Towards a Chaotic Probability Model for Frequentist Probability: The Univariate Case*

P.I. FIERENS
Cornell University, U.S.A.

T.L. FINE
Cornell University, U.S.A.

Abstract

We adopt the same mathematical model of a set **M** of probability measures as is central to the theory of coherent imprecise probability. However, we endow this model with an **objective, frequentist interpretation** in place of a behavioral subjective one. We seek to use **M** to model **stable physical sources of time series data that have highly irregular behavior** and not to model states of belief or knowledge that are assuredly imprecise. The approach we present in this paper is to understand a set of measures model **M** not as a traditional compound hypothesis, in which one of the measures in **M** is a true description, but rather as one in which none of the individual measures in **M** provides an adequate description of the potential behavior of the physical source as actualized in the form of a long time series.

We provide an **instrumental interpretation** of random process measures consistent with **M** and the highly irregular physical phenomena we intend to model by **M**. This construction provides us with the basic tools for simulation of our models.

We present a method to estimate **M** from data which studies any given data sequence by analyzing it into subsequences selected by a set of computable rules. We prove results that help us to choose an adequate set of rules and evaluate the performance of the estimator.

Keywords

imprecise probability, sets of measures, objective, frequentist interpretation

*Research of both authors partially supported by AFOSR Grant F30602-00-2-0558

1 Introduction

1.1 Orientation

We adopt the same mathematical model of a set $\mathbf{M} = \{\nu\}$ of probability measures as is central to the theory of coherent imprecise probability (e.g., see Walley [18]). However, we endow this model with an **objective, frequentist interpretation** in place of a behavioral subjective one, and ask completely different questions of this model. While the mathematical model **M** is the same in the two theories of probability (as it is in a variety of interpretations that have been offered for conventional probability), on our account there is no focus on imprecision as is appropriate in the behavioral account. In order to signal the distinction between the two theories sharing the same mathematical model, we do not use the descriptor "imprecise" and instead use **"chaotic"**. Although we remain interested in alternatives to this term, it does connote a highly irregular sequence of physical (typically mechanical) origin. We seek to use **M** to model **stable** (although not stationary in the traditional stochastic sense) **physical sources of time series data that have highly irregular behavior** and not to model states of belief or knowledge that are assuredly imprecise. Support for the existence of such chaotic sources is lent by the following quotation from Kolmogorov [9]:

> *In everyday language we call random those phenomena where we cannot find a regularity allowing us to predict precisely their results. Generally speaking, there is no ground to believe that random phenomena should possess any definite probability. Therefore, we should distinguish between randomness proper (as absence of any regularity) and stochastic randomness (which is the subject of probability theory). There emerges the problem of finding reasons for the applicability of the mathematical theory of probability to the real world.*

1.2 Previous Work

Previous work focused on asymptotics or laws of large numbers for interval-valued probability models can be found in Fine et al. [10][12][7][15]. Cozman and Chrisman [1] estimate credal sets by looking at the limiting relative frequencies along several subsequences of a time series. Our current work focussed on modelling finite length time series.

Our previous attempt at supplying an objective frequentist interpretation for a set of measures **M**, reported at ISIPTA '01 in Fierens and Fine [3], was based upon the use of Kolmogorov complexity to enable us to simulate highly complex time series data from the model and then to estimate the model from such data through the sequence of alternating minima and maxima of relative frequencies calculated along a given sequence. The underlying motivation was an attempt at an analog of the $i.i.d.$ standard probability model; the model **M** gave us the marginal or univariate description and the high complexity was meant to ensure that there was

no further exploitable structure in the time evolution. We subsequently judged this approach to be inadequate, in part after considering the performance of martingale betting systems on such time series as advocated by the then newly-published Shafer and Vovk [16].

1.3 Overview

As in our previous work, we focus on a description of univariate or marginal events and not on descriptions of k-tuples of outcomes. This restriction is intended only to simplify our search for a meaningful interpretation and not because we deny the importance of an extension to k-tuples. In Section 2.1, we provide an **instrumental interpretation** of random process measures consistent with **M** and the highly irregular physical phenomena we intend to model by **M**. Although we do not offer this description as an explanation for real world data, we develop it because it helps us to better understand chaotic probability models by reference to well-known standard stochastic processes, and, at the same time, this description provides us with the basic tools for simulation of our models (see Section 2.2). Essentially, our instrumental interpretation consists of a decision mechanism that at each time instant chooses a probability measure $\nu \in \mathbf{M}$ from which the next outcome of a sequence will be generated. This measure selection function has both properties of being highly complex so that it is difficult to discover it from any given data sequence, and having enough simple structure to allow for the estimation of **M** (see Theorems 1-3). The approach we present in this paper is to understand a set of measures model **M** not as a traditional compound hypothesis, in which one of the measures in **M** is a true description, but rather as one in which none of the individual measures in **M** provides an adequate description of the potential behavior of the physical source as actualized in the form of a long time series. Instead, it is the whole set **M** that describes the potential behavior, and this distinction has operational significance in terms of the time series data that is anticipated from the physical source.

As explained in Section 3, **we estimate M from a data sequence by computing the relative frequencies along some of its subsequences**. Subsequence selection is a well-entrenched method of exposing behavioral patterns in time series. It formed the basis of Richard von Mises' pioneering definition of randomness ([17],[4],[11],[13]) for infinitely long sequences and the seminal work of A.N. Kolmogorov on randomness of finite strings ([8]). Cozman and Chrisman [1] estimate credal sets by looking at the relative frequencies along several subsequences. In a similar way, we also study a given sequence by analyzing it into subsequences selected by rules in some set Ψ. Technically, we use **causal subsequence selection rules**, also known as **Church place selection rules** (see Definition 1 and also Li and Vitányi [11]). For any given model **M**, we expect to find some set of rules Ψ_V for which **M** becomes **"visible"**, that is, a set of rules such that all measures in **M** can be estimated by the relative frequencies along the

selected subsequences (see Definition 2 and Theorems 2 and 3). Although such a set Ψ_V may exist, identifying it will not be easy. Furthermore, there are sets of rules Ψ_T for which a chaotic source may appear to be **"temporally homogeneous"**, that is, for a certain set Ψ_T there may exist a chaotic source generating sequences such that the relative frequencies along subsequences selected by rules in Ψ_T cannot expose more than a small neighborhood of a single measure contained in the convex hull of $\mathbf{M}$ (see Definition 3, Lemma 1 and Theorem 4).

Proofs have been omitted in what follows. However, they are available in the appendices of Fierens [2].

2 From the Model to Data

2.1 An Instrumental Interpretation of the Model

Let $\mathbf{X} = \{z_1, z_2, \cdots, z_\xi\}$ be a finite sample space. We denote by $\mathbf{X}^*$ the set of all finite sequences of elements taken in $\mathbf{X}$. A particular sequence of n samples from $\mathbf{X}$ is denoted by $x^n = \{x_1, x_2, \cdots, x_n\}$. $\mathbf{P}$ denotes the set of all measures on the power set of $\mathbf{X}$. A chaotic probability model $\mathbf{M}$ is a subset of $\mathbf{P}$ and models the "marginals" of some process generating sequences in $\mathbf{X}^*$. In this section, we present an instrumental (that is, without commitment to reality) interpretation of such a process.

Consider the generation of a sequence x^n by the following pseudo-algorithm:

```
FOR k = 1 TO k = n

 1. Choose ν ∈ M.
 2. Generate x_k according to ν.
```

If the decision mechanism in `1` is very complex[1], say, random, with decisions made in an *i.i.d.* manner according to some distribution on $\mathbf{M}$, we would not be able to distinguish whether x^n was produced by an *i.i.d.* process according to some measure in $ch(\mathbf{M})$, the convex hull of $\mathbf{M}$, or by the algorithm in question. On the other hand, if the decision rule were very simple and deterministic, we would possibly be able to make such a distinction. For example, consider the simple choice mechanism that alternates between two measures $\nu_1, \nu_2 \in \mathbf{M}$. In this case, for sufficiently large n, we expect to discover the alternating-measure rule and to be able to estimate ν_1 and ν_2. However, if the choice mechanism in `1` were neither too complex (as in the first example) nor too simple (as in the second example), we may still be able to estimate $\mathbf{M}$ (or part of it), but we would probably

[1] Although Kolmogorov complexity captures part of the complexity to which we make reference here, it seems not to suffice. Thus, the discussion in this paragraph follows at a more intuitive level.

find it difficult (if not impossible given our computational resources) to discover the choice mechanism itself. It is in this case that we believe chaotic probability models to be useful: when dealing with chaotic sources, the measure selection function F has both properties of being highly complex so that it is difficult to discover it from any given data sequence, and having enough simple structure to allow for the estimation of $\mathbf{M}$.

We formalize the decision in `1` of the previous algorithm by means of a function $F : \mathbf{X}^* \to \mathbf{M}$. Furthermore, we restrict ourselves to *causally* made decisions, ones dependent only upon the past:

FOR k = 1 TO k = n

1. Choose $\nu = F(x^{k-1}) \in \mathbf{M}$.
2. Generate x_k according to ν.

Let $\nu_k = F(x^{k-1})$. For any $k \leq n$, F determines the probability distribution of the *potential* kth outcome X_k of the sequence,

$$(\forall \mathbf{A} \subseteq \mathbf{X})\ P(X_k \in \mathbf{A} | X^{k-1} = x^{k-1}) = \nu_k(X_k \in \mathbf{A}).$$

An actual data sequence x^n is assessed by the graded potential of the realization of a sequence of random variables X^n described by

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{k=1}^{n} \nu_k(X_k = x_k).$$

We denote by $\mathbf{M}^*$ the family of all such process measures P. From the analysis of data, we do not expect in general to be able to pinpoint a single $P \in \mathbf{M}^*$ or even a small subset of $\mathbf{M}^*$, what we call a **fine-grained picture** of the source. On the contrary, we expect our knowable **operational quantities to be (large) subsets of $\mathbf{M}^*$** which provide an appropriate **coarse-grained** description of the source. These ideas are related to those of **coarse grainedness** and **fine grainedness** in physics. For example, in classical physics we commonly have situations, say, kinetic theory, in which a coarse description suffices even though we have access in principle to a more detailed quantum mechanical one. Unlike the case of classical physics, there need be no more than instrumental reality in the fine details of our model $\mathbf{M}^*$. A similar situation may be found in quantum mechanics where there are fine-grained pictures that have no empirical reality (see Gell-Mann [6], Chapter 11, especially pp. 143-147).

2.2 Simulation

Simulation of sequences coming from a source modelled by a set of measures $\mathbf{M}$ can be achieved by simply choosing an appropriate function F and applying

the algorithm presented above. Since we expect not to know F in general, the choice of the measure selection functions used for simulation depends on our judgment, based on our knowledge of the physical phenomenon being modelled, the intended use of the simulated sequences, etc.

In the typical case where $\mathbf{M}$ has infinite cardinality, we need a notion of approximation to the measures in $\mathbf{M}$ by finitely many other measures in (or close to) $\mathbf{M}$. Given a distance or metric d on $\mathbf{P}$, a particular form of approximation is provided by an **ε-covering** of $\mathbf{M}$, that is, by a covering of the set $\mathbf{M}$ by open balls of radius ε (according to d) and centers in some set $\mathbf{M}_\varepsilon \subset \mathbf{P}$ (perhaps a subset of $\mathbf{M}$). Note that, if $\mathbf{P}$ is compact with respect to d, we can find a **finite** ε-covering of $\mathbf{M}$. Choose a minimal set $\mathbf{M}_\varepsilon$ so that each ball has a non-empty intersection with $\mathbf{M}$ and call $B(\varepsilon,\nu)$ the ball with center $\nu \in \mathbf{M}_\varepsilon$ and radius ε. Then, given an appropriate measure selection function $F : \mathbf{X}^* \to \mathbf{M}_\varepsilon$, the following algorithm can be used for simulation.

```
FOR k = 1 TO k = n
```

1. Choose $\nu = F(x^{k-1}) \in \mathbf{M}_\varepsilon$.
2. Choose any $\nu' \in B(\varepsilon,\nu) \cap \mathbf{M}$.
3. Use a pseudo-random number generator to generate x_k according to ν'.

Since we want to expose all of $\mathbf{M}$ in a single, but sufficiently long, simulated sequence, we require F to visit, many times, each measure in $\mathbf{M}_\varepsilon$. Theorems 1-2 in Section 3 can help us choose the minimum number of times that each measure should be visited. Examples of simulation algorithms based on the basic strategy presented above are available in the appendices of Fierens [2] (see, e.g., the proof of Theorem 4) and in Section 3.5

3 From Data to the Model

3.1 Subsequence Analysis

We begin the study of a sequence $x^n \in \mathbf{X}^*$ by analyzing it into several subsequences. These subsequences are selected by rules that satisfy the following

Definition 1 (Causal Subsequence Selection Rule)
An effectively computable function ψ is a **causal subsequence selection rule** *(also known as a Church place selection rule) if*

$$\psi : \mathbf{X}^* \to \{0,1\},$$

and, for any $x^n \in \mathbf{X}^*$, x_k is the j-th term in the generated subsequence $x^{\psi,n}$, of length $\lambda_{\psi,n}$, if

$$\psi(x^{k-1}) = 1, \quad \sum_{i=1}^{k} \psi(x^{i-1}) = j, \quad \lambda_{\psi,n} = \sum_{k=1}^{n} \psi(x^{k-1}).$$

Let $\Psi = \{\psi_\alpha\}$ be a set of causal subsequence selection rules. For each $\psi \in \Psi$, we study the behavior of the relative frequency of (only) marginal events along the chosen subsequence. That is, given x^n and a selection rule $\psi \in \Psi$ we determine the **frequentist empirical (relative frequency) measure** $\bar{\mu}_{\psi,n}$ along the subsequence $x^{\psi,n}$ through

$$(\forall \mathbf{A} \subset \mathbf{X})\ \bar{\mu}_{\psi,n}(\mathbf{A}) = \frac{1}{\lambda_{\psi,n}} \sum_{k=1}^{n} I_{\mathbf{A}}(x_k)\psi(x^{k-1}),$$

where $I_{\mathbf{A}}(\cdot)$ is the $\{0,1\}$-valued indicator function of the event $\mathbf{A}$. In a similar manner, for any such rule ψ, we may compute the **time average conditional measure** $\bar{\nu}_{\psi,n}$ defined by

$$(\forall \mathbf{A} \subset \mathbf{X})\ \bar{\nu}_{\psi,n}(\mathbf{A}) = \frac{1}{\lambda_{\psi,n}} \sum_{k=1}^{n} \mathrm{E}\left[I_{\mathbf{A}}(X_k) \middle| X^{k-1} = x^{k-1} \right] \psi(x^{k-1}).$$

Rewritten in terms of our instrumental understanding of the measure selection function F,

$$\bar{\nu}_{\psi,n}(\mathbf{A}) = \frac{1}{\lambda_{\psi,n}} \sum_{k=1}^{n} \nu_k(\mathbf{A})\psi(x^{k-1}),$$

where $\nu_k = F(x^{k-1})$.

Since we want to expose some of the structure of the chaotic probability model $\mathbf{M}$ by means of the rules in Ψ, we are interested in how good an estimator of $\bar{\nu}_{\psi,n}$ is $\bar{\mu}_{\psi,n}$. Introduce the norm-based metric

$$(\forall \mu, \mu' \in \mathbf{P})\ d(\mu,\mu') = \max_{z \in \mathbf{X}} \left| \mu(z) - \mu'(z) \right|,$$

which quantifies the "closeness" between two probability measures on $\mathbf{X}$. We call a rule ψ applied to x^n **causally faithful** if the resulting subsequence yields a small value of $d(\bar{\nu}_{\psi,n}, \bar{\mu}_{\psi,n})$. The existence of such rules is guaranteed by

Theorem 1 *Let ξ be the cardinality of $\mathbf{X}$ and denote the cardinality of Ψ by $\|\Psi\|$. Let $m \le n$. If $\|\Psi\| \le t_n$, then for any process measure $P \in \mathbf{M}^*$*

$$P\left(\max_{\psi \in \Psi} \left\{ d(\bar{\mu}_{\psi,n}, \bar{\nu}_{\psi,n}) :\ \lambda_{\psi,n} \ge m \right\} \ge \varepsilon \right) \le 2\xi t_n \mathrm{e}^{-\frac{\varepsilon^2 m^2}{2n}}.$$

Hence, so long as we restrict to a family of causal selection rules of size t_n and examine discrepancies of size ε only over subsequences of length at least m, with m large, we can with high probability avoid uncontrollably imposing our own patterns through some of the selected subsequences and instead exhibit only the patterns that have inductive validity. If, to the contrary, we allow the set of subsequence selection rules to be too large, we will observe with non-negligible probability measures that are outside the convex hull of $\mathbf{M}$. For example, if we enlarge the set of subsequence selection rules by including all possible subsequences, then we will observe measures that concentrate all the mass on a single atom (outcome in $\mathbf{X}$).

3.2 Visibility and Estimation

The possibility of exposing all of $\mathbf{M}$ by means of the rules in Ψ is expressed in the following

Definition 2 (Visibility)
(a) $\mathbf{M}$ *is* **made visible** (Ψ,θ,δ,m,n) *by* $P \in \mathbf{M}^*$ *if*

$$P\left(\bigcap_{\mu\in\mathbf{M}}\bigcup_{\psi\in\Psi}\{X^n : \lambda_{\psi,n}(X^n) \geq m, d(\bar{\mu}_{\psi,n},\mu) \leq \theta\}\right) \geq 1-\delta.$$

(b) *A subset of* $\mathbf{M}^*$ **renders M uniformly visible** (Ψ,θ,δ,m,n) *if* $\mathbf{M}$ *is made visible* (Ψ,θ,δ,m,n) *by each of its elements. The maximal such subset is denoted* $\mathbf{M}_V(\Psi)$ *and* $\mathbf{M}_V(\Psi)$ *may be empty.*

The non-triviality of Definition 2**(a)**, and, hence, of Definition 2**(b)**, is asserted in

Theorem 2 *Let* $0 < 2\varepsilon < \theta$ *and* $\mathbf{M}_\varepsilon \subseteq \mathbf{M}$ *be the centers of a minimal covering of* $\mathbf{M}$ *by* N_ε *balls of radius* ε *(according to the metric* d *as defined above*[2]*). Then, for large n, there exists a process measure P and a family* Ψ *of size* N_ε *such that* $\mathbf{M}$ *is made visible* (Ψ,θ,δ,m,n) *with*

$$\delta = 2(\xi+1)N_\varepsilon e^{-\frac{(\theta-2\varepsilon)^2 m^2}{2n}}.$$

Theorem 2 asserts the existence of a set of rules Ψ such that $\mathbf{M}_V(\Psi)$ is not empty.

The following theorem shows how it is possible to estimate $\mathbf{M}$ by means of an appropriate set of rules Ψ.

[2]According to our choice of d, although $\mathbf{M}$ is not necessarily compact, $\mathbf{P}$ certainly is. Therefore, as a subset of compact $\mathbf{P}$, $\mathbf{M}$ will always have a finite open covering.

Theorem 3 *Let* $\mathbf{M}_\alpha$ *be a subset of the non-empty maximal* $\mathbf{M}_V(\Psi) \subseteq \mathbf{M}^*$ *that renders* $\mathbf{M}$ *uniformly visible* $(\Psi, \theta, \delta, m, n)$. *Let* $[\mathbf{A}]^\varepsilon$ *denote the* **ε-enlargement** *of a set* $\mathbf{A}$ *defined by*

$$(\forall \mathbf{A} \subseteq \mathbf{P})\ (\forall \varepsilon > 0)\ [\mathbf{A}]^\varepsilon = \{\mu : (\exists \mu' \in \mathbf{A}) d(\mu, \mu') < \varepsilon\}.$$

Let $\hat{\mathbf{M}}_{\theta,\Psi}$ *be an estimator of* $\mathbf{M}$ *defined by*

$$(\forall x^n \in \mathbf{X}^*)\ \hat{\mathbf{M}}_{\theta,\Psi}(x^n) = \bigcup_{\{\psi : \psi \in \Psi,\ \lambda_{\psi,n}(x^n) \geq m\}} B(\theta, \bar{\mu}_{\psi,n}).$$

Then the estimator $\hat{\mathbf{M}}_{\theta,\Psi}$ *satisfies*

$$(\forall P \in \mathbf{M}_\alpha) \quad P\left([ch(\mathbf{M})]^{\theta+\varepsilon} \supset \hat{\mathbf{M}}_{\theta,\Psi} \supset \mathbf{M}\right) \geq 1 - \delta - \tau_n,$$

where $ch(\mathbf{M})$ *is the convex hull of* $\mathbf{M}$ *and*

$$\tau_n = 2\xi \|\Psi\| \mathrm{e}^{-\frac{\varepsilon^2 m^2}{2n}}.$$

3.3 Temporal Homogeneity

Not every set of rules Ψ can expose all of $\mathbf{M}$. The following definition deals with some sets of rules that can only expose a small neighborhood of a single probability measure in $ch(\mathbf{M})$.

Definition 3 (Temporal Homogeneity)

(a) $P \in \mathbf{M}^*$ *is* **temporally homogeneous** $(\Psi, \theta, \delta, m, n)$ *if*

$$P\left(\max_{\psi_1, \psi_2 \in \Psi} \left\{d(\bar{\mu}_{\psi_1,n}, \bar{\mu}_{\psi_2,n}) : \lambda_{\psi_1,n}(X^n), \lambda_{\psi_2,n}(X^n) \geq m\right\} \leq \theta\right) \geq 1 - \delta.$$

(b) *A subset of* $\mathbf{M}^*$ *is* **uniformly temporally homogeneous** $(\Psi, \theta, \delta, m, n)$ *if each of its elements is temporally homogeneous* $(\Psi, \theta, \delta, m, n)$. *The maximal such subset is denoted* $\mathbf{M}_T(\Psi)$.

The non-triviality of Definition 3**(a)**, and, hence, of Definition 3**(b)**, is established by

Lemma 1 *Choose* $\mu_0 \in \mathbf{P}$ *and* $0 < 2\varepsilon < \theta$ *and constrain the measure selection mechanism* F *so that*

$$(\forall x^* \in \mathbf{X}^*)\ F(x^*) \in B(\varepsilon, \mu_0),$$

where $B(\varepsilon,\mu_0)$ is a ball with center μ_0 and radius ε; that is, every P induced by F is approximately i.i.d. μ_0. Then each such P is temporally homogeneous (Ψ,θ,δ,m,n) provided that δ satisfies

$$\delta = 2\xi t_n \mathrm{e}^{-\frac{[(\theta-2\varepsilon)m]^2}{8n}},$$

where $t_n = \|\Psi\|$.

3.4 Consistency between Visibility and Temporal Homogeneity

We can better appreciate the difficulty of choosing an appropriate set of rules for estimation of $\mathbf{M}$ by means of the next theorem, which in some sense complements Theorem 2 and Lemma 1.

Theorem 4 *Let $\varepsilon > \frac{1}{m}$. Assume that there is an ε-cover of $\mathbf{M}$ by N_ε open balls with centers in a set $\mathbf{M}_\varepsilon = \{\mu_1,\mu_2,\cdots,\mu_{N_\varepsilon}\}$ such that, for each μ_i, there is a* **recursive probability measure** *$\nu \in B(\varepsilon,\mu_i)\cap\mathbf{M}$. Let Ψ_0 be a set of (causal deterministic) place selection rules. Then, there are a process measure P and a family Ψ_1 such that, for large enough n, P will both render $\mathbf{M}$ visible $(\Psi_1,3\varepsilon,\delta,m,n)$ and ensure temporal homogeneity $(\Psi_0,6\varepsilon,\delta,m,n)$ with*

$$\delta = 2\xi t_n \mathrm{e}^{-\frac{\varepsilon^2 m^2}{2n}},$$

where

$$t_n = \max\{\|\Psi_0\|,\|\Psi_1\|\}.$$

A more transparent version of Theorem 4, given in terms of an analyzing set Ψ_0 formed by **finite history rules** defined as follows

Definition 4 Finite History Rules

We say that ψ is a **finite history rule** *if there is a positive integer L, called* **the history length of** ψ*, and a function $\Gamma : \mathbf{X}^L \to \{0,1\}$ such that for all $x^n \in \mathbf{X}$ we have*

$$\psi(x^{k-1}) = \begin{cases} \Gamma(x^{k-L},x^{k-L+1},\cdots,x^{k-1}) & \text{if } k > L, \\ 0 & \text{otherwise.} \end{cases}$$

The next theorem is similar to Theorem 4:

Theorem 5 *Assume that $\mathbf{M}$ makes all atoms possible, i.e., there is $\phi > 0$ such that*

$$\inf_{\mu\in\mathbf{M}} \min_{z\in\mathbf{X}} \mu(\{z\}) \geq \phi.$$

Let Ψ_0 consist of finite history rules with length smaller than a given L. Then, for $\varepsilon > 0$, there are a process measure P and a family Ψ_1 such that P will both render $\mathbf{M}$ *visible* $(\Psi_1, 2\varepsilon, \delta, m, n)$ *and ensure temporal homogeneity* $(\Psi_0, 4\varepsilon, \delta, m, n)$ *with*

$$\delta = 4\xi t_n \mathrm{e}^{-\frac{\varepsilon^2 m^2}{2n}},$$

where

$$t_n = \max\{\|\Psi_0\|, \|\Psi_1\|\}.$$

Although we do not present a complete proof here (it can be found in Fierens [2]), we give the basic idea behind the construction of P in Section 3.5 because it provides a simple example of several ideas in this paper.

Put picturesquely, the results in this section show that Ψ determines the **resolving power of the analytical microscope** with which we examine $\mathbf{M}$. When one prepares a sample to be put under the lenses of the microscope, little or nothing is seen of the structure of the sample, e.g., it may just look like some watery solution. Similarly, in the case of a chaotic probability model, the temporal homogeneity property tells us that $\mathbf{M}$ looks just like the traditional single measure. As we explore $\mathbf{M}$ with a large numbers of more complex selection rules, say, under the more powerful lenses of the microscope, we begin to see or isolate different relative frequency measures and begin to see $\mathbf{M}$ as a set of measures. However, we do not know in advance the final scale at which $\mathbf{M}$ exhibits all of its structure and do not know in advance how to choose Ψ to render all of $\mathbf{M}$ visible. Our abilities at progressive exploration are, of course, limited both by the increasing computational burden, and by considerations of extracting faithful subsequences. Preserving the faithful subsequence property requires a relation between $\|\Psi\|$ and the resulting confidence level $1-\delta$. As $\|\Psi\|$ increases, maintaining confidence levels requires longer subsequences (larger m) and in turn more data (larger n). These considerations make good traditional statistical sense.

3.5 Simulation Example

Let $\mathbf{M}_\varepsilon = \{\mu_1, \mu_2, \cdots, \mu_{N_\varepsilon}\} \subseteq \mathbf{M}$ be the centers of a finite ε-cover of $\mathbf{M}$ by N_ε open balls. Let γ be defined as

$$\gamma = \left\lceil \log_\xi N_\varepsilon \right\rceil.$$

Let $\mathbf{B}_1, \cdots, \mathbf{B}_{N_\varepsilon}$ be a partition of $\mathbf{X}^\gamma$, the histories of length γ, into N_ε subsets and consider the memory-γ Markov process defined by the following transition probabilities:

$$(\forall \mathbf{A} \subseteq \mathbf{X})\ P\left(X_k \in \mathbf{A} | X_{k-1} = x_{k-1}, \cdots, X_{k-\gamma} = x_{k-\gamma}\right) = \mu_i\left(X_k \in \mathbf{A}\right), \qquad (1)$$

iff $(x_{k-\gamma}, \cdots, x_{k-1}) \in \mathbf{B}_i$. It can be proved that this Markov process has a unique stationary probability measure μ_S (see Fierens [2]).

Let R be an integer greater than a given L and consider the construction of a process measure $P \in \mathbf{M}^*$ by an algorithm that: a) initializes R *i.i.d.* Markov processes (as described by Eqn. 1) at the stationary measure; b) generates the sequence x^n by choosing outcomes from the R Markov processes in a round-robin fashion. A more detailed description of this algorithm follows.

FOR l = 1 TO l = R

1. Generate $(x_{l,1}, x_{l,2}, \cdots, x_{l,\gamma})$ according to μ_S.
2. FOR k = $\gamma+1$ TO k = $\lceil n/R \rceil$
   (a) Find the set $\mathbf{B}_i$ such that $(x_{k-\gamma}, \cdots, x_{k-1}) \in \mathbf{B}_i$.
   (b) Generate x_k according to μ_i.

Set R counters $i_1, i_2, \cdots, i_R$ to 1.

FOR k = 1 TO k = n

1. Let $l = [(k-1) \bmod R] + 1$.
2. Let $x_k = x_{l,i_l}$.
3. Let $i_l = i_l + 1$.

We now sketch the proof that the previous algorithm succeeds in constructing a process measure $P \in \mathbf{M}^*$ satisfying the conditions stated in Theorem 5. By the previous algorithm, for $k > R\gamma$, the outcome X_k depends on $X_{k-R\gamma}$, $X_{k-R(\gamma-1)}$, $\cdots$, X_{k-R}, but it does not depend on X_{k-R+1}, X_{k-R+2}, $\cdots$, X_{k-1}. Let ψ be any rule in Ψ_0. Since ψ has a limited time horizon L which is strictly smaller than R, we have for all $\mathbf{A} \subset \mathbf{X}$

$$\mathrm{E} \sum_{k=1}^{n} \psi(X^{k-1}) \left[\mathrm{E} \left[I_{\mathbf{A}}(X_k) \middle| X^{k-1} \right] - \mu_S(\mathbf{A}) \right] =$$

$$= \sum_{k=1}^{n} P\left(\psi(X^{k-1}) = 1\right) \mathrm{E} \left[I_{\mathbf{A}}(X_k) - \mu_S(\mathbf{A}) \middle| \psi(X^{k-1}) = 1 \right] = \text{ (by memory } L\text{)}$$

$$= \sum_{k=1}^{n} P\left(\psi(X^{k-1}) = 1\right) \mathrm{E} \left[I_{\mathbf{A}}(X_k) - \mu_S(\mathbf{A}) \middle| \psi(X^{k-L:k-1}) = 1 \right] = \text{ (by indep.)}$$

$$= \sum_{k=1}^{n} P\left(\psi(X^{k-1}) = 1\right) \mathrm{E} \left[I_{\mathbf{A}}(X_k) - \mu_S(\mathbf{A}) \right] = 0.$$

It can be shown, by means of the same techniques used in the proof of Theorem 1, that this fact implies that

$$P\left(\max_{\psi \in \Psi_0} \left\{ d(\bar{\nu}_{\psi,n}, \mu_S) : \lambda_{\psi,n} \geq m \right\} \geq \varepsilon \right) \leq 2\xi \|\Psi_0\| \mathrm{e}^{-\frac{\varepsilon^2 m^2}{2n}}.$$

Finally, this statement together with Theorem 1 imply that $\mathbf{M}$ is Ψ_0-temporal homogeneous.

Let $\Psi_1 = \{\psi_1, \cdots, \psi_{N_\varepsilon}\}$ be a set of rules such that

$$\psi_i(x^{k-1}) = \begin{cases} 1 & \text{if } k > R\gamma \text{ and } (x_{k-R\gamma}, x_{k-R(\gamma-1)}, \cdots, x_{k-R}) \in \mathbf{B}_i, \\ 0 & \text{otherwise.} \end{cases}$$

Then, it is easy to see that

$$(\forall \mathbf{A} \subseteq \mathbf{X})\ \bar{\nu}_{\psi_i,n}(\mathbf{A}) = \frac{1}{\lambda_{\psi_i,n}} \sum_{k=1}^{n} \mathrm{E}\left[I_{\mathbf{A}}(X_k) \,\middle|\, X^{k-1} = x^{k-1} \right] \psi_i(x^{k-1}) = \mu_i(\mathbf{A}).$$

This fact together with Theorem 1 ensure Ψ_1-visibility.

4 Conclusions and Future Work

As is well known in cognitive psychology (see, e.g., [5]), perception is intimately related to expectation: in many cases, we see what we expect to see. In a similar manner, our capacity to recognize new phenomena is conditioned by our existing mathematical constructs (see Fierens and Fine [3]). In the words of Meno to Socrates[3]:

> *And how will you enquire, Socrates, into that which you do not know? What will you put forth as the subject of enquiry? And if you find what you want, how will you ever know that this is the thing which you did not know?* (From [14]).

We have presented here a new way of "seeing" time series by introducing chaotic probability models. Although we have not shown real-world data supporting our models, we have provided the basic tools needed to recognize and study such data. We have developed a basic understanding of chaotic sources by means of the instrumental interpretation in Section 2 and we have presented methods to estimate the model from data and to simulate it given the model in Section 3.

Bridge-building provides a metaphor for our approach to the development of an objective theory based on sets of probability models $\mathbf{M}$. The two piers of the bridge are: the model $\mathbf{M}$ as a set of probability measures on all subsets of $\mathbf{X}$ (see Section 2) representing potential; time series data in the form of a sequence x^n of finite length n with terms in the sequence all drawn from a finite sample space $\mathbf{X}$ (see Section 3) representing the actualization of potential. Our models need to show consistent descriptions of both piers and methods to traverse this bridge in both directions. In estimation we have many ways to proceed from a unique data sequence to an approximate model. In simulation we have many ways to proceed from a model to multiple data sequences that are typical of the model.

[3] We owe this quote to an anonymous referee.

Structural soundness of the bridge amounts to self-consistency of estimation and simulation, in the sense that a model $\hat{\mathbf{M}}$ estimated from a simulated sequence $\hat{x}^n$ must be similar to the original model $\mathbf{M}$ being simulated:

$$\mathbf{M} \xrightarrow[\text{source gen.}]{} x^n \xrightarrow[\text{estimation}]{} \hat{\mathbf{M}}(x^n)(\approx \mathbf{M}) \xrightarrow[\text{simulation}]{} \hat{x}^n \xrightarrow[\text{estimation}]{} \hat{\mathbf{M}}(\hat{x}^n) \approx \mathbf{M}.$$

More work remains to be done on estimation and simulation before being able to evaluate fairly this kind of consistency in our chaotic probability models. Also, we need to find a way of quantifying such consistency. Do the models obtained from simulated sequences look similar to the models used for simulation? How do we quantify these similarities? These questions need an answer if we want the framework of chaotic probability models to be consistent.

In view of the instrumental interpretation in Section 2.1, it may be argued that a set of probability measures $\mathbf{M}$ provides only an unfinished picture of a chaotic source, the description of F being needed for a complete model. However, we believe that $\mathbf{M}$ provides, not an incomplete picture of the source, but a coarse grained one. As thermodynamics in physics provides good (complete) enough models of gases for many practical purposes, we believe sets of measures $\mathbf{M}$ may be good (complete) enough models of chaotic sources in many cases. Although examples of the successful use of chaotic models in applied probability have yet to be provided, the main elements needed for the application of our models have been given in this paper.

References

[1] COZMAN, F., AND CHRISMAN, L. Learning convex sets of probability from data. Tech. Rep. CMU-RI-TR-97-25, Robotics Institute, Carnegie Mellon University, 1997.

[2] FIERENS, P. I. *Towards a Chaotic Probability Model for Frequentist Probability*. PhD thesis, Cornell University, August 2003. Available at http://www.people.cornell.edu/pages/pif1/thesis.pdf.

[3] FIERENS, P. I., AND FINE, T. L. Towards a frequentist interpretation of sets of measures. In *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications* (The Netherlands, 2001), G. de Cooman, T. L. Fine, and T. Seidenfeld, Eds., Shaker Publishing.

[4] FINE, T. L. *Theories of Probability: An Examination of Foundations*. Academic Press, 1973.

[5] GALOTTI, K. M. *Cognitive Science In and Out of the Laboratory*. Brooks/Cole, 1994.

[6] GELL-MANN, M. *The Quark and The Jaguar*. W. H. Freeman and Company, 1994.

[7] GRIZE, Y.-L., AND FINE, T. L. Continuous lower probability-based models for stationary processes with bounded and divergent time averages. *Annals of Probability 15* (1987), 783–803.

[8] KOLMOGOROV, A. N. On tables of random numbers. *Sankhya: The Indian Journal of Statistics, Series A* (1963), 369–376.

[9] KOLMOGOROV, A. N. On logical foundations of probability theory. In *Probability Theory and Mathematical Statistics*, K. Ito and J. Prokhorov, Eds., vol. 1021 of *Lecture Notes in Mathematics*. Springer-Verlag, 1983.

[10] KUMAR, A., AND FINE, T. L. Stationary lower probabilities and unstable averages. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 69* (1985), 1–17.

[11] LI, M., AND VITÁNYI, P. *An Introduction to Kolmogorov Complexity and Its Applications*, second ed. Graduate Texts in Computer Science. Springer, 1997.

[12] PAPAMARCOU, A., AND FINE, T. L. A note on undominated lower probabilities. *Annals of Probability 14* (1986), 710–723.

[13] PAPAMARCOU, A., AND FINE, T. L. Unstable collectives and envelopes of probability measures. *Annals of Probability* (1991), 893–906.

[14] PLATO. Meno. Translated by Bengamin Jowett. Available at http://textkit.com.

[15] SADROLHEFAZI, A., AND FINE, T. L. Finite-dimensional distribution and tail behavior in stationary interval-valued probability models. *Annals of Statistics 22* (1994), 1840–70.

[16] SHAFER, G., AND VOVK, V. *Probability and Finance. It's Only a Game!* Wiley Series in Probability and Statistics. John Wiley & Sons, 2001.

[17] VON MISES, R. *Probability, Statistics and Truth*. Dover, 1981.

[18] WALLEY, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.