The Annals of Statistics
1994, Vol. 22, No. 4, 1840–1870

FINITE-DIMENSIONAL DISTRIBUTIONS AND TAIL BEHAVIOR IN STATIONARY INTERVAL-VALUED PROBABILITY MODELS[1]

BY AMIR SADROLHEFAZI AND TERRENCE L. FINE

Morgan Stanley & Co. and Cornell University

We consider the relationship between the finite-dimensional distributions of a stationary time series model and its asymptotic behavior in the framework of interval-valued probability (IVP), a simple generalization of additive probability measures. By Caratheodory's theorem, the specification of a countably additive probability measure on the algebra of cylinders $\mathcal{C}$ uniquely defines its behavior on $\sigma(\mathcal{C})$ (containing the tail events). If the measure is stationary, then the ergodic theorem indicates that its extension assigns zero probability to the tail event consisting of all sequences for which the time averages diverge (the divergence event). This link between the marginals and the tail behavior is no longer valid in IVP, and we can reconcile arbitrary finite-dimensional distributions and tail behavior for stationary IVP-based models.

The linking mechanism between the marginals of a time series model and its asymptotic behavior turns out to be continuity, not stationarity or even additivity. We prove that any stationary finitely additive probability (charge) defined on cylinders has a stationary charge extension that can assign the divergence event any prescribed probability. Moreover, on the space of binary sequences, we consider IVP models that incorporate:

(i) Stationarity.
(ii) Continuity along $\mathcal{C}$.
(iii) Almost sure support for divergence.
(iv) Estimability of the divergence event from cylinders.
(v) Nearly additive finite-dimensional distributions.

We enhance the previous constructions of IVP's satisfying (i)–(iv) so that they satisfy (v) by agreeing with a stationary measure either exactly on one-dimensional cylinders or arbitrarily closely on a given class of bounded-dimensional cylinders.

Our time series constructions follow from the observation that the algebra of cylinders and the tail σ-algebra are mutually nonsingular. We use the same idea to prove the existence of joints for general marginal IVP's. These constructions have implications for frequentist interpretations of probability.

1. Introduction. This paper is the culmination of a series of papers by Grize (1987), Kumar (1985), Papamarcou (1986, 1991), and Walley (1982), written in association with Fine, that have the goal of employing lower probability

Received April 1992; revised November 1993.
[1]Prepared with partial support from NSF Grant DIR-89-09873.
AMS 1991 *subject classifications*. Primary 62M10; secondary 60G05.
Key words and phrases. Interval-valued probability, upper and lower probability, time series models.

to extend the scope of frequentist-based, probability-like modeling of nondeterministic phenomena. Current probabilistic methodology imposes constraints on our ability to describe reality and such constraints should come from the empirical disciplines concerned with the phenomena and not from the mathematical methodology. To be specific, we have been concerned with the implication of the stationarity convergence theorem that time series having bounded random variables and that are deemed to be stationary on the basis of observations and theoretical considerations (e.g., physical theory leading us to believe that the system generating the time series is time invariant) must perforce be assumed to have convergent time averages. There is, for example, empirical evidence in the study of the so-called flicker or $1/f$ noises that these processes are stationary but possess unstable long-term time averages. Hence, they cannot be accommodated within conventional probability theory without some contradiction of the empirical and theoretical evidence. The lower probability-based theory developed in these papers allows us to reconcile stationarity with divergent time averages or, better yet, to remain noncommittal about this tail event.

In order to introduce our time series models based upon interval-valued probability and having the ability to model phenomena excluded from consideration by standard probability, we need to develop first the fundamentals of specification of an interval-valued probability on a (infinite) product space. In particular, we wish to render compatible certain properties of lower-order joint experiments ("marginals"), having operationally observable outcomes, with assertions about not directly observable properties of asymptotically determined tail events. In Section 2 we introduce the "kernel" representation of a lower probability and a notion of "nonsingularity" that enables us in Section 3 to provide a complete answer to the question of when marginal lower probability experiments admit of a joint lower probability on the product space. Of course, in conventional probability theory the existence of a joint experiment is trivially answered with the product/independence construction. We then proceed, in Sections 4 and 5, to construct stationary time series having bounded coordinate random variables. The advantage of the construction provided here over that presented in previous work is that we can now reconcile support for asymptotic divergence with an ability to arbitrarily closely approximate any given conventional probability measure on a collection of cylinder sets of bounded dimension (e.g., Proposition 5.8). In effect, lower probability time series models allow us to closely approximate conventional probability on given collections of cylinder sets while still preserving our modeling flexibility insofar as certain tail events are concerned. Issues that should be left to the scientific disciplines concerned with the particular nondeterministic phenomena are no longer settled by probabilistic methodology. Furthermore, these results have implications for frequentist interpretations of probability in that they open the possibility that a frequentist account of probability might be made for interval-valued probability that would avoid unsubstantiated commitments to long-run convergence.

The theory of interval-valued probability (IVP) is a simple generalization of the theory of additive probabilities in which the likelihood of an event is associated with a subinterval of [0, 1]. The lower and upper endpoints of the interval are called the *lower* and *upper probability* and denoted by $\underline{P}$ and $\overline{P}$, respectively. In contrast to additive probabilities, the domain of definition of $\underline{P}$, $\overline{P}$ is not critical and they can in general be defined on an arbitrary class of events. However, for the purpose of comparison with additive probabilities, we will define them on an algebra. The formal definition of IVP is as follows.

1.1. DEFINITION. Given an algebra $\mathcal{A}$ of subsets of Ω, IVP $(\underline{P}, \overline{P})$ is a pair of conjugate real-valued set functions on $\mathcal{A}$ related by $\underline{P}(A) + \overline{P}(A^c) = 1$ and satisfying:

(IVP1) Normalization: $\underline{P}(\Omega) = 1$.
(IVP2) Nonnegativity: $\underline{P} \geq 0$.
(IVP3) Superadditivity of $\underline{P}$: $(\forall A, B \in \mathcal{A}, A \perp B)$ $\underline{P}(A + B) \geq \underline{P}(A) + \underline{P}(B)$.
(IVP4) Subadditivity of $\overline{P}$: $(\forall A, B \in \mathcal{A}, A \perp B)$ $\overline{P}(A + B) \leq \overline{P}(A) + \overline{P}(B)$.

It follows easily from the definition that $\underline{P}(\emptyset) = \overline{P}(\emptyset) = 0$, and $\underline{P}$, $\overline{P}$ are monotone. Also, $0 \leq \underline{P} \leq \overline{P} \leq 1$, hence justifying the term *upper and lower probability*. A finitely additive probability (a *charge*) π is a degenerate case of IVP with $\underline{P} = \overline{P} = \pi$, while a countably additive (a *measure*) μ is a charge with the additional continuity condition

$$(\forall A_n \in \mathcal{A}) \qquad A_n \downarrow \emptyset \quad \Rightarrow \quad \mu(A_n) \downarrow 0.$$

A nonobvious consequence of the axioms is the following:

$$(\forall A, B \in \mathcal{A}) \qquad \underline{P}(A) + \underline{P}(B) \leq \underline{P}(AB) + \overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B).$$

Using this property and the fact that $\underline{P}(A) + \overline{P}(A') = 1$, we can rewrite axiom IVP4 in terms of $\underline{P}$, and arrive at the following equivalent axiom:

(IVP4′) Conjugacy: $(\forall A, B \in \mathcal{A})$ $1 + \underline{P}(AB) \geq \underline{P}(A) + \underline{P}(B)$.

We take axioms IVP1–IVP3 and IVP4′ as the definition of IVP, and work solely with lower probabilities (LP's), implicitly understanding their conjugate upper probabilities. The triple $(\Omega, \mathcal{A}, \underline{P})$ is referred to as a *lower probability (LP) space*.

1.2. PROPOSITION. *The family of all LP's on an algebra $\mathcal{A}$ is closed under forming*:

(i) *Arbitrary setwise infima*: $\inf_\alpha \underline{P}_\alpha$.
(ii) *Finite or countable convex combinations*: $\sum_i \lambda_i \underline{P}_i$ *where* $\lambda_i > 0$, $\sum_i \lambda_i = 1$.
(iii) *Finite setwise multiplications*: $\Pi_{i=1}^n \underline{P}_i$,
(iv) *Minimal combinations*: $(\underline{P}_1 + \underline{P}_2 - 1)^+$.
(v) *Scales*: *Given on LP* $\underline{P}$, *and* $\alpha \in [0, 1]$, *form the "scaled" version* $\underline{P}^{(\alpha)}$ *by*

$$(\forall A \in \mathcal{A},\ A \neq \Omega) \qquad \underline{P}^{(\alpha)}(A) = \alpha \underline{P}(A).$$

PROOF. Routine verification. □

Property 1.2(i) is commonly used to generate LP's from charges and measures. As noted before, a charge is an LP, therefore, the setwise infimum of a family of charges is a (nonadditive) LP. This method of generating LP's from charges gives rise to a class of LP's, the *lower envelopes*, that have been the main object of study in Walley (1981, 1991). However, not every LP can be generated in this manner, and there is an important class of LP's—the *undominated* LP's—that have no relationship to additive probabilities.

1.3. DEFINITION. Given an algebra $\mathcal{A}$ of subsets of Ω, let $\mathcal{P}(\mathcal{A})$ denote the family of all charges defined on $\mathcal{A}$. Given an LP $\underline{P}$ on $\mathcal{A}$, its family of *dominating charges*, $\mathcal{M}(\underline{P})$, is defined as

$$\mathcal{M}(\underline{P}) \equiv \{\pi \in \mathcal{P}(\mathcal{A}): \pi(A) \geq \underline{P}(A) \text{ for all } A \in \mathcal{A}\}.$$

Since every charge on an algebra has a charge extension to 2^{Ω} [Rao and Rao (1983), Corollary 3.3.5], we can drop the reference to the algebra $\mathcal{A}$ and only consider $\mathcal{P} \equiv \mathcal{P}(2^{\Omega})$.

If $\mathcal{M}(\underline{P}) = \emptyset$, $\underline{P}$ is said to be *undominated*, *dominated* otherwise. A dominated LP $\underline{P}$ is called a *lower envelope* if

$$(\forall A \in \mathcal{A}) \qquad \underline{P}(A) = \inf\{\pi(A): \pi \in \mathcal{M}(\underline{P})\}.$$

Its corresponding upper probability $\overline{P}$ is the *upper envelope* of $\mathcal{M}(\underline{P})$.

We note that there exist dominated LP's that are not lower envelopes; for an example see Walley (1981), page 35. An LP $\underline{P}$ is said to be *vacuous* on a *nontrivial* set A—a set other than the trivial sets $\emptyset$, Ω—if $\underline{P}(A) = 0$, $\overline{P}(A) = 1$. The *vacuous LP* $\underline{P}$ is defined to be the LP that is vacuous on all nontrivial sets:

$$(\forall A \in \mathcal{A}^{+} \equiv \mathcal{A}\backslash\{\emptyset, \Omega\}) \qquad \underline{P}(A) = 0, \qquad \overline{P}(A) = 1.$$

An equivalent definition is an LP $\underline{P}$ for which $\mathcal{M}(\underline{P}) = \mathcal{P}$. The vacuous LP is a good model for "total ignorance," since it is noncommittal on all events except the trivial ones.

Different classes of LP's (measures, capacities, belief functions, etc.) can be obtained by imposing various regularity conditions on $\underline{P}$ and by endowing $\mathcal{A}$ with topological structures; see Choquet (1953), Dempster (1967), Shafer (1976) and Anger (1977).

2. Kernels. As noted before, the domain of an LP need not be an algebra. The following extension procedure is standard in the literature of LP's.

2.1. MINIMAL EXTENSION THEOREM. *On a class $\mathcal{A}$ of subsets of Ω containing $\emptyset$, Ω and closed under finite intersections and finite disjoint unions, let $\underline{P}_0$ be a monotone real-valued set function satisfying the axioms of an LP. Then $\underline{P}_0$ can be extended to an LP $\underline{P}$ on 2^{Ω} given by*

$$(\forall A \subset \Omega) \qquad \underline{P}(A) = \sup\{\underline{P}_0(B): B \subset A,\ B \in \mathcal{A}\}.$$

Moreover, $\underline{P}$ is the minimal such extension; that is, if $\underline{Q}$ is another LP on 2^Ω such that $\underline{Q} = \underline{P}_0$ on $\mathcal{A}$, then $\underline{Q} \geq \underline{P}$ on 2^Ω.

Therefore an LP defined on an algebra can be extended to an LP on the power set. Conversely, we observe that the restriction of an LP (defined on the power set) to any algebra is still an LP. From now on, unless specifically noted, *we will assume that our LP's are defined on the power set*. In this case, we will also omit 2^Ω and call the pair $(\Omega, \underline{P})$ an LP space.

The above extension procedure suggests other ways of generating LP's from a small class of sets. For this, we need a set function defined on a class of events that can approximate an LP from below. This gives rise to the notion of a *kernel*, a "concise" description of an LP.

2.2. DEFINITION. A *kernel* $(\mathcal{K}, \rho)$ is a class $\mathcal{K}$ of subsets of Ω, where $\emptyset, \Omega \in \mathcal{K}$, together with a monotone real-valued set function ρ on $\mathcal{K}$, that satisfy:

(i) Normalization: $\rho(\Omega) = 1$.
(ii) Nonnegativity: $\rho \geq 0$.
(iii) Closure under superadditivity: $(\forall A, B \in \mathcal{K}, A \perp B)$

$$\sup\{\rho(C) \colon C \subset A + B, C \in \mathcal{K}\} \geq \rho(A) + \rho(B).$$

(iv) Closure under conjugacy: $(\forall A, B \in \mathcal{K})$

$$1 + \sup\{\rho(C) \colon C \subset AB,\ C \in \mathcal{K}\} \geq \rho(A) + \rho(B).$$

We note that closure under superadditivity (conjugacy) is a weaker notion than superadditivity (conjugacy), since it involves both the set function ρ *together* with its domain $\mathcal{K}$—for example, $\mathcal{K}$ need not be closed under intersections or disjoint unions. Note also that it follows from the definition that $\rho(\emptyset) = 0$.

The following proposition indicates how the kernels can be used to construct LP's.

2.3. PROPOSITION. *Given a kernel $(\mathcal{K}, \rho)$, define $\underline{P}$ as follows:*

$$(1) \qquad (\forall A \subset \Omega) \qquad \underline{P}(A) = \sup\{\rho(B) \colon B \subset A,\ B \in \mathcal{K}\}.$$

Then $\underline{P}$ is the minimal LP extension of ρ to 2^Ω. Conversely, if for a given LP $\underline{P}$, (1) *holds for some pair $(\mathcal{K}, \rho)$ where $\emptyset, \Omega \in \mathcal{K}$, and ρ is monotone, then $(\mathcal{K}, \rho)$ is a kernel.*

The LP so constructed is said to be *generated* by the kernel $(\mathcal{K}, \rho)$, or we say that $(\mathcal{K}, \rho)$ *generates* $\underline{P}$.

PROOF. Clearly, $\underline{P}(\Omega) = 1$, and $\underline{P} \geq 0$. Given two disjoint sets A, B and $\varepsilon > 0$, we can find two disjoint $\mathcal{K}$-sets K_1, K_2, where $K_1 \subset A$, $K_2 \subset B$, and

$$\underline{P}(A) \leq \rho(K_1) + \varepsilon, \qquad \underline{P}(B) \leq \rho(K_2) + \varepsilon.$$

By definition, there is a $\mathcal{K}$-set C where $C \subset K_1 + K_2 \subset A + B$, and

$$\rho(C) \geq \rho(K_1) + \rho(K_2) - \varepsilon.$$

Hence $\underline{P}(A + B) \geq \underline{P}(A) + \underline{P}(B) - 3\varepsilon$. Since ε was arbitrary, this show that $\underline{P}$ is superadditive.

The proof of conjugacy is similar: Given two sets A, B and $\varepsilon > 0$, there are $\mathcal{K}$-sets K_1, K_2 such that $K_1 \subset A$, $K_2 \subset B$, and

$$\underline{P}(A) \leq \rho(K_1) + \varepsilon, \qquad \underline{P}(B) \leq \rho(K_2) + \varepsilon.$$

Again by definition, there exists a $\mathcal{K}$-set C where $C \subset K_1K_2$,

$$1 + \rho(C) \geq \rho(K_1) + \rho(K_2) - \varepsilon.$$

Hence $1 + \underline{P}(AB) \geq \underline{P}(A) + \underline{P}(B) - 3\varepsilon$, thereby establishing conjugacy. Since ρ is monotone, $\underline{P} = \rho$ on $\mathcal{K}$. The minimality of $\underline{P}$ comes from the fact that every LP is monotone.

For the converse, assume that (1) holds for some monotone set function ρ defined on a class $\mathcal{K}$. Nonnegativity and normalization follow from the corresponding properties of $\underline{P}$. Given disjoint $\mathcal{K}$-sets A, B, we have:

$$\begin{aligned} \rho(A) + \rho(B) &= \underline{P}(A) + \underline{P}(B) \\ &\leq \underline{P}(A + B) \\ &= \sup\{\rho(C)\colon C \subset A + B, C \in \mathcal{K}\}, \end{aligned}$$

since by monotonicity of ρ, we have $\underline{P} = \rho$ on $\mathcal{K}$. A similar argument shows that $(\mathcal{K}, \rho)$ is closed under conjugacy. □

While every kernel generates a unique LP, an LP may be generated by different kernels. For example, for a given kernel $(\mathcal{K}, \rho)$, if we strip away from $\mathcal{K}$ all nonempty sets for which ρ is 0, the truncated kernel generates the same LP as the original. Similarly, if we augment $\mathcal{K}$ by arbitrary sets, we still end up with the same LP. The importance of kernels lies in the flexibility that they afford us in constructing an LP. Our goal is to use this flexibility in a judicious way to provide tractable kernels.

Trivially, given an LP space (Ω, P), we can set $\mathcal{K} = 2^{\Omega}$ and $\rho = \underline{P}$. Then $(\mathcal{K}, \rho)$ generates $\underline{P}$. In this case, the kernel is called the *maximal kernel*. A charge (defined on a finite space) that assigns positive mass to all singletons can only be generated by the maximal kernel. On the other hand, the *trivial kernel* $(\mathcal{K}, \rho)$ defined as

$$\mathcal{K} = \{\emptyset, \Omega\}, \qquad \rho(\emptyset) = 0, \qquad \rho(\Omega) = 1,$$

generates the vacuous LP. Other nontrivial examples of kernels follow.

2.4. NOTATION. Given classes $\mathcal{A}, \mathcal{B}$ of subsets of Ω, $\mathcal{A} \odot \mathcal{B}$ denotes the class of sets that are obtained by intersecting an $\mathcal{A}$-set and a $\mathcal{B}$-set: $\mathcal{A} \odot \mathcal{B} \equiv \{AB\colon A$

$\in \mathcal{A}$, $B \in \mathcal{B}\}$. When $\mathcal{A} = \mathcal{B}$, we write $\mathcal{A}^{(2)} = \mathcal{A} \odot \mathcal{A}$, and we define $\mathcal{A}^{(n)} \equiv \bigodot_{i=1}^{n} \mathcal{A}$ for $n \geq 1$, $\mathcal{A}^{(0)} \equiv \{\Omega\}$. By $\mathcal{A}^{(c)}$ we mean the class of all complements of $\mathcal{A}$-sets: $\mathcal{A}^{(c)} \equiv \{A^{(c)}: A \in \mathcal{A}\}$.

2.5. Example. Given a nontrivial subset A of Ω and a number $p \in [0,1]$, let $\mathcal{K} = \{\emptyset, A, \Omega\}$, and define ρ by $\rho(\emptyset) = 0$, $\rho(A) = p$ and $\rho(\Omega) = 1$. The kernel $(\mathcal{K}, \rho)$ generates an LP $\underline{P}$ where $\underline{P}(A) = p$.

2.6. Example [Grize and Fine (1987), Lemma 5]. Let $\mathcal{A}$ be a class of subsets of Ω where for some integer $n \geq 1$, the intersection of any $2n$ sets is nonempty, that is, $\emptyset \notin \mathcal{A}^{(2n)}$. Let $\mathcal{K} = \{\emptyset, \Omega\} \cup \mathcal{A}^{(n)}$, and define ρ by $\rho(\emptyset) = 0$,

$$(\forall A \in \mathcal{K}) \qquad \rho(A) = \sup\left\{1 - \frac{m}{n+1}: B \subset A, B \in \mathcal{A}^{(m)}, 0 \leq m \leq n\right\}.$$

Then $(\mathcal{K}, \rho)$ is a kernel. The LP $\underline{P}$ generated by this kernel satisfies

$$(\forall m \geq 0)\big(\forall A \in \mathcal{A}^{(m)}\big) \qquad \underline{P}(A) \geq 1 - \frac{m}{n+1},$$

or consequently for every $\mathcal{A}$-set A, $\underline{P}(A) \geq 1 - 1/(n+1)$.

3. Common extensions and joint LP's. We consider the following problem in this section: given two LP's $\underline{P}_1, \underline{P}_2$ on a space Ω and two classes $\mathcal{A}_1, \mathcal{A}_2$ of subsets of Ω, under what conditions can we construct an LP $\underline{P}$ that extends (to the power set) both $\underline{P}_1$ from $\mathcal{A}_1$ and $\underline{P}_2$ from $\mathcal{A}_2$? The key property for this turns out to be the following.

3.1. Definition. Two classes $\mathcal{A}, \mathcal{B}$ of subsets of Ω are said to be mutually *nonsingular*, $\mathcal{A}!\mathcal{B}$, if

$$(\forall A \in \mathcal{A}, B \in \mathcal{B}) \qquad AB = \emptyset \quad \Rightarrow \quad A = \emptyset \text{ or } B = \emptyset,$$

or equivalently $\emptyset \notin \mathcal{A}^{+} \odot \mathcal{B}^{+}$.

3.2. Proposition. *Let $\mathcal{K}_1, \mathcal{K}_2$ be two class of subsets of Ω that contain $\emptyset$, Ω. On $\mathcal{K}_1 \times \mathcal{K}_2$ define σ by $\sigma(K_1, K_2) = K_1 K_2$, and let $\mathcal{K}$ denote the range of σ: $\mathcal{K} = \mathcal{K}_1 \odot \mathcal{K}_2$. If $(\mathcal{K}_1 \cup \mathcal{K}_1^{(c)})^{(2)}!\mathcal{K}_2^{(2)}$, we have:*

(i) *If $\emptyset \neq \sigma(A_1, A_2) \subset \sigma(B_1, B_2)$, then $A_1 \subset B_1$. Hence*

$$\emptyset \neq \sigma(A_1, A_2) = \sigma(B_1, B_2) \quad \Rightarrow \quad A_1 = B_1.$$

(ii) *If $\sigma(A_1, A_2) \perp \sigma(B_1, B_2)$, then $A_1 \perp B_1$ or $A_2 \perp B_2$.*

Moreover, if $\mathcal{K}_1^{(2)}$ and $(\mathcal{K}_2 \cup \mathcal{K}_2^{(c)})^{(2)}$ are also nonsingular, then:

(iii) *σ^{-1} exists on $\mathcal{K} - \{\emptyset\}$.*

PROOF. We will only prove (i), since (ii) follows from definition, and (i) implies (iii).

(i) We have $\emptyset \neq A_1A_2 \subset B_1B_2$. Intersect both sides by $A_1 \setminus B_1$ to get $A_1B_1^c A_2 = \emptyset$. Note that $A_1B_1^c \in (\mathcal{K}_1 \cup \mathcal{K}_1^{(c)})^{(2)}$ and $A_2 \neq \emptyset$. By the nonsingularity assumption, $A_1B_1^c = \emptyset$ or $A_1 \subset B_1$. □

The following proposition indicates that nonsingularity is the essential requirement for the construction of common extensions or common pseudo-extensions—LP's that dominate both of the original LP's—for two given LP's.

3.3. MINIMAL COMMON EXTENSION. *Let $(\mathcal{K}_1, \rho_1)$, $(\mathcal{K}_2, \rho_2)$ be given kernels on Ω, and let $\underline{P}_1, \underline{P}_2$ be their respective generated LP's. On $\mathcal{K}_1 \times \mathcal{K}_2$, define σ, η by*

$$\sigma(K_1, K_2) \equiv K_1K_2,$$
$$\eta(K_1, K_2) \equiv \big(\rho_1(K_1) + \rho_2(K_2) - 1\big)^+.$$

Let $\mathcal{K}$ be the range of σ, that is, $\mathcal{K} = \mathcal{K}_1 \odot \mathcal{K}_2$, and on $\mathcal{K}$, define ρ by

$$\rho(A) \equiv \sup\{\eta(A_1, A_2): \sigma(A_1, A_2) \subset A,\ (A_1, A_2) \in \mathcal{K}_1 \times \mathcal{K}_2\}. \tag{2}$$

(i) *If $\mathcal{K}_1^{(2)} ! \mathcal{K}_2^{(2)}$, then $(\mathcal{K}, \rho)$ is a kernel and $\rho \geq \rho_i$ on $\mathcal{K}_i$. The generated LP $\underline{P}$ is the minimal common pseudo-extension of $\underline{P}_i$'s.*

(ii) *If $(\mathcal{K}_1 \cup \mathcal{K}_1^{(c)})^{(2)} ! \mathcal{K}_2^{(2)}$, then $(\mathcal{K}, \rho)$ is a kernel, $\rho = \rho_1$ on $\mathcal{K}_1$ and $\rho \geq \rho_2$ on $\mathcal{K}_2$. The generated LP $\underline{P}$ is the minimal pseudo-extension of $\underline{P}_2$ that equals $\underline{P}_1$ on $\mathcal{K}_1$.*

(iii) *If $(\mathcal{K}_1 \cup \mathcal{K}_1^{(c)})^{(2)} ! \mathcal{K}_2^{(2)}$ and $\mathcal{K}_1^{(2)} ! (\mathcal{K}_2 \cup \mathcal{K}_2^{(c)})^{(2)}$, then* (2) *simplifies to*

$$\rho\big(\sigma(A_1, A_2)\big) = \big(\rho_1(A_1) + \rho_2(A_2) - 1\big)^+. \tag{3}$$

In this case, $\rho = \rho_i$ on $\mathcal{K}_i$ for $i = 1, 2$, and the generated LP $\underline{P}$ is the minimal common extension of $\underline{P}_1, \underline{P}_2$ from $\mathcal{K}_1, \mathcal{K}_2$.

PROOF. (i) Since $\mathcal{K}_1 \cup \mathcal{K}_2 \subset \mathcal{K}$, $\emptyset, \Omega \in \mathcal{K}$. Clearly, $0 \leq \rho \leq 1$, $\rho(\Omega) = 1$, and since $A_1 = \sigma(A_1, \Omega_2)$, $A_2 = \sigma(\Omega_1, A_2)$ for all $(A_1 \times A_2) \in \mathcal{K}_1 \times \mathcal{K}_2$, $\rho \geq \rho_i$ on $\mathcal{K}_i$. The proof of closure under conjugacy being straightforward, we will just prove closure under superadditivity. For this, we only have to consider disjoint $\mathcal{K}$-sets A, B where $\rho(A)\rho(B) > 0$. In this case, for any $\varepsilon > 0$, we can find $(A_1, A_2), (B_1, B_2) \in \mathcal{K}_1 \times \mathcal{K}_2$ such that

$$\sigma(A_1, A_2) \subset A, \qquad \sigma(B_1, B_2) \subset B,$$

and

$$0 < \eta(A_1, A_2) \leq \rho(A) \leq \eta(A_1, A_2) + \varepsilon,$$
$$0 < \eta(B_1, B_2) \leq \rho(B) \leq \eta(B_1, B_2) + \varepsilon.$$

Since $A \perp B$, $\sigma(A_1, A_2) \perp \sigma(B_1, B_2)$, which by the nonsingularity assumption implies $A_1 \perp A_2$ or $B_1 \perp B_2$. Assume $A_1 \perp A_2$. Since $(\mathcal{K}_i, \rho_i)$'s are kernels, we can find $C_1 \in \mathcal{K}_1$, $C_2 \in \mathcal{K}_2$ such that

$$C_1 \subset A_1 + B_1, \qquad \rho_1(C_1) \geq \rho_1(A_1) + \rho_1(B_1) - \varepsilon,$$
$$C_2 \subset A_2 B_2, \quad 1 + \rho_2(C_2) \geq \rho_2(A_2) + \rho_2(B_2) - \varepsilon.$$

Let $C \equiv \sigma(C_1, C_2)$. Then $C \subset (A_1 + B_1)A_2B_2 \subset AB$, and

$$\begin{aligned}\rho(C) &\geq \eta(C_1, C_2)\\ &\geq \eta(A_1, A_2) + \eta(B_1, B_2) - 1 - 2\varepsilon\\ &\geq \rho(A) + \rho(B) - 4\varepsilon.\end{aligned}$$

Hence

$$\sup\{\rho(C)\colon C \subset A + B, C \in \mathcal{K}\} \geq \rho(A) + \rho(B) - 4\varepsilon.$$

Since ε was arbitrary, this proves the closure under superadditivity.

By the conjugacy and nonnegativity axioms, any LP that dominates $\underline{P}_i$ on $\mathcal{K}_i$ must dominate ρ on $\mathcal{K}$. Since $\rho \geq \rho_i$ on $\mathcal{K}_i$ the generated LP $\underline{P}$ is the *minimal* LP that dominates both $\underline{P}_1$ and $\underline{P}_2$.

(ii) The only new assertion here is that $\rho = \rho_1$ on $\mathcal{K}_1$—the rest of the assertions follow from part (i). We have already shown that $\rho \geq \rho_1$, hence it suffices to show $\rho \leq \rho_1$ on $\mathcal{K}_1$. Given $A_1 \in \mathcal{K}_1$, let $\sigma(B_1, B_2) \subset A_1$ for some $(B_1, B_2) \in \mathcal{K}_1 \times \mathcal{K}_2$. If $\eta(B_1, B_2) = 0$, we are all set. Otherwise, $\eta(B_1, B_2) > 0$ implies

$$\begin{aligned}\rho_1(B_1) + \rho_2(B_2) - 1 &> 0\\ \Rightarrow \quad B_1, B_2 &\neq \emptyset\\ \Rightarrow \quad \sigma(B_1, B_2) &\neq \emptyset,\end{aligned}$$

by the nonsingularity condition. We have

$$\begin{aligned}\emptyset \neq \sigma(B_1, B_2) &\subset \sigma(A_1, \Omega_2) = A_1\\ \Rightarrow \quad B_1 &\subset A_1,\end{aligned} \tag{4}$$

by Proposition 3.2(i). Since ρ_1 is monotone, (4) implies $\eta(B_1, B_2) \leq \rho_1(A_1)$. Therefore

$$\rho(A_1) = \{\eta(B_1, B_2)\colon \sigma(B_1, B_2) \subset A_1\} \leq \rho_1(A_1).$$

(iii) We just have to show that (3) is valid; the rest of the assertions follow from part (ii). By the nonsingularity assumptions and Proposition 3.2(iii), σ is invertible on $\mathcal{K} - \{\emptyset\}$; hence (3) is well defined for all nonempty $\mathcal{K}$-sets. Furthermore, if $\sigma(A_1, A_2) = \emptyset$, then $A_1 = \emptyset$ or $A_2 = \emptyset$. Therefore, for all different representations of the empty set, ρ as defined in (3) is 0. □

We find the first application of Proposition 3.3 in proving the existence of joint LP's on product spaces. Given two LP spaces $(\Omega_1, \underline{P}_1)$, $(\Omega_2, \underline{P}_2)$, we are interested in forming a joint LP $\underline{P}$ on $\Omega = \Omega_1 \times \Omega_2$ that preserves the marginals

$$(\forall A_1 \subset \Omega_1, A_2 \subset \Omega_2) \qquad \underline{P}(A_1 \times \Omega_2) = \underline{P}_1(A_1), \qquad \underline{P}(\Omega_1 \times A_2) = \underline{P}_2(A_2).$$

While the existence of joint LP's is immediate in additive probability (product measures), the absence of adaptivity in LP structures significantly complicates this issue. Previous attempts [e.g., Papamarcou (1983) and Papamarcou and Fine (1986)] have only partially solved this for certain classes of marginal LP's, and the existence of a joint for general marginal LP's has been an open question.

The question of finding a joint LP is equivalent to finding a common extension for $(\Omega_1 \times \Omega_2, \mathcal{A}_1, \underline{Q}_1)$, $(\Omega_1 \times \Omega_2, \mathcal{A}_2, \underline{Q}_2)$, where

$$\mathcal{A}_1 = \{A_1 \times \Omega_2 : A_1 \subset \Omega_1\}, \qquad \mathcal{A}_2 = \{\Omega_1 \times A_2 : A_2 \subset \Omega_2\},$$
$$\underline{Q}_1(A_1 \times \Omega_2) = \underline{P}_1(A_1), \qquad \underline{Q}_2(\Omega_1 \times A_2) = \underline{P}_2(A_2).$$

In Proposition 3.2, we observe that if $\mathcal{K}_1, \mathcal{K}_2$ are algebras, then the condition for the existence of σ^{-1} is simply the nonsingularity of $\mathcal{K}_1$ and $\mathcal{K}_2$. Since $\mathcal{A}_1, \mathcal{A}_2$ are nonsingular algebras, Proposition 3.3(iii) completely resolves the joint question by constructing the joint LP for arbitrary marginals. This construction has the additional desirable property of being the minimal such joint LP.

The next proposition is essentially Proposition 3.3(iii) applied to the joint problem for general product spaces. Given LP spaces $(\Omega_\alpha, \underline{P}_\alpha)$, where α's can belong to an arbitrary index set I, let $\Omega = \Pi_\alpha \Omega_\alpha$. By a rectangle, we mean the Cartesian product of subsets of Ω_α's; that is, $\Pi_\alpha A_\alpha$ where $A_\alpha \subset \Omega_\alpha$. A *finite rectangle* is a rectangle $\Pi_\alpha A_\alpha$ where $A_\alpha = \Omega_\alpha$ for all but finitely many α's. We let $\mathcal{R}_f$ denote the class of all finite rectangles, and note the following properties of rectangles:

(i) If $\emptyset \neq \Pi_\alpha A_\alpha \subset \Pi_\alpha B_\alpha$, then $A_\alpha \subset B_\alpha$ for all $\alpha \in I$. Hence

$$\emptyset \neq \prod_\alpha A_\alpha = \prod_\alpha B_\alpha \quad \Rightarrow \quad A_\alpha = B_\alpha \quad \text{for all } \alpha \in I. \tag{5}$$

(ii) If $(\Pi_\alpha A_\alpha) \perp (\Pi_\alpha B_\alpha)$, then $A_\alpha \perp B_\alpha$ for some $\alpha \in I$.

(iii) $(\Pi_\alpha A_\alpha) \cap (\Pi_{\alpha}, B_\alpha) = \Pi_\alpha A_\alpha B_\alpha$, hence the class of (finite) rectangles is closed under (finite) intersections.

3.4. MINIMAL JOINT LP. *Given LP spaces $(\Omega_\alpha, \underline{P}_\alpha)$, let $\Omega = \Pi_\alpha \Omega_\alpha$, and set $\mathcal{K} = \mathcal{R}_f$, the class of all finite rectangles. On $\mathcal{K}$ define ρ by*

$$\rho\left(\prod_\alpha A_\alpha\right) \equiv \left(1 + \sum_\alpha (\underline{P}_\alpha(A_\alpha) - 1)\right)^+. \tag{6}$$

Then $(\mathcal{K}, \rho)$ is a kernel, and ρ preserves the marginals:

$$(\forall \beta)(\forall A_\beta \subset \Omega_\beta) \qquad \underline{P}\left(\prod_\alpha A_\alpha\right) = \underline{P}_\beta(A_\beta),$$

where $A_\alpha = \Omega_\alpha$ for all $\alpha \neq \beta$. Moreover, it generates the minimal LP on Ω that preserves the marginals.

PROOF. Since we are only considering *finite* rectangles, there are only finitely many nonzero terms in the sum in (6), and the sum is well defined. Also, (5) indicates that ρ is unambiguously defined, since for all different representations of the empty set, ρ is 0. Clearly, $0 \leq \rho \leq 1$, $\rho(\Omega) = 1$ and ρ preserves the marginals. For conjugacy, we only have to consider rectangles $A = \Pi_\alpha A_\alpha$, $B = \Pi_\alpha A_\alpha$ where $\rho(A) + \rho(B) > 1$, hence $\rho(A) > 0$, $\rho(B) > 0$. We have

$$\begin{aligned}
\rho(A) + \rho(B) &= \Big[1 + \sum_\alpha (\underline{P}_\alpha(A_\alpha) - 1)\Big] + \Big[1 + \sum_\alpha (\underline{P}_\alpha(B_\alpha) - 1)\Big] \\
&= 1 + 1 + \sum_\alpha \Big[(\underline{P}_\alpha(A_\alpha) + \underline{P}_\alpha(B_\alpha) - 1) - 1\Big] \\
&\leq 1 + 1 + \sum_\alpha (\underline{P}_\alpha(A_\alpha \cap B_\alpha) - 1) \qquad (7)\\
&\leq 1 + \Big(1 + \sum_\alpha (\underline{P}_\alpha(A_\alpha B_\alpha) - 1)\Big)^+ \\
&= 1 + \rho(AB).
\end{aligned}$$

The inequality (7) follows from the fact that all $\underline{P}_\alpha$'s must satisfy conjugacy. For superadditivity, we only have to consider disjoint finite rectangles $A = \Pi_\alpha A_\alpha$, $B = \Pi_\alpha B_\alpha$ where $\rho(A)\rho(B) > 0$. By observation (ii) above, $A_\beta \perp B_\beta$ for some β. Consider the finite rectangle $C = \Pi_\alpha C_\alpha$ where $C_\beta = A_\beta + B_\beta$ and $C_\alpha = A_\alpha B_\alpha$ for $\alpha \neq \beta$. We have $C \subset A + B$, and

$$\begin{aligned}
\rho(A) + \rho(B) &= 1 + 1 + \sum_\alpha \Big[(\underline{P}_\alpha(A_\alpha) - 1) + (\underline{P}_\alpha(B_\alpha) - 1)\Big] \\
&= 1 + \sum_{\alpha \neq \beta} \Big[(\underline{P}_\alpha(A_\alpha) + \underline{P}_\alpha(B_\alpha) - 1) - 1\Big] \\
&\quad + \Big[(\underline{P}_\beta(A_\beta) + \underline{P}_\beta(B_\beta) - 1)\Big] \\
&\leq 1 + \sum_{\alpha \neq \beta} (\underline{P}_\alpha(A_\alpha B_\alpha) - 1) + (\underline{P}_\beta(A_\beta + B_\beta) - 1) \\
&\leq \Big(1 + \sum_\alpha (\underline{P}_\alpha(C_\alpha) - 1)\Big)^+ = \rho(C).
\end{aligned}$$

Any LP that preserves the marginals, by conjugacy and nonnegativity has to dominate ρ. Therefore, the LP generated by the kernel is the minimal such LP. □

Note that by the conjugacy and nonnegativity axioms, any LP $\underline{P}$ must satisfy

$$\underline{P}\Big(\bigcap_{i=1}^n A_i\Big) \geq 1 + \sum_{i=1}^n (\underline{P}(A_i) - 1)$$

for any *finite* collection of sets. This is why we only considered finite rectangles. Proposition 3.4 still holds if the index set I is countable, and we consider the class of *all* (countable) rectangles, but the generated LP will not necessarily be minimal. (In this case, it would be the minimal *continuous* LP that preserves the marginals; see Definition 5.1.)

As stated before, given a kernel $(\mathcal{K}, \rho)$, we can strip away from $\mathcal{K}$ all nonempty sets for which ρ is 0. The next proposition is an application of this flexibility for the case of identical marginal spaces.

3.5. PROPOSITION. *Given an LP space* (Ω_0, P_0), *let* $\Omega = \prod_{\alpha \in I} \Omega_0 = \Omega_0^I$ *where* I *is an arbitrary index set. Let* $\underline{P}_0$ *satisfy the following range constraint*:

$$0 < h \equiv \sup\{\underline{P}_0(A_0)\colon A_0 \neq \Omega_0\} < 1.$$

Set $H = \lceil 1/(1-h) \rceil - 1$, *and let* $\mathcal{K}$ *be the set of all finite rectangles* $\prod_{\alpha \in I} A_\alpha$ *such that at most* H *of the* A_α*'s are different from* Ω_0. *On* $\mathcal{K}$ *define* ρ *by*

$$\rho\left(\prod_{\alpha \in I} A_\alpha\right) \equiv \left(1 + \sum_{\alpha \in I} (\underline{P}_0(A_\alpha) - 1)\right)^+.$$

Then $(\mathcal{K}, \rho)$ *is a kernel, and it generates the minimal LP* $\underline{P}$ *that has* $\underline{P}_0$ *as marginals*:

$$(\forall \beta \in I)(\forall A_\beta \subset \Omega_0) \qquad P\left(\prod_{\alpha \in I} A_\alpha\right) = \underline{P}_0(A_\beta),$$

where $A_\alpha = \Omega_0$ *for* $\alpha \neq \beta$.

PROOF. Combining the range constraint and (1), we observe that the ρ-assignment of any rectangle with more that H nontrivial *sides* is 0. □

In the next section, we will concentrate on the case where all $(\Omega_\alpha, \underline{P}_\alpha)$'s are the same. In this case, the joint LP can be used as a *stationary* model for a given time series. If the time series is generated by repetitions of the same probabilistic experiment, then the construction in Proposition 3.4 gives us the minimal joint model, and does not incorporate any dependence structure between the experiments. In this sense, it can be argued to be a good representation of the *unlinkedness* of the experiments.

We have shown that nonsingularity is sufficient for the construction of common extensions. The next proposition indicates that it is in a sense also necessary.

3.6. PROPOSITION. *Let* $\mathcal{A}_1, \mathcal{A}_2$ *be algebras of subsets of* Ω. $\mathcal{A}_1$ *and* $\mathcal{A}_2$ *are nonsingular iff any pair of LP's* $\underline{P}_1, \underline{P}_2$ *defined respectively on* $\mathcal{A}_1, \mathcal{A}_2$ *admit a common extension. The minimal such extension for* $\underline{P}_1, \underline{P}_2$ *when* $\mathcal{A}_1 ! \mathcal{A}_2$ *is given by* $(\forall A \subset \Omega)$

$$\underline{P}(A) = \sup\left\{ (\underline{P}_1(A_1) + \underline{P}_2(A_2) - 1)^+ \colon A_1 A_2 \subset A, A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \right\}.$$

PROOF. If $\mathcal{A}_1 ! \mathcal{A}_2$, then since $\mathcal{A}_i$'s are algebras, the nonsingularity condition in 3.3(iii) is satisfied. Therefore, any pair of LP's $\underline{P}_1, \underline{P}_2$ defined respectively on $\mathcal{A}_1, \mathcal{A}_2$, admit a common extension.

Conversely, let $A_1A_2 = \emptyset$, where $A_i \in \mathcal{A}_i$. Assume, by way of contradiction that $A_1, A_2 \neq \emptyset$. For $i = 1, 2$, let $\underline{P}_i$ be an LP on $\mathcal{A}_i$ such that $\underline{P}_i(A_i) = 3/4$ (see Example 2.5). By assumption, $\underline{P}_i$'s admit a common extension $\underline{P}$. We then arrive at the contradiction

$$\begin{aligned} 0 = \underline{P}(\emptyset) = \underline{P}(A_1A_2) &\geq \underline{P}(A_1) + \underline{P}(A_2) - 1 \\ &= \underline{P}_1(A_1) + \underline{P}_2(A_2) - 1 = 1/2. \end{aligned} \qquad \square$$

4. Time series modeling with IVP. In this section we explore IVP models that can incorporate the "contradictory" properties of stationarity and support for bounded yet divergent time averages. The contradiction is due to the ergodic theorems which indicate that a stationary stochastic process must have almost surely convergent time averages. Granted that divergence and convergence are infinitary issues and cannot be (operationally) verified/refuted in finite time, we are still surprised by such an a priori statement about the physical world, one based upon methodology rather than upon empirical experience. In this vein, we explore IVP in search of models that can characterize stationary processes with apparently divergent time averages. The need for such models has been argued for in the words of Kumar (1982), Grize (1984), Papamarcou (1987) and Fine (1988). In these works, the class of flicker ($1/f$) noise phenomena is identified as a potential example of real-world empirical processes that would require such new modeling methodology. For a more extensive discussion of flicker noise and related issues, see Grize (1984), Andrews (1985) and Papamarcou (1987).

Mathematical setup. We start with a nonempty subset X of $\mathbf{R}$ as our "marginal" space and let $\Omega \equiv \mathbf{X} \equiv X^{\mathbf{Z}}$, the set of all double-sided sequences of X. We denote a generic element of $\mathbf{X}$, a *sample path*, by $\mathbf{x} \equiv (x_n)_{n \in \mathbf{Z}}$. By an *index set*, we mean a *nonempty* subset $I \subset \mathbf{Z}$. For every index set I, the projection operator $B_I: \mathbf{X} \mapsto X^I$ is defined by

$$(\forall \mathbf{x} = (x_n)_{n \in \mathbf{Z}}) \qquad B_I\big((x_n)_{n \in \mathbf{Z}}\big) \equiv (x_n)_{n \in I}.$$

Following Papamarcou, we call an index set I a *sufficient set* for a set $A \subset \mathbf{X}$, or "sufficient for A," if

$$A = \big\{\mathbf{x} \in \mathbf{X}: (x_n)_{n \in I} \in B_I(A)\big\}. \tag{8}$$

We use the shorthand notation $A \leftrightarrow I$ to denote that I is sufficient for A.

A *cylinder* C is a set that has a *finite* sufficient set. We denote the algebra of all cylinders by $\mathcal{C}$. In general, the sufficient set of a set A is not unique, but for nontrivial cylinders there exists a *minimal* sufficient set.

4.1. DEFINITION. The *dimension* of a set A, $\dim(A)$, is the minimum number of coordinates required to specify A, and is defined by

$$(\forall A \subset \mathbf{X}) \qquad \dim(A) \equiv \min\{\|I\|: A \leftrightarrow I\}.$$

Note that $\dim(A) < \infty$ iff $A \in \mathcal{C}$. For all sets $A \notin \mathcal{C}$, we let $\dim(A) = \infty$. A related concept is the *diameter* of a set A, $\operatorname{diam}(A)$, defined as

$$\operatorname{diam}(A) \equiv \begin{cases} \min\left\{ \max\limits_{i \in I} i - \min\limits_{i \in I} i + 1: A \leftrightarrow I, \|I\| < \infty \right\}, & \text{if } A \in \mathcal{C}, \\ \infty, & \text{otherwise.} \end{cases}$$

Note that $\dim(\cdot) \le \operatorname{diam}(\cdot)$. For $N \ge 1$, we let $\mathcal{C}_N$ ($\mathcal{D}_N$) denote the class of cylinders whose dimension (diameter) does not exceed N:

$$(\forall N \ge 1) \qquad \mathcal{C}_N \equiv \{C \in \mathcal{C}: \dim(C) \le N\},$$
$$\mathcal{D}_N \equiv \{C \in \mathcal{C}: \operatorname{diam}(C) \le N\}.$$

Clearly, $\mathcal{D}_N \subset \mathcal{C}_N$.

For a nonempty class of sets $\mathcal{A}$, we define $\dim(\mathcal{A}) \equiv \sup\{\dim(A): A \in \mathcal{A}\}$. If $\dim(\mathcal{A}) < \infty$, we say that $\mathcal{A}$ has *bounded* dimension, or "$\mathcal{A}$ is bounded."

It is easy to show that if I, J are sufficient for a nontrivial cylinder C, where $\|I\| = \|J\| = \dim(C)$, then $I = J$. Therefore, a nontrivial cylinder has a *unique* minimal sufficient set. For a nontrivial cylinder C, we denote this minimal sufficient set by $I(C)$ and define its *base* by $B(C) \equiv B_{I(C)}(C)$.

Some of the properties of $\dim(\cdot)$ are summarized below. It is interesting that $\dim(\cdot)$ satisfies the "dual" properties of an LP.

4.2. PROPOSITION. $\dim(\cdot)$ *satisfies the following*:

(i) $\dim(\emptyset) = \dim(X) = 1$.
(ii) $\dim(\cdot) \ge 1$.
(iii) $\dim(AB) \le \dim(A) + \dim(B)$.
(iv) $\dim(A + B) \le \dim(A) + \dim(B) - 1$; $\dim(A \cup B) \le \dim(A) + \dim(B)$.
(v) $\dim(A) = \dim(A^c)$.

Topological issues/invariance and stationarity. If X is compact under a suitable topology, then by Tychonoff's theorem, $\mathbf{X} = X^{\mathbf{Z}}$ is compact under the product topology. Since we only consider *finite* marginal spaces, X is compact under any topology. Therefore, without any loss of generality, we endow X with the discrete topology. In this case, the class of cylinders $\mathcal{C}$ is a base for the product topology. We denote the class of all countable unions and intersections of cylinders respectively by $\mathcal{C}_\sigma$ and $\mathcal{C}_\delta$. Since the complement of a cylinder is itself a cylinder, cylinders are both closed and open (clopen). Moreover, since X is finite, $\mathcal{C}$ is countable. Hence $\mathcal{C}_\sigma(\mathcal{C}_\delta)$ is exactly the class of all open (closed) sets.

The algebra of cylinders $\mathcal{C}$ is identified with the class of all observable sets, namely, events whose occurrence or nonoccurrence can be determined in finite time. We refer to the specification of an LP on $\mathcal{C}$ as its set of *finite-dimensional distributions*. Another algebra of interest is the *tail σ-algebra*, $\mathcal{T}$, containing the *divergence* event D^*:

$$D^* \equiv \{\mathbf{x} \in \mathbf{X}: \liminf_n r_n(\mathbf{x}) \neq \limsup_n r_n(\mathbf{x})\}$$
$$= \bigcup_{k \geq 1} \bigcap_{n \geq 1} \bigcup_{m \geq n} \left\{\mathbf{x} \in \mathbf{X}: |r_n(\mathbf{x}) - r_m(\mathbf{x})| > \frac{1}{k}\right\},$$

where $r_n(\mathbf{x}) \equiv (1/n)\sum_{i=1}^n x_i$. The complement of D^*, the *convergence* event, is denoted by C^*.

We borrow the concept of stationarity from measure theory and define the left-shift operator $T: \mathbf{X} \mapsto \mathbf{X}$ by $(\forall i \in \mathbf{Z})(T\mathbf{x})_i \equiv x_{i+1}$. Since we are considering the double-sided sequence, T is invertible, and both T and its inverse T^{-1} preserve all the set-theoretic operations.

A set A is called *invariant* if $TA = T^{-1}A = A$. Invariance is preserved by all the set-theoretic operations, and the class of all invariant sets forms the invariant σ-algebra. Similarly, a class of sets $\mathcal{A}$ is called invariant if it is closed under T. A set function ρ on an invariant domain $\mathcal{A}$ is called *stationary* if

$$(\forall A \in \mathcal{A}) \qquad \rho(TA) = \rho(T^{-1}A) = \rho(A).$$

We observe the following:

(i) $C^*, D^*, \mathcal{C}, \mathcal{T}$ are invariant.

(ii) An invariant kernel, that is, kernel $(\mathcal{K}, \rho)$ with invariant $\mathcal{K}$ and stationary ρ, generates a stationary LP.

(iii) An LP $\underline{P}$ is stationary iff its conjugate upper probability $\overline{P}$ is stationary.

(iv) In Proposition 3.4, if all the LP spaces $(\Omega_\alpha, \underline{P}_\alpha)$'s are the same, and $I = \mathbf{Z}$, then the constructed kernel is invariant. Therefore, the generated minimal point LP is stationary.

Finite-dimensional distributions and support for tail events. In measure theory, the finite-dimensional distributions of a stochastic process uniquely determine its behavior on $\sigma(\mathcal{C}) \supset \mathcal{T}$. By the stationarity convergence theorem, any measure with *stationary* finite-dimensional distributions supports C^* almost surely. As expected—in the absence of additivity and continuity constraints—this is no longer true for IVP structures. From now on, we will concentrate on the issue, of finite-dimensional distributions and support for tail events in the context of stationary IVP models. Since we only consider finite marginal spaces, without loss of generality, we set $X = \{0, 1\}$.

If for an event A, we associate the interval $[\underline{P}(A), \overline{P}(A)]$ with the "imprecision" about the likelihood of A, then we observe that an additive IVP—a charge—and the vacuous IVP are extreme cases. The former allows only one point in [0, 1], while the latter assigns the whole interval. We also observed this antipodal behavior in their generating kernels: the vacuous LP can be generated from the trivial kernel ($\mathcal{K} = \{\emptyset, \Omega\}$), while a charge in general requires the maximal kernel ($\mathcal{K} = 2^{\Omega}$). We are mostly interested in the interplay of these two extremes: Can a stationary IVP model have additive finite-dimensional distributions and yet be vacuous on $\mathcal{T}$? Or in the other direction, can it be additive on $\mathcal{T}$ and vacuous on $\mathcal{C}$? In what follows, we experiment with variations of these questions. Our answers will establish the logical independence that has long been obscured by conventional probability theory, and thereby provide a new perspective on the frequentist interpretation of probability.

Most of our results are based on the following key observation.

4.3. PROPOSITION. *C and $\mathcal{T}$ are mutually nonsingular*, $\mathcal{C}!\mathcal{T}$.

PROOF. Follows from the observation that if $A \in \mathcal{C} \leftrightarrow I$, $B \in \mathcal{T} \leftrightarrow J$, where I and J are disjoint, we have

$$A \cap B = \emptyset \quad \Leftrightarrow \quad A = \emptyset \text{ or } B = \emptyset. \qquad \square$$

Therefore, by an application of 3.3(iii), *any* pair of LP's $\underline{P}_1$ and $\underline{P}_2$ defined respectively on $\mathcal{C}$ and $\mathcal{T}$ admit a common extension. Moreover, we note that in 3.3, if $\underline{P}_i$'s are stationary, then the resultant common extension is also stationary. Therefore, any pair of stationary LP's $\underline{P}_1, \underline{P}_2$ defined on $\mathcal{C}, \mathcal{T}$, respectively, admit a stationary common extension, with the minimal such extension given by

$$(\forall A \subset \mathbf{X}) \qquad P(A) = \sup\left\{ \left(\underline{P}_1(C) + \underline{P}_2(D) - 1\right)^+ : C \in \mathcal{C}, D \in \mathcal{T}, CD \subset A \right\}.$$

This indicates that *within the framework of stationary IVP models, the issues of finite-dimensional distributions and support for tail events are unrelated.*

In terms of support for divergence, the situation is not changed if one requires that the LP's be additive. Indeed, there are well-known examples of stationary charges that support divergence; see Kumar and Fine (1985) and Ramakrishnan and Sudderth (1988). The next theorem generalizes these examples by showing that *any* stationary charge on $\mathcal{C}$ has a stationary charge extension with a prescribed behavior on invariant tail events.

4.4. THEOREM. *Let $\{D_1, \dots, D_n\}$ be a finite partition of $\mathbf{X}$ into nontrivial invariant tail events, and let $\mathcal{D}$ be the algebra generated by this partition. For any pair of stationary charges π_1, π_2 defined on $\mathcal{C}, \mathcal{D}$, there exists a stationary charge π on the algebra generated by $\mathcal{C} \cup \mathcal{D}$ that is a common extension of π_1, π_2.*

PROOF. The proof consists of verifying the following:

(i) The invariant class $\mathcal{A} = \{\sum_{i=1}^n C_i D_i : C_i \in \mathcal{C}\}$ is the algebra generated by $\mathcal{C} \cup \mathcal{D}$.

(ii) The representation of $\mathcal{A}$-sets as $\sum_{i=1}^n A_i D_i$ is unique; that is,

$$\sum_{i=1}^{n} A_i D_i = \sum_{i=1}^{n} B_i D_i \quad \Rightarrow \quad A_i = B_i \quad \text{for } i = 1, \dots, n.$$

(iii) The set function π defined by $\pi(\sum_{i=1}^{n} A_i D_i) = \sum_{i=1}^{n} \pi_1(A_i) \cdot \pi_2(D_i)$ is the desired stationary common extension of π_1, π_2.

We will only prove (ii), as the rest of them follow from routine verification. By intersecting both sides by D_j, we get

$$A_j D_j = B_j D_j \quad \Rightarrow \quad (A_j \backslash B_j) D_j = \emptyset, \qquad (B_j \backslash A_j) D_j = \emptyset \quad \text{for } j = 1, \dots, n.$$

Since $\mathcal{C} \,!\, \mathcal{D}$, this shows that $A_j \backslash B_j = B_j \backslash A_j = \emptyset$, hence $A_j = B_j$ for all j. □

Hence, we can form the algebra $\mathcal{D} = \{\emptyset, \mathbf{X}, C^*, D^*\}$, and define the charge π_2 by setting $\pi_2(D^*) = \lambda$ for a given $\lambda \in [0,1]$. By Theorem 4.4, π_2 and any stationary charge on $\mathcal{C}$ admit a stationary (common) charge extension.

5. Continuity, estimability and the Basic Problem. What is lacking from the previous constructions is a notion of *estimability* from observable events. Indeed, two stationary LP's may agree on all cylinder sets, but differ significantly on events not in $\mathcal{C}$, for example, tail events. Therefore, based on a finite data set from a stochastic source, one cannot distinguish between two such competing LP's, and hence one cannot learn from data. This is in sharp contrast to measures, where because of countable additivity the extension of a measure from an algebra to the generated σ-algebra is unique (Caratheodory's theorem). This unicity of extension in measure theory suggests that by imposing some form of a continuity constraint on our models, we might be able to uniquely extend an LP from $\mathcal{C}$ to a class of events outside of $\mathcal{C}$.

5.1. DEFINITION. An LP P is *monotonely continuous along* an algebra $\mathcal{A}$ if

$$(\forall A_i \in \mathcal{A}) \qquad \underline{P}\left(\bigcup_{i=1}^{n} A_i\right) \uparrow \underline{P}\left(\bigcup_{i \geq 1} A_i\right), \qquad \underline{P}\left(\bigcap_{i=1}^{n} A_i\right) \downarrow \underline{P}\left(\bigcap_{i \geq 1} A_i\right).$$

$\underline{P}$ is monotonely continuous along $\mathcal{A}$ iff $\overline{P}$ is. The class of continuous LP's along $\mathcal{C}$ is denoted by $\mathcal{S}$.

Given an LP $\underline{P}$, its class of *estimable* sets, $\mathcal{E}(\underline{P})$ is defined as

$$\mathcal{E}(\underline{P}) = \{A \subset \mathbf{X}\colon (\forall \underline{Q} \in \mathcal{S})\, \underline{Q} = \underline{P} \text{ on } \mathcal{C} \Rightarrow \underline{Q}(A) = \underline{P}(A)\},$$

namely, $\mathcal{E}(\underline{P})$ is the collection of all sets A such that for any given continuous LP $\underline{Q}$ that agrees with $\underline{P}$ on $\mathcal{C}$, we have $\underline{Q}(A) = \underline{P}(A)$. Clearly, if $\underline{P}$ is continuous, then $\mathcal{C}_\delta$ and $\mathcal{C}_\sigma$ are both subsets of $\mathcal{E}(\underline{P})$.

For modeling time series that exhibit the properties of stationarity and unstable time averages, we seek an LP $\underline{P}$ that satisfies the following:

(BP1) $\underline{P}$ is stationary.
(BP2) $\underline{P}$ is monotonely continuous along $\mathcal{C}$: $\underline{P} \in \mathcal{S}$.
(BP3) $\underline{P}$ almost surely supports the divergence event: $\underline{P}(D^*) = \overline{P}(D^*) = 1$.

(BP4) The divergence event is estimable: $D^* \in \mathcal{E}(\underline{P})$.

We refer to the LP $\underline{P}$ as a solution to the basic problem. The original formulation of the basic problem in these terms is due to Grize (1984). We briefly review the previous work on this problem.

Kumar and Fine (1985) were the first to consider stationary and continuous LP's that support divergence. Their major result is that such an LP is necessarily undominated. Grize and Fine (1987) provide a specific construction of stationary and continuous LP's that support divergence with positive probability (but not almost surely) and satisfy a weaker notion of estimability than the above.

Papamarcou and Fine (1991) extend Grize's construction to provide a solution to the basic problem. They also show that a solution to the basic problem exists iff there exist invariant families $\mathcal{F}_1, \mathcal{F}_2, \dots$ of closed ($\mathcal{C}_\delta$) subsets of D^* such that

$$(\forall i \geq 1)(\forall F_i^1, \dots, F_i^{2i} \in \mathcal{F}_i) \qquad \bigcap_{i \geq 1} \bigcap_{j \leq 2i} F_i^j \neq \emptyset.$$

We have been interested in the marginals of these constructions: what is the behavior of a solution to the basic problem on a given class of cylinders, for example, $\mathcal{C}_N$, or $\mathcal{D}_N$? Is there a solution to the basic problem that mimics the behavior of a given (competing) time series model—maybe a measure—on a class of cylinders, but (as opposed to measures) would go on to support divergence?

The next proposition provides sufficient conditions for the existence of such a solution, and is in fact a minor modification of Papamarcou and Fine's construction.

5.2. Proposition. *Let $\underline{P}_1$ be generated by an invariant kernel $(\mathcal{K}_1, \rho_1)$ where $\dim(\mathcal{K}_1) < \infty$. There exists a solution to the basic problem $\underline{P}$ such that $\underline{P} = \underline{P}_1$ on $\mathcal{K}_1$ (and $P \geq P_1$ on $2^{\mathbf{X}}$.)*

Proof. See the Appendix. □

Consider the kernel $(\mathcal{K}_1, \rho_1)$ defined by $\mathcal{K}_1 = \mathcal{C}_N, \rho_1(A) = 0 (\forall A \in \mathcal{K}_1^+)$, which generates an LP $\underline{P}_1$ that is vacuous on $\mathcal{C}_N$. Proposition 5.2 implies that for any positive integer N, we can construct a solution to the basic problem that is vacuous on $\mathcal{C}_N$. This suggests that the issues of support for divergence and the specification of probability on cylinders are still unrelated in the family of stationary and continuous LP's, provided that we restrict our attention to bounded-dimensional cylinders.

In fact, the combination of the next proposition and the previous one allows us to show the stronger result that: for a given stationary LP $\underline{P}, N \geq 1$ and $\varepsilon > 0$, there exists a solution to the basic problem that has the same (within ε) N-dimensional distribution as $\underline{P}$.

5.3. Proposition. *Let $\underline{P}_0$ be a stationary LP. Given any integer $N \geq 1$ and $0 < \varepsilon < 1$, there exists an invariant kernel $(\mathcal{K}, \rho)$ where $\mathcal{K}$ is bounded, $\mathcal{D}_N \subset \mathcal{K}$,* *and*

$$(9) \qquad (\forall C \in \mathcal{D}_N) \qquad |\rho(C) - \underline{P}_0(C)| \le \varepsilon.$$

$(\mathcal{K}, \rho)$ *generates an LP* $\underline{Q}$ *that is stationary and satisfies*

$$(10) \qquad (\forall C \in \mathcal{D}_N) \qquad |\underline{Q}(C) - \underline{P}_0(C)| \le \varepsilon, \qquad |\overline{Q}(C) - \overline{P}_0(C)| \le \varepsilon.$$

PROOF. See the Appendix. □

Consider the following consequence of Proposition 5.3: given a stationary measure μ, an integer $N \ge 1$ and $0 < \varepsilon < 1$, there exists a stationary and continuous IVP $(\underline{P}, \overline{P})$ such that $\underline{P}(D^*) = \overline{P}(D^*) = 1$, D^* is estimable and

$$(\forall C \in \mathcal{D}_N) \qquad \mu(C) - \varepsilon \le \underline{P}(C) \le \overline{P}(C) \le \mu(C) + \varepsilon.$$

Assume that we are given a time series of N data points, and we are asked to propose a probabilistic model P for this time series. Based on these N data points, we can give (in the strict sense) statistical validation to at most the N-dimensional distributions of P, that is, its behavior on $\mathcal{D}_N$. The specification of P on other events is more a function of methodology and modeling assumptions than the evidence conveyed by the data. For instance, the absence of a law-like trend in the data suggests the adoption a stationary model. However, stationarity is a property of the model and not the data: it makes claims about the M-dimensional distributions of P (for all M) which clearly cannot be verified on the basis of a finite data record. Continuity is also a property of the model, as it delimits the choice of models to tractable ones and provides a mechanism to infer from data (go from $\mathcal{C}$ to events outside of $\mathcal{C}$).

If we let P be a measure μ, then by stationarity convergence theorems, we are bound to accept that the time averages of the time series are (almost surely) convergent. But what if we observe a persistent instability in the time averages of the data? It is exactly this scenario for which we argue that our construction $\underline{P}$ rather than μ is a better model for the data.

The construction in Proposition 5.2 and Example 2.6 provide support for D^* through events that exhibit "apparent divergence": Given levels $0 < \alpha < \beta < 1$ and a *rate sequence* $\mathbf{M} = (M_n)_{n \ge 1}$ where $1 = M_1 < M_2 < \cdots$, the P-assignment of the $\mathcal{C}_\varepsilon$-sets $F_i(\mathbf{M}, \alpha, \beta) \equiv \bigcap_{j \ge i} E_j(\mathbf{M}, \alpha, \beta)$ where

$$(\forall j \ge 1) \qquad E_j(\mathbf{M}, \alpha, \beta) \equiv \Big\{\mathbf{x} \in \mathbf{X}\colon \big(\exists p, q \in [M_j, M_{j+1})\big)\, r_p(\mathbf{x}) < \alpha < \beta < r_q(\mathbf{x})\Big\}$$

satisfies the constraint

$$\underline{P}\big(F_i(\mathbf{M}, \alpha, \beta)\big) \ge 1 - \frac{1}{i+1} = \frac{i}{i+1}$$
$$\Rightarrow \quad \underline{P}(D^*) = 1.$$

Note that the cylinder $E_j(\mathbf{M}, \alpha, \beta)$ consists of sequences whose time-averages between times M_j and M_{j+1} cross $[\alpha, \beta]$, and its $\underline{P}$-assignment exceeds $i/(i+1)$, since $F_j(\mathbf{M}, \alpha, \beta) \subset E_j(\mathbf{M}, \alpha, \beta)$ for all $j \geq 1$. Therefore, support for divergence is not achieved by indefinitely delaying it, and there is an explicit rate and level of divergence that can be used to test the claim of unstable time averages. This can be used as a basis for hypothesis testing between our constructions and a proposed measure μ. For example, on the measurable space $(\{0,1\}^{\mathbf{z}}, \sigma(\mathcal{C}))$, let μ be an ergodic measure such that its coordinate process $\{\xi_i\}, \xi_i(\mathbf{x}) \equiv x_i$, has mean m_ξ and variance σ^2. Given a rate sequence $\mathbf{M} = (M_i)$ and a divergence threshold $0 < \tau < \min(m_\xi, 1 - m_\xi)$, consider the following cylinders D_i:

$$D_i \equiv \left\{ \mathbf{x} : \max_{k \in [M_i, M_{i+1})} |A_k \xi(\mathbf{x}) - m_\xi| \geq \tau \right\},$$

where $A_k \xi(\mathbf{x}) \equiv (1/k)\sum_{j=1}^{k} \xi_j(\mathbf{x}) = r_k(\mathbf{x})$. For any of our constructions $\underline{P}$, we have

$$\begin{aligned} & F_i(\mathbf{M}, \alpha, \beta) \subset E_i(\mathbf{M}, \alpha, \beta) \subset D_i \\ \Rightarrow \quad & \underline{P}(D_i) \geq \underline{P}\big(E_i(\mathbf{M}, \alpha, \beta)\big) \geq \underline{P}\big(F_i(\mathbf{M}, \alpha, \beta)\big) \geq 1 - \frac{1}{i+1}, \end{aligned}$$

where $\alpha = m_\xi - \tau$, $\beta = m_\xi + \tau$. On the other hand, by the ergodic theorem,

$$\begin{aligned} & \mu\big[\mathbf{x} \colon A_n \xi(\mathbf{x}) \to m_\xi\big] = 1 \\ \Rightarrow \quad & (\forall \varepsilon > 0) \qquad \lim_n \mu\left[\sup_{m \geq n} |A_m \xi(\mathbf{x}) - m_\xi| \geq \varepsilon\right] = 0 \\ \Rightarrow \quad & \mu(D_i) \to 0. \end{aligned}$$

Therefore, $\lim_i \underline{P}(D_i) = 1$, while $\lim_i \mu(D_i) = 0$. If for the given time series D_i is observed, that is, the time averages between M_i and M_{i+1} cross $[m_\xi - \tau, m_\xi + \tau]$, we are encouraged to choose $\underline{P}$ over μ. Our confidence in this decision is reinforced by the observation of persistent τ-deviation (more D_i's) as more data is collected (*consistency*).

In some instances, we can give bounds on the rate of decrease of $\mu(D_i)$'s. For instance, if $\{\xi_i\}$ is i.i.d., we have

$$\begin{aligned} D_i & \subset \left\{ \mathbf{x} : \max_{k \in [M_i, M_{i+1})} \left| \frac{\sum_{j=1}^{k} \xi_j(\mathbf{x}) - k m_\xi}{M_i} \right| > \tau \right\} \\ & \subset \left\{ \mathbf{x} : \max_{k \in [1, M_{i+1})} \left| \sum_{j=1}^{k} \xi_j(\mathbf{x}) - k m_\xi \right| \geq M_i \tau \right\}, \end{aligned}$$

where by Kolmogorov's inequality, we have

$$\mu\left[\max_{k \in [1, M_{i+1}]} \left| \sum_{j=1}^{k} \xi_j - k m_\xi \right| \geq M_i \tau \right] \leq \frac{\operatorname{Var}\big(\sum_{j=1}^{M_{i+1}} \xi_j\big)}{M_i^2 \tau^2} = \frac{M_{i+1}\sigma^2}{M_i^2 \tau^2}.$$

Therefore, $\mu(D_i) \leq M_{i+1}\sigma^2/(M_i^2\tau^2)$. With the rate sequence $\mathbf{M}$ defined in the proof of Proposition 5.2 [see (11) and (14) in A.1], $\{M_{i+1}/M_i^2\}$ is a vanishing sequence and can be used as a bound on the rate of decrease of $\mu(D_i)$.

In the proof of Proposition 5.3, we observe that if $N > 1$, then $M = \lceil 2(N-1)/\varepsilon \rceil$ and $\dim(\mathcal{K}) \geq M$. Hence, we cannot let ε tend to 0, and still maintain a bounded kernel. However, if $N = 1$, then $M = 1$ for all ε, and we can set $\varepsilon = 0$. In this case, we can construct an LP that has the same marginals as $\underline{P}_0$, that is, agrees *exactly* with $\underline{P}_0$ on $\mathcal{D}_1$. We have the following result.

5.4. PROPOSITION. *Let $\underline{P}_0$ be a stationary LP that satisfies*

$$\sup\{\underline{P}_0(C)\colon C \in \mathcal{D}_1^+\} < 1.$$

There exists an invariant kernel $(\mathcal{K}, \rho)$ where $\mathcal{K}$ is bounded, $\mathcal{D}_1 \subset \mathcal{K}$ and the generated LP $\underline{Q}$ satisfies

$$(\forall \mathcal{C} \in \mathcal{D}_1) \qquad \underline{Q}(C) = \underline{P}_0(C), \qquad \overline{Q}(C) = \overline{P}_0(C).$$

For instance, there exists a solution to the basic problem that has the same marginals as a (nondegenerate) Bernoulli measure.

5.5. EXAMPLE. Let $\underline{P}_0 = \overline{P}_0 = \mu$ where μ is a stationary measure such that for some $0 < p < 1$, we have

$$(\forall i \in \mathbf{Z}) \qquad \mu[\mathbf{x}\colon x_i = 1] = p, \qquad \mu[\mathbf{x}\colon x_i = 0] = q = 1 - p$$

(e.g., let μ be the Bernoulli measure with parameter p). Without loss of generality, let $p \geq q$. Let $H = \lceil 1/q \rceil - 1$. Define the cylinders $\mathbf{0}_i$ and $\mathbf{1}_{i_1,\ldots,i_l}$, $1 \leq l \leq H$, as

$$(\forall\, i \in \mathbf{Z}) \qquad \mathbf{0}_i \equiv \{\mathbf{x}\colon x_i = 0\},$$
$$\mathbf{1}_{i_1, i_2,\ldots, i_l} \equiv \{\mathbf{x}\colon (x_{i_1}, x_{i_2}, \ldots, x_{i_l}) = (1, 1, \ldots, 1)\},$$

where $i_1, \ldots, i_l$ are distinct integers. The following bounded and invariant kernel $(\mathcal{K}, \rho)$ generates the minimal stationary LP that has the same marginals as μ:

$$\mathcal{K} = \{\emptyset, \mathbf{X}\} + \{\mathbf{0}_i\colon i \in \mathbf{Z}\} + \{\mathbf{1}_{i_1,\ldots,i_l}\colon 1 \leq l \leq H\},$$
$$\rho(\mathbf{0}_i) = q, \qquad \rho(\mathbf{1}_{i_1,\ldots,i_l}) = 1 - l_q.$$

We have also been able to construct a bounded kernel that generates a stationary LP agreeing *exactly* with the symmetric Bernoulli measure ($p = 1/2$) on $\mathcal{D}_2$, see Sadrolhefazi (1990). This has led us to conjecture that Proposition 5.4 is true for general N, subject to some nondegeneracy conditions on the target $\underline{P}_0$. It is easy to show that these conditions cannot be dropped altogether: consider the case of a (degenerate) Bernoulli measure μ where the μ-assignment of "getting a zero" is 1. Any LP $\underline{P}$ that agrees with μ on $\mathcal{D}_1$ must by conjugacy assign unit lower probability to the cylinders $Z_n = \{\mathbf{x}: (x_1, \ldots, x_n) = (0, \ldots, 0)\}$, and hence cannot have a bounded generating kernel. We have, however, been unable to obtain a complete proof for our conjecture.

Since all the constructions in Propositions 5.3 and 5.4 and Example 5.5 have a bounded kernel, they are vacuous on $\mathcal{T}$. However, they are not necessarily continuous. While it seems that boundedness should be sufficient for continuity, we have had to resort to the stronger condition of rectangularity.

5.6. DEFINITION. A nonempty finite subset $I \subset \mathbf{Z}$ is said to be an *interval* if it consists of consecutive integers:

$$I = \{i, i+1, \ldots, i + \|I\| - 1\}.$$

A nontrivial cylinder C is said to be *rectangular* if for some integer $M \geq 1$ and disjoint intervals $\{I_i\}$, we have

$$I(C) = \sum_{i=1}^{M} I_i, \qquad B(C) = \prod_{i=1}^{M} B_{I_i}(C).$$

Similarly, a class of cylinders is said to be rectangular if all of its nontrivial members are rectangular.

5.7. PROPOSITION. *Let $(\mathcal{K}, \rho)$ be a kernel where $\mathcal{K}$ is rectangular and* $\dim(\mathcal{K}) = L < \infty$. *Then the generated LP $\underline{P}$ is continuous.*

PROOF. See the Appendix.

We observe that each of the cyclostationary LP's $\underline{Q}_i$ in the proof of Proposition 5.3 is rectangular, and hence continuous. Therefore, the LP $\underline{Q} \equiv (1/M)\sum_{i=1}^{M} \underline{Q}_i$ is continuous. Moreover, since $\underline{Q}$ is generated by a bounded kernel, it is vacuous on $\mathcal{T}$. We have the following result.

5.8. PROPOSITION. *Given a stationary LP $\underline{P}_0$, an integer $N \geq 1$ and $0 < \varepsilon < 1$, there exists a stationary and continuous LP $\underline{Q}$ that is vacuous on $\mathcal{T}$ and satisfies*

$$(\forall C \in \mathcal{D}_N) \qquad |\underline{Q}(C) - \underline{P}_0(C)| \leq \varepsilon, \qquad |\overline{Q}(C) - \overline{P}_0(C)| \leq \varepsilon.$$

One should be reluctant to adopt a model restricting the potential asymptotic behavior of a time series when we have little more evidence than is provided by a finite number of observations. The LP $\underline{Q}$ in Proposition 5.8 captures this intuitive notion in that it has nearly the same N-dimensional distributions—suggested by the data—as the competing LP $\underline{P}_0$, it is stationary and continuous, but is noncommittal (vacuous) on tail events. Unlike conventional probabilistic and frequentist approaches, we are able to avoid unfounded commitments.

Moreover, if based on the data we come to believe that the process has unstable time-averages, then by Proposition 5.2 we can form an LP model that has the same N-dimensional distributions as $\underline{P}_0$, but goes on to support divergence. It is our hope that this type of flexibility inherent in IVP structures would lead to their acceptance as a viable modeling tool for empirical processes requiring such new modeling methodology.

APPENDIX

A.1. PROOF OF PROPOSITION 5.2 Let $N = \dim((\mathcal{K} \cup \mathcal{K}_1^{(c)})^{(2)})$. A sufficient condition for the construction of $\underline{P}$ is the existence of an invariant kernel $(\mathcal{K}_2, \rho_2)$ that satisfies:

(I) $\mathcal{K}_2^{(2)}!\mathcal{C}_N$.
(II) $\mathcal{K}_2^+ \subset \mathcal{C}_\delta \cap 2^{D^*}$, that is, $\mathcal{K}_2^+$ consists solely of closed subsets of D^*.
(III) $\sup\{\rho_2(A): A \in \mathcal{K}_2\} = 1$.

If such a kernel exists, since $(\mathcal{K}_1 \cup \mathcal{K}_1^{(c)})^{(2)} \subset \mathcal{C}_N$, by Proposition 3.3(ii) there exists an LP $\underline{Q}$ such that $\underline{Q} = \underline{P}_1$ on $\mathcal{K}_1$ and $\underline{Q} \geq \rho_2$ on $\mathcal{K}_2$. It is easy to see that $\underline{Q}$ is stationary and

$$\sup\{\underline{Q}(A): A \in \mathcal{K}_2\} = 1.$$

Let $\underline{P}_0$ be the restriction of $\underline{Q}$ to $\mathcal{C}$. By applying the next proposition to $\underline{P}_0$, we obtain $\underline{P}$, the minimal continuous extension of $\underline{P}_0$ to $2^{\mathbf{X}}$.

A.1.1. PROPOSITION [Grize and Fine (1987), Lemma 3]. *Given an LP space* $(\Omega, \mathcal{A}, \underline{P}_0)$, *define* P_* *on* $\mathcal{A}_\delta$ *by*

$$(\forall A_\delta \in \mathcal{A}_\delta) \qquad P_*(A_\delta) = \inf\{\underline{P}_0(A): A \in \mathcal{A}, A_\delta \subset A\},$$

and let $\underline{P}$ *be the minimal extension of* P_* *to the power set*

$$(\forall A \subset \Omega) \qquad \underline{P}(A) = \sup\{P_*(A_\delta): A_\delta \in \mathcal{A}_\delta, A_\delta \subset A\}.$$

Then $\underline{P}$ *is the minimal continuous extension of* $\underline{P}_0$.

$\underline{P}$ is the desired LP in Proposition 5.2, since:

(i) $\underline{P}$ is stationary, since $\underline{Q}$ and hence $\underline{P}_0$ are stationary.
(ii) $\underline{P}$ is by construction continuous.
(iii) $\underline{P}(D^*) = 1$, since $\underline{P} \geq \underline{Q}$ on $\mathcal{C}_\delta$ and

$$\underline{P}(D^*) \geq \sup\{\underline{P}(A): A \in \mathcal{C}_\delta \cap 2^{D^*}\} \geq \sup\{\underline{Q}(A): A \in \mathcal{K}_2\} = 1.$$

(iv) $D^* \in \mathcal{E}(\underline{P})$, since by construction $\underline{P}$ is the *minimal* continuous extension of $\underline{P}_0$. Therefore, any other LP that agrees with $\underline{P}$ on $\mathcal{C}(\underline{P}_0)$ has to dominate $\underline{P}$ on $2^{\mathbf{X}}$.
(v) $\underline{P} = \underline{P}_0 = \underline{Q}$ on $\mathcal{C}$, hence $\underline{P} = \underline{P}_1$ on $\mathcal{K}_1 (\subset \mathcal{C})$.

Papamarcou and Fine (1991) construct a kernel $(\mathcal{K}_2, \rho_2)$ as follows: given two levels $0 < \alpha < \beta < 1$, they define a *rate sequence* $\mathbf{M} = (M_n)_{n \geq 1}$ where $1 = M_1 < M_2 < \cdots$ by

$$(11) \qquad (\forall j \geq 1) \qquad \lambda_{j+1} = \left\lceil \frac{1 + j(j+1)}{\alpha \wedge (1-\beta)} + 1 \right\rceil, \qquad M_{j+1} = \lambda_{j+1}^3 M_j.$$

Next they define cylinders $E_j(\mathbf{M}, \alpha, \beta)$ $(\forall j \geq 1)$

$$E_j(\mathbf{M}, \alpha, \beta) \equiv \left\{ \mathbf{x} \in \mathbf{X}\colon \left(\exists p, q \in [M_j, M_{j+1})\right) r_p(\mathbf{x}) < \alpha < \beta < r_q(\mathbf{x}) \right\}$$

and closed subsets of $D^*, F_i(\mathbf{M}, \alpha, \beta)$, by

$$F_i(\mathbf{M}, \alpha, \beta) \equiv \bigcap_{j \geq i} E_j(\mathbf{M}, \alpha, \beta),$$

to get invariant families $\mathcal{F}_i$:

$$(12) \qquad (\forall i \geq 1) \qquad \mathcal{F}_i \equiv \left\{ T^k F_i(\mathbf{M}, \alpha, \beta)\colon k \in \mathbf{Z} \right\}.$$

These families give $\mathcal{K}_2$:

$$\mathcal{K}_2 \doteq \{\emptyset, \mathbf{X}\} + \bigodot_{i \geq 1} \mathcal{F}_i^{(i)}.$$

The above rate sequence $\mathbf{M}$ is selected so that

$$(13) \qquad (\forall i \geq 1)(\forall F_i^1, \ldots, F_i^{2i} \in \mathcal{F}_i) \qquad \bigcap_{i \geq 1} \bigcap_{j \leq 2i} F_i^j \neq \emptyset.$$

This condition is sufficient to construct the kernel $(\mathcal{K}_2, \rho_2)$ so that (II) and (III) above are satisfied. We claim that by modifying the rate sequence in (11) to

$$(14) \qquad (\forall j \geq 1) \qquad \lambda_{j+1} = \left\lceil \frac{1 + j(j+1) + N/M_j}{\alpha \wedge (1-\beta)} + 1 \right\rceil,$$

we get invariant families $\mathcal{F}_i$ that satisfy (I), which is a stronger condition than (12).

The first part of the next proposition follows directly from Papamarcou and Fine (1991), Proposition 7, while the second part is their Propositions 8 and 9 recast in our notation.

A.1.2. PROPOSITION [Papamarcou and Fine (1991), Propositions 7–9].

(i) *The classes* $\mathcal{F}_1, \mathcal{F}_2, \ldots$ *defined in* (12) *satisfy the nonempty intersection property*

$$(\forall i \geq 1)(\forall F_i^1, \ldots, F_i^{2i} \in \mathcal{F}_i)(\forall C \in \mathcal{C}_N^+) \qquad C \cap \bigcap_{i \geq 1} \bigcap_{j \leq 2i} F_i^j \neq \emptyset$$

iff for every collection $\{K_i: i \geq 1\}$ *of nonnegative indices such that* $(\forall j)\|K_j\| = 2j$,

$$(\forall C \in \mathcal{C}_N^+)(\forall n \geq 1) \qquad C \cap \bigcap_{j \leq n} \bigcap_{k \in K_1 \cup \cdots \cup K_j} T^{-k}E_j \neq \emptyset.$$

(ii) *Let* $K_1, \dots, K_n$ *be a class of nonnegative index sets* $\|K_j\| = 2j$. *Then there exist nontrivial* (*thin*) *cylinders* $C_1, \dots, C_n$ *with disjoint sufficient sets such that*

$$(1 \leq j \leq n) \qquad C_j \subseteq \bigcap_{k \in K_1 \cup \cdots \cup K_j} T^{-k}E_j.$$

For a given $n \geq 1$ and $K_1, \dots, K_n$, the cylinders $C_1, \dots, C_n$ are defined as follows:

$$C_j = \left\{\mathbf{x} \in \mathbf{X}: \mathbf{x} = T^{-k_t}\mathbf{z} \text{ on } \left[K_t + m, (k_t + M) \wedge k_{t-1}\right), \ 1 \leq t \leq j(j+1)\right\},$$

where

$$z_i = \begin{cases} 0, & \text{if } i < M_j \lambda_{j+1}, \\ 1, & \text{if } i \geq M_j \lambda_{j+1}, \end{cases}$$

and $k_1 > \cdots > k_{j(j+1)}$ are elements of $K_1 \cup \cdots \cup K_j$ in decreasing order ($k_0 = +\infty$). (The intervals above should be read as *integers* in the specified range.) It is then shown inductively that for all $\mathbf{x} \in C_j$ $(1 \leq t \leq j(j+1))$

$$(15) \qquad \begin{aligned} (\exists p_t, q_t \in [M_j, M_{j+1})) \qquad r_{p_t}\left(T^{k_t}\mathbf{x}\right) &\leq \frac{t+1}{\lambda} < \alpha < \beta \\ &< 1 - \frac{t+1}{\lambda} \leq r_{q_t}\left(T^{k_t}\mathbf{x}\right). \end{aligned}$$

We claim that by modifying λ_{j+1} and $\mathbf{M}$ in (11) to (14), for a given $C \in \mathcal{C}_N^+$, we can find nontrivial cylinders D_j with disjoint sufficient sets so that $(\forall i \geq 1) I(D_i) \perp I(C)$ and

$$\emptyset \neq D_j \subset \bigcap_{k \in K_1 \cup \cdots \cup K_j} t^{-k}E_j \quad \text{for } j = 1, \dots, n.$$

Therefore,

$$\emptyset \neq C \cap \bigcap_{j=1}^{n} D_j \subset C \cap \bigcap_{j \leq n} \bigcap_{k \in K_1 \cup \cdots \cup K_n} T^{-k}E_j,$$

as desired. Using the sequence $\mathbf{z}$ above, we define the cylinder D_j as

$$D_j = \left\{\mathbf{x} \in \mathbf{X}: \mathbf{x} = T^{-k_t}\mathbf{z} \text{ on } I(C)^c \cap \left[k_t + m, (k_t + M) \wedge k_{t-1}\right), 1 \leq t \leq \Gamma\right\}.$$

Note that D_j is different from the cylinder C_j in at most N positions. Therefore,

$$(16) \qquad (\forall \mathbf{y} \in D_j)(\forall n \geq 1)(\exists \mathbf{x} \in C_j)(\forall n \geq 1) \qquad |r_n(\mathbf{y}) - r_n(\mathbf{x})| \leq 1 \wedge \frac{N}{n}.$$

Combining (15) and (16), we get $(\forall \mathbf{y} \in D_j)(\forall 1 \le t \le \Gamma)$,

$$r_{p_t}(T^{k_t}\mathbf{y}) \le \frac{t+1}{\lambda} + \frac{N}{p_t} < \alpha < \beta < 1 - \frac{t+1}{\lambda} - \frac{N}{q_t} \le r_{q_t}(T^{k_t}\mathbf{y}).$$

Hence

$$\begin{aligned} &(\forall \mathbf{y} \in D_j)(\forall 1 \le t \le \Gamma) \qquad T^{k_t}\mathbf{y} \in E_j \\ &\Rightarrow \quad D_j \subseteq \bigcap_{k \in K} T^{-k} E_j. \end{aligned}$$

□

A.2. Proof of Proposition 5.3. Equation (10) follows from (9), since $\underline{Q} = \rho$ on $\mathcal{D}_N$ and $\mathcal{D}_N$ is closed under complementation. In order to construct the kernel, select an integer $M \ge N$ such that $(M-N+1)/M \ge 1-\varepsilon/2$, for example,

$$M = \left\lceil \frac{N-1}{\varepsilon/2} \right\rceil \vee 1,$$

and let $\mathcal{A} \equiv \{C \in \mathcal{C}: C \leftrightarrow \{0, 1, \ldots, M-1\}\}$, that is, the class of sets that depend on only the first M coordinates. Note that $\mathcal{A}$ is an algebra, and $(\mathbf{X}, \mathcal{A}, \underline{P}_0)$ is an LP space. Form a new LP $\underline{P}_0$ as follows: if $\sup\{\underline{P}(A): A \in \mathcal{A}^+\} < 1$, let $\underline{P}_0 = \underline{P}$, otherwise let $\underline{P}_0$ be $\underline{P}$ scaled by $1-\varepsilon/2$, that is, $\underline{P}_0 = \underline{P}^{(1-\varepsilon/2)}$. In either case, we have

$$(\forall A \subset \mathbf{X}) \qquad |\underline{P}(A) - \underline{P}_0(A)| \le \varepsilon/2$$

and

$$h \equiv \sup\{\underline{P}_0(A): A \in \mathcal{A}^+\} < 1.$$

If $h > 0$, set $H \equiv \lceil 1/(1-h) \rceil - 1$, otherwise set $H = 1$. Define $\mathcal{K}_0$ by

$$\mathcal{K}_0 = \left\{ \bigcap_{i=1}^{l} T^{Mk_i} A_i \colon 1 \le l \le H, A_i \in \mathcal{A}, k_1, \ldots, k_l \text{ distinct integers} \right\},$$

and on $\mathcal{K}_0$, define ρ_0 by

$$(\forall A \in \mathcal{K}_0) \qquad \rho_0\left(\bigcap_{i=1}^{l} T^{Mk_i} A_i \right) = \left[1 + \sum_{i=1}^{l} \left(\underline{P}_0(A_i) - 1 \right) \right]^+.$$

It is easy to show that $(\mathcal{K}_0, \rho_0)$ is a kernel. The idea is to identify $\mathbf{X} = X^{\mathbf{Z}}$ and $Y^{\mathbf{Z}}$ where $Y = X^{\{0,1,\ldots,M-1\}}$, and apply the construction in Proposition 3.5.

We observe the following:

(i) $(\forall k \in \mathbf{Z}) T^{kM}\mathcal{A} \subset \mathcal{K}_0$, $\rho_0 = \underline{P}_0$ on $T^{kM}\mathcal{A}$.

(ii) $(\mathcal{K}_0, \rho_0)$ is *cyclostationary* with period M:

$$\begin{aligned} T^M \mathcal{K}_0 = T^{-M} \mathcal{K}_0 = \mathcal{K}_0, \\ (\forall A \in \mathcal{K}_0) \qquad \rho_0(T^M A) = \rho_0(T^{-M} A) = \rho_0(A). \end{aligned}$$

(iii) The generated LP $\underline{Q}_0$ is cyclostationary:

$$(\forall A \subset \mathbf{X}) \qquad \underline{Q}_0(T^M A) = \underline{Q}_0(T^{-M} A) = \underline{Q}_0(A),$$

$\underline{Q}_0 = \underline{P}_0$ on $T^{kM}\mathcal{A}$ for all $k \in \mathbf{Z}$, and $\underline{Q}_0 \le \underline{P}_0$.

(iv) $\mathcal{K}_0$ is bounded: $\dim(\mathcal{K}_0) \le H \cdot M$.

For $1 \le j \le M-1$, define the kernels $(\mathcal{K}_j, \rho_j)$ by $\mathcal{K}_j = T^{-j}\mathcal{K}_0$,

$$(\forall A \in \mathcal{K}_j) \qquad \rho_j(A) = \rho_0(T^j A),$$

and let $\underline{Q}_j$'s be the respective generated LP's. We observe that $\underline{Q}_j$ is cyclostationary with period M, $\underline{Q}_j \le \underline{P}_0$ and $\underline{Q}_j = \underline{P}_0$ on $T^{kM-j}\mathcal{A}$ for all $k \in \mathbf{Z}$. Also, we have

$$\begin{aligned}(\forall A \subset \mathbf{X}) \qquad & \underline{Q}_{(j+1)}(A) = \underline{Q}_j(TA) \\ \Rightarrow \quad & \underline{Q}_j(A) = \underline{Q}_0(T^j A),\end{aligned}$$

where the subscript $(j+1)$ denotes addition modulo M.

The desired LP $\underline{Q}$ is defined as

$$\underline{Q} \equiv \frac{1}{M}\sum_{j=0}^{M-1} \underline{Q}_j.$$

For every $A \subset \mathbf{X}$, we have

$$\underline{Q}(TA) = \frac{1}{M}\sum_{j=0}^{M-1} \underline{Q}_j(TA) = \frac{1}{M}\sum_{j=0}^{M-1} \underline{Q}_{(j+1)}(A) = \underline{Q}(A),$$

and similarly $\underline{Q}(T^{-1}A) = \underline{Q}(A)$. Therefore, $\underline{Q}$ is stationary.

Let $\mathcal{K}\{\cup_{j=0}^{M-1} A_j : A_j \in \mathcal{K}_j\}$, and on $\mathcal{K}$ define ρ by setting $\rho = \underline{Q}$. It can be shown that $(\mathcal{K}, \rho)$ generates $\underline{Q}$. Note that $\mathcal{D}_N \subset \mathcal{K}$ and $\mathcal{K}$ is bounded: $\dim(\mathcal{K}) \le (H \cdot M) \cdot M$. The only thing remaining is to show that the ρ approximates $\underline{P}_0$ on $\mathcal{D}_N$. Since for any cylinder $C \in \mathcal{D}_N$, $C \leftrightarrow \{i+1, \dots, i+N\}$ for some $i \in \mathbf{Z}$, C must belong to at least $M-N+1$ shifts of $\mathcal{A}$, and therefore to at least $M-N+1$ of the $\mathcal{K}_j$'s,

$$(\forall C \in \mathcal{D}_N) \qquad \|\{j : C \in \mathcal{K}_j\}\| \ge M - N + 1.$$

Since $\underline{Q}_j = \underline{P}_0$ on $T^{kM-j}\mathcal{A}$, we have

$$\begin{aligned}(\forall A \in \mathcal{D}_N) \qquad \underline{Q}(A) &= \frac{1}{M}\sum_{j=0}^{M-1} \underline{Q}_j(A) \\ &\ge \frac{M-N+1}{M}\underline{P}_0(A) \\ &\ge \underline{P}_0(A) - \varepsilon/2.\end{aligned}$$

Also, since $\underline{Q}_j \leq \underline{P}_0$, we have $\underline{Q} \leq \underline{P}_0$. Therefore,

$$(\forall A \in \mathcal{D}_N) \qquad |\underline{Q}(A) - \underline{P}_0(A)| \leq \varepsilon/2.$$

By construction, $|\underline{P} - \underline{P}_0| \leq \varepsilon/2$. Therefore,

$$(\forall A \in \mathcal{D}_N) \qquad |\underline{Q}(A) - \underline{P}(A)| \leq \varepsilon/2.$$

Since $\mathcal{D}_N \subset \mathcal{K}$ and $\rho = \underline{Q}$ on $\mathcal{K}$, this establishes (9). □

A.3. PROOF OF PROPOSITION 5.7.

Continuity from below. Given cylinders $C_1 \subset C_2 \subset \cdots$, let $C_n \uparrow C_\sigma$. By monotonicity, we have $0 \leq \lim_n \underline{P}(C_n) \leq \underline{P}(C_\sigma)$. Hence, if $\underline{P}(C_\sigma) = 0$, we are all set. Otherwise, if $\underline{P}(C_\sigma) > 0$, for any $\varepsilon > 0$, we can find a $\mathcal{K}$-set K where $\emptyset \neq K \subset C_\sigma$, and

$$0 < \rho(K) \leq \underline{P}(C_\sigma) \leq \rho(K) + \varepsilon.$$

Since $K \in \mathcal{K} \subset C$, K is closed and hence compact. Therefore, $K \subset C_M$ for some $M \geq 1$. We have

$$\begin{aligned} &\underline{P}(C_M) \geq \rho(K) \geq \underline{P}(C_\sigma) - \varepsilon \\ \Rightarrow \quad &\lim_n \underline{P}(C_n) \geq \underline{P}(C_\sigma) - \varepsilon. \end{aligned}$$

Continuity from above. Given cylinders $C_n \downarrow C_\delta$, let $\lambda = \lim_n \underline{P}(C_n)$. By monotonicity, $\underline{P}(C_\delta) \leq \lambda$. Hence, if $\lambda = 0$, we are all set. For the case $\lambda > 0$, we may assume without loss of generality that C_n's are nontrivial. Given $0 < \varepsilon < \lambda$, set

$$\mathcal{K}_{\lambda,\varepsilon} \equiv \{K \in \mathcal{K}: |\rho(K) - \lambda| \leq \varepsilon\}.$$

Assume, by way of contradiction, that

$$\underline{P}(C_\delta) < \lambda - \varepsilon.$$

We can find $N_1 \geq 1$ such that

$$(\forall n \geq N_1)(\exists K_n \in \mathcal{K}_{\lambda,\varepsilon}) \qquad K_n \subset C_n, \qquad \underline{P}(C_n) = \rho(K_n) \geq \lambda - \varepsilon,$$

where K_n's are nontrivial, since C_n's are nontrivial and $\lambda - \varepsilon > 0$. By assumption, all nontrivial K_n's are rectangular, that is,

$$(\forall n \geq N_1) \qquad I(K_n) = \sum_{i=1}^{m_n} I_{n,i}, \qquad B(K_n) = \prod_{i=1}^{m_n} B_{I_{n,i}}(K_n),$$

where $I_{n,i}$'s are mutually disjoint intervals. Define

$$(\forall n \geq N_1) \qquad \mathcal{I}_n \equiv \{I_{n,i} : 1 \leq i \leq m_n\}.$$

Since $\dim(\mathcal{K}) = L$,

$$(\forall n \geq N_1) \qquad \|\mathfrak{I}_n\| \leq L.$$

For all $n \geq N_1$, define

$$\mathfrak{J}_n^1 \equiv \left\{ I \in \mathfrak{I}_n \colon I \cap I(C_{N_1}) \neq \emptyset \right\}.$$

Clearly, all $\mathfrak{J}_n^1$'s are nonempty. Since there can be at most finitely many distinct $\mathfrak{J}_n^1$'s, we can find a sequence of integers $\{n_{1,m}\}$ and an interval I^1 such that

$$N_1 < n_{1,1} < n_{1,2} < \cdots$$

and

$$(\forall m \geq 1) \qquad I^1 \in \mathfrak{J}_{n_{1,m}}^1 \subset \mathfrak{I}_{n_{1,m}}, \qquad B_{I^1}(K_{n_{1,m}}) \equiv B^1.$$

Define the rectangular cylinder K^1 by

$$K^1 \equiv B^1 \times X^{(I^1)^c}.$$

We have

$$(\forall m \geq 1) \qquad K_{n_{1,m}} \subset K^1.$$

We can recursively obtain $K^l, I^l, \{n_{l,m}\}$ with the following properties:

(i) $\{n_{l,m}\} \subset \{n_{l-1,m}\}$, that is, $\{n_{l,m}\}$ is a subsequence of $\{n_{l-1,m}\}$.
(ii) $I^1, \dots, I^l \in \mathfrak{I}_{n_{l,m}}$ for all $m \geq 1$.
(iii) $I^l \perp I^j$ for all $1 \leq j \leq l-1$.
(iv) $I(K^l) = I^1 + \cdots + I^l$, and

$$(\forall m \geq 1) \qquad K_{n_{l,m}} \subset K^l \quad \Rightarrow \quad \underline{P}(K^l) \geq \lambda - \varepsilon.$$

We obtain our desired contradiction by observing that after $L+1$ steps, we have found a subsequence $\{n_{L+1,m}\}$, and $I^1, \dots, I^{L+1}$, where

$$(\forall m \geq 1) \qquad I^1, \dots, I^{L+1} \in \mathfrak{I}_{n_{L+1,m}},$$

and all I^j's are distinct (nonempty and mutually disjoint). Therefore,

$$(\forall m \geq 1) \qquad \|\mathfrak{I}_{n_{L+1,m}}\| \geq L + 1!.$$

Hence

$$\begin{aligned}(\forall\, 0 < \varepsilon < \lambda) \qquad & \lambda - \varepsilon \leq \underline{P}(C_\delta) \leq \lambda \\ \Rightarrow \quad & \underline{P}(C_\delta) = \lambda = \lim_m \underline{P}(C_n).\end{aligned}$$

The recursion proceeds as follows:

At step $l+1$, since $\underline{P}(C_\delta) < \lambda - \varepsilon \le \underline{P}(K^l)$, K^l may be a subset of $C_{n_{l,m}}$ for only finitely many m's. Therefore,

$$(\exists N_{l+1} \in \{n_{l,m}\}) \qquad K^l \not\subset C_{N_{l+1}}.$$

Since

$$(\forall m \ge 1) \qquad K_{n_{l,m}} \subset K^l, \qquad K_{n_{l,m}} \subset C_{n_{l,m}} \subset C_{N_{l+1}},$$

we conclude that for all $n_{l,m} \ge N_{l+1}, K_{n_{l,m}}$ is a *proper* subset of K^l. Define

$$(\forall n_{l,m} \ge N_{l+1}) \qquad \mathfrak{J}_{n_{l,m}}^{l+1} \equiv \left\{ I \in \mathfrak{I}_{n_{l,m}} : I \cap I(C_{n_{l+1}}) \ne \emptyset; I \ne I^1, \dots, I^l \right\}.$$

We claim that all $\mathfrak{J}_{n_{l,m}}^{l+1}$'s are nonempty. Otherwise, if $\mathfrak{J}_{n_{l,m}}^{l+1} = \emptyset$ for some $n_{l,m}$, since $K_{n_{l,m}} \subset C_{N_{l+1}}$, we have

$$B_{(I^1+\cdots+I^l)\cup I(C_{N_{l+1}})}(K_{n_{l,m}}) \subset B_{(I^1+\cdots+I^l)\cup I(C_{N_{l+1}})}(C_{N_{l+1}}).$$

Hence

$$B_{I^1+\cdots+I^l}(K_{n_{l,m}}) \times X^{I(C_{N_{l+1}})\setminus(I^1+\cdots+I^l)} \subset B(C_{N_{l+1}}) \times X^{(I_1+\cdots+I^l)\setminus I(C_{N_{l+1}})}.$$

By forming the Cartesian product of both sides of the above with $X^{[(I^1+\cdots+I^l)\cup I(C_{N_{l+1}})]^c}$, we obtain

$$K^l \subset C_{N_{l+1}}!.$$

Since there can be at most finitely many distinct $\mathfrak{J}_{n_{l,m}}^{l+1}$'s (with m varying), we can find a subsequence $\{n_{l+1,m}\}$ of $\{n_{l,m}\}$ and an interval I^{l+1} such that

$$N_{l+1} \le n_{l+1,1} < n_{l+1,2} < \cdots$$

and

$$(\forall m \ge 1) \qquad I^{l+1} \in \mathfrak{J}_{n_{l+1,m}}^{l+1} \subset \mathfrak{I}_{n_{l,m}}, \qquad B_{I^{l+1}}(K_{n_{l+1,m}}) \equiv B^{l+1}.$$

We define the rectangular cylinder K^{l+1} by

$$K^{l+1} \equiv \prod_{i=1}^{l+1} B^i \times X^{(\sum_{i=1}^{l+1} I^i)^c},$$

which satisfies

$$\begin{aligned} (\forall m \ge 1) \qquad & K_{n_{l+1,m}} \subset K^{l+1} \\ \Rightarrow \quad & \underline{P}(K^{l+1}) \ge \lambda - \varepsilon, \end{aligned}$$

thereby concluding the recursion step. □

REFERENCES

ANDREWS, S. (1985). Flicker noise: parameter estimation and data analysis. M.S. dissertation, Cornell Univ.

ANGER, B. (1977). Representation of capacities. *Math. Ann.* **229** 245–258.

CHOQUET, G. (1953/54). Theory of capacities. *Ann. Inst. Fourier* **9** 83–89.

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

FINE, T. L. (1988). Lower probability models for uncertainty and nondeterministic processes. *J. Statist. Plann. Inference* **20** 389–411.

GRIZE, Y.-L. (1984). Towards a stationary continuous lower probability-based model for flicker noise. Ph.D. dissertation, Cornell Univ.

GRIZE, Y.-L. and FINE, T. L. (1987). Continuous lower probability-based models for stationary processes with bounded and divergent time averages. *Ann. Probab.* **15** 783–803.

KUMAR, A. (1982). Lower probabilities on infinite spaces and instability of stationary sequences. Ph.D. dissertation, Cornell Univ.

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

PAPAMARCOU, A. (1983). Some results on undominated lower probabilities. M.S. dissertation, Cornell Univ.

PAPAMARCOU, A. (1987). Unstable random sequences as an objective basis for interval-valued probability models. Ph.D. dissertation, Cornell Univ.

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

PAPAMARCOU, A. and FINE, T. L. (1991). Stationarity and almost sure divergence of time averages in interval valued probability. *J. Theoret. Probab.* **4** 239–260.

RAMAKRISHNAN, S. and SUDDERTH, W. D. (1988). A sequence of coin toss variables for which the strong law fails. *Amer. Math. Monthly* **95** 939–941.

RAO, K. P. S. and RAO, M. (1983). *Theory of Charges.* Academic, London.

SADROLHEFAZI, A. (1990). Interval-valued probability based time series models with measure-like marginals. Ph.D. dissertation, Cornell Univ.

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

WALLEY, P. (1981). Coherent lower and upper probabilities. Dept. Statistics, Univ. Warwick, Coventry.

WALLEY, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

WALLEY, P. and FINE, T. L. (1982). Towards a frequentist theory of upper and lower probability. *Ann. Statist.* **10** 742–761.

MORGAN STANLEY & CO.
NEW YORK, NEW YORK 10020

SCHOOL OF ELECTRICAL ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853