The Annals of Probability
1987, Vol. 15, No. 2, 783–803

CONTINUOUS LOWER PROBABILITY-BASED MODELS FOR STATIONARY PROCESSES WITH BOUNDED AND DIVERGENT TIME AVERAGES[1]

BY YVES L. GRIZE AND TERRENCE L. FINE

Ciba-Geigy and Cornell University

We have undertaken to develop a new type of stochastic model for nondeterministic empirical processes that exhibit paradoxical characteristics of stationarity, bounded variables, and unstable time averages. By the well-known ergodic theorems of probability theory there is no measure that can model such processes. Hence we are motivated to broaden the scope for mathematical stochastic models. The emerging theory of upper and lower probability, a simple generalization of the theory of finitely additive probability, seems to provide a locus for this new modelling methodology. We focus our attention on the problem of the existence and construction of a lower probability $\underline{P}$ on the power set 2^X of a countably infinite product X of a finite set of reals X_0, that is shift invariant, monotonely continuous along some class $\mathbf{M}$ of sets that includes the cylinder sets $\mathbf{C}$ and such that $\underline{P}(D^*) > 0$ where $D^* = \{\mathbf{x} = (x_i)_{i \in Z} \in X: (1/n)\sum_{i=0}^{n-1} x_i \text{ diverges as } (n \to \infty)\}$. We show that these constraints are incompatible when $\mathbf{M} = 2^X$, but when $\mathbf{M} = \mathbf{C}$ we are able to construct such a lower probability. Most of our results extend to the case of a compact marginal space X_0.

1. Introduction: The Basic Problem. The same notation is used throughout the paper. Denote by Z the set of integers and let X be the space of doubly infinite sequences from a subset X_0 of the reals, i.e.,

$$X = X_0^Z = \{\mathbf{x} = (x_i)_{i \in Z}: x_i \in X_0\}.$$

Let $\mathbf{C}$ be the algebra of cylinder sets of X, $\sigma(\mathbf{C})$ the σ-algebra generated by $\mathbf{C}$, T the left shift on X (i.e., $(\forall i)\ (T\mathbf{x})_i = x_{i+1}$) and T^{-1} the right shift on subsets of X (i.e., $(\forall A \subset X)\ T^{-1}A = \{\mathbf{x}: T\mathbf{x} \in A\}$). Consider the following subset D^* of X:

$$D^* = \left\{\mathbf{x} \in X: \frac{1}{n}\sum_{i=0}^{n-1} x_i \text{ diverges as } (n \to \infty)\right\}.$$

$$= \bigcup_{k=1}^{\infty} \bigcap_{n=0}^{\infty} \bigcup_{m=n}^{\infty} \left\{\mathbf{x} \in X: |\bar{x}_n - \bar{x}_m| > \frac{1}{k}\right\},$$

where

$$\bar{x}_n = \frac{1}{n}\sum_{i=0}^{n-1} x_i.$$

Received March 1985.

[1]This research was supported in part by DOE Contract #DE-AC01-80RA5022.

AMS 1980 *subject classifications*. 60G05, 60A05.

Key words and phrases. Upper and lower probability, interval-valued probability, nonadditive probability, law of large numbers, ergodic theorems, stationarity, unstable relative frequencies, $1/f$ noise, flicker noise.

The problem we discuss in this paper, and that we shall refer to as the *Basic Problem*, is the existence and the construction of a set function $\underline{P}$ under the following constraints:

(BP1) $\underline{P}$ is a *lower probability* on the power set 2^X of X, i.e., $\underline{P}$ is a nonnegative, unit-normed set function with $\underline{P}(\varnothing) = 0$ and which satisfies:

$$\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B) + \underline{P}(A \cap B)$$

for all pairs (A, B) of sets in X such that $A \cup B = X$ or $A \cap B = \varnothing$;

(BP2) X_0 is a finite set;

(BP3) $\underline{P}$ is stationary, i.e., $(\forall A \subset X)\ \underline{P}(T^{-1}A) = \underline{P}(A)$;

(BP4) $\underline{P}$ supports the divergence of averages, i.e., $\underline{P}(D^*) > 0$;

(BP5) $\underline{P}$ is continuous along monotone sequences of sets that belong to a class $\mathbf{M}$ such that $2^X \supset \mathbf{M} \supset \mathbf{C}$, i.e.,

$$(\forall n, A_n \in \mathbf{M}) \qquad A_{n+1} \supset A_n \Rightarrow \underline{P}\Big(\bigcup_n A_n\Big) = \lim_n \underline{P}(A_n)$$

(continuity from below along $\mathbf{M}$),

$$(\forall n, A_n \in \mathbf{M}) \qquad A_{n+1} \subset A_n \Rightarrow \underline{P}\Big(\bigcap_n A_n\Big) = \lim_n \underline{P}(A_n)$$

(continuity from above along $\mathbf{M}$).

The motivation for the study of this particular problem arises in part from certain experimental results that are briefly discussed in Section 2. In Section 3 we summarize the fundamentals of the theory of (upper and) lower probability. In Section 4 the five constraints that constitute the Basic Problem are examined with particular attention devoted to justifying the continuity requirement (BP5). We then proceed to show that the constraints of the Basic Problem are incompatible when $\mathbf{M} = 2^X$ (Section 5) but that, if $\mathbf{M} = \mathbf{C}$, solutions to the Basic Problem exist (Section 6). Moreover these solutions can be chosen to assign a lower probability to D^* that is arbitrarily close to unity. An explicit example of a solution to the Basic Problem is presented in Section 7. We conclude in Section 8 by indicating some possible generalizations of our results, in particular to the case of a compact marginal space X_0.

2. Motivation. The Basic Problem addresses the question of the existence and construction of stochastic models for nondeterministic empirical processes that exhibit paradoxical characteristics of stationarity and bounded but divergent (i.e., fluctuating) time averages. By the celebrated Birkoff ergodic theorem no probability measure can model processes having these properties, under the assumption of the existence of the mean of the process. In fact, it turns out that even without this assumption, stationarity and almost surely bounded but divergent averages are incompatible (Kalikow (1984)). The inability of probability theory to reconcile these properties might tempt one to assert the nonexistence of such processes. However, although claims of stationarity or divergence of averages are mathematical and not directly empirical assertions, a number of compelling experimental results suggest the existence of processes with such paradoxical combination of properties.

More specifically our attention has been drawn to the empirical and well-studied process of the frequency fluctuations of high quality quartz crystal oscillators (see, e.g., Howe, Allan, and Barnes (1981), Kroupa (1983)). The fluctuations have long been modelled as a flicker (or $1/f$) noise process. As noted in Walsh (1981), "although numerous models have been proposed to account for flicker noise, none seems to have been generally accepted; it has often been explained but never understood."

In examining the Allan Variance, a measure developed at the National Bureau of Standards to characterize oscillator instability (and flicker noise), and relating it to the behavior of the time averages of the process, we found that extensive experimental results strongly support the hypothesis that flicker noise is characterized by divergent time averages. From this, under the reasonable physical assumption of the existence of the mean of the process, we must conclude that flicker noise is a nonstationary process. However, this not only contradicts the controversial but widely maintained hypothesis of stationarity of flicker noise (see, e.g., Dutta and Horn (1981) and Keshner (1982)) but also appears hardly tenable in the specific case considered; e.g., high quality oscillators are state-of-the-art nonevolving systems. It is easily shown that allowing for small departures from stationarity in the model does not resolve the contradiction. We also showed that deterministic stationary (i.e., shift invariant) flicker noise-like sequences with bounded (and of course divergent) time averages could be constructed. These results (Grize (1984)) motivated our desire to model flicker noise as a (1) stationary process with (2) bounded averages. Taking into account the experimental evidence of (3) divergent averages, we were led to explore the possibility of constructing stochastic models for processes having the three aforementioned properties. The emerging theory of upper and lower probability seems to provide a conceptual and mathematical framework broad enough to support such models.

3. Upper and lower probability structures. The theory of upper and lower probability (also called interval-valued probability) is a simple generalization of the theory of finitely additive probability. We present only the definitions and results that we shall need in the sequel. Further details of the theory of upper and lower probability can be found, for example, in Walley and Fine (1982) and the references cited therein.

If $\mathbf{A}$ denotes a collection of events modelled as subsets of a set Ω, then the propensity for an event A in $\mathbf{A}$ to occur may be described by the interval $(\underline{P}(A), \overline{P}(A))$ where the two set functions $\underline{P}$ and $\overline{P}$ satisfy the following axioms:

(I) Normalization: $\underline{P}(\Omega) = 1$;
(II) Nonnegativity: $(\forall A \in \mathbf{A})\ \underline{P}(A) \geq 0$;

(III) Conjugacy: $(\forall A \in \mathbf{A})\ \underline{P}(A) + \overline{P}(A^c) = 1$;
(IV) Super- and Subaddivivity:

$$(\forall A, B \in \mathbf{A}) \qquad A \cap B = \varnothing \Rightarrow \begin{cases} \underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B) \\ \qquad \text{(superadditivity (IVa))} \\ \overline{P}(A) + \overline{P}(B) \geq \overline{P}(A \cup B) \\ \qquad \text{(subadditivity (IVb))}. \end{cases}$$

Elementary consequences of these axioms include:

(A) $(\forall A \in \mathbf{A})$

$$\underline{P}(\varnothing) = \overline{P}(\varnothing) = 0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1 = \underline{P}(\Omega) = \overline{P}(\Omega);$$

(B) $(\forall A, B \in \mathbf{A})$

$$A \subset B \Rightarrow \underline{P}(A) \leq \underline{P}(B) \quad \text{and} \quad \overline{P}(A) \leq \overline{P}(B) \quad \text{(monotonicity)};$$

(C) $(\forall A, B \in \mathbf{A})$

$$\underline{P}(A \cup B) \leq \underline{P}(A) + \overline{P}(B).$$

The case where $\underline{P} = \overline{P}$ corresponds to a finitely additive probability structure. If $\overline{P}$ happens to be countably subadditive then it is an outer measure but the converse is in general false.

In view of the conjugacy relation (III) between $\underline{P}$ and $\overline{P}$, the theory of upper and lower probability can be reexpressed in terms of the lower probability $\underline{P}$ by replacing Axiom (IVb) with

(IVc) $(\forall A, B \in \mathbf{A})$

$$A \cup B = \Omega \Rightarrow \underline{P}(A) + \underline{P}(B) \leq 1 + \underline{P}(A \cap B).$$

DEFINITION 1. A *lower probability* $\underline{P}$ on class $\mathbf{A}$ of subsets of a set Ω, which contains Ω and is closed under finite intersections and unions, is a set function that satisfies (I), (II), (IVa), and (IVc). The complementary set function $\overline{P}$ defined by $(\forall A \in \mathbf{A})\ \underline{P}(A) + \overline{P}(A^c) = 1$ is called the upper probability associated with $\underline{P}$.

Henceforth we shall only deal with lower probabilities. The class of lower probabilities can be partitioned into two subclasses:

DEFINITION 2. Let $\mathbf{A}$ be an algebra of subsets of a set Ω and $\underline{P}$ a lower probability on $\mathbf{A}$.

(i) The set $m(\underline{P})$ of all (finitely or countably additive) probability measures μ such that $(\forall A \in \mathbf{A})\ \mu(A) \geq \underline{P}(A)$ is called the *set of dominating measures of* $\underline{P}$.

(ii) If $m(\underline{P}) \neq \varnothing$ then $\underline{P}$ is called a *dominated* lower probability, otherwise it is called *undominated*.

Kumar and Fine (1985) have shown that solutions of the Basic Problem must be sought in the class of undominated lower probabilities.

We conclude this section by observing that a lower probability can always be extended to the power set provided that its domain of definition satisfies some simple conditions:

LEMMA 1. *Let* $\mathbf{F}$ *be a class of subsets of a set* Ω *that contains* $\varnothing$ *and* Ω, *and that is closed under finite unions and finite intersections. Any lower probability* $\underline{P}$ *on* $\mathbf{F}$ *can be extended to a lower probability* $\underline{Q}$ *on* 2^{Ω} *by defining*

$$(\forall A \subset \Omega) \qquad \underline{Q}(A) = \sup\{\underline{P}(B) \colon B \in \mathbf{F},\, B \subset A\}.$$

Further if $\underline{R}$ *is another lower probability on* 2^{Ω} *that extends* $\underline{P}$ *then* $(\forall A \subset \Omega)$ $\underline{R}(A) \geq \underline{Q}(A)$. *For this reason* $\underline{Q}$ *is called the least-committal extension of* $\underline{P}$ *from* $\mathbf{F}$.

PROOF. Routine verification. □

4. Discussion of the constraints in the Basic Problem. Our original problem, as motivated in Section 2, is the construction of stochastic models for stationary random processes with bounded but divergent time averages. As a first step towards the construction of such models, we choose to restrict ourselves to the simpler case where the infinite product space X is generated by a finite marginal space X_0. This finiteness assumption is also entertained so as to treat a case that clearly lies outside the scope of countably additive probability theory. (If X_0 is infinite the measure corresponding to a sequence of independent identically Cauchy-distributed random variables provides a solution to the Basic Problem.)

To justify the imposition of the continuity requirement on our model (constraint (BP5)) we observe the following:

LEMMA 2. *If* $\underline{P}$ *is a stationary lower probability on* 2^X *then there exists a stationary lower probability* $\underline{Q}$ *on* 2^X *such that*

$$(1) \qquad (\forall C \in \mathbf{C}) \qquad \underline{Q}(C) = \underline{P}(C);$$

$$(2) \qquad \underline{Q}(D^*) = 0.$$

PROOF. Define $\underline{Q}$ as the least-committal extension of the restriction of $\underline{P}$ to $\mathbf{C}$ (Lemma 1), and observe that D^* does not contain any nonempty cylinder set. The stationarity of $\underline{Q}$ follows easily from the stationarity of $\underline{P}$. □

Bearing in mind that the cylinder sets are often viewed as the only genuinely observable events, Lemma 2 shows that if a statistician attempts to fit a lower probability model to some real data, his cylinder set observations will yield a lower probability on $\mathbf{C}$ but not enable him to distinguish between models that support D^* and those that do not. Hence we need to restrict the class of desirable models so that models that agree on the observable sets also agree on other sets, such as D^*, of interest.

DEFINITION 3. A monotone set function Ψ on 2^X is *monotonely continuous along* a class $\mathbf{F}$ of subsets of X if

$$(\forall n, A_n \in \mathbf{F}) \qquad A_n \subset A_{n+1} \Rightarrow \Psi\Big(\bigcup_n A_n\Big) = \lim_n \Psi(A_n)$$

(continuity from below),

$$(\forall n, A_n \in \mathbf{F}) \qquad A_n \supset A_{n+1} \Rightarrow \Psi\Big(\bigcap_n A_n\Big) = \lim_n \Psi(A_n)$$

(continuity from above).

If Ψ is continuous only along those monotone sequences $\{A_n\}$ in $\mathbf{F}$ that converge to a set in $\mathbf{F}$ (i.e., such that $\bigcup_n A_n \in \mathbf{F}$ or $\bigcap_n A_n \in \mathbf{F}$) then Ψ is said to be *continuous on* $\mathbf{F}$ (there is of course no distinction if $\mathbf{F}$ is a monotone class).

REMARKS.

1. Observe that if $X = X_0^{\mathbf{Z}}$ with X_0 finite, then X is a compact topological space in the product topology induced by endowing X_0 with the discrete topology, and the cylinder sets $\mathbf{C}$ form the class of clopen (closed and open) sets. Using a compactness argument it is easy to see that any monotone set function on 2^X is continuous on $\mathbf{C}$ (but not necessarily along $\mathbf{C}$).
2. Denote by $\mathbf{C}_\sigma$ and $\mathbf{C}_\delta$, respectively, the class of countable unions and countable intersections of elements of $\mathbf{C}$. For a monotone set function, continuity along $\mathbf{C}$ is equivalent to continuity from above on $\mathbf{C}_\delta$ and from below on $\mathbf{C}_\sigma$ (to see this, use for example an argument similar to the one in Neveu (1965), Proposition I.5.1(d)).

The constraint (BP5) of continuity along a class $\mathbf{M}$ of sets that contain $\mathbf{C}$ will enable us to discriminate between models that disagree in their probability assignments to sets such as D^*.

DEFINITION 4. A lower probability $\underline{P}$ on 2^X, continuous along a class $\mathbf{M}$ of subsets of X, that supports divergent averages is *strongly* (respectively, *weakly*) *distinguishable on* $\mathbf{C}$ (from models with the same continuity properties that do not support divergent averages) if

$$\left.\begin{array}{ll}\text{(i)} & \underline{Q} \text{ is a lower probability} \\ \text{(ii)} & \underline{Q} \text{ is continuous along } \mathbf{M} \\ \text{(iii)} & (\forall C \in \mathbf{C}) \qquad \underline{Q}(C) = \underline{P}(C)\end{array}\right\} \Rightarrow \underline{Q}(D^*) = \underline{P}(D^*)$$

(respectively $\underline{Q}(D^*) > 0$).

Indeed, suppose that a lower probability $\underline{P}$, continuous along a class $\mathbf{M}$ of sets, supports divergent averages. If $\mathbf{M}$ contains $\sigma(\mathbf{C})$, then by virtue of the monotone class theorem (e.g., Loève (1977), page 60) $\underline{P}$ is strongly distinguishable on $\mathbf{C}$. If, however, $\underline{P}$ is only continuous from above along C and if D^* contains a $\mathbf{C}_\delta$-subset of positive lower probability, then $\underline{P}$ is weakly distinguishable on $\mathbf{C}$. The larger the domain of continuity of $\underline{P}$, the stronger the distinguishability of $\underline{P}$.

The problem we wish to solve is therefore the construction of a set function $\underline{P}$ defined on 2^X that satisfies the five constraints ((BP1) through (BP5)) of the Basic Problem and that is (at least weakly) distinguishable on $\mathbf{C}$.

5. Stationarity, continuity on the power set, and support of divergent averages. In light of Lemma 1 the natural domain of definition of a lower probability on X is the power set of X. This motivates the question of the existence of solutions to the Basic Problem with a domain of continuity $\mathbf{M}$ equal to 2^X.

As noted in Section 3, the work of Kumar and Fine (1985) has shown that solutions of the Basic Problem must belong to the class of undominated lower probabilities, excluding finitely additive probabilities. This essentially follows from the requirement of monotone continuity along $\mathbf{C}$ (see the same reference for an example of a finitely additive probability that satisfies (BP1) through (BP4). Until recently (Papamarcou (1983), Papamarcou and Fine (1986)) the only continuous lower probabilities known in the literature were all of the dominated type (e.g., two-monotone capacities, as defined in Huber and Strassen (1973)). Stationary lower probabilities that are undominated and continuous on the power set of X are now known to exist (examples can be found in Grize (1984)). However, we prove that such set functions cannot support divergent averages and thus that the Basic Problem is unsolvable when $\mathbf{M} = 2^X$.

THEOREM 1. *Let $X = X_0^{\mathbf{Z}}$ where X_0 is an arbitrary set. Denote by A the set of aperiodic sequences, i.e., $A = \{\mathbf{x} \in X: (\forall k \in \mathbf{Z})\ (k \neq 0)\ T^k\mathbf{x} \neq \mathbf{x}\}$. If Ψ is a nonnegative superadditive set function such that*

$$(1) \qquad (\forall E \subset X) \qquad \Psi(T^{-1}E) = \Psi(E),$$

$$(2) \qquad \Psi(X) < \infty,$$

$$(3) \qquad \Psi \text{ is continuous from below on } 2^X,$$

then

$$\Psi(A) = 0.$$

COROLLARY 1. *The Basic Problem cannot be solved by a lower probability that is continuous on 2^X.*

PROOF OF THE COROLLARY. D^* is contained in the set of aperiodic sequences. □

Proof of Theorem 1. Define on X an equivalence relation by

$$(\forall \mathbf{x}, \mathbf{y} \in X)\mathbf{x} \approx \mathbf{y} \Leftrightarrow (\exists n \in \mathbf{Z})\mathbf{x} = T^n\mathbf{y}.$$

By employing the axiom of choice, form a set B by choosing exactly one element in each of the equivalence classes that partition A. Note that $(\forall n \in \mathbf{Z})\ T^nA = A$ and since $B \subset A$: $(\forall n \in \mathbf{Z})\ T^nB \subset A$. Thus $\bigcup_{n=-\infty}^{+\infty} T^nB \subset A$. Conversely,

$$\mathbf{x} \in A \Rightarrow (\exists \mathbf{y} \in B)\mathbf{x} \approx \mathbf{y} \Rightarrow (\exists n \in \mathbf{Z})\mathbf{x} \in T^nB;$$

hence,

$$A \subset \bigcup_{n=-\infty}^{+\infty} T^nB.$$

Therefore

$$A = \bigcup_{n=-\infty}^{+\infty} T^nB.$$

Observe now that B is a wandering set (i.e., $\{T^nB,\ n \in \mathbf{Z}\}$ is a disjoint collection of sets). Indeed

$$\begin{aligned} \mathbf{x} \in T^nB \cap T^mB &\Rightarrow (\exists! \mathbf{y} \in B)(\exists! \mathbf{z} \in B) \qquad (\mathbf{x} \approx \mathbf{y}, \mathbf{x} \approx \mathbf{z}) \\ &\Rightarrow \mathbf{y} = \mathbf{z} \\ &\Rightarrow T^n\mathbf{y} = \mathbf{x} = T^m\mathbf{y} \\ &\Rightarrow n = m \quad (\text{by } \mathbf{y} \in A). \end{aligned}$$

Thus $n \neq m \Rightarrow T^nB \cap T^mB = \varnothing$. Using successively the continuity from below, superadditivity, and stationarity of Ψ we have

$$\begin{aligned} (\forall n \in N) \qquad \Psi(A) &= \lim_{k \to \infty} \Psi\left(\bigcup_{i=-k}^{k} T^{i(2n+1)}\left(\bigcup_{j=-n}^{n} T^jB \right)\right) \\ &\geq \lim_{k \to \infty} \sum_{i=-k}^{k} \Psi\left(T^{i(2n+1)} \bigcup_{j=-n}^{n} T^jB \right) \\ &= \lim_{k \to \infty} (2k+1)\Psi\left(\bigcup_{j=-n}^{n} T^jB \right). \end{aligned}$$

Since $\Psi(A)$ is finite we conclude that $(\forall n \in N)\ \Psi(\bigcup_{j=-n}^{n} T^jB) = 0$ and by continuity from below it follows that $\Psi(A) = 0$. □

Remark. There obviously exist stationary probability measures μ such that $\mu(A) > 0$, which proves that the set B in the proof of Theorem 1 cannot be measurable nor can any of its shifts. Up to the set of periodic sequences we have invoked a partition of X into nonmeasurable sets. This result is closely related to the classical construction of nonmeasurable sets (e.g., Halmos (1974), page 69).

6. The Basic Problem when M = C. Theorem 1 obliges us to search for solutions of the Basic Problem that possess a domain **M** of continuity strictly smaller than 2^X. As mentioned in Section 4, continuity along monotone sequences of cylinder sets might be sufficient to ensure weak distinguishability on **C**, which suggests that we seek for solutions with $\mathbf{M} = \mathbf{C}$. Our main result is:

THEOREM 2. *Let $X = X_0^{\mathbf{Z}}$ where X_0 is a finite set. For any p in $[0,1)$ there exists a stationary lower probability $\underline{P}$ continuous along the cylinder sets* **C** *such that $\underline{P}(D^*) > p$. Moreover, $\underline{P}$ is weakly distinguishable on* **C**.

We begin the proof of Theorem 2 by introducing the following lemma, which can be viewed as a partial generalization of the classical extension theorem for probability measures.

LEMMA 3. *Let $\underline{P}$ be a lower probability defined on an algebra* **A** *of subsets of a set Ω. If $\underline{P}$ is continuous on* **A**, *then $\underline{P}$ can be extended to a lower probability $\underline{Q}$ on 2^Ω that is continuous along* **A** *(i.e., from above on $\mathbf{A}_\delta$ and from below on $\mathbf{A}_\sigma$) by letting*

$$(\forall A \in \mathbf{A}_\delta) \qquad P_*(A) = \inf\{\underline{P}(B)\colon B \in \mathbf{A}, B \supset A\},$$
$$(\forall A \subset X) \qquad \underline{Q}(A) = \sup\{P_*(B)\colon B \in \mathbf{A}_\delta, B \subset A\}.$$

REMARK. Unlike the extension of a probability measure, the extension $\underline{Q}$ need not be continuous on $\sigma(\mathbf{A})$ (see part (D) of Section 8).

PROOF. It is easy to see that if Φ is a monotone set function continuous from above on an algebra **A** of subsets of a set Ω, then

$$(\forall A_n, B_m \in \mathbf{A}) \qquad \bigcap_n A_n = \bigcap_m B_m \Rightarrow \lim_n \Phi\Big(\bigcap_{i\le n} A_i\Big) = \lim_m \Phi\Big(\bigcap_{j\le m} B_j\Big)$$

(for example see Neveu (1965), Lemma I.5.1). Therefore we can define a new set function P_* on $\mathbf{A}_\delta$ by

$$(\forall A \in \mathbf{A}_\delta) P_*(A) = \lim_n \underline{P}\Big(\bigcap_{i\le n} A_i\Big)$$

when $\{A_n\}$ is any sequence of **A** such that $\bigcap_n A_n = A$. Evidently

$$(\forall A \in \mathbf{A}_\delta) \qquad P_*(A) = \inf\{P(B)\colon B \in \mathbf{A}, B \supset A\}.$$

Clearly P_* extends $\underline{P}$ to $\mathbf{A}_\delta$ and is monotonely continuous from above along **A**. Since P_* was obtained through a simple limit operation, P_* is also a lower probability on $\mathbf{A}_\delta$. But $\mathbf{A}_\delta$ is closed under finite unions and finite intersections and we can apply Lemma 1 to P_*. Denote by $\underline{Q}$ the least-committal extension of P_* from $\mathbf{A}_\delta$, i.e.,

$$(\forall A \subset X) \qquad \underline{Q}(A) = \sup\{P_*(B)\colon B \in \mathbf{A}_\delta, B \subset A\}.$$

$\underline{Q}$ is monotonely continuous from above along **A** and it remains only to show that it is also monotonely continuous from below along **A**. Let $A \in \mathbf{A}_\sigma$ and

$A = \bigcup_n A_n$ for some sequence $\{A_n\}$ in **A**. Clearly

$$(\forall n) \qquad \underline{Q}\Big(\bigcup_{i\le n} A_i\Big) = \underline{P}\Big(\bigcup_{i\le n} A_i\Big) \le \underline{Q}(A).$$

Thus

$$\lim_n \underline{Q}\Big(\bigcup_{i\le n} A_i\Big) \le \underline{Q}(A).$$

Suppose that

$$\lim_n \underline{Q}\Big(\bigcup_{i\le n} A_i\Big) < \underline{Q}(A).$$

Then $(\exists B \subset \mathbf{A}_\delta)\ B \subset A$ and

$$\lim_n \underline{P}\Big(\bigcup_{i\le n} A_i\Big) = \lim_n \underline{Q}\Big(\bigcup_{i\le n} A_i\Big) < \underline{Q}(B) = \lim_n \underline{P}\Big(\bigcap_{j\le n} B_j\Big)$$

for some sequence $\{B_j\}$ in **A** with $\bigcap B_j = B$. By the properties of upper and lower probabilities (Section 3 (C)) we have

$$\underline{P}\bigg(\Big(\bigcup_{i\le n} A_i\Big) \cup \Big(\bigcap_{j\le n} B_j\Big)^c\bigg) \le \underline{P}\Big(\bigcup_{i\le n} A_i\Big) + \overline{P}\bigg(\Big(\bigcap_{j\le n} B_j\Big)^c\bigg)$$

$$= \underline{P}\Big(\bigcup_{i\le n} A_i\Big) + 1 - \underline{P}\Big(\bigcap_{j\le n} B_j\Big).$$

But $\{(\bigcup_{i\le n} A_i) \cup (\bigcap_{j\le n} B_j)^c\}$ is a monotonely increasing sequence in **A** that converges to Ω. Using the continuity from below on **A** of $\underline{P}$ we have

$$1 \le \lim_n \underline{P}\Big(\bigcup_{i\le n} A_i\Big) + 1 - \lim_n \underline{P}\Big(\bigcap_{j\le n} B_j\Big),$$

i.e.,

$$\lim_n \underline{P}\Big(\bigcap_{j\le n} B_j\Big) \le \lim_n \underline{P}\Big(\bigcup_{i\le n} A_i\Big),$$

which contradicts the hypothesis. We conclude that

$$\lim_n \underline{Q}\Big(\bigcup_{i\le n} A_i\Big) = \underline{Q}(A)$$

and thus $\underline{Q}$ is continuous from below along **A**. Equivalently, by Remark (B) of Section 4, $\overline{Q}$ is continuous from above on $\mathbf{A}_\delta$ and from below on $\mathbf{A}_\sigma$. □

LEMMA 4. *If there exists a stationary lower probability* $\underline{P}$ *on* 2^X *whose restriction to* **C** *is continuous on* **C** *and such that* $\underline{P}(E) > 0$ *for some* $\mathbf{C}_\delta$*-subset* E *of* D^*, *then there exist a weakly distinguishable solution of the Basic Problem (with* $\mathbf{M} = \mathbf{C}$*) and a stationary lower probability* $\underline{Q}$ *continuous along* **C** *such that* $\underline{Q}(E) \ge \underline{P}(E)$.

PROOF. Suppose that there exists a stationary lower probability $\underline{P}$ such that

$$(\exists E \in \mathbf{C}_\delta) \qquad E \subset D^* \quad \text{and} \quad \underline{P}(E) > 0.$$

Consider the restriction $\underline{P}_0$ of $\underline{P}$ to C. By applying Lemma 3 to $\underline{P}_0$ we obtain an extension $\underline{Q}$ of $\underline{P}_0$ on 2^X that is continuous along $\mathbf{C}$. We note that the stationarity of $\underline{P}$ implies the stationarity of $\underline{Q}$. Observe that

$$(\forall C \in \mathbf{C}) \qquad E \subset C \Rightarrow \underline{P}_0(C) = \underline{P}(C) \geq \underline{P}(E) > 0.$$

Let $E = \bigcap_n C_n$ for some sequence $\{C_n\}$ of cylinder sets. In view of the definition of $\underline{Q}$ we have

$$\underline{Q}(E) = \lim_n \underline{P}_0\Big(\bigcap_{i \leq n} C_i\Big) \geq \underline{P}(E) > 0.$$

Therefore $\underline{Q}(D^*) > 0$ and $\underline{Q}$ satisfies (BP2). Clearly $\underline{Q}$ is also weakly distinguishable on $\mathbf{C}$. □

The next two lemmas address the question of the existence of a lower probability that satisfies the conditions of Lemma 4.

LEMMA 5. *Let n be an integer greater than* 1 *and* $\mathbf{F}$ *a collection of subsets of a set* Ω *such that all collections of* $2n - 2$ *elements of* $\mathbf{F}$ *have a nonempty intersection. Set* $\mathbf{F}_0 = \{\Omega\}$ *and for* $(1 \leq k \leq n-1)$ *define* $\mathbf{F}_k$ *by*

$$\mathbf{F}_k = \{G\colon (\exists G_1, G_2, \ldots, G_k \in \mathbf{F}) G = G_1 \cap G_2 \cap \cdots \cap G_k\}.$$

Then the set function $\underline{P}_n$ *defined by*

$$\underline{P}_n(A) = \begin{cases} 1, & \text{if } A = \Omega, \\ 1 - \dfrac{k}{n}, & \text{if } (\forall G \in \mathbf{F}_{k-1})\ G \not\subset A \text{ and } (\exists G \in \mathbf{F}_k)\ G \subset A, \\ 0, & \text{otherwise}, \end{cases}$$

is a lower probability on 2^Ω.

PROOF. Clearly $\underline{P}_n(\Omega) = 1$, $\underline{P}_n$ is nonnegative and monotone. By the property of the collection $\mathbf{F}$ no two sets of positive lower probability can be disjoint, and since $\underline{P}_n$ is monotone, it must be superadditive. Finally consider a pair (A, B) of sets $(A \neq \Omega, B \neq \Omega)$ of positive lower probability, say:

$$\underline{P}_n(A) = 1 - \frac{r}{n} \quad \text{and} \quad \underline{P}_n(B) = 1 - \frac{s}{n}$$

for some r and s $(1 \leq r, s \leq n-1)$. Then there exist $G_i \in \mathbf{F}$ $(1 \leq i \leq r)$ and $H_j \in \mathbf{F}$ $(1 \leq j \leq s)$ such that

$$G_1 \cap \cdots \cap G_r \subset A \quad \text{and} \quad H_1 \cap \cdots \cap H_s \subset B$$

(and $G_i \neq G_k$ if $i \neq k$ and $H_j \neq H_l$ if $j \neq l$). But then

$$\varnothing \neq G_1 \cap \cdots \cap G_r \cap H_1 \cap \cdots \cap H_s \subset A \cap B.$$

Thus

$$\underline{P}_n(A \cap B) \geq 1 - \frac{r+s}{n}$$

and

$$\underline{P}_n(A) + \underline{P}_n(B) = 1 + \left(1 - \frac{r+s}{n}\right) \leq 1 + \underline{P}(A \cap B).$$

This concludes the proof that $\underline{P}_n$ is a lower probability on **A**. □

Note that if the collection **F** in Lemma 5 (where $\Omega = X$) is shift-invariant (i.e., $(\forall G \in \mathbf{F})\ G \in \mathbf{F} \Leftrightarrow T^{-1}G \in \mathbf{F}$), then the induced lower probability is stationary. We can apply Lemma 5 to generate a lower probability that satisfies the conditions of Lemma 4 by taking for **F** the collection $\{T^k E, k \in \mathbf{Z}\}$ where E is a $\mathbf{C}_\delta$-subset of D^*, provided that **F** has the adequate nonempty intersection property. The final step required for our proof of Theorem 2 is:

LEMMA 6. *Let K be an arbitrary positive integer and ε a real number such that $0 < \varepsilon < 1$. Consider any sequence of positive integers $\{b_j, j \in \mathbf{N}\}$ such that*

$$(\forall i) \qquad K\frac{\sum_{j<i} b_j}{\sum_{j \leq i} b_j} < \delta,$$

where $\delta = (1 - \varepsilon)/2$, and define

$$r_0 = 0, \qquad r_i = \sum_{j=1}^{3i} b_j \qquad (\forall i, i \geq 1).$$

Finally, define for all i ($i \geq 0$) the cylinder set

$$D_i = \{\mathbf{x} \in X: (\exists n, m)\ r_i < n < m \leq r_{i+1}, |\bar{x}_n - \bar{x}_m| \geq \varepsilon\},$$

where $\bar{x}_n = (1/n)\sum_{i=0}^{n-1} x_i$.

The $\mathbf{C}_\delta$-set E defined by $E = \bigcap_{i=0}^{\infty} D_i$ is a subset of D^* and enjoys the property

$$(\forall k_1, \dots, k_K \in \mathbf{Z}) \qquad T^{-k_1}E \cap \cdots \cap T^{-k_K}E \neq \varnothing.$$

PROOF. E is obviously a subset of D^*. The proof of the second part of the conclusion is elementary but lengthy, and only the main line is presented here (the detailed proof can be found in Grize (1984), pages 88–92 and 118–122, and is appended).

Define a sequence $\mathbf{y} = (y_n)_{n \in \mathbf{Z}}$ by

$$y_n = \begin{cases} 1, & \text{if } \sum_{j=1}^{2m-1} b_j < n + 1 \leq \sum_{j=1}^{2m} b_j, \\ 0, & \text{otherwise,} \end{cases}$$

where m is a positive integer. Note that for nonnegative n, $\mathbf{y}$ is composed of alternating blocks of zeros and ones of length b_i ($i \geq 1$) and clearly $\mathbf{y}$ belongs to E. The case $K = 1$ is trivial. By induction on K ($K \geq 2$) one can prove that for all $k_1, \dots, k_{K-1}$ such that $0 \le k_i < k_{i+1}$ the sequence $\mathbf{z}$ defined by

$$(\forall n \in \mathbf{Z}) \qquad z_n = \begin{cases} y_n, & \text{if } n \le k_1 - 1, \\ y_{n-k_1}, & \text{if } k_1 - 1 < n \le k_2 - 1, \\ y_{n-k_2}, & \text{if } k_2 - 1 < n \le k_3 - 1, \\ \vdots & \\ y_{n-k_{K-1}}, & \text{if } k_K < n, \end{cases}$$

belongs to $E \cap T^{-k_1}E \cap \cdots \cap T^{-k_{K-1}}$. The conclusion of the lemma follows then easily. □

PROOF OF THEOREM 2. Let p in $[0,1)$ be given and choose a nonnegative integer n such that $1 - 1/n > p$. Let $\mathbf{F} = \{T^kE, k \in \mathbf{Z}\}$ where E is a set as in Lemma 6 for $K = 2n - 2$. $\mathbf{F}$ is shift-invariant and has the nonempty intersection property required for the application of Lemma 5. Using this lemma we obtain a stationary lower probability $\underline{P}_n$ on 2^X such that $\underline{P}_n(E) \ge p$. By Remark (A) of Section 4 the restriction of $\underline{P}_n$ to $\mathbf{C}$ is continuous on $\mathbf{C}$ and Theorem 2 follows now from Lemma 4. □

7. An explicit example of a solution to the Basic Problem when $\mathbf{M} = \mathbf{C}$. We conjecture that, for all n greater than 1, the lower probability $\underline{P}_n$ of Lemma 5 when induced by the family $\mathbf{F} = \{T^kE, k \in \mathbf{Z}\}$ where E is a set as in Lemma 6 with $K = 2n - 2$, is always continuous along $\mathbf{C}$ and therefore solves the Basic problem when $\mathbf{M} = \mathbf{C}$. This means that, in this case, the extension $\underline{Q}$ of Lemma 3 is identical to $\underline{P}_n$. We prove this conjecture for $n = 2$.

Let E be the set defined in Lemma 6 for $K = 2$ (and say $\varepsilon = \frac{1}{2}$ and $b_j = 2^{2^{100+j}}$). Denote by $\hat{\underline{P}}$ the lower probability induced by $\mathbf{F} = \{T^kE, k \in \mathbf{Z}\}$, i.e.,

$$(\forall A \subset X) \qquad \hat{\underline{P}}(A) = \begin{cases} 1, & \text{if } A = X, \\ \frac{1}{2}, & \text{if } A \ne X \text{ and } (\exists k \in \mathbf{Z})\ T^kE \subset A, \\ 0, & \text{otherwise.} \end{cases}$$

Obviously $\hat{\underline{P}}$ supports divergent averages and is stationary. It is possible to see directly that $\hat{\underline{P}}$ is continuous along $\mathbf{C}$ and therefore also weakly distinguishable on $\mathbf{C}$ (by $E \in \mathbf{C}_\delta$).

Continuity from below. Consider X as a compact topological space as in Remark (A) of Section 4. Continuity from below along $\mathbf{C}$ follows easily from the fact that cylinder sets are open and the elements of $\mathbf{F}$ are closed and hence compact. Indeed, suppose that $(\forall n, A_n \in \mathbf{C})$ $A_n \subset A_{n+1}$ and $\bigcup_n A_n = A$. If $\hat{\underline{P}}(A) = 0$ there is nothing to prove by monotonicity of $\hat{\underline{P}}$. If $\hat{\underline{P}}(A) = \frac{1}{2}$ then

$$(\exists F \in \mathbf{F}) \qquad F \subset A = \bigcup_n A_n.$$

Therefore $(\exists N)\ \mathbf{F} \subset A_N$ and thus

$$\lim_n \underline{\hat{P}}(A_n) = \underline{\hat{P}}(A_N) = \tfrac{1}{2},$$

and similarly if $\underline{\hat{P}}(A) = 1$.

Continuity from above.

LEMMA 7. *If the family* $\mathbf{F}$ *inducing* $\underline{\hat{P}}$ *is such that any cylinder set other than* X *contains at most finitely many elements of* $\mathbf{F}$, *then* $\underline{\hat{P}}$ *is continuous from above along* $\mathbf{C}$.

PROOF. Let $\{A_n\}$ be a decreasing sequence of cylinder sets and let $A = \bigcap_n A_n$. If $\underline{\hat{P}}(A) \geq \frac{1}{2}$ then continuity from above is obvious. Assume that $\underline{\hat{P}}(A) = 0$. Suppose that for some n_0 $(\forall n \geq n_0)$ $\underline{\hat{P}}(A_n) = \frac{1}{2}$. Then

$$(\forall n \geq n_0)\ (\exists F_n \in \mathbf{F}) \qquad A_n \supset F_n.$$

By the monotony of $\{A_n\}$: $(\forall n \geq n_0)$ $A_{n_0} \supset F_n$. Then by hypothesis there must exist a set F in $\mathbf{F}$ contained in A_n for all n, thus also contained in A. This contradicts $\underline{\hat{P}}(A) = 0$ and concludes the proof. □

We shall show that the collection $\mathbf{F} = \{T^k E, k \in \mathbf{Z}\}$ where E is as in Lemma 6 satisfies the condition of Lemma 7. The following definition is useful.

DEFINITION 5. Let $m \leq n$ be two integers. Two elements $\mathbf{x}$ and $\mathbf{y}$ of X *agree modulo* (m, n) (write $\mathbf{x} \equiv \mathbf{y} \bmod (m, n)$) if

$$(\forall i, m \leq i \leq n) \qquad x_i = y_i.$$

LEMMA 8. (i) *If a subset* F *of* X *is such that*

$$(\forall \mathbf{x} \in X)\ (\forall m, n \in \mathbf{Z})\ (\exists K \geq 0)\ (\forall |k| \geq K)\ (\exists \mathbf{y} \in T^k F) \qquad \mathbf{y} \equiv \mathbf{x} \bmod (m, n),$$

then

$$(\forall C \in \mathbf{C})\ (C \neq \varnothing)\ (\exists K)\ (\forall |k| > K) \qquad C \cap T^k F \neq \varnothing$$

and the family $\{T^k F, k \in \mathbf{Z}\}$ *satisfies the hypothesis of Lemma* 7.

(ii) *The set* E *defined in Lemma* 6 *satisfies the condition in* (i) *above.*

PROOF. (i) Let $C \in \mathbf{C}$, $C \neq \varnothing$. By the characteristic finitary property of cylinder sets there exist $m \leq n$ such that

$$\forall \mathbf{x} \in \mathbf{C} \quad \text{and} \quad \mathbf{x} \equiv \mathbf{y} \bmod (m, n) \Rightarrow \mathbf{y} \in \mathbf{C}.$$

Now choose $\mathbf{x}$ in C. By hypothesis:

$$\exists K(m, n)\ (\forall |k| > K)\ (\exists \mathbf{y} \in T^k F) \qquad \mathbf{y} \equiv \mathbf{x} \bmod (m, n).$$

But then $\mathbf{y} \in T^k F \cap C$, which proves the first part of (i). Again let C be an arbitrary cylinder set such that $C \neq X$. Then

$$(\exists K)\,(\forall |k| > K) \qquad T^k F \cap C^c \neq \varnothing;$$

hence, $T^k F \not\subset C$ and C cannot contain more than finitely many sets of the form $T^k F$ $(k \in \mathbf{Z})$.

(ii) Recall that $E = \bigcap_{i=1}^{\infty} D_i$ where

$$D_i = \{\mathbf{x} \in X\colon (\exists n, n')\; 0 \le r_i < n < n' \le r_{i+1}, |\bar{x}_n - \bar{x}_{n'}| \ge \varepsilon\}$$

and (r_i) diverges as $(i \to \infty)$ (Lemma 6). Fix $\mathbf{x}$ in X and $m \le n$ two integers.

For all k, $k \ge n$, the sequence $\mathbf{y}$ defined by

$$y_i = \begin{cases} x_i, & \text{if } i \le k, \\ z_{n-i}, & \text{if } i > k, \end{cases}$$

where $\mathbf{z}$ is an arbitrary element of E, belongs clearly to $T^{-k}E$ and agrees with $\mathbf{x}$ modulo (m, n).

On the other hand, we can choose r_{i_0} such that $(\forall i \ge i_0)$ $r_{i+1} - r_i$ is large enough to ensure for any k greater than r_{i_0} the existence of $\mathbf{z}$ in E with

$$(\forall j, m \le j \le n) \qquad z_{k+j} = x_j.$$

Let now $\mathbf{y} = T^k \mathbf{z}$ (where $k \ge r_{i_0}$). Clearly $\mathbf{y} \in T^k E$ and $\mathbf{y} \equiv \mathbf{x} \bmod (m, n)$.

Therefore by taking for K the maximum of n and r_{i_0} we have

$$(\forall |k| > K)\,(\exists \mathbf{y} \in T^k E) \qquad \mathbf{y} \equiv \mathbf{x} \bmod (m, n). \qquad \square$$

The continuity from above along $\mathbf{C}$ of $\underline{\hat{P}}$ follows from the application of Lemmas 8 and 7.

REMARK. Examples with a continuous range can also be constructed (Grize (1984)).

8. Generalizations. Generalizations of our results can be sought in the following directions:

(A) by dealing with the noninvertible transformation T in the case of single-sided sequences $X = X_0^N$;
(B) by relaxing the condition that X_0 be finite;
(C) by imposing on the solutions of the Basic Problem the stronger condition that $\underline{P}(D^*) = 1$;
(D) by requiring a domain of continuity larger than $\mathbf{C}$ (and so possibly achieving a stronger form of distinguishability on $\mathbf{C}$).

We shall briefly discuss each of these four possible generalizations:

(A) $X = X_0^N$. All of our results extend without difficulties to that case and we refer to Grize (1984) for the details.

(B) *More general marginal space.* Our results are summarized in the following theorem:

THEOREM 3. *Let $X = X_0^{\mathbf{Z}}$ or $X_0^{\mathbf{N}}$ where X_0 is a compact subset of* $\mathbf{R}$ *(in the usual topology). There exists a stationary lower probability $\underline{P}$ on 2^X that is continuous from below on the open sets and from above on the $\mathbf{C}_\delta$-sets, $\underline{P}(D^*) > 0$, and $\underline{P}$ is weakly distinguishable on* $\mathbf{C}$ *for models having these continuity properties.*

PROOF. Consider the example of Section 7 and note that E is a closed, hence compact, subset of D^*. The argument of Section 7 remains valid. □

(C) $\underline{P}(D^*) = 1$. Theorem 2 asserts the existence of solutions to the Basic Problem that assign to D^* a lower probability arbitrarily close to unity. However, the approach taken fails to produce a lower probability that satisfies the conditions of the Basic Problem with $\underline{P}(D^*) = 1$ and we are unable at this point to exhibit an example of such a lower probability. (Incidentally, $\underline{P} = \liminf_n \underline{P}_n$, where $\underline{P}_n$ solves the Basic Problem with $\underline{P}_n(D^*) \geq 1 - 1/n$, is not continuous from below along $\mathbf{C}$.)

(D) *Larger domain of continuity.* The most natural domain $\mathbf{M}$ of continuity such that $\mathbf{C} \subset \mathbf{M} \subset 2^X$ is evidently $\sigma(\mathbf{C})$. However, the lower probability $\hat{\underline{P}}$ of Section 7 is not continuous from below on $\sigma(\mathbf{C})$. Indeed, let $\{\mathbf{x}^i,\ i = 1, 2, \dots\}$ be a countable set of elements of D^{*c} and set

$$(\forall n) \qquad A_n = X - \bigcup_{j \geq n} \{\mathbf{x}^j\} \in \mathbf{C}_{\sigma\delta}.$$

Observe that A_n converges to X but $(\forall n)\ \hat{\underline{P}}(A_n) = \frac{1}{2}$.

Although we are aware that many questions still remain unanswered, we think that we have taken the first step toward the solution of the important problem of modelling stationary stochastic processes with unstable bounded averages. We also hope to have supported the applicability of the concept of interval-valued probability.

APPENDIX

PROOF OF LEMMA 6. E is obviously a subset of D^*. The second part of the conclusion is proven by induction on K. The proof is elementary but lengthy and we first consider here the case $K = 2$.

We proceed in a number of steps.

STEP 1. Observe that it is enough to show

$$(\forall k \geq 1) \qquad E \cap T^{-k}E \neq \varnothing.$$

STEP 2. Consider the binary sequence $\mathbf{y} = (y_n)_{n \in \mathbf{Z}}$ composed for $n > 0$ of alternating blocks of zeros and ones of length b_i $(i \geq 1)$, i.e.,

$$(\forall n \in \mathbf{Z}) \qquad y_n = \begin{cases} 1, & \text{if } \sum_{j=1}^{2m-1} b_j < n \leq \sum_{j=1}^{2m} b_j \text{ for some } m \in N, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\bar{y}_n = (1/n)\sum_{i=0}^{n-1} y_i$ and $c_m = \sum_{j=1}^{m} b_j$ and note that, by the property of the sequence $(b_j)_{j \in \mathbf{N}}$,

$$(\forall m \in \mathbf{N}) \qquad \bar{y}_{c_{2m}} > 1 - \delta \quad \text{and} \quad \bar{y}_{c_{2m-1}} < \delta.$$

Clearly $(\bar{y}_n)_{n \in \mathbf{N}}$ fluctuates with an amplitude larger than ε and of course $\mathbf{y}$ belongs to E.

We shall now prove that

$$(\forall k \in \mathbf{N}) \qquad E \cap T^{-k}E \neq \varnothing.$$

To this end, for any given k $(k \in \mathbf{N})$, define a sequence $\mathbf{z} = (z_n)_{n \in \mathbf{Z}}$ by

$$(\forall n \in \mathbf{Z}) \qquad z_n = \begin{cases} y_n, & \text{if } n < k, \\ y_{n-k}, & \text{if } n \geq k, \end{cases}$$

i.e.,

$$\mathbf{z} = (\ldots, y_0, y_1, \ldots, y_{k-1}, y_0, y_1, y_2, \ldots).$$

We shall prove that $\mathbf{z} \in E \cap T^{-k}E$. Note that

$$(\forall i \geq 0) \qquad (T^k \mathbf{z})_i = y_i.$$

Thus $T^k \mathbf{z} \in E$, hence $\mathbf{z} \in T^{-k}E$. It remains to show that $\mathbf{z} \in E$, i.e., that $(\forall i \geq 0)$ $\mathbf{z} \in D_i$.

STEP 3. Fix i $(i \geq 0)$ and consider the interval $(r_i, r_{i+1}]$. Writing for simplicity $b_{3i+1} = \beta_1$, $b_{3i+2} = \beta_2$, $b_{3i+3} = \beta_3$ we get

$$r_i + \beta_1 + \beta_2 + \beta_3 = r_{i+1}.$$

Observe that

$$2(r_i + \beta_1 + \beta_2) < \delta(r_i + \beta_1 + \beta_2 + \beta_3) < r_{i+1}.$$

W.l.o.g. we can assume that

$$(r_i < n \leq r_{i+1}) y_n = \begin{cases} 1, & \text{if } r_i < n \leq r_i + \beta_1, \\ 0, & \text{if } r_i + \beta_1 < n \leq r_i + \beta_1 + \beta_2, \\ 1, & \text{if } r_i + \beta_1 + \beta_2 < n \leq r_{i+1}. \end{cases}$$

CASE I. $k > r_{i+1}$. This case is simple. For all $n \leq k$ we have $y_n = z_n$; therefore,

$$\mathbf{y} \in \bigcap_{i=0}^{\infty} D_i \Rightarrow \mathbf{z} \in D_i.$$

CASE II. $k \le r_{i+1}$.
(A) If $k > r_i + \beta_1$, then

$$\bar{z}_{r_i+\beta_1} > \frac{\beta_1}{r_i + \beta_1} = 1 - \frac{r_i}{r_i + \beta_1} > 1 - \delta$$

and

$$r_i < r_i + \beta_1 < k \le r_{i+1}.$$

If $k \le r_i + \beta_1$, then

$$\bar{z}_{k+r_i+\beta_1} > 1 - \frac{2r_i}{k + r_i + \beta_1}$$
$$> 1 - \frac{2r_i}{r_i + \beta_1} > 1 - \delta$$

and

$$r_i < k + r_i + \beta_1 < 2(r_i + \beta_1) < r_{i+1}.$$

In both situations we have found an index n between r_i and r_{i+1} s.t.

$$\bar{z}_n \ge 1 - \delta.$$

(B) If $r_i + \beta_1 + \beta_2 \le k$, then

$$\bar{z}_{r_i+\beta_1+\beta_2} < \frac{r_i + \beta_1}{r_i + \beta_1 + \beta_2} < \delta \quad \text{and} \quad r_i < r_i + \beta_1 + \beta_2 < r_{i+1}.$$

If $r_i + \beta_1 + \beta_2 > k$, then

$$\bar{z}_{k+r_i+\beta_1+\beta_2} < \frac{2(r_i + \beta_1)}{k + r_i + \beta_1 + \beta_2} < \frac{2(r_i + \beta_1)}{r_i + \beta_1 + \beta_2} < \delta$$

and

$$r_i < k + r_i + \beta_1 + \beta_2 < 2(r_i + \beta_1 + \beta_2) < r_{i+1}.$$

In both situations we have found an index m between r_i and r_{i+1} s.t.

$$\bar{z}_m \le \delta.$$

From observations A and B we conclude that $|\bar{z}_n - \bar{z}_m| \ge 1 - 2\delta = \varepsilon$ for some m, n in $(r_i, r_{i+1}]$; hence, $\mathbf{z} \in D_i$.

In both Cases I and II we have $\mathbf{z} \in D_i$. We conclude that $\mathbf{z} \in D_i$ $(\forall i \ge 0)$, i.e., $\mathbf{z} \in E$.

We now turn to the case of $K > 2$. We proceed in a number of steps.

STEP 1. Again it is enough to prove that

$$(*) \qquad (\forall k_1, \ldots, k_{K-1} \ge 0) \qquad E \cap T^{-k_1}E \cap \cdots \cap T^{-k_{K-1}}E \ne \varnothing.$$

STEP 2. It remains to show that E satisfies $(*)$. Let $\mathbf{y}$ denote the same sequence as in the case $K = 2$. Set $L = k - 1$. For any given set of L natural numbers $k_1 < k_2 < \cdots < k_L$ we define a sequence $\mathbf{z}$ by

$$(\forall n \in \mathbf{Z}) \qquad z_n = \begin{cases} y_n, & \text{if } n \le k_1, \\ y_{n-k_1}, & \text{if } k_1 < n \le k_2, \\ y_{n-k_2}, & \text{if } k_2 < n \le k_3, \\ \vdots & \\ y_{n-k_L}, & \text{if } k_L < n, \end{cases}$$

i.e.,

$$\mathbf{z} = (\ldots, y_1, y_2, \ldots, y_{k_1}, y_1, \ldots, y_{k_2-k_1}, y_1, \ldots, y_{k_L-k_{L-1}}, y_1, y_2, \ldots).$$

We shall prove by induction on L that such a sequence $\mathbf{z}$ belongs to

$$E \cap T^{-k_1}E \cap \cdots \cap T^{-k_L E}.$$

The case $L = 1$ (i.e., $K = 2$) has been treated above. We assume now that the proposition is true for $L - 1$. Note that

$$T^{k_1}\mathbf{z} = (\ldots, y_1, y_2, \ldots, y_{k_2-k_1}, y_1, \ldots, y_{k_3-k_2}, \ldots, y_{k_L-k_{L-1}}, y_1, y_2, \ldots).$$

By the induction hypothesis,

$$T^{k_1}\mathbf{z} \in E \cap T^{-(k_2-k_1)}E \cap T^{-(k_3-k_1)}E \cap \cdots \cap T^{-(k_L-k_{L-1})}E,$$

i.e.,

$$\mathbf{z} \in T^{-k_1}E \cap T^{-k_2}E \cap \cdots \cap T^{-k_L}E.$$

It remains only to show that $\mathbf{z} \in E$, i.e., that $\mathbf{z} \in D_i$ $(\forall i)$.

STEP 3. As in Step 3 of the case $K = 2$ we fix i $(i \ge 0)$ and write

$$r_i + \beta_1 + \beta_2 + \beta_3 = r_{i+1}$$

and assume w.l.o.g. that β_1 is a block of ones (β_2 a block of zeros).

CASE I. $k_L > r_{i+1}$. Then we are in a situation of dimension lower than L and by the induction hypothesis $\mathbf{z} \in D_i$.

CASE II. $k_L \le r_{i+1}$. Then there is a p $(p \ge 1)$ s.t.

$$k_1 < k_2 < k_{p-1} < r_i < k_p < \cdots < k_L \le r_{i+1}$$

(assume if necessary that $k_0 = 0$).

(a) Consider the first index j s.t.

$$p \le j \le L \quad \text{and} \quad k_{j-1} + r_i + \beta_1 < k_j.$$

If such an index exists, then

$$\bar{z}_{k_{j-1}} + r_i + \beta_1 > 1 - \frac{j \cdot r_i}{r_i + \beta_1 + k_{j-1}} > 1 - j\frac{r_i}{r_i + \beta_1} > 1 - \delta$$

and

$$r_i < k_{j-1} + r_i + \beta_1 < j(r_i + \beta_1) < (L+1)(r_i + \beta_1 + \beta_2)$$
$$< \delta \cdot r_{i+1} < r_{i+1}.$$

If no such index exists then

$$\bar{z}_{k_L + r_i + \beta_1} > 1 - (L+1)\frac{r_i}{r_i + \beta_1} > 1 - \delta$$

and

$$k_L + r_i + \beta_1 < (L+1)(r_i + \beta_1 + \beta_2) \le r_{i+1}.$$

In both situations we have found an index n in $(r_i, r_{i+1}]$ where $\bar{z}_n \ge 1 - \delta$.

(b) Consider now the first index l s.t.

$$P \le l \le L \quad \text{and} \quad k_{l-1} + r_i + \beta_1 + \beta_2 \le k_l.$$

Note that necessarily $l \ge j$. If such an index l exists then

$$\bar{z}_{k_{l-1} + r_i + \beta_1 + \beta_2} < \frac{l(r_i + \beta_1)}{k_{l-1} + r_i + \beta_1 + \beta_2} < l\frac{r_i + \beta_1}{r_i + \beta_1 + \beta_2} < \delta$$

and

$$r_i < k_{l-1} + r_i + \beta_1 + \beta_2 < l(r_i + \beta_1 + \beta_2) < r_{i+1}.$$

If no such index l exists then

$$\bar{z}_{k_L + r_i + \beta_1 + \beta_2} < (L+1)\frac{r_i + \beta_1}{r_i + \beta_1 + \beta_2} < \delta$$

and

$$r_i < k_L + r_i + \beta_1 + \beta_2 < (L+1)(r_i\beta_1 + \beta_2) < r_{i+1}.$$

In both situations we have found an index m between r_i and r_{i+1} s.t. $\bar{z}_m \le \delta$. From the observations (a) and (b) we conclude that $z \in D_i$. Therefore $\mathbf{z} \in E$. □

REFERENCES

DUTTA, P. and HORN, P. H. (1981). Low frequency fluctuation in solids: $1/f$ noise. *Rev. Mod. Phys.* **53** 497–516.

GRIZE, Y. L. (1984). Towards a stationary continuous lower probability-based model for flicker noise. Ph.D. thesis, Cornell University, Ithaca, N.Y.

HALMOS, P. R. (1974). *Measure Theory*. Springer, New York.

HOWE, A. D., ALLAN, D. W. and BARNES, T. A. (1981). Properties of signal sources and measurement methods. *Proc. 35th Annual Freq. Control Symposium*. USAERADCOM, Ft. Monmouth, N.J.

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

KALIKOW, S. (1984). Private communication.

KESHNER, M. S. (1982). $1/f$ noise. *Proc. IEEE* **70** 212–218.

KROUPA, V. (1983). *Frequency Stability Fundamentals and Measurement.* IEEE Press, New York.

KUMAR, A. and FINE, T. L. (1985). Stationary lower probabilities and unstable averages. *Z. Wahrsch. verw. Gebiete* **69** 1–17.

LOÈVE, M. (1977). *Probability Theory* **1**. Springer, New York.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

PAPAMARCOU, A. (1983). Some results on undominated lower probabilities. M.S. thesis, Cornell University, Ithaca, N.Y.

PAPAMARCOU, A. and FINE, T. L. (1986). A note on undominated lower probabilities. *Ann. Probab.* **14** 710–723.

WALLEY, P. and FINE, T. L. (1982). Towards a frequentist theory of upper and lower probabilities. *Ann. Statist.* **10** 741–761.

WALSH, J. B. (1981). A stochastic model of neural response. *Adv. Appl. Probab.* **13** 231–281.

CIBA-GEIGY
STATISTICS DEPARTMENT
CH-4002 BASEL
SWITZERLAND

SCHOOL OF ELECTRICAL ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853