $k \geqq 1$, by T^k the set function on 2^{Ω^∞} defined by $(\forall A \subset \Omega^\infty)\, T^k A = \{T^k \underset{\sim}{\omega}: \underset{\sim}{\omega} \in A\}$. For all $k \geqq 1$, T^{-k} denotes the set function on 2^{Ω^∞} defined by $(\forall A \subset \Omega^\infty)\, T^{-k} A = \{\underset{\sim}{\omega}: T^k \underset{\sim}{\omega} \in A\}$. Note that T^{-k} cannot be thought of as a point function in the same way as T^k can. Also observe that

$$(\forall A \subset \Omega^\infty)(\forall k \geqq 1)\, T^k(T^{-k} A) = A$$

but $T^{-k}(T^k A) \supseteq A$; for instance

$$T^{-1}(T\{\underset{\sim}{\omega}: \omega_1 = 0\}) = T^{-1}(\Omega^\infty) = \Omega^\infty \supset \{\underset{\sim}{\omega}: \omega_1 = 0\}.$$

As usual we define the *relative frequency function* by

$$(\forall A \in \mathcal{A})(\forall n \geqq 1)\, r_n(A): \Omega^\infty \to [0, 1],$$

where $r_n(A)(\underset{\sim}{\omega}) = \frac{1}{n} \sum_{i=1}^{n} I_A(\omega_i)$.

Let D be the set in $\mathcal{A}^\infty$ defined by

$$D = \{\underset{\sim}{\omega} \in \Omega^\infty: (\exists A \in \mathcal{A}) \limsup_{n \to \infty} r_n(A)(\underset{\sim}{\omega}) > \liminf_{n \to \infty} r_n(A)(\underset{\sim}{\omega})\}.$$

D will be called the *divergence* (of relative frequency) event, and $C \triangleq D^C$ is the *convergence* event.

4.1. Lemma. (i) $\forall(\{C_i, C_i \in \mathscr{C}_\infty, i \geqq 1\}).\ \bigcap_{i\geqq 1} C_i = \emptyset \Rightarrow \exists N \geqq 1$ s.t. $\bigcap_{i \leqq N} C_i = \emptyset$.

(ii) Every finitely additive probability measure on $\mathscr{C}_\infty$ is countably additive.

Proof. (i) $\mathscr{C}_\infty$ is the algebra generated by the semialgebra of measurable rectangles which is easily seen to possess the property asserted for $\mathscr{C}_\infty$. Since $\mathscr{C}_\infty$ is just the collection of finite disjoint unions of rectangles, Lemma I.6.1. of Neveu, 1965, p. 26, now applies.

This result also follows from a topological compactness argument. Since Ω is finite, Ω^∞ can be shown to be isomorphic to a compact subset of $[0, 1]$ in $\mathbb{R}$ (e.g., for $\|\Omega\| = 2$, Ω^∞ is isomorphic to the ternary Cantor set). Measurable rectangles then correspond to closed subsets of this compact set. Since cylinder sets are finite unions of rectangles, they too are closed. The result then follows from the definition of compactness.

(ii) Follows from (i). □

We say that a set function $\underline{P}$ on $(\Omega^\infty, \mathscr{A}^\infty)$ is (right shift) *stationary* when

$$(\forall A \in \mathscr{A}^\infty)(\forall k \geqq 1)\, \underline{P}(T^{-k}A) = \underline{P}(A).$$

We seek stationary models which support divergence in the strong sense that $\underline{P}(D) = 1$.

4.2. Proposition. *Let $\underline{P}$ be a stationary, dominated set function on $(\Omega^\infty, \mathscr{C}_\infty)$. Then there exists a right-shift stationary probability measure ψ on $(\Omega^\infty, \mathscr{C}_\infty)$* s.t.

$$\psi \geqq_{\mathscr{C}_\infty} \underline{P}$$

i.e., there is a stationary dominating measure on $\mathscr{C}_\infty$.

Proof. Let μ be an hypothesized probability measure on $(\Omega^\infty, \mathscr{C}_\infty)$ s.t. $\mu \geqq_{\mathscr{C}_\infty} \underline{P}$. So

$$(\forall A \in \mathscr{C}_\infty)(\forall k \geqq 1)\, \mu(T^{-k+1}A) \geqq \underline{P}(T^{-k+1}A) = \underline{P}(A).$$

Hence

$$(\forall A \in \mathscr{C}_\infty)(\forall n \geqq 1) \frac{1}{n} \sum_{k=1}^{n} \mu(T^{-k+1}A) \geqq \underline{P}(A).$$

Since $(\forall A \in \mathscr{C}_\infty) \left\{ \frac{1}{n} \sum_{k=1}^{n} \mu(T^{-k+1}A) \right\}$ is a uniformly bounded sequence it has convergent subsequences. By the countability of $\mathscr{C}_\infty$ and a diagonalization argument, there exists a subsequence indexed by $\{n_i\}$ s.t.

$$(\forall A \in \mathscr{C}_\infty) \frac{1}{n_i} \sum_{k=1}^{n_i} \mu(T^{-k+1}A) \text{ converges to } \psi(A) \text{ (say).}$$

Observe now that:

(i) $\psi \geqq_{\mathscr{C}_\infty} \underline{P}$,

(ii) $(\forall A \in \mathscr{C}_\infty)(\forall k \geqq 1)\, \psi(A) = \psi(T^{-k}A)$,

(iii) $(\forall A \in \mathscr{C}_\infty)\, \psi(A) \geqq 0$ and $\psi(\Omega^\infty) = 1$,

(iv) Since $A \cap B = \emptyset \Rightarrow (\forall k \geqq 1)(T^{-k+1}A) \cap (T^{-k+1}B) = \emptyset$

$$\psi(A \cup B) = \lim_{i\to\infty} \frac{1}{n_i} \sum_{k=1}^{n_i} \mu(T^{-k+1}(A \cup B)) = \lim_{i\to\infty} \frac{1}{n_i} \sum_{k=1}^{n_i} (\mu(T^{-k+1}A) + \mu(T^{-k+1}B))$$
$$= \psi(A) + \psi(B).$$

Thus ψ is a right shift stationary, finitely additive probability measure on$(\Omega^\infty, \mathscr{C}_\infty)$ which dominates $\underline{P}$ on $\mathscr{C}_\infty$. Lemma 4.1 ensures that ψ is also countably additive on $\mathscr{C}_\infty$ and hence is a probability measure in the usual sense. □

4.3. Theorem. *If $\underline{P}$ is a monotone set function on* $(\Omega^\infty, \mathscr{A}^\infty)$ s.t.

(i) $\exists$ *a probability measure* μ, $\mu \geqq_{\mathscr{C}_\infty} \underline{P}$,

(ii) $\underline{P}$ *is stationary on* $\mathscr{C}_\infty$,

(iii) $\underline{P}$ *is continuous from below on* $\mathscr{C}_\infty$ *(i.e.* $\forall \{C_i, C_i \in \mathscr{C}_\infty, i \geqq 1\}$ $C_i \uparrow C \in (\mathscr{C}_\infty)_\sigma \Rightarrow \underline{P}(C_i) \uparrow \underline{P}(C)$), *then there exists a stationary probability measure* ν *on* $(\Omega^\infty, \mathscr{A}^\infty)$ s.t. $\nu \geqq_{\mathscr{A}^\infty} \underline{P}$.

Proof. Proposition 4.2 $\Rightarrow \exists$ stationary probability measure ψ on $(\Omega^\infty, \mathscr{C}_\infty)$ s.t. $\psi \geqq_{\mathscr{C}_\infty} \underline{P}$. Let ν be the unique extension of ψ to a (countably additive) probability measure on $(\Omega^\infty, \mathscr{A}^\infty)$. Then ν is stationary and Proposition 3.3 $\Rightarrow \nu \geqq_{\mathscr{A}^\infty} \underline{P}$. □

4.4. Corollary. *If $\underline{P}$ is a monotone set function on* $(\Omega^\infty, \mathscr{A}^\infty)$ *which is dominated, continuous from below on* $\mathscr{C}_\infty$ *and stationary then* $\underline{P}(D) = 0$ *and* $\underline{P}$ *cannot support diverging relative frequency.*

Proof. Theorem 4.3 $\Rightarrow \exists$ a stationary probability measure ν s.t. $\nu \geqq_{\mathscr{A}^\infty} \underline{P}$. But clearly by the finiteness of Ω and the strong ergodic theorem $\nu(D) = 0$. Hence $\underline{P}(D) = 0$. □

Thus stationary set functions, on $(\Omega^\infty, \mathscr{A}^\infty)$, which satisfy regularity conditions much weaker than countable additivity cannot support divergence. Hence stationary set functions that do support divergence are going to be strongly constrained.

5. Stationary Lower Probabilities That Support Diverging Relative Frequency

In this section we give several examples of lower probabilities which support diverging relative frequency and are (right-shift) stationary on $\mathscr{A}^\infty$. In constructing stationary lower probability models on $(\Omega^\infty, \mathscr{A}^\infty)$ that support divergence, we have not proceeded by specifying marginals, constructing a lower probability on $(\Omega^\infty, \mathscr{C}_\infty)$ and then extending it to $(\Omega^\infty, \mathscr{A}^\infty)$. One of the problems in adopting this procedure is that the lower probability on $(\Omega^\infty, \mathscr{C}_\infty)$ will in general have several extensions to $(\Omega^\infty, \mathscr{A}^\infty)$; the natural extensions (e.g. the canonical extension of a lower envelope) do not support divergence, and it is not clear how to extend so that the extension does support divergence. Instead we start with some kind of infinitary specification on $(\Omega^\infty, \mathscr{A}^\infty)$; e.g. a class of measures on $(\Omega^\infty, \mathscr{A}^\infty)$ or an infinite sequence in Ω^∞ on which relative frequencies diverge. From this specification a lower probability is then determined (Examples 5.2, 5.3, 5.4). The specification is chosen so as to ensure stationarity and support of divergence in the resulting lower probability. We have obtained several classes of models using this procedure and have, specifically, shown the existence of stationary finitely additive models that support divergence.

Let $\Omega=\{0,1\}$ and $\mathscr{A}=2^{\{0,1\}}$. Denote by $\prod_{i=1}^{\infty}\pi_i$ the product probability measure on $(\Omega^\infty, \mathscr{A}^\infty)$ for which the probability of 1 on the i^{th} "trial" is π_i.

5.1. Example. Let $0\leqq \underline{p}\leqq \overline{p}\leqq 1$ and let $\mathscr{M}=\Big\{\mu\colon \mu=\prod_{i=1}^{\infty}\pi_i, (\forall i\geqq 1, \underline{p}\leqq\pi_i\leqq\overline{p})$ and

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_i=\underline{p},$$

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_i=\overline{p}\Big\}.$$

Let $\underline{P}$ be the lower envelope on $(\Omega^\infty, \mathscr{A}^\infty)$ of the class of measures $\mathscr{M}$.

It is easily seen that $\underline{P}$ is stationary on $\mathscr{A}^\infty$. Also since $(\forall\mu\in\mathscr{M})\,\mu(D)=1$, therefore $\underline{P}(D)=1$. Corollary 4.4 will then have us conclude that $\underline{P}$ is not continuous from below on $\mathscr{C}_\infty$. □

5.2. Example. Let $\underset{\sim}{\omega}^{\cdot}\in\Omega^\infty$ be a sequence of alternating blocks of 0's and 1's such that the block length increases rapidly (e.g., i^{th} block has length 2^{2^i}).

Let $S=\{\underset{\sim}{\omega}^{\cdot}, T\underset{\sim}{\omega}^{\cdot}, T^2\underset{\sim}{\omega}^{\cdot}, \ldots, T^k\underset{\sim}{\omega}^{\cdot}, \ldots\}$. $(\forall A\subset\Omega^\infty)$ define

$$\underline{P}(A)=\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}I_A(T^{k-1}\underset{\sim}{\omega}^{\cdot})$$

where $I_A(\cdot)$ is the indicator function of A. Because of the properties of lim inf, it is easily seen that $\underline{P}$ is a lower probability on 2^{Ω^∞}. Also

$$(\forall A\subset\Omega^\infty)(\forall k\geqq 1)\,\underline{P}(A)=\underline{P}(T^{-k}A)$$

i.e., $\underline{P}$ is stationary, and since $S\subset D$ and $\underline{P}(S)=1$, therefore $\underline{P}(D)=1$. □

The following example shows that it is in fact possible to contruct a finitely additive probability measure on $(\Omega^\infty, 2^{\Omega^\infty})$ which is stationary and supports diverging relative frequency.

5.3. Example. Let $\underset{\sim}{\omega}^{\cdot}$ and S be as in Example 5.2.

Index the elements of S by the positive integer sequence $\{j\}$, the correspondence being $j\leftrightarrow T^{j-1}\underset{\sim}{\omega}^{\cdot}$.

Consider now $(\forall A\subset\Omega^\infty)$ the sequence $\frac{1}{n}\sum_{i=1}^{n}I_A(T^{i-1}\underset{\sim}{\omega}^{\cdot})$. Let $\mathscr{E}\subset 2^{\Omega^\infty}$ be the class of sets for which this sequence converges and denote by P the limit function on $\mathscr{E}$. Our objective is now to extend P on $\mathscr{E}$ to a stationary, finitely additive probability measure on 2^{Ω^∞}. This extension is achieved by using a generalization of the Hahn-Banach theorem (Royden, 1968, Chap. 10, Proposition 5). In order to bring the problem into a form suitable for the application of this result an intermediate extension step is required.

Denote by $\mathscr{L}(\mathscr{E})$ the linear space of all finite, linear combinations of indicator functions of sets in $\mathscr{E}$, i.e. $X \in \mathscr{L}(\mathscr{E})$ iff $\exists E_1, \ldots, E_m$, $E_i \in \mathscr{E}$ and real numbers $\alpha_1, \ldots, \alpha_m$ s.t. $X = \sum_{i=1}^{m} \alpha_i I_{E_1}$. Similarly let $\mathscr{L}$ denote the linear space of all finite, linear combinations of indicator functions of sets in Ω^∞. $(\forall X \in \mathscr{L})(\forall k \geq 1)$ denote by $T^{-k}X$ the right-shift by k of X, defined by $(\forall \underset{\sim}{\omega} \in \Omega^\infty)(T^{-k}X)(\underset{\sim}{\omega}) = X(T^k \underset{\sim}{\omega})$. Clearly $\{T^{-k}: k \geq 0\}$ forms an Abelian semi-group of linear operators on $\mathscr{L}$.

Lemma 5.4 a (below) allows us to extend P on $\mathscr{E}$ to a linear functional e on $\mathscr{L}(\mathscr{E})$. Through e we define a function $\underline{E}$ on $\mathscr{L}$. Lemma 5.4 b shows that e on $\mathscr{L}(\mathscr{E})$ and $\underline{E}$ on $\mathscr{L}$ (and the semi-group of right-shift operators) satisfy the conditions of the generalization of the Hahn-Banach theorem mentioned above. Application of the theorem yields a right-shift invariant linear functional E on $\mathscr{L}$ which agrees with e on $\mathscr{L}(\mathscr{E})$.

Define the real value functional e on $\mathscr{L}(\mathscr{E})$ as follows: If $X = \sum_{i=1}^{m} \alpha_i I_{Eq}$ (where $E_i \in \mathscr{E}$, $1 \leq i \leq m$) then $e(X) \triangleq \sum_{i=1}^{m} \alpha_i P(E_i)$.

5.4 a. Lemma. (i) *If* $(\forall \underset{\sim}{\omega} \in \Omega^\infty) \sum_{i=1}^{m} \alpha_i I_{Eq}(\underset{\sim}{\omega}) \geq \sum_{j=1}^{n} \beta_j I_{Fr}(\underset{\sim}{\omega})$ *then*

$$\sum_{i=1}^{m} \alpha_i P(E_i) \geq \sum_{j=1}^{n} \beta_j P(F_j).$$

(ii) *e on $\mathscr{L}(\mathscr{E})$ is well-defined i.e. if $X (\in \mathscr{L}(\mathscr{E}))$ is* s.t. $X = \sum_{i=1}^{m} \alpha_i I_{Eq}$ *and* $X = \sum_{j=1}^{n} \beta_j I_{Fr}$, *then*

$$\sum_{i=1}^{m} \alpha_i P(\mathscr{E}_i) = \sum_{j=1}^{n} \beta_j P(\mathscr{F}_j)$$

(iii) *e is monotone on $\mathscr{L}(\mathscr{E})$ i.e.* $(\forall X, Y \in \mathscr{L}(\mathscr{E}))\, X \geq Y \Rightarrow eX \geq eY$.

Proof. (i) From the definition of the set function P on $\mathscr{E}$, and indexing the elements of S by k in the following expressions, we have

$$\sum_{i=1}^{m} \alpha_i P(E_i) = \sum_{i=1}^{m} \alpha_i \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} I_{E_i}(k) = \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} \sum_{i=1}^{m} \alpha_i I_{E_i}(k)$$

which, by hypothesis,

$$\begin{aligned} &\geq \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} \sum_{j=1}^{n} \beta_j I_{F_j}(k) \\ &= \sum_{j=1}^{n} \beta_j \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} I_{F_j}(k) \\ &= \sum_{j=1}^{n} \beta_j P(F_j). \end{aligned}$$

(ii) and (iii) follow immediately from (i). □

Clearly e is a linear functional on the linear space $\mathscr{L}(\mathscr{E})$. Further we observe that, $(\forall k \geqq 1)\,\mathscr{E}$ is closed under T^{-k}, $\mathscr{L}(\mathscr{E})$ is closed under T^{-k}, $(\forall E \in \mathscr{E})\, P(T^{-k} E) = P(E)$ and, consequently, $(\forall X \in \mathscr{L}(\mathscr{E}))\, e(T^{-k} X) = e(X)$.

Define the function $\underline{E}$ on $\mathscr{L}$ as follows.

$$(\forall X \in \mathscr{L})\, \underline{E}X = \sup\{e(X') \colon X' \in \mathscr{L}(\mathscr{E}), X' \leqq X\}.$$

5.4 b. Lemma. *$\underline{E}$ defined above has the following properties.*

(i) $(\forall \lambda \geqq 0)\, \underline{E}(\lambda X) = \lambda \underline{E}X$,

(ii) $(\forall X, Y \in \mathscr{L})\, \underline{E}(X+Y) \geqq \underline{E}X + \underline{E}Y$,

(iii) $(\forall X \in \mathscr{L}(\mathscr{E}))\, \underline{E}X \leqq eX$,

(iv) $(\forall X \in \mathscr{L})(\forall k \geqq 1)\, \underline{E}(T^{-k}X) \geqq \underline{E}X$.

Proof. (i) Obvious.

(ii) $(\forall \varepsilon > 0)\, \exists X' \in \mathscr{L}(\mathscr{E})$, $X' \leqq X$, $eX' \geqq \underline{E}X - \varepsilon/2$ and $\exists Y' \in \mathscr{L}(\mathscr{E})$, $Y' \leqq Y$, $eY' \geqq \underline{E}Y - \varepsilon/2$ then $(X' + Y') \in \mathscr{L}(\mathscr{E})$ and $(X' + Y') \leqq (X + Y)$. Hence $\underline{E}(X+Y) \geqq e(X' + Y')$

(iii) Follows from the monotonicity of e on $\mathscr{L}(\mathscr{E})$ (cf. Lemma 5.4 a).

(iv) $(\forall X \in \mathscr{L})(\forall k \geqq 1)$

$$\begin{aligned} \underline{E}(T^{-k}X) &= \sup\{e(X') \colon X' \in \mathscr{L}(\mathscr{E}), X' \leqq T^{-k}X\} \\ &\geqq \sup\{e(T^{-k}X') \colon X' \in \mathscr{L}(\mathscr{E}), T^{-k}X' \leqq T^{-k}X\} \\ &= \sup\{e(X') \colon X' \in \mathscr{L}(\mathscr{E}), X' \leqq X\} \end{aligned}$$

where the last step follows because of the stationarity of e and because $T^{-k}X' \leqq T^{-k}X$ iff $X' \leqq X$.

Thus $\underline{E}(T^{-k}X) \geqq \underline{E}X$. □

We now state and use the following result from Royden, 1968. We use Royden's notation in stating this result.

5.4 c. Proposition. *X is a linear space and S is a linear subspace of X. G is an Abelian semigroup of linear operators on X* s.t. $(\forall s \in S)(\forall A \in G)\, As \in S$. *If p is a real-valued function on X* s.t. $(\forall x, y \in X)$,

(i) $(\forall \alpha \geqq 0)\, p(\alpha x) = \alpha p(x)$,

(ii) $p(x+y) \geqq p(x) + p(y)$,

(iii) $(\forall A \in G)\, p(Ax) \geqq p(x)$

and f is real-valued linear functional on S s.t. $(\forall s \in S)$

(i) $f(s) \geqq p(s)$,

(ii) $(\forall A \in G)\, f(As) = f(s)$

then $\exists$ a real-valued linear functional F on X s.t.

(i) $(\forall s \in S)\, F(s) = f(s)$,

(ii) $(\forall x \in X)\, F(x) \geqq p(x)$,

(iii) $(\forall A \in G)(\forall x \in X)\, F(Ax) = F(x)$.

Proof. See Royden 1968, pp. 188–190.

Returning to our notation in the present example.

5.4 d. Lemma. *The function $\underline{E}$ on the linear space $\mathscr{L}$, the linear functional e on the subspace $\mathscr{L}(\mathscr{E})$ of $\mathscr{L}$, and the Abelian semigroup of linear operators*

$\{T^{-k}, k \geqq 0\}$ *on* $\mathscr{L}$ *satisfy the conditions of Proposition* 5.4 c. *Hence there exists a linear functional* E *on* $\mathscr{L}$ s.t.

(i) $E \geqq_{\mathscr{L}} \underline{E}$,

(ii) $E =_{\mathscr{L}(\mathscr{E})} e$,

(iii) $(\forall X \in \mathscr{L})(\forall k \geqq 1)\, E(T^{-k}X) = EX$.

Proof. Direct application of Proposition 5.4 c. □

So now we have a linear functional E on $\mathscr{L}$ s.t. $(\forall X \in \mathscr{L})(\forall k \geqq 1)\, E(T^{-k}X) = EX$. Observe that

(i) $(\forall X \in \mathscr{L})\, X \geqq 0 \Rightarrow EX \geqq 0$, which follows because $(\forall X \in \mathscr{L})\, X \geqq 0 \Rightarrow \underline{E}X \geqq 0$ which in turn follows because $(\forall X \in \mathscr{L}(E))\, X \geqq 0 \Rightarrow e(X) \geqq 0$ which follows by the monotonicity of e,

(ii) (i) $\Rightarrow$ monotonicity of E,

(iii) $EI_{\Omega^\infty} = eI_{\Omega^\infty} = 1$.

Thus the function $m: 2^{\Omega^\infty} \to [0, 1]$ defined by $m(A) = E(I_A)$, is a finitely additive probability measure. m is stationary because of the stationarity of E and since $S \subset D$ (S as in Example 5.3) and $m(S) = 1$, therefore $m(D) = 1$. □

6. Estimability

Having demonstrated the existence of stationary models that support diverging relative frequency, we turn now to the important question of estimability; i.e., a rational selection of a model from a given family on the basis of observations. The most general question that can be framed in this context is whether there exists a notion of estimability for the entire class of stationary models (i.e., including countably additive, finitely additive and non-additive structures). If we restrict considerations to the class of countably additive, stationary and ergodic models then, at least in principle, the ergodic theorem provides a notion of estimability with an asymptotic justification. The general problem is, however, fraught with difficulties, many of which are induced by a rather intractable subclass of models. If such a notion of estimability for the entire class of stationary models existed, then it would allow us, to estimate a model on the basis of observable events (i.e., cylinder sets). Thus, a fortiori, it would allow us to induce support of divergence or convergence of relative frequency in the estimated model. That such a notion of estimability with respect to the entire class of stationary models is impossible is shown by the following example.

6.1. Example. Let $\underline{p}, \bar{p}, M$ and $\underline{P}$ be as in Example 5.1. Let

$$\mathscr{M}' = \left\{\mu: \mu = \sum_{i=1}^{\infty} \pi_i, (\forall i \geqq 1, \underline{p} \leqq \pi_i \leqq \bar{p}) \text{ and} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \pi_i = \frac{\underline{p} + \bar{p}}{2}\right\} \quad \text{(i)}$$

Let $\underline{P}'$ be the lower envelope of $\mathscr{M}'$ on $(\Omega^\infty, \mathscr{A}^\infty)$. It is then easily seen that

(i) $\underline{P}'$ and $\underline{P}$ are stationary on $\mathscr{A}^\infty$.

(ii) $\underline{P}' =_{\mathscr{C}_\infty} \underline{P}$.

(iii) $\underline{P}'(D^C) = 1$ and $\underline{P}(D) = 1$.

Thus $\underline{P}$ and $\underline{P}'$ are stationary lower probabilities that agree on $\mathscr{C}_\infty$ and yet one of them supports divergence whereas the other supports convergence. □

However, obviously, no finite collection of observations from an infinite sequence can convince us of the convergence or divergence of that sequence. It is thus hardly surprising that one is unable to base a notion of estimability, for the entire class of stationary models, just on observable events. In the classical theory, by choosing to model all stochastic phenomena with countably additive measures, it is as if one independently imposes the constraint of convergence of long-run averages on all stationary models. What then if we have independent reasons to expect divergence in a stationary model? Thus it is interesting to examine whether certain sub-classes of stationary models that support divergence are estimable in terms of specific notions of estimability.

The following is a direct extension of weak-law type estimability to the framework of lower probability.

6.2. Definition. We say that a class of models $\mathscr{E}$ on $(\Omega^\infty, \mathscr{A}^\infty)$ is *strongly estimable* if

$$(\forall C \in \mathscr{C}_\infty)(\forall n \geqq 1)(\exists \widehat{\underline{P}_n}(C))\, \widehat{\underline{P}_n}(C): \Omega^\infty \to [0,1] \quad \text{and} \quad \widehat{\underline{P}_n}(C)$$

is measurable w.r.t. $(\Omega^\infty, \mathscr{C}_n)$ s.t.

$$(\forall \underline{P} \in \mathscr{E})(\forall C \in \mathscr{C}_\infty)(\forall \varepsilon > 0)\, \underline{P}(|\widehat{\underline{P}_n}(C) - \underline{P}(C)| < \varepsilon) \to 1. \quad \square$$

Note that since the models we are considering are in general not determined by their restrictions to $(\Omega^\infty, \mathscr{C}_n)$ for any n, an infinite number of observations will be needed to estimate the complete model on $(\Omega^\infty, \mathscr{C}_\infty)$. Further, since these models are in general not determined by their restrictions to $(\Omega^\infty, \mathscr{C}_\infty)$, this notion is estimability cannot distinguish between models that agree on the cylinder sets (cf. Example 6.1).

Consider, for instance, the models discussed in Example 5.1. We might expect that at least the marginal lower probabilities for these models can be estimated through an estimator of the form

$$\widehat{\underline{P}_n}(1) = \min_{n \leqq j \leqq n+k(n)} r_j(1),$$

where $k(n) \to \infty$ as $n \to \infty$. However, no matter what the choice of $k(n)$ it can easily be shown that for $0 < \varepsilon < \bar{p} - \underline{p}$

$$\underline{P}(|\min_{n \leqq j \leqq n+k(n)} r_j(1) - \underline{P}(\underline{\omega}: \omega_1 = 1)| < \varepsilon) \to 0.$$

This last result holds even if we were to impose a natural choice of $k(n)$ on the class of models by defining $\underline{P}$ as the lower envelope of: ($k(n)$ is fixed)

$$\mathscr{M}'' = \Bigg\{ \mu: \mu = \prod_{i=1}^{\infty} \pi_i: (\forall i \geqq 1)\, \underline{p} \leqq \pi_i \leqq \bar{p}$$
$$\min_{n \leqq j \leqq n+k(n)} \frac{1}{j} \sum_{i=1}^{j} \pi_i \to \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \pi_i = \underline{p}$$
$$\max_{n \leqq j \leqq n+k(n)} \frac{1}{j} \sum_{i=1}^{j} \pi_i \to \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \pi_i = \bar{p} \Bigg\}.$$

In the latter case the following strong law type result holds:

$$\underline{P}\{\underset{\sim}{\omega}: \min_{n \leqq j \leqq n+k(n)} r_j(1) \rightarrow \underline{P}(\omega: \omega_1 = 1) = \underline{p}\} = 1.$$

But lack of continuity in the models denies us the weak law. This phenomenon seems surprising since we have borrowed the 'strong/weak' terminology from the theory of countably additive measures, where a strong law is indeed strong in the sense that it implies the weak law.

A very similar situation arises for Example 5.3 where it can be shown that

$$(\forall C \in \mathscr{C}_\infty)(\exists k(n))\, \underline{P}\{\underset{\sim}{\omega}: \min_{n \leqq j \leqq n+k(n)} r_j(C)(\underset{\sim}{\omega}) \rightarrow \underline{P}(C)\} = 1$$

where $r_j(C)(\underset{\sim}{\omega}) \triangleq \frac{1}{j} \sum_{i=1}^{j} I_C(T^{i-1}\underset{\sim}{\omega})$. However, it also turns out that

$$(\forall \varepsilon, 1 > \varepsilon > 0)\, \underline{P}\{|\min_{n \leqq j \leqq n+k(n)} r_j(\underset{\sim}{\omega}: \omega_1 = 0) - \underline{P}(\underset{\sim}{\omega}: \omega_1 = 0)| < \varepsilon\} = 0$$

for all n. (Note that in this example $\underline{P}(\underset{\sim}{\omega}: \omega_1 = 0) = 0$.)

Thus if we adopt strong estimability as the relevant notion then in all the examples we have considered, the models are not estimable through the naturally identifiable estimators.

That such situations, as have been described in the above examples, can arise is due to the lack of continuity of the probability models considered therein. All the examples considered in Sect. 5 involve dominating measures on $(\Omega^\infty, \mathscr{C}_\infty)$ and hence, according to Corollary 4.4, cannot be continuous from below on $\mathscr{C}_\infty$ if they are to support diverging relative frequency. So, if we confine ourselves to dominated lower probabilities, the natural question that arises is whether there is some restricted notion of continuity from below which when coordinated with a notion of estimability, such as strong estimability (Definition 6.2), will lead to an estimable class of models.

Given our present strategy for constructing lower probabilities that support diverging relative frequency, it appears that such a notion of continuity cannot exist. We start with some kind of infinitary specification (i.e., do not specify marginals) and then the lower probability ($\underline{P}$) is determined from this specification in one way or another (see Sect. 5). This method of specification enables us to assure stationarity and support of divergence, and the way in which divergence is assured motivates an estimator (say $\widehat{\underline{P}_n}$). A weak law type result (and hence estimability) would now hold if, through some kind of continuity, it is possible to approximate the divergence event through events of the form $\{|\widehat{\underline{P}_n}(A) - \underline{P}(A)| < \varepsilon\}$. However, if there is a dominating measure on $(\Omega^\infty, \mathscr{C}_\infty)$, its continuous extension to $(\Omega^\infty, \mathscr{A}^\infty)$ will clearly be able to approximate the divergence event in the same way, and the limit under the dominating measure will dominate the limit under $\underline{P}$. Since the dominating measure can always be taken to be stationary (cf. Corollary 4.4) this will imply that $\underline{P}(D) = 0$ which is a contradiction.

To illustrate this argument we provide the following example.

6.3. Example. Let $\underline{P}$ be specified through a class of measures $\mathscr{M}''$ defined earlier in this section. $\underline{P}$ is continuous from above by virtue of being a lower envelope (Proposition 3.1). Suppose $\underline{P}$ is also continuous from below on $(\mathscr{C}_\infty)_\delta \sim \mathscr{C}_\infty$. Now for $\varepsilon > 0$ (small enough)

$$1 = \underline{P}\left(\bigcup_{m \geqq 1} \bigcap_{n \geqq m} \left(\bigcup_{j=n}^{n+k(n)} |r_n - r_j| > \varepsilon\right)\right)$$

$$= \lim_{m \to \infty} \underline{P}\left(\bigcap_{n \geqq m} \left(\bigcup_{j=n}^{n+k(n)} |r_n - r_j| > \varepsilon\right)\right) \quad \text{(by continuity from below on } (\mathscr{C}_\infty)_\delta \sim \mathscr{C}_\infty)$$

$$= \lim_{m \to \infty} \lim_{l \to \infty} \underline{P}\left(\bigcap_{l \geqq n \geqq m} \left(\bigcup_{j=n}^{n+k(n)} |r_n - r_j| > \varepsilon\right)\right) \quad \text{(by continuity from above)}$$

$$\leqq \lim_{m \to \infty} \lim_{l \to \infty} \mu\left(\bigcap_{l \geqq n \geqq m} \left(\bigcup_{j=n}^{n+k(n)} |r_n - r_j| > \varepsilon\right)\right)$$

where $\mu \geqq_{\mathscr{C}_\infty} \underline{P}$ and μ is stationary

$$= \mu\left(\bigcup_{m \geqq 1} \bigcap_{n \geqq m} \left(\bigcup_{j=n}^{n+k(n)} |r_n - r_j| > \varepsilon\right)\right)$$

$$= 0 \quad \text{since } \mu(D^C) = 1.$$

Hence we have a contradiction and $\underline{P}$ cannot be continuous from below on $(\mathscr{C}_\infty)_\delta \sim \mathscr{C}_\infty$. □

Thus, if we confine ourselves to the present class of stationary models that support divergence, it seems futile to look for an estimator under which this class of models (or a subclass of it) is strongly estimable. Absence of continuity disallows derivation of a "weak law" type result (which is what strong estimability amounts to), even through we have forced these models to satisfy the corresponding "strong law". Instead we turn to the following notion of estimability which has its obvious counterpart in the classical theory of hypothesis testing. We restrict to finite classes of models.

6.4. Definition. We say that the class of models $\mathscr{E} = \{\underline{P}_1, \ldots, \underline{P}_k\}$ is estimable if

$$\exists \{\{E_i^1\}_{i \geqq 1} \ldots \{E_i^k\}_{i \geqq 1} (\forall i, j, 1 \leqq j \leqq k, i \geqq 1, E_1^j \in \mathscr{C}_i)$$
$$(\forall j_1, j_2, 1 \leqq j_1 < j_2 \leqq k \text{ and } \forall i \geqq 1\ E_i^{j_1} \cap E_i^{j_2} = \emptyset)\}$$
$$\text{s.t. } \forall j, 1 \leqq j \leqq k\ \underline{P}_j (E_1^j \text{ all but f.o. in } i) = 1,$$

i.e., in words, there exists a sequence of (disjoint) acceptance regions (or critical regions) such that the lower probability of accepting the "correct" hypothesis all but finitely often is equal to one. □

6.5. Example. Let $\underline{P}_{1,\varepsilon}$ and $\underline{P}_2$ be lower envelopes of $\mathscr{M}_{1,\varepsilon}$ and $\mathscr{M}_2$ defined as follows.

Let $0 \leqq \underline{p} < \bar{p} \leqq 1$ and $o < \varepsilon < \bar{p} - \underline{p}$, $k(n) \to \infty$ as $n \to \infty$

$$\mathscr{M}_{1,\varepsilon} = \Big\{ \mu : \mu = \prod_{i=1}^{\infty} \pi_i, \forall i \geqq 1, \underline{p} \leqq \pi_i \leqq \bar{p} : \min_{n \leqq j \leqq n+k(n)} \frac{1}{j} \sum_{i=1}^{j} i_i \quad \text{converges} \quad \max_{n \leqq j \leqq n+k(n)} \frac{1}{j} \sum_{i=1}^{j} \pi_i \quad \text{converges and the limits differ by more than } \varepsilon \Big\}$$

$$\mathscr{M}_2 = \Big\{ \mu : \mu = \prod_{i=1}^{\infty} \pi_i, \forall i \geqq 1, \underline{p} \leqq \pi_i \leqq \bar{p}, \frac{1}{n} \sum_{i=1}^{n} \pi_i \text{ converges} \Big\}.$$

Let $E_n^1 = \Big\{ \bigcup_{j=n}^{n+k(n)} |r_n - r_j| > \frac{\varepsilon}{2} \Big\}$ and $E_n^2 = (E_n^1)^C$.

Then clearly

(i) $\underline{P}_{1,\varepsilon}(\bigcup_{m \geqq 1} \bigcap_{n \geqq m} E_n^1) = 1$

$\underline{P}_2(\bigcup_{m \geqq 1} \bigcap_{n \geqq m} E_n^2) = 1$

(ii) $\underline{P}_{1,\varepsilon} =_{\mathscr{C}_\infty} \underline{P}_2$, and

(iii) $\underline{P}_{1,\varepsilon}$ and $\underline{P}_2$ are stationary on $\mathscr{A}^\infty$.

Now according to the testing procedure in Definition 6.4 $\underline{P}_{1,\varepsilon}$ and $\underline{P}_2$ are distinguishable, yet as far as observable events are concerned they are indistinguishable! □

7. Conclusions

Our principal result is that if a model for a stationary stochastic phenomenon, (on a finite sample space) with unstable sample averages, is a monotone set function, dominated by a probability measure, then it cannot be continuous from below on the cylinder sets. Thus, in particular, an upper/lower probability model must satisfy these necessary constraints. We have, however, demonstrated several reasonable stationary lower probability models which do support divergence. In particular we have constructed such a finitely additive model on $(\{0,1\}^\infty, 2^{\{0,1\}^\infty})$.

The question of estimability for these classes of models reveals some anomalous behavior in them, anomalous because such behavior is not displayed by the classical models based on countably additive probability measures. The models are not estimable in terms of the obvious counterparts of some classical notions of estimability. Models which satisfy a "strong law" type result fail to satisfy the corresponding "weak law". Models which are asymptotically distinguishable in terms of an hypothesis testing procedure are indistinguishable in terms of observable events (i.e., cylinder sets).

All models considered involve dominating measures and hence, owing to the result stated above, cannot be continuous from below. The classical notions of estimability depend on continuity from below and no restricted notion of continuity seems suitable. Thus it appears that interesting, estimable classes of models should be sought outside the class of dominated lower probabilities.

References

1. Good, I.J.: Subjective probability as the measure of a non-measurable set. In: Nagel, E., Suppes, P., Tarski, A. (eds.), Logic, methodology and philosophy of science. Stanford: University Press 1962
2. Grize, Y.L.: Towards a stationary continuous lower probability-based model for Flicker noise. Ph. D. dissertation. N.Y.: Cornell University, Ithaca 1984
3. Huber, P.J.: Kapazitäten statt Wahrscheinlichkeiten? Gedanken zur Grundlegung der Statistik. Jber. d. Dt. Math.-Verein **78**, 81–92 (1976)
4. Kalikow, S.: Private communication (1984)
5. Kroupa, V.: Frequency stability: Fundamentals and measurement. New York: Inst. of Electric and Electrical Engineering 1983

5a. Kumar, A.: Lower Probabilities on Infinite Spaces and Instabilitity of Stationary Sequences. Ph.D. dissertation. N.Y.: Cornell University, Ithaca 1982

6. Neveu, J.: Mathematical foundations of the calculus of probability. San Francisco: Holden-Day 1965
7. Papamarcou, A.: Some results on undominated lower probabilities. M.S. dissertation, Cornell University, Ithaca, N.Y. 1983
8. Royden, H.L.: Real analysis. New York: Macmillan 1968
9. Shafer, G.: A mathematical theory of evidence. Princeton: Princeton University Press 1976
10. Smith, C.A.B.: Consistency in statistical inference and decision. J. Royal Statist. Soc., Ser. B **23**, 1–25 (1961)
11. Suppes, P.: The measurement of belief. J. Royal Stat. Soc., Ser. B **36**, 160–175 (1974)
12. Walley, P.: Coherent lower (and upper) probabilities. Technical report. Department of Statistics, University of Warwick, England (1981)
13. Walley, P., Fine, T.L.: Towards a frequentist theory of upper and lower probability. Annal. Statist. **10**, 741–761 (1982)
14. Williams, P.M.: Indeterminate probabilities, In: M. Przelecki, K. Szaniawski, R. Wojcicki (eds.). Formal methods in the methodology of empirical sciences. Dordrecht: Reidel 1976

Received April 15, 1983