The Annals of Probability
1991, Vol. 19, No. 2, 893–906

UNSTABLE COLLECTIVES AND ENVELOPES OF PROBABILITY MEASURES[1]

BY ADRIAN PAPAMARCOU AND TERRENCE L. FINE

University of Maryland and Cornell University

We discuss issues of existence and stochastic modeling in regard to sequences that exhibit combined features of independence and instability of relative frequencies of marginal events. The concept of independence used here is borrowed from the frequentist account of numerical probability advanced by von Mises: A sequence is independent if certain salient asymptotic properties are invariant under the causal selection of subsequences. We show that independence (in the above sense) and instability of relative frequency are indeed compatible and that sequences with such features support stochastic models expressed in terms of envelopes of probability measures.

1. Introduction. The concept of the collective first appeared in 1919 in the frequentist theory of numerical probability proposed by von Mises (1957, 1964). The empirical basis for that theory was the long-run stability of the relative frequencies of events in unlinked repetitions of nondeterministic experiments. von Mises postulated that the outcomes of such repetitions could be modeled by numerical sequences with specific asymptotic properties. These sequences, called *collectives*, facilitated a coherent definition of finitely additive probability in terms of limits of relative frequencies and led to an illuminating frequentist interpretation of key ideas in probability theory.

Our aim here is to explore parallels to the von Mises account, in pursuit of a frequentist justification for an alternative aleatory concept. In particular, we are interested in identifying numerical sequences that support *upper* and *lower envelopes* much in the same way that von Mises collectives support numerical probabilities. Upper and lower envelopes are set functions derived from a given class $\mathscr{M}$ of probability measures by taking the eventwise supremum (resp., infimum) of the measures in $\mathscr{M}$. They are subadditive and superadditive, respectively, and belong to the broader class of upper and lower (or interval-valued) probabilities.

Upper and lower envelopes have been primarily studied from a subjectivist viewpoint. In Smith (1961), Williams (1976) and Walley (1981), these structures are introduced as extremal betting rates subject to axioms of rational, or coherent, behavior. A statistical theory based on the above interpretation is developed in Walley (1990). Dempster (1968) and Shafer (1976) have employed

---

Received July 1989.
[1]Research supported in part by NSF Grant ECS-88-09325.
AMS 1980 *subject classifications*. 60G05, 60A05.
Key words and phrases. Interval-valued probability, upper and lower probability, upper and lower envelopes, unstable relative frequencies, frequentist probability, von Mises collectives, independence, subsequence selections.

the subclass of lower envelopes known as belief functions to represent degrees of belief and their dependence on evidence. Upper envelopes have also been used to specify model uncertainties in robust hypothesis testing [Huber and Strassen (1973), Huber (1981)].

As in Walley and Fine (1982), this paper departs from the subjectivist standpoint to consider upper and lower envelopes in an objective setting: namely, that of a hypothetical experiment which is repeated indefinitely under causally independent conditions, yet fails to produce outcomes whose relative frequencies converge. This may at first seem paradoxical, in view of the apparent convergence of relative frequencies encountered in practice. Nevertheless, the distinction between apparent and eventual (or actual) should be borne in mind: Eventual convergence or divergence cannot be established on the basis of any finite observation record, whereas apparent trends are always observable and quantifiable. Thus apparent convergence and eventual divergence of relative frequency are in principle reconcilable, especially when the fluctuations associated with eventual divergence occur at a deceptively slow rate. We also suspect that there are instances in which nondeterministic phenomena (e.g., speech) have exhibited unstable relative frequencies and that these instabilities have been casually explained away by invocations of nonstationarities in the process that had little independent substantiation.

A salient connection between envelopes and relative frequencies was established in Walley and Fine [(1982), Theorem 4.2, page 750]. Given a finite marginal sample space X, an algebra $\mathscr{A}$ of subsets of X and a sequence $\mathbf{x} = (x_1, x_2, \dots)$ in $X^{\mathbb{N}}$, denote by $r_n(A; \mathbf{x})$ the relative frequency of an event $A \in \mathscr{A}$ in the first n terms of $\mathbf{x}$,

$$r_n(A; \mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} I_A(x_i).$$

Then the set functions $\limsup_n r_n(\cdot\,; \mathbf{x})$ and $\liminf_n r_n(\cdot\,; \mathbf{x})$ are upper (resp., lower) envelopes on $(X, \mathscr{A})$. In other words, there exists a class $\mathscr{M}_{\mathbf{x}}$ of probability measures on that measurable space such that for every event A,

$$\limsup_n r_n(A; \mathbf{x}) = \sup\{\mu(A) : \mu \in \mathscr{M}_{\mathbf{x}}\},$$

$$\liminf_n r_n(A; \mathbf{x}) = \inf\{\mu(A) : \mu \in \mathscr{M}_{\mathbf{x}}\}.$$

Furthermore, the above envelopes are finitely generated, that is, the class $\mathscr{M}_{\mathbf{x}}$ can be chosen to be finite.

In the degenerate case $\mathscr{M}_{\mathbf{x}} = \{\mu_{\mathbf{x}}\}$, we have almost sure convergence of relative frequencies of all marginal events in $\mathscr{A}$, so that

$$\forall A \in \mathscr{A}, \qquad \lim_n r_n(A; \mathbf{x}) = \mu_{\mathbf{x}}(A).$$

By the strong law of large numbers, the above relationship will hold for a set of sequences $\mathbf{x}$ of probability 1 under the i.i.d. measure with marginal $\mu_{\mathbf{x}}$. It is natural to ask whether a similar statement can be made for the nondegenerate case, where $\mathscr{M}_{\mathbf{x}}$ contains more than one measure and where the relative frequency of at least one event diverges. In other words, does there exist an i.i.d. *envelope* model that strongly supports the oscillation of relative frequencies with (asymptotically) extremal values specified by the marginals of that model? If so, then one could envisage a series of unlinked repetitions of a nondeterministic experiment with relative frequencies that fail to stabilize eventually.

The above question was one of the main issues addressed in Walley and Fine (1982). In that work, a defensible method of extending an arbitrary pair of upper and lower envelopes defined on $(X, \mathscr{A})$ to an i.i.d. envelope model on the algebra $\mathscr{C}$ of cylinder events of $X^{\mathbb{N}}$ was proposed. It was shown, however, that the resulting i.i.d. model only supported the (apparent) divergence of relative frequencies in a weak sense. Kumar and Fine (1985) further examined the support that a stationary model defined directly on the power set of $X^{\mathbb{N}}$ could lend to eventual divergence. Their study concluded that any monotone set function possessing basic monotone continuity properties and dominated on $\mathscr{C}$ by a measure must fail to support the divergence of relative frequencies.

Our work was motivated by an inquiry into the above results regarding the construction of an i.i.d. envelope model and the notion of support for divergence. The following two questions arose: How does one construe independence in the nonadditive case in which the model is specified through two set functions, an upper and a lower envelope? And what is the objective meaning of the (possibly distinct) values assumed by these two functions on any given event?

We have found the von Mises approach to similar interpretational issues in numerical probability particularly helpful. The primary stochastic entity in that theory is the collective, which represents the outcomes of unlinked repetitions of an experiment. By taking limits of relative frequencies of cylinder sets along the collective, one *obtains* the i.i.d. stochastic model P for the repeated trials. In that framework, the role of P is limited to summarizing the asymptotic relative frequency characteristics of the collective and probability statements are strictly interpreted in terms of frequencies. In the same vein, the independence relationships derived from P reflect those built-in frequency properties of the collective that are consistent with an intuitive notion of causal unlinkedness.

In this paper we seek answers to the following questions. First, do there exist sequences that resemble von Mises collectives in possessing features typical of causal independence, yet differ from collectives by exhibiting unstable relative frequency behavior? Second, to what extent do such unstable collectives support, in a frequentist sense, an i.i.d. envelope model whose values coincide with the limits inferior and superior of the relative frequencies? We show that the answer to the first question is affirmative and we propose an answer to the second question assuming the existence of sequences with additional asymptotic relative frequency properties.

2. Invariance of asymptotic properties under the selection of subsequences. We begin this section with a brief discussion of von Mises collectives. For more details, see von Mises (1957, 1964), Martin-Löf (1969), Schnorr (1971) and Kolmogorov and Uspenskii (1987).

Intuitively, unlinked trials of a nondeterministic experiment should be mutually noninformative. Thus given an infinite sequence $\mathbf{x}$ of outcomes of such trials, any subsequence of $\mathbf{x}$ selected in a causal, or nonpredictive, fashion, should itself share the same salient properties as $\mathbf{x}$. In the von Mises theory, these properties are identified as the limits of relative frequencies of marginal events and sequences $\mathbf{x}$ that represent the outcomes of unlinked trials are termed collectives. Thus informally, a von Mises collective is a sequence along which the relative frequencies of events converge to limits that are invariant under the causal selection of subsequences [von Mises (1957)]. We note here that invariance under causal subsequence selection does not suffice to guarantee all of the properties commonly held by probabilistically random sequences (e.g., law of iterated logarithm); however, this invariance is a necessary first step in characterizing such sequences. For a review of the main contributions to the definition of a random sequence, see Kolmogorov and Uspenskii (1987).

Mathematically, causal selection can be defined in terms of a binary $\{0, 1\}$-valued function f on the space X^* of finite strings from X (including the empty string $\mathbf{e}$). Given any sequence $\mathbf{x}$ in $X^{\mathbb{N}}$, one can generate a finite or infinite subsequence $\mathbf{x}_\tau$ of $\mathbf{x}$ by taking $x_0 = \mathbf{e}$ and applying the following recursion: For every n, append x_n to the current string of $\mathbf{x}_\tau$ if and only if $f(x_0, \dots, x_{n-1}) = 1$. Such functions are commonly called *place selection rules*.

Henceforth we will assume that X is a finite space and $\mathscr{A}$ is the power set of X. $\mathscr{C}$ denotes the algebra of cylinder sets of $X^{\mathbb{N}}$.

DEFINITION. Let $\mathscr{H}$ be a family of place selection rules which includes the identity $f \equiv 1$ and μ be a probability measure on $(X, \mathscr{A})$. A sequence $\mathbf{x}$ in $X^{\mathbb{N}}$ is a *$(\mathscr{H}, \mu)$-collective* if every infinite subsequence $\mathbf{x}_\tau = (x_{\tau_1}, x_{\tau_2}, \dots)$ generated by a place selection rule f in $\mathscr{H}$ satisfies

$$\forall A \subset X, \qquad \lim_n r_n(A; \mathbf{x}_\tau) = \mu(A).$$

The existence of sequences with the above invariance property depends on the choice of the family $\mathscr{H}$. In the least restrictive case where $\mathscr{H}$ contains every binary function on X^*, such sequences do not exist except for degenerate measures μ. This is so because for every sequence $\mathbf{x}$, there exists a place selection rule that yields a *constant* subsequence $\mathbf{x}_\tau$. To remove such obstacles, one may impose a further requirement that all rules in $\mathscr{H}$ represent systematic, rather than haphazard, selection procedures. This intuitively natural restriction on $\mathscr{H}$ can be made precise by utilizing the concept of computability: A rule f is *computable* if there exists a Turing machine T_f that can determine $f(\mathbf{s})$ in a finite number of iterations for every $\mathbf{s} \in X^*$. As there are only countably many computable rules, any family $\mathscr{H}$ solely consisting of such rules is countable and the following theorem of Wald (1936) applies to $\mathscr{H}$.

THEOREM 2.1. *There exist $(\mathscr{H}, \mu)$-collectives for every countable family $\mathscr{H}$ of place selection rules on X^* and probability measure μ on $(X, \mathscr{A})$. Furthermore, the set of such sequences has probability* 1 *under the i.i.d. product of μ on $(X^{\mathbb{N}}, \sigma(\mathscr{C}))$.*

Thus collectives exist for any countable family of place selection rules and the ensuing development of the von Mises theory (based solely on computable rules) is nonvacuous.

In what follows we prove an analogue of the previous existence result for sequences with unstable relative frequencies. The asymptotic properties of interest here are the limits superior and inferior of the relative frequencies of the marginal events. As discussed in the introduction, for any sequence $\mathbf{x}$, these limit points are given by the upper and lower envelopes (respectively) of a class $\mathscr{M}_{\mathbf{x}}$ of probability measures on the marginal space. Our question is whether there exist sequences $\mathbf{x}$ for which the limits superior and inferior of the relative frequencies differ, yet are preserved under a countable family of place selection rules. The answer is affirmative and the precise statement of our result is given in:

THEOREM 2.2. *Let $\mathscr{H}$ be a countable family of place selection rules on X^* and $\mathscr{M}$ be a family of probability measures on $(X, \mathscr{A})$. Then there exists a sequence $\mathbf{x}$ such that for every infinite subsequence $\mathbf{x}_\tau$ generated by a rule in $\mathscr{H}$,*

$$\forall A \subset X, \qquad \begin{aligned} \limsup_n r_n(A; \mathbf{x}_\tau) &= \sup\{\mu(A) : \mu \in \mathscr{M}\}, \\ \liminf_n r_n(A; \mathbf{x}_\tau) &= \inf\{\mu(A) : \mu \in \mathscr{M}\}. \end{aligned}$$

REMARK. The two relationships in the above statement are equivalent, since $r_n(A; \mathbf{y}) + r_n(A^c; \mathbf{y}) = 1$ and $\sup\{\mu(A): \mu \in \mathscr{M}\} + \inf\{\mu(A^c): \mu \in \mathscr{M}\} = 1$.

In the argument that follows, we will construct a probability measure P on $(X^{\mathbb{N}}, \sigma(\mathscr{C}))$ such that every rule in $\mathscr{H}$ almost surely selects a subsequence with the required asymptotic properties. Then by the countability of $\mathscr{H}$, the set of sequences satisfying the statement of Theorem 2.2 will have P-measure 1. This technique was used by Wald to prove Theorem 2.1 with a particularly simple choice of P, namely the i.i.d. product of μ. In our case, the construction of P will be somewhat more complex.

PROOF OF THEOREM 2.2. (A) *Preliminaries*. Consider the probability space $(X^{\mathbb{N}}, \sigma(\mathscr{C}), P)$, where P is specified by the family $\{P(\cdot|\mathbf{s}), \mathbf{s} \in X^*\}$ of conditional distributions on $(X, \mathscr{A})$. Denote the generic element of $X^{\mathbb{N}}$ by $\mathbf{x}$ (as before) and the tth coordinate function on $X^{\mathbb{N}}$ by Z_t, that is, $Z_t(\mathbf{x}) = x_t$. In keeping with earlier notation, set Z_0 equal to the empty string $\mathbf{e}$. For every $t \geq 0$, define $\mathscr{F}_t = \sigma(Z_0, \dots, Z_t)$ and let $\mathscr{F}_\infty = \sigma(\mathscr{C})$.

Consider a fixed place selection rule $f \in \mathscr{H}$. The random sequence of indices selected by f (in increasing order) is denoted by $\tau = (\tau_i, i \in \mathbb{N})$, and is given by the recursion

$$\tau_i = \min\{t: t > \tau_{i-1} \vee 0 \text{ and } f(Z_0, \dots, Z_{t-1}) = 1\}.$$

Here we take $\min \varnothing = +\infty$, so that $\tau_k(\mathbf{x}) = +\infty$ if and only if f selects a total of fewer than k terms from the sequence $\mathbf{x}$. It is easy to show that for every i and every t (including $+\infty$), the event $\{\tau_i - 1 = t\}$ lies in $\mathscr{F}_t$, so that $\sigma_i =_{\text{def}} \tau_i - 1$ is a stopping time of the coordinate process.

We are interested in the asymptotic behavior of the time average

$$r_n(A; \mathbf{Z}_\tau) = \frac{1}{n} \sum_{i=1}^{n} I_A(Z_{\tau_i})$$

over the event $D_f = \bigcap_i \{\tau_i < \infty\}$, for every $A \subset X$. This behavior is linked to the conditional distributions $P(\cdot|\mathbf{s})$ through the following variant of the strong law of large numbers [Theorem 32.1.E, Loève (1978)]. For every increasing sequence $(\mathscr{B}_i, i \in \mathbb{N})$ of σ-fields in $\sigma(\mathscr{C})$ and every uniformly bounded adapted process $(Y_i, i \in \mathbb{N})$,

$$\lim_n \frac{1}{n} \sum_{i=1}^{n} \left(Y_{i+1} - E[Y_{i+1}|\mathscr{B}_i]\right) = 0 \quad \text{a.s.}$$

To apply this result to the problem in hand, we take $\mathscr{B}_i = \mathscr{F}_{\sigma_i}$, $Y_i = I_A(Z_{\tau_{i-1}})$ and for definiteness, let Z_∞ be constant over $X^{\mathbb{N}}$. Then by the above theorem,

$$\liminf_n \frac{1}{n} \sum_{i=1}^{n} I_A(Z_{\tau_i}) = \liminf_n \frac{1}{n} \sum_{i=1}^{n} P\left[Z_{\tau_i} \in A | \mathscr{F}_{\sigma_i}\right] \quad \text{a.s.}$$

In other words, there exists a set O_f of P measure 1 such that for every sequence in $\mathbf{x} \in O_f \cap D_f$ and every marginal event A,

$$\liminf_n r_n(A; \mathbf{x}_\tau) = \liminf_n \frac{1}{n} \sum_{i=1}^{n} P\left(A | x_0, \dots, x_{\tau_i - 1}\right).$$

Now consider the entire family $\mathscr{H}$. If we can specify $\{P(\cdot|\mathbf{s}): \mathbf{s} \in X^*\}$ in such a way that every infinite subsequence $\mathbf{y}_\tau$ generated by any rule f in $\mathscr{H}$ satisfies

$$\liminf_n \frac{1}{n} \sum_{i=1}^{n} P\left(A | y_0, \dots, y_{\tau_i - 1}\right) = \inf\{\mu(A): \mu \in \mathscr{M}\},$$

then the disjunction

$$\mathbf{x}_\tau \text{ finite} \quad \text{or} \quad \liminf_n r_n(A; \mathbf{x}_\tau) = \inf\{\mu(A): \mu \in \mathscr{M}\}$$

will hold for every $\mathbf{x}$ in $\bigcap_{f \in \mathscr{H}} O_f$. Since $\mathscr{H}$ is assumed countable, the last intersection will have probability 1 and the proof will be complete.

(B) *The construction of* P. Let $X = \{1, \dots, J\}$. It is convenient to think of X^* as a J-ary tree, with nodes corresponding to strings $\mathbf{s} \in X^*$. Specifying the family $\{P(\cdot|\mathbf{s}): \mathbf{s} \in X^*\}$ then amounts to assigning a measure $P(\cdot|\mathbf{s})$ to each node $\mathbf{s}$. We shall do so by recursion on a suitable ordering of X^*.

Let $2^X = \{A_1, \dots, A_K\}$. Since $\mathscr{M}$ is isomorphic to a bounded subset of $\mathbb{R}^{J-1}$, there exists a finite collection $\mathscr{M}' = \{\mu_1, \dots, \mu_K\} \subset \mathrm{cl}(\mathscr{M})$ such that

$$\forall\, 1 \le k \le K, \qquad \mu_k(A_k) = \inf\{\mu(A_k) : \mu \in \mathscr{M}\} = \min\{\mu_j(A_k) : \mu_j \in \mathscr{M}'\}.$$

In what follows, the conditional distributions $P(\cdot|\mathbf{s})$ will be drawn exclusively from the collection $\mathscr{M}'$.

Underlying our construction is the following idea. Suppose that for a certain rule f and an infinite subsequence $\mathbf{x}_\tau$ selected by f, the sequence of measures

$$\left(P\left(\cdot|x_0, \dots, x_{\tau_i - 1}\right), i \in \mathbb{N}\right)$$

consists of successive blocks of identical elements of $\mathscr{M}'$; for example, a block of N_1 μ_1's, followed by a block of N_2 μ_2's, and so on, continuing in a cyclical fashion. Then the sought relationship

$$(1) \qquad \forall\, 1 \le k \le K, \qquad \liminf_n \frac{1}{n} \sum_{i=1}^{n} P\left(A_k|x_0, \dots, x_{\tau_i - 1}\right) = \mu_k(A_k)$$

will hold provided the block lengths grow fast enough, for example, if

$$\lim_k N_k(N_1 + \cdots + N_{k-1})^{-1} = +\infty.$$

Although the above effect is easily achieved for all infinite subsequences drawn from any single rule f, it is clearly infeasible in the general case of two or more rules with conflicting selection patterns. Our construction resolves such conflicts by allowing contaminations of blocks of identical measures: For fixed f, the sequence $(P(\cdot|x_0, \dots, x_{\tau_i - 1}), i \in \mathbb{N})$ may exhibit deliberate irregularities which serve to satisfy the demands of competing rules without affecting the limit inferior of the time averages.

Briefly, in recursively assigning conditional measures to the nodes corresponding to the initial substrings of a sequence $\mathbf{x}$, we are guided by a countable hierarchy of constraints we call *phases*. Each such phase has associated with it a characteristic place selection rule f, a measure μ and a bin of size N. Initially, all phases have empty bins; at any given step of the recursion on $\mathbf{x}$, a phase is said to be *complete* if the corresponding bin is full.

To determine the conditional distribution at node $(x_0, \dots, x_n)$, we consider all incomplete phases whose corresponding rules f satisfy $f(x_0, \dots, x_n) = 1$. The highest such phase in the hierarchy wins node $(x_0, \dots, x_n)$: it sets $P(\cdot|x_0, \dots, x_n)$ equal to the corresponding measure μ and also places a copy of that node in its bin. By suitable choice of the triplets (f, μ, N), one can obtain the desired effect, that is, time averages satisfying (1) above for *all* infinite subsequences of $\mathbf{x}$ drawn by rules from $\mathscr{H}$. The details are as follows.

For $q \in \mathbb{N}$, let $b(q) \in \mathbb{N}$, $c(q) \in \{1, \dots, K\}$ and $N_q \in \mathbb{N}$ be such that

(i) for every (b, c) in $\mathbb{N} \times \{1, \dots, K\}$, the pair $(b(q), c(q))$ equals (b, c) for infinitely many values of q; and

(ii) the N_q's satisfy

$$\lim_q \frac{N_q}{N_1 + \cdots + N_{q-1}} = +\infty. \tag{2}$$

Here $f_{b(q)}$ is the place selection rule, $\mu_{c(q)}$ the measure and N_q the bin size corresponding to the qth phase in the hierarchy.

Given below is the recursion for $\{P(\cdot|\mathbf{s}), \mathbf{s} \in X^*\}$. We assume a linear ordering of J-ary strings by increasing size, with ties broken arbitrarily.

R1. Initially let $\mathbf{s} = \mathbf{e}$ and $m_q(\mathbf{s}) = 0$ for all $q \in \mathbb{N}$.
R2. Let $\Theta(\mathbf{s}) = \{q: m_q(\mathbf{s}) < N_q\}$.
R3. Let $r(\mathbf{s})$ be the least index q in $\Theta(\mathbf{s})$ such that $f_{b(q)}(\mathbf{s}) = 1$.
R4. Let $P(\cdot|\mathbf{s}) \equiv \mu_{c(r(\mathbf{s}))}(\cdot)$.
R5. For every $a \in X$, let $m_{r(\mathbf{s})}(\mathbf{s}a) = m_{r(\mathbf{s})}(\mathbf{s}) + 1$.
R6. Go to the next string $\mathbf{s}$.

The recursion will be valid iff every $\mathbf{s} \in X^*$ is selected by some rule in $\mathscr{H}$; this is ensured by assuming w.l.o.g. that the identity rule $f \equiv 1$ is in $\mathscr{H}$.

(C) *The validity of* (1). Let $f_b \in \mathscr{H}$ and $\mathbf{x}$ be such that the subsequence $\mathbf{x}_\tau$ generated by f_b is infinite. Again, we write $\sigma_i = \tau_i - 1$.

Consider any q for which $b(q) = b$, and take L to exceed the sum of the bin sizes of all phases up to and including the qth one, for example, $L = 1 + N_1 + \cdots + N_q$. We claim that $m_q(x_0, \dots, x_{\sigma_L}) = N_q$, that is, the qth phase is complete at node $(x_0, \dots, x_{\sigma_L})$.

Indeed, assume the contrary, namely $m_q(x_0, \dots, x_{\sigma_L}) < N_q$. Since for every $i < L$ we have that

$$f_{b(q)}(x_0, \dots, x_{\sigma_i}) = f_b(x_0, \dots, x_{\tau_i - 1}) = 1,$$

our assumption implies that q lies in $\Theta(x_0, \dots, x_{\sigma_i})$ for all $i \le L$. But then for each such value of i, $r(x_0, \dots, x_{\sigma_i}) \le q$. As there are at most $N_1 + \cdots + N_q = L - 1$ substrings $\mathbf{s}$ of $\mathbf{x}$ such that $r(\mathbf{s}) \le q$, the sought contradiction is obtained.

Now consider a marginal event A_k. By construction of the sequences $b(\cdot)$ and $c(\cdot)$, for any q_0 there exists $q > q_0$ such that $b(q) = b$, $c(q) = k$. Denoting by T the least index i such that $m_q(x_0, \dots, x_{\sigma_i}) = N_q$, we have by the foregoing discussion that

(i) $N_q \le T \le 1 + N_1 + \cdots + N_q$;

(ii) at least N_q of the summands in $\sum_{i=1}^T P(\cdot|x_0, \dots, x_{\sigma_i})$ are equal to $\mu_{c(q)} = \mu_k$. Thus

$$\left| \frac{1}{T} \sum_{i=1}^{T} P\big(A_k | x_0, \dots, x_{\sigma_i}\big) - \mu_k(A_k) \right| \le \frac{1 + N_1 + \cdots + N_{q-1}}{N_q},$$

where by (2), the right-hand side can be made arbitrarily small by choosing q_0 sufficiently large. Since $\mu_k(A_k) \le \mu_j(A_k)$ for $j \ne k$, we also have

$$\forall\, n, \qquad \frac{1}{n} \sum_{i=1}^{n} P\big(A_k | x_0, \ldots, x_{\sigma_i}\big) \ge \mu_k(A_k),$$

whence we conclude that

$$\liminf_n \frac{1}{n} \sum_{i=1}^{n} P\big(A_k | x_0, \ldots, x_{\sigma_i}\big) = \mu_k(A_k).$$

Hence we have shown that for almost all (P-measure) $\mathbf{x}$, if $f \in \mathscr{H}$ yields an infinite subsequence $\mathbf{x}_\tau$, then

$$\forall\, 1 \le k \le K, \qquad \liminf_n r_n(A_k; \mathbf{x}_\tau) = \mu_k(A_k). \qquad \square$$

Thus there exist sequences that exhibit unstable relative frequencies and possess invariance characteristics akin to those of von Mises collectives. In the following section, we discuss the possibility of using such sequences to represent the outcomes of unlinked trials governed by i.i.d. envelope models.

3. On the derivation of i.i.d. envelope models. As pointed out in the introduction, the collective is the primary entity in the von Mises theory and the concept of probability is essentially a derived one. Thus given a random experiment $\mathscr{E}$ with sample space X, the probability of an event $A \subset X$ is given by the limit of the relative frequency of A along a collective $\mathbf{x}$ that represents the outcomes of unlinked trials of $\mathscr{E}$.

The derivation of the i.i.d. model on $\mathscr{E}$ is based on the following idea. Let $\mathbf{x}$ be the sequence of outcomes of unlinked repetitions of $\mathscr{E}$. Then a partition of $\mathbf{x}$ into successive blocks of length k should yield a sequence representing the outcomes of unlinked repetitions of the experiment $\mathscr{E}^k$, where $\mathscr{E}^k$ itself consists of k unlinked trials of $\mathscr{E}$. In other words, if $\mathbf{x}$ is a collective for $\mathscr{E}$, then the sequence $\mathbf{x}^k \in (X^k)^{\mathbb{N}}$ defined by

$$\mathbf{x}^k = \big((x_1, \ldots, x_k), (x_{k+1}, \ldots, x_{2k}), \ldots\big)$$

should also be a collective for $\mathscr{E}^k$.

It turns out that the above is true for appropriate place selection systems $\mathscr{H}$ and $\mathscr{H}^k$ on X^* and $(X^k)^*$, respectively (the simplest such choices are the classes of computable functions on the respective spaces). More importantly, it can be shown [von Mises (1957, 1964)] that the probability measure governing $\mathscr{E}^k$ is the k-fold product of the measure $\mu_{\mathbf{x}}$ governing $\mathscr{E}$. Thus from a single collective $\mathbf{x}$, it is possible to establish the i.i.d. model on $\mathscr{E}$ as the *unique* stochastic model for the unlinked repetitions of $\mathscr{E}$.

One might expect that a parallel derivation is feasible for i.i.d. envelope models, using sequences such as those in Theorem 2.2. Unfortunately, this does not seem to be the case: The fact that $\liminf_n r_n(A, \mathbf{x}_\tau)$ is the same for all subsequences $\mathbf{x}_\tau$ selected from $\mathbf{x}$ does not readily imply that $\mathbf{x}^k$ possesses similar invariance properties. To escape this impasse, we discuss i.i.d. envelope models issuing from sequences whose invariance properties encompass the asymptotics of relative frequencies of *all* cylinder sets and which we term *unstable collectives*. In what follows, $T^i(\mathbf{x})$ denotes the left shift of $\mathbf{x}$ by i positions.

DEFINITION. Let $\mathscr{H}$ be a family of place selection rules which includes the identity $f \equiv 1$ and $\underline{P}$ be a set function on the algebra $\mathscr{C}$ of cylinder sets of $X^{\mathbb{N}}$. A sequence $\mathbf{x}$ in $X^{\mathbb{N}}$ is an *unstable* $(\mathscr{H}, \underline{P})$*-collective* if for every infinite sequence $\tau = (\tau_1, \tau_2, \dots)$ generated by a place selection rule in $\mathscr{H}$,

$$\forall\, C \in \mathscr{C}, \qquad \liminf_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{\tau_i - 1}\mathbf{x}) = \underline{P}(C).$$

Note that in the previous definition, relative frequency is understood in the context of a sequence of translations of $\mathbf{x}$. Also, if we let $\overline{P}(C) = 1 - \underline{P}(C^c)$, the defining relationship becomes equivalent to

$$\forall\, C \in \mathscr{C}, \qquad \limsup_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{\tau_i - 1}\mathbf{x}) = \overline{P}(C).$$

Theorem 3.1 effectively states that an unstable collective representing the outcomes of an experiment $\mathscr{E}$ gives rise to a unique envelope model for unlinked repetitions of $\mathscr{E}$ (such as $\mathscr{E}^k$).

THEOREM 3.1. *Let $\mathscr{H}$ and $\mathscr{H}^k$ denote the classes of computable binary functions on X^* and $(X^k)^*$, respectively. Then for every unstable $(\mathscr{H}, \underline{P})$-collective, the following statements are true*:

(i) *$\underline{P}$ is the lower envelope of a class of stationary probability measures on $\mathscr{C}$.*

(ii) *There exists on the algebra $\mathscr{C}^k$ of cylinder sets of $(X^k)^{\mathbb{N}}$ a unique lower envelope $\underline{P}^k$ whose marginal on X^k agrees with that of $\underline{P}$ and for which the sequence $\mathbf{x}^k$ is an unstable $(\mathscr{H}_k, \underline{P}_k)$-collective.*

PROOF. (i) The first assertion follows from the lemma given in the Appendix, which states that for every sequence $\mathbf{y}$ in $X^{\mathbb{N}}$, there exists a countable family $\mathscr{Q}_{\mathbf{y}}$ of stationary measures on $\mathscr{C}$ such that

$$\forall\, C \in \mathscr{C}, \qquad \liminf_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{i-1}\mathbf{y}) = \min\{\mu(C), \mu \in \mathscr{Q}_{\mathbf{y}}\}.$$

For the case in hand, we take $f \equiv 1$ in the definition of the unstable collective and note that

$$\liminf_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{i-1}\mathbf{x}) = \liminf_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{\tau_i - 1}\mathbf{x}) = \underline{P}(C).$$

(ii) Let f^k be a computable place selection rule on $(X^k)^*$ that selects infinitely many terms (i.e., k-tuples) from $\mathbf{x}^k$ and let $\nu = (\nu_i, i \in \mathbb{N})$ be the sequence of selected indices. To prove that $\mathbf{x}^k$ is an unstable collective, we must show that the relationship

$$\underline{P}^k(\cdot) = \liminf_n \frac{1}{n} \sum_{i=1}^{n} I_{\cdot}(T^{\nu_i - 1}\mathbf{x}^k)$$

defines a unique set function $\underline{P}^k$ on $\mathscr{C}^k$ which is independent of the choice of f^k.

Consider an arbitrary set $C^k \in \mathscr{C}^k$ and let m be the dimension of C^k. Then for some $S \in X^{km}$, we have

$$C^k = \{\mathbf{y}^k : \mathbf{y} \in X^{\mathbb{N}}, (y_1, \ldots, y_{km}) \in S\}.$$

Also let C be the (km)-dimensional cylinder in $\mathscr{C}$ comprising the sequences with initial segments in S,

$$C = \{\mathbf{y} : \mathbf{y} \in X^{\mathbb{N}}, (y_1, \ldots, y_{km}) \in S\}.$$

Noting that the ν_ith k-tuple of $\mathbf{x}^k$ begins with $x_{k(\nu_i - 1)+1}$, we have

$$I_{C^k}(T^{\nu_i - 1}\mathbf{x}^k) = I_S(x_{k(\nu_i - 1)+1}, \ldots, x_{k(\nu_i + m - 1)}) = I_C(T^{k(\nu_i - 1)}\mathbf{x}).$$

Thus with $\underline{P}^k$ defined as above,

$$\underline{P}^k(C^k) = \liminf_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{k(\nu_i - 1)}\mathbf{x}).$$

The sequence $(\tau_i = k(\nu_i - 1) + 1, i \in \mathbb{N})$ of indices is generated from $\mathbf{x}$ by the place selection rule f on X^*, defined by

$$f(x_0, \ldots, x_{tk}) = \begin{cases} f^k\big(x_0, (x_1, \ldots, x_k), \ldots, (x_{(t-1)k+1}, \ldots, x_{tk})\big), & \text{if } n = tk, \\ 0, & \text{otherwise}. \end{cases}$$

One easily sees that the computability of f^k implies that of f, hence $f \in \mathscr{H}$. Since $\mathbf{x}$ is a $(\mathscr{H}, \underline{P})$ unstable collective, we have

$$\liminf_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{\tau_i - 1}\mathbf{x}) = \underline{P}(C),$$

and therefore $\underline{P}^k(C^k) = \underline{P}(C)$ regardless of the choice of f^k. This proves that $\mathbf{x}^k$ is an unstable $(\mathscr{H}^k, \underline{P}^k)$ collective. By taking $m = 1$ (so that $S \subset X^k$), we also see that the marginals of $\underline{P}$ and $\underline{P}^k$ on X^k coincide. □

4. Discussion. The results of the preceding sections suggest that the frequentist interpretation of numerical probability may be adapted to support an alternative stochastic concept based on envelopes of probability measures. The framework for this frequentist account of upper and lower envelopes is borrowed from the von Mises theory and comprises

1. a nondeterministic experiment $\mathscr{E}$ with finite outcome space X;
2. a system $\mathscr{H}$ of causal subsequence selection schemes;
3. a quantitative property $\Pi(\mathscr{E})$ that manifests itself asymptotically in the outcomes of $\mathscr{E}$ and is invariant under subsequence selection through $\mathscr{H}$.

The notion of unlinkedness, or causal independence, is critical here. Our experience with a diverse body of aggregate phenomena, and in particular with experiments repeated under carefully controlled conditions, substantiates a link between causal independence and the apparent convergence of relative frequencies. For most such phenomena, a priori claims of causal independence can be made on the basis of detailed knowledge of the mechanisms generating the repetitions. This raises the question as to what happens in situations where such knowledge is unavailable and where the relative frequencies do not appear to converge. Does the asymptotic instability of relative frequency always signify causal dependence? Or is it possible for the time series in question to exhibit other traits that still exemplify causal independence?

An affirmative answer to the latter question is conceivable within the above theoretical framework, where unlinkedness is construed independently of any extrinsic empirical input. Thus a sequence of repetitions of $\mathscr{E}$ is unlinked if the resulting sequence of outcomes $\mathbf{x}$ exhibits property $\Pi(\mathscr{E})$ and if the same is true for any subsequence of $\mathbf{x}$ selected by a rule in $\mathscr{H}$. By postulating convergence of relative frequencies along $\mathbf{x}$ to a limit given by $\Pi(\mathscr{E})$, one can retrace the familiar von Mises account to obtain the complete stochastic model $\{\Pi(\mathscr{E}^k), k \in \mathbb{N}\}$.

In this paper we have suggested an alternative avenue suited to the case in which unlinked repetitions yield unstable relative frequencies. We have identified $\Pi(\mathscr{E})$ with the limit inferior of the relative frequency of the marginal events and shown the existence of sequences $\mathbf{x}$ that satisfy the invariance constraints with respect to any countable (hence including computable) system $\mathscr{H}$ (Theorem 2.2). Our transition from $\Pi(\mathscr{E})$ to the unique i.i.d. envelope model $\{\Pi(\mathscr{E}^k), k \in \mathbb{N}\}$ (Theorem 3.1) presupposed the existence of sequences $\mathbf{x}$ with additional invariance properties, which we termed unstable collectives.

Two issues that remain open are the existence of unstable collectives and the representation of the class $\mathscr{L}$ of lower envelopes $\underline{P}$ on $\mathscr{C}$ associated with unstable collectives. To date, our attempts at resolving the former issue by modifying the proof of Theorem 2.2 so as to accommodate all cylinder sets have yielded little. With regard to the problem of representation of $\mathscr{L}$, our knowledge is limited to the first statement in Theorem 3.1, namely that $\mathscr{L}$ is a subclass of the stationary lower envelopes on $\mathscr{C}$.

We hope that this account will motivate further inquiry into the possibilities of stochastic modeling based on envelopes of probability measures, as well as other unconventional probabilistic structures.

APPENDIX

LEMMA. *Let* $\mathbf{x} \in X^{\mathbb{N}}$, *where* X *is any finite set. Then there exists a countable family* $\mathscr{Q}_{\mathbf{x}}$ *of stationary probability measures on* $\mathscr{C}$ *such that for every* $C \in \mathscr{C}$,

$$\limsup_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{i-1}\mathbf{x}) = \max\{\mu(C), \mu \in \mathscr{Q}_{\mathbf{x}}\},$$

$$\liminf_n \frac{1}{n} \sum_{i=1}^{n} I_C(T^{i-1}\mathbf{x}) = \min\{\mu(C), \mu \in \mathscr{Q}_{\mathbf{x}}\}.$$

PROOF. For $A \subset X^{\mathbb{N}}$, we define

$$\rho_n(A) = \frac{1}{n} \sum_{i=1}^{n} I_A(T^{i-1}\mathbf{x}),$$

and note that $\rho_n(\cdot)$ is finitely additive on $\mathscr{C}$. We shall construct the family $\mathscr{Q}_{\mathbf{x}}$ with the aid of a Banach limit L on the space l_∞ of bounded real sequences.

Consider a fixed cylinder set C and an increasing sequence $(t_n, n \in \mathbb{N})$ of indices such that $\rho_{t_n}(C)$ converges to $\liminf_n \rho_n(C)$. We define the set function μ on $\mathscr{C}$ by

$$\forall\, C' \in \mathscr{C}, \qquad \mu(C') = L\big(\rho_{t_n}(C'), n \in \mathbb{N}\big).$$

Since L is nonnegative on l_∞ and $L(1, 1, \dots) = 1$, we have that $\mu \geq 0$, $\mu(X^{\mathbb{N}}) = 1$. From the linearity of L on l_∞ and the finite additivity of $\rho_{t_n}(\cdot)$ on $\mathscr{C}$, we deduce that μ is a finitely additive probability measure on $\mathscr{C}$; it is also countably additive since every cylinder set is both open and compact in the product topology induced by the discrete topology on X. Furthermore, the property

$$\forall\, (\mathbf{a}) \in l_\infty, \qquad \liminf_n a_n \leq L(\mathbf{a}) \leq \limsup_n a_n$$

implies both

$$\liminf_n \rho_n(C) = \mu(C) \quad \text{and} \quad \forall\, C' \quad \liminf_n \rho_n(C') \leq \mu(C').$$

Finally, stationarity of μ follows from the stated properties of L and the relationship

$$\rho_{t_n}(C') - \rho_{t_n}(T^{-1}C') \to 0.$$

By letting C range over $\mathscr{C}$, we obtain a collection $\mathscr{Q}_{\mathbf{x}}$ of probability measures μ on $\mathscr{C}$ such that

$$\forall\, C \in \mathscr{C}, \qquad \liminf_n \rho_n(C) = \min\{\mu(C), \mu \in \mathscr{Q}_{\mathbf{x}}\}.$$

The corresponding identity for lim sup follows from

$$\liminf_n \rho_n(A) + \limsup_n \rho_n(A^c) = 1. \qquad \square$$

REFERENCES

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

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York.

Huber, P. J. and Strassen, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263. [Correction (1974) **2** 233.]

Kolmogorov, A. N. and Uspenskii, V. A. (1987). Algorithms and randomness. *Theory Probab. Appl.* **32** 389–412.

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

Loève, M. (1978). *Probability Theory* **2**, 4th ed. Springer, New York.

Martin-Löf, P. (1969). The literature on von Mises' kollektivs revisited. *Theoria* **35** 12–37.

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**(2).

Schnorr, C. P. (1971). *Zufälligkeit und Wahrscheinlichkeit. Lecture Notes in Math.* **218**. Springer, New York.

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

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

von Mises, R. (1957). *Probability, Statistics and Truth*. Macmillan, London.

von Mises, R. and Geiringer, H. (1964). *The Mathematical Theory of Probability and Statistics*. Academic, New York.

Wald, A. (1936). Sur la notion de collectif dans le calcul des probabilités. *C. R. Acad. Sci. Paris Sér. A–B* **202** 180–183.

Walley, P. (1981). Coherent lower and upper probabilities. Research Report, Univ. Warwick, Coventry.

Walley, P. (1990). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London. To appear.

Walley, P. and Fine, T. L. (1982). Towards a frequentist theory of upper and lower probability. *Ann. Statist.* **10** 741–761.

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

Department of Electrical Engineering
and Systems Research Center
University of Maryland
College Park, Maryland 20742

School of Electrical Engineering
Cornell University
Ithaca, New York 14853