Conditioning in Chaotic Probabilities Interpreted as a Generalized Markov Chain

Leandro Chaves Rêgo
Statistics Department
Federal University of Pernambuco
Recife-PE Brazil
Email: leandro@de.ufpe.br

Abstract

We propose a new definition for conditioning in the Chaotic Probability framework. We show that the Conditional Chaotic Probability model that we propose can be given the interpretation of a generalized Markov chain. Chaotic Probabilities were introduced by Fine et al. as an attempt to model chance phenomena with a usual set of measures $\mathcal{M}$ endowed with an *objective, frequentist interpretation* instead of a compound hypothesis or behavioral subjective one. We follow the presentation of the univariate case chaotic probability model and provide an instrumental interpretation of random process measures consistent with a conditional chaotic probability source, which can be used as a tool for simulation of our model. Given a finite time series, we also present a universal method for estimation of conditional chaotic probability models that is based on the analysis of the relative frequencies taken along a set of subsequences chosen by a given set of rules.

Keywords: Imprecise Probabilities, Foundations of Probability, Church Place Selection Rules, Probabilistic Reasoning, Conditioning, Complexity

1 Introduction

1.1 What is Chaotic Probability About?

Unlike the standard theory of real valued probability which since its beginning was *Janus faced* having to deal with both objective and subjective phenomena, sets of measures are mainly used to model behavior and subjective beliefs. Chaotic Probabilities were developed by Fine et al. [2] [4] [12] as an attempt to make sense of an objective, frequentist interpretation of a usual set of probability measures $\mathcal{M}$. In this setting, $\mathcal{M}$ is intended to model stable (although not stationary in the standard stochastic sense) physical sources of finite time series data that are highly irregular. The work was in part inspired in the following quotation from Kolmogorov 1983 [7]:

"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."

Despite that fact pointed out by Kolmogorov, the idea of models of physical chance phenomena sharing the precision of real number system is so well-entrenched that identifications of chaotic probability phenomena are difficult to make and hard to defend.

1.2 Previous Work and Overview

A large portion of the literature on imprecise probabilities gives a behavioral, subjective interpretation of this model Walley 1991 [14]. But some work has been done on the development of a frequentist interpretation of imprecise probabilities. Fine et al. have worked on asymptotics or laws of large numbers for interval-valued probability models [9] [13] [11] [6].

The work of Cozman and Chrisman 1997 [1] studying estimation of credal sets by analyzing limiting relative frequencies along a set of subsequences of a time series is very similar to the approach taken by Fierens and Fine, except that the latter restrict themselves to studying finite time series data. Another quote from Kolmogorov 1963 [8] explains the reason for such a restriction:

"The frequency concept based on the notion of limiting frequency as the number of trials increases to infinity, does not contribute anything to substantiate the ap-

plicability of the results of probability theory to real practical problems where we have always to deal with a finite number of trials."

In their work on Chaotic Probability models [2] [4], Fierens and Fine provided an instrumental interpretation of the model, a method for simulation of a random sequence given the model, and a method for estimation of the model given a finite random sequence. They have worked both on the univariate case and in the conditional case. This paper will parallel their approach on a different conditional setting that can be interpreted as a generalized Markov chain. We discuss the differences between their approach to conditioning and ours in Section 3.5. Roughly speaking, in our setting we have that a conditional chaotic probability model $\mathcal{M}_{|K}$ is a function that associates for each possible sequence y of K-previous outcomes a univariate chaotic probability model $\mathcal{M}_{|K}(y)$, i.e., a set of probability measures.

Section 2.1 provides an instrumental interpretation of conditional chaotic probability models. Although we do not claim this interpretation for explaining real world data, it is useful to develop it because it provides a means to understand the behavior of conditional chaotic probabilities using standard well-known tools of probability theory, it also provides the basis for simulation of these models, and finally it extends the interpretation proposed by Fierens and Fine for univariate chaotic probabilities. With that interpretation in mind we also provide a method for simulation of a data sequence given the Conditional Chaotic Probability model in Section 2.2.

In Section 3, we analyze the problem of estimating conditional chaotic probabilities from data. As in the univariate setup, we do that by studying the relative frequency taken along selected subsequences. We define three properties of a set of subsequence selection rules: *Conditional Causal Faithfulness, Conditional Homogeneity and Conditional Visibility.* By Conditional Causally Faithful rules we mean rules that, for each fixed sequence of past K outcomes, select subsequences such that the empirical and theoretical time averages along the selected subsequence are sufficiently close together. A set of rules renders $\mathcal{M}_{|K}$ conditionally visible if, for each fixed sequence y of past K outcomes, all measures in $\mathcal{M}_{|K}(y)$ can be estimated by relative frequencies along the selected subsequences. Finally, a set of rules is conditionally homogeneous if, for each fixed sequence y of past K outcomes, it cannot expose more than a small neighborhood of a single measure contained in the convex hull of $\mathcal{M}_{|K}(y)$, intuitively a set of rules is conditionally homogenous if the relative frequencies taken along the terms selected by the rules and that have y as the previous K outcomes are all close to a single measure in the convex hull of $\mathcal{M}_{|K}(y)$. We then prove the existence of families of causal subsequence selection rules that can make $\mathcal{M}_{|K}$ conditionally visible. Following the steps of Rêgo and Fine 2005 [12], in Section 4 we describe a universal methodology for finding a family of causal subsequence selection rules that can make $\mathcal{M}_{|K}$ conditionally visible, and in Section 5, we strengthen this result by assuring that the relative frequency taken along every subsequence analyzed is close to some measure in $\cup_y \mathcal{M}_{|K}(y)$ with high probability. In Section 6, we give the interpretation of conditional chaotic probabilities as a Generalized Markov Chain that instead of a single transition probability measure has a set of transition probabilities. We conclude in Section 7.

2 From Model to Data

2.1 Instrumental Interpretation

Let $\mathcal{X} = \{z_1, z_2, ..., z_\xi\}$ be a finite sample space.[1] We denote by $\mathcal{X}^*$ the set of all finite sequences of elements taken from $\mathcal{X}$. A particular sequence of n samples from $\mathcal{X}$ is denoted by $x^n = \{x_1, x_2, ..., x_n\}$. $\mathcal{P}$ denotes the set of all measures on the power set of $\mathcal{X}$ and $x^{i:j} = \{x_i, x_{i+1}, ..., x_{j-1}, x_j\}$. A conditional chaotic probability model given the past K outcomes $\mathcal{M}_{|K} : \mathcal{X}^K \to 2^{\mathcal{P}}$ is a function associating for each sequence of past K outcomes a subset of $\mathcal{P}$. Intuitively, $\mathcal{M}_{|K}$ models the "marginals" of the next outcome of some process generating sequences in $\mathcal{X}^*$ given the previous K outcomes. This section provides an interpretation of such a process.

Let F be a conditional chaotic selection function, $F : \mathcal{X}^* \to \cup_{y \in \mathcal{X}^K} \mathcal{M}_{|K}(y)$. At each instant i, a measure $\nu_i = F(x^{i-1})$ is chosen according to this selection function F. We require that the complexity of F be neither too complex, so that $\mathcal{M}_{|K}$ can not be exposed on the basis of a finite time series, nor too simple so that a standard stochastic process can be used to model the phenomena. We also require that F satisfies the following restriction

$$F(x^{i-1}) \in \mathcal{M}_{|K}(x^{i-K:i-1}),\ \forall i > K. \qquad (1)$$

Let $\mu_F \in \mathcal{P}^K$ be the initial probability distribution over the first K symbols.

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

[1] Recently, Fierens 2007 [3] extended the univariate Chaotic Probability Model to be defined on any subset of the reals. For ease of exposition, we focus on the finite case here.

variables X^n described by:

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) =$$
$$= \mu_F(X_1 = x_1, ..., X_K = x_K) \prod_{l=K+1}^{n} v_l(x_l) \qquad (2)$$

where $\nu_l \in M_{|K}(x^{l-K:l-1})$

We denote by $\mathcal{M}^*_{|K}$ the family of all such process measures P. This interpretation is usually considered instrumental (i.e., without commitment to reality), although theory also applies if there is empirical reality in the description of F, but F is simply unknown.

In the unconditional chaotic probability model, it is understood that all relevant information produced by the source is captured by the coarse-grained description provided by the set of measures $\mathcal{M}$ and the further information contained in the fine-grained description, F, has no empirical reality (Rêgo and Fine 2005 [12] emphasize that a similar situation occurs in quantum mechanics, see Gell-Mann 1994 [5], pg. 144–146). With this conditional model, we try to develop a model that does not discard all information provided by F, it considers the fact that for different previous K outcomes the source behaves differently. That is, it allows for the existence of a simple structure in the choice selection function.

Notice also, that in this account, for each sequence y of previous K outcomes, it is the whole set $\mathcal{M}_{|K}(y)$ that models the chance phenomena and not a "true" individual measure in $\mathcal{M}_{|K}(y)$ that is unknown to us, as in the usual compound hypothesis modeling.

Like in the unconditional case, no matter how complex the conditional selection function is, the process measure P is a standard stochastic process, the issue is whether it reflects the reality of the underlying phenomena. In the unconditional case, if the selection function is chaotic,[2] then all we can hope to learn and therefore predict for future terms in the sequence is the coarse-grained description of the model given by $\mathcal{M}$, a subset of $\mathcal{P}$. However, in the conditional case that we present here, the conditional selection function satisfies (1), and one can hope to learn $\mathcal{M}_{|K}(y)$ for each $y \in \mathcal{X}^K$. Therefore, in the conditional chaotic probability model, there is some structure in the chaotic behavior of the conditional selection function so that, as we will see in Section 3, the fact that the previous K outcomes were equal to some sequence y allow us to have a finer description of the model than just the coarse-grained description of the model given by $\mathcal{M}_{|K}$.

The next subsection digresses on a new statistical model that gives an application for the mathematical tools developed here. Fierens 2007 [3] also provides a motivation for the mathematical tools developed in the theory of chaotic probabilities using it for robust stochastic simulation.

2.1.1 Digression on a New Statistical Model

While our primary interest is not in a statistical compound hypothesis, the results of this paper do bear on a new statistical estimation model.

We can partially specify any stochastic process model $P \in \mathcal{P}$ by specifying the following set of conditional measures for the individual times for all possible $y \in \mathcal{X}^K$:

$$\mathcal{M}^P_{|K}(y) = \{\nu : (\exists j \geq K)(\exists x^j),\ x^{j-K+1:j} = y,$$

$$\nu(X_{j+1} \in A) = P(X_{j+1} \in A | X^j = x^j),\ \forall A \subset \mathcal{X}\}$$

Note that we do not keep track of the full conditioning event, only of the measure ν and of the previous K outcomes. We then wish to estimate $\mathcal{M}^P_{|K}(y),\ \forall y \in \mathcal{X}^K$, from data x^n. Also note that, in general, the process is not Markovian in the traditional sense, as the conditional selection function F depends on the whole history. Although, as we show in Section 6, it can be given the interpretation of a generalized Markov chain.

The model can be used in the following situation. Suppose we have an opponent in a game who can decide whether or not to act upon a trial t after examining the history of outcomes x^{t-1} prior to that trial. Certain distributions $P(X_t \in A | X^{t-1} = x^{t-1})$ for the trial at time t are favorable to us and others to our opponent. An assessment of the range of consequences to us from choices made by an intelligent opponent can be calculated from $\mathcal{M}^P_{|K}(y),\ \forall y \in \mathcal{X}^K$.

2.2 Simulation

In this section, we provide a method for sequence generation according to a source that is modeled by a conditional chaotic probability. First of all, we define a distance metric between probability measures as:

[2] As in the original work of Fierens and Fine [2] [4], the adjective "chaotic" is not used in the traditional technical sense of the mathematical literature on chaos, rather it is used in the sense of the selection function F being neither too simple nor too complex, where the complexity of the selection function can be measured, for example, in terms of Kolmogorov Complexity [10]. As well stated by an anonymous referee, *"the term 'chaotic probabilities' refers to viewpoint in which there is no true measure that is the model, because models for individual outcomes vary unpredictably while remaining in a given set $\mathcal{M}$"*.

$$(\forall \mu, \mu' \in \mathcal{P})\ d(\mu, \mu') \doteq \max_{z \in \mathcal{X}} |\mu(z) - \mu'(z)|$$

Note that $\mathcal{P}$ is compact with respect to d, so for all $\epsilon > 0$ we can find a minimal finite covering of it by Q_ϵ balls of radius ϵ, $\{B(\epsilon, \mu_i)\}$, where μ_i are computable measures. Let N_ϵ be the size of the smallest subset of the above covering of the simplex that covers the actual set of probabilities that can be selected by the conditional chaotic selection function, $\cup_{y \in \mathcal{X}^K} \mathcal{M}_{|K}(y)$, and denote this subset by $\mathcal{M}_\epsilon$. Let $\mathcal{M}_\epsilon(y)$ be the smallest subset of $\mathcal{M}_\epsilon$ that covers the set of probabilities that can be selected after the string of outcomes y, $\mathcal{M}_{|K}(y)$. Then, given an appropriate chaotic selection function $F : \mathcal{X}^* \to \mathcal{M}_\epsilon$, where $\mathcal{M}_\epsilon = \cup_{y \in \mathcal{X}^K} \mathcal{M}_\epsilon(y)$, satisfying $F(x^{i-1}) \in \mathcal{M}_\epsilon(x^{i-K:i-1})$, $\forall i > K$, and an appropriate initial probability distribution μ_F, the following algorithm can be used for simulation:

- Use a pseudo-random number generator to generate x^K according to μ_F
- For $i = K + 1$ to n
  - Choose $\nu_i = F(x^{i-1}) \in \mathcal{M}_\epsilon(x^{i-K:i-1})$
  - Choose any $\nu_i' \in B(\epsilon, \nu_i) \cap \mathcal{M}_{|K}(x^{i-K:i-1})$
  - Use a pseudo-random number generator to generate x_i according to ν_i'

Since we want to expose all of $\mathcal{M}_{|K}(y)$, $\forall y \in \mathcal{X}^K$, in a single but sufficiently long simulated sequence, we require F to visit several times each measure in $\mathcal{M}_\epsilon$. In the following sections, we study the problem of estimating a conditional chaotic probability model given a long enough but finite data sequence.

3 From Data to Model

3.1 Subsequence Analysis

The estimation process in the conditional chaotic probability framework uses a finite time series and analyzes it calculating ξ^K sets of relative frequencies taken along subsequences selected by causal subsequence selection rules (also known as Church place selection rules). These rules are called causal because the next choice is a function only of past values in the sequence and not, say, of the whole sequence. These rules satisfy the following:

Definition 3.1: An effectively computable function φ is a causal subsequence selection rule if:

$$\varphi : \mathcal{X}^* \to \{0, 1\}$$

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

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

Given a set of causal subsequence selection rules, Ψ, for each $\varphi \in \Psi$ and $y \in \mathcal{X}^K$, define the empirical and theoretical conditional time averages along a chosen subsequence by:

$$(\forall \mathbf{A} \subset \mathcal{X}),$$

$$\overline{\mu}_{\varphi,n,y}(\mathbf{A}) \doteq \sum_{i=K+1}^{n} \frac{I_{\mathbf{A}}(x_i) I_{\{y\}}(x^{i-K:i-1}) \varphi(x^{i-1})}{\lambda_{\varphi,n,y}}$$

$$\overline{\nu}_{\varphi,n,y}(\mathbf{A}) \doteq \frac{1}{\lambda_{\varphi,n,y}} \sum_{i=K+1}^{n} E[I_{\mathbf{A}}(X_i) | X^{i-1} = x^{i-1}] \times$$

$$\times I_{\{y\}}(x^{i-K:i-1}) \varphi(x^{i-1})$$

where $I_{\mathbf{A}}$ is the $\{0, 1\}$-valued indicator function of the event $\mathbf{A}$ and $\lambda_{\varphi,n,y} \doteq \sum_{i=K+1}^{n} I_{\{y\}}(x^{i-K:i-1}) \varphi(x^{i-1})$.

$\overline{\nu}_{\varphi,n}(.|y)$ can be rewritten in terms of the instrumental understanding as:

$$\overline{\nu}_{\varphi,n,y}(\mathbf{A}) \doteq \frac{1}{\lambda_{\varphi,n,y}} \sum_{i=K+1}^{n} \nu_i(\mathbf{A}) I_{\{y\}}(x^{i-K:i-1}) \varphi(x^{i-1})$$

A rule φ applied to x^n is said to be *conditionally causally faithful* if $\forall y \in \mathcal{X}^K$, $d(\overline{\nu}_{\varphi,n,y}, \overline{\mu}_{\varphi,n,y})$ is small. Essentially, φ is conditionally faithful if it does not extract an arbitrary pattern. The existence of such rules is shown by the following theorem.

Theorem 3.2: *Let ξ be the cardinality of $\mathcal{X}$ and denote the cardinality of Ψ by $||\Psi||$. Let $m \leq n$. If $||\Psi|| \leq t$, then for any process measure $P \in \mathcal{M}^*_{|K}$ and $y \in \mathcal{X}^K$:*

$$P(\max_{\varphi \in \Psi} \{d(\overline{\mu}_{\varphi,n,y}, \overline{\nu}_{\varphi,n,y}) : \lambda_{\varphi,n,y} \geq m\} \geq \epsilon) \leq$$

$$\leq 2\xi t \exp\{\frac{-\epsilon^2 m^2}{2(n-K)}\}$$

Proof: Follows immediately from Theorem 1 of Fierens and Fine 2003 [4], considering

$$I_{\{y\}}(x^{i-K:i-1}) \varphi(x^{i-1})$$

to be a selection rule, $\varphi'(x^{i-1})$, for the original sequence x^n. ∎

Note that, *as long as the size of the family of selection rules is not too big, conditional faithfulness is guaranteed with high probability if the subsequence selected is*

long enough. Note that the restriction on the size of Ψ is necessary, since if we allow all possible selection rules, we will get all the measures giving probability 1 to each one of the elements of the sample space $\mathcal{X}$.

Note also that, if we take $m = \alpha(n-K)$, for $\alpha \in (0,1)$, the size t of the family of selection rules can be as large as $e^{\rho(n-K)}$, for $\rho < \frac{\alpha^2\epsilon^2}{2}$; conditional faithfulness of the rules is guaranteed with high probability for large n.

3.2 Conditional Visibility and Estimation

The property that a set of rules, Ψ, must satisfy in order to expose all of $\mathcal{M}_{|K}(y)$, $\forall y \in \mathcal{X}^K$, is given by the following definition:

Definition 3.3: (Conditional Visibility) $\mathcal{M}_{|K}$ is **conditionally made visible** $(\Psi, \theta, \delta, m, n)$ by $P \in \mathcal{M}^*_{|K}$ if $\forall y \in \mathcal{X}^K$:

$$P(\bigcap_{\nu \in M_{|K}(y)} \bigcup_{\varphi \in \Psi} \{X^n : \lambda_{\varphi,n,y} \geq m, \\ d(\nu, \overline{\mu}_{\varphi,n,y}) \leq \theta\}) \geq 1 - \delta$$

Let $\hat{\mathcal{M}}^{\theta,\Psi,y}_{|K}$ be an estimator of $\mathcal{M}_{|K}(y)$ defined by:

$$\forall x^n \in \mathcal{X}^*,\ \hat{\mathcal{M}}^{\theta,\Psi,y}_{|K}(x^n) = \bigcup_{\varphi \in \Psi : \lambda_{\varphi,n,y} \geq m} B(\theta, \overline{\mu}_{\varphi,n,y})$$

where, $B(\theta, \overline{\mu}_{\varphi,n}) \doteq \{\mu \in \mathcal{P} : d(\mu, \overline{\mu}_{\varphi,n}) < \theta\}$.

Let $[\mathbf{A}]^\epsilon$ denote the ϵ-enlargement of a set $\mathbf{A}$ defined by:

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

The next theorem shows that for an appropriate set of rules Ψ, it is possible to conditionally expose $\mathcal{M}_{|K}$.

Theorem 3.4: (Estimability) *Let P render $\mathcal{M}_{|K}$ conditionally visible $(\Psi, \theta, \delta, m, n)$. Then, $\forall y \in \mathcal{X}^K$:*

$$P[[ch(\mathcal{M}_{|K}(y))]^{\theta+\epsilon} \supset \hat{\mathcal{M}}^{\theta,\Psi,y}_{|K} \supset \mathcal{M}_{|K}(y)] \geq 1 - \delta - \tau_n$$

where $\tau_n = 2\xi||\Psi|| \exp(\frac{-\epsilon^2 m^2}{2(n-K)})$ and $ch(\mathcal{M})$ is the convex hull of $\mathcal{M}$.

Proof: Follows immediately from Theorem 3 of Fierens and Fine 2003 [4] and Theorem 3.2 above, considering each fixed $y \in \mathcal{X}^K$. ∎

3.3 Conditional Homogeneity

There are some families of causal subsequence selection rules that are too simple to expose the structure underlying the conditional chaotic probability model, such families have the following property:

Definition 3.5: (Conditional Homogeneity) $P \in \mathcal{M}^*_{|K}$ is **conditionally homogeneous** $(\Psi, \theta, \delta, m, n)$ if $\forall y \in \mathcal{X}^K$:

$$P(\max_{\varphi_1,\varphi_2 \in \Psi} \{d(\overline{\mu}_{\varphi_1,n,y}, \overline{\mu}_{\varphi_2,n,y}) : \\ \lambda_{\varphi_1,n,y}, \lambda_{\varphi_2,n,y} \geq m\} \leq \theta) \geq 1 - \delta$$

3.4 Consistency Between Conditional Visibility and Conditional Homogeneity

Theorem 3.6: (Consistency) *Let $\epsilon > 1/m$. Assume that $\forall y \in \mathcal{X}^K$, there is an ϵ-cover of $\mathcal{M}_{|K}(y)$ by $N_\epsilon(y)$ open balls with centers in a set $\mathcal{M}_\epsilon(y) \doteq \{\mu^y_1, \mu^y_2, ..., \mu^y_{N_\epsilon(y)}\}$ such that, for each μ^y_i, there is a recursive probability measure $\nu \in B(\epsilon, \mu^y_i) \cap \mathcal{M}_{|K}(y)$. Let Ψ_0 be a set of causal subsequence selection rules. Assume also:*

$$\underline{p} \doteq \inf_{\nu \in \cup_{y \in \mathcal{X}^K} \mathcal{M}_{|K}(y)} \min_{z \in \mathcal{X}} \nu(z) > 0$$

Then, there are a process measure P and a family Ψ_1 such that, for large enough n, P will both render $\mathcal{M}_{|K}$ conditionally visible $(\Psi_1, 3\epsilon, \delta, m, n)$ and ensure conditional homogeneity $(\Psi_0, 6\epsilon, \delta, m, n)$ with

$$\delta = 2(\xi t_n + 1) \exp(\frac{-\epsilon^2 m^2}{2(n-K)})$$

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

Proof: It follows closely proofs contained in the Appendix C and D of [2]; we omit details here. ∎

The importance of this result is that there are conditional chaotic sources for which analysis by simple selection rules would give us the impression that the phenomena can be modeled by a standard probability model (indeed, it will look like a Markov chain where the set of states is $\mathcal{X}^K$). But if we further analyze the source with a set of more complex selection functions we can expose the underlying structure of the model. In this way, as pointed out by Fierens and Fine 2003, *the family of causal subsequence selection rules determines the power of the resolution of the model we see.*

3.5 Fierens and Fine's Approach to Conditioning

Fierens and Fine 2003 [2] also provided a model for Conditional Chaotic Probabilities, where the conditioning events are the previous K outcomes in the sequence. In their approach, they define

$$\mathbf{P}_{|K} = \{\nu : (\forall A \subseteq \mathcal{X}) \\ \nu(A, X^K) = E_\mu(I_A(X_{K+1})|X^K), \mu \in \mathcal{P}^{K+1}\}.$$

For them, a conditional chaotic probability model $\mathbf{M}_{|K}$ is any subset of $\mathbf{P}_{|K}$. They also provide an instrumental understanding of the model, by defining a selection function $\mathbf{F} : \mathcal{X}^* \to \mathbf{M}_{|K}$. It is easy to see that there is a one-to-one correspondence between their model and the one presented here. Given $\mathbf{M}_{|K}$, a conditional chaotic probability model according to our definition is given by:

$$\mathcal{M}_{|K}(y) = \{\mu \in \mathcal{P} : \forall z \in \mathcal{X}, \mu(z) = \nu(z, X^K = y), \nu \in \mathbf{M}_{|K}\}, \forall y \in \mathcal{X}^K.$$

For the converse, given $\mathcal{M}_{|K}$, a conditional chaotic probability model according to Fierens and Fine's definition is given by:

$$\mathbf{M}_{|K}(y) = \{\nu \in \mathbf{P}_{|K} : \exists y \in \mathcal{X}^K, \forall z \in \mathcal{X}, \nu(z, X^K = y) = \mu(z), \mu \in \mathcal{M}_{|K}(y)\}.$$

The major difference between both approaches is the estimation procedure; the set of subsequence selection rules Fierens and Fine allow for estimating the conditional chaotic probability model is a subset of the set we allow. Unlike us, for each fixed sequence of K outcomes y, Fierens and Fine analyze the subsequence $x^{y,n}$ of x^n, that is formed by all terms in x^n whose previous K outcomes are equal to y, using causal subsequence selection rules that depend only on past terms that appear in $x^{y,n}$, not on all past terms of the whole original sequence x^n, as we do in our approach. As the chaotic selection function both in their approach and in ours is allowed to depend on all past symbols of the sequence x^n, we believe that it is more appropriate to allow the more general set of selection rules we allow.

Although Fierens and Fine were able to prove results analogous to Theorems 3.2, 3.4, and 3.6 using their restricted set of selection rules, they did not provide a procedure for finding a family of selection rules Ψ that renders $\mathbf{M}_{|K}$ conditionally visible. We will now extend the result of Rêgo and Fine 2005 [12] providing a procedure for finding a family of selection rules Ψ that renders $\mathcal{M}_{|K}$ conditionally visible. In the next section, we provide a methodology for finding such a family of rules Ψ that works for any conditional chaotic probability source, and we call it a *universal family of selection rules*. As we see in the next section, for finding such a universal family it is crucial that we allow the more general set of subsequence selection rules that depend on the whole past terms in the sequence x^n. Unfortunately, as in the univariate case, such a family may "extract" more than $\cup_{y \in \mathcal{X}^K} \mathcal{M}_{|K}(y)$. We return to this point in Section 5.

4 Universal Family of Selection Rules

In this section we prove that there exists a universal family, which depends basically on the precision we want our estimator to have, that is able to conditionally expose all measures of any set of probabilities $\mathcal{M}_{|K}$.

Let $\lambda_{y,n} \doteq \sum_{i=K+1}^{n} I_{\{y\}}(x^{i-K:i-1})$.

Define for each family of causal selection rules, Ψ, and each $y \in \mathcal{X}^K$ the estimator based on this family as:

$$\hat{\mathcal{M}}^{\Psi,y}_{|K} \doteq \{\overline{\mu}_{\varphi,n,y} : \varphi \in \Psi, \lambda_{y,n} \geq m_0, \lambda_{\varphi,n,y} \geq m\}$$

Approximate $F(x^{j-1})$ by $F_\epsilon(x^{j-1}) = \mu_j$ if μ_j is the closest measure to $F(x^{j-1})$ among all μ_i's that belongs to $\mathcal{M}_\epsilon(x^{j-K:j-1})$. Let $F_{\epsilon,n}$ be the restriction of F_ϵ to $\mathcal{X}^{1:n}$ (all sequences of length not greater than n). The following theorem provides the desired method of finding a universal family of selection rules for conditional chaotic probability sources.

Intuitively, Theorem 4.1 states that as long as the Kolmogorov Complexity [10] of the conditional chaotic measure selection function is not too high, and we have a long enough data sequence, then for every given sequence $y \in \mathcal{X}^K$ of past K symbols that appeared frequently enough, we are able to make visible with high probability all measures in $\mathcal{M}_{|K}(y)$ that were selected frequently enough in the sequence.

Theorem 4.1: *Choose $f, f_0 \geq 1$, $\alpha_0 = (f_0 \xi^K)^{-1}$, $\alpha = (fN_\epsilon)^{-1}$ and let $m_0 = \alpha_0(n-K)$ and $m = \alpha\lambda_{y,n}$. Define $\mathcal{M}^f_{|K}(y) \doteq \{\nu : \nu \in \mathcal{M}_{|K}(y)$ and $\exists \mu_i \in \mathcal{M}_\epsilon(y)$ such that $d(\nu, \mu_i) < \epsilon$ and μ_i is selected at least m times by $F_{\epsilon,n}$ when the previous K outcomes were equal to y and $\lambda_{y,n} \geq m_0\}$. Given β smaller than $\frac{\alpha_0^2 \alpha^2 \epsilon^2}{2}$, choose $\epsilon' \in (0, \beta \log_2 e)$ and assume the Kolmogorov complexity, $K(F_{\epsilon,n})$, of $F_{\epsilon,n}$ satisfies the following condition:*

$$\exists \kappa \geq 0,\ \exists L_{\epsilon',\kappa} \text{ such that } \forall n \geq L_{\epsilon',\kappa}, \quad \frac{K(F_{\epsilon,n})}{n} < \beta \log_2 e + \frac{\kappa \log_2 n}{n} - \epsilon' \qquad (3)$$

Define $\mathcal{M}^*_{|K,R} \doteq \{P : P \in \mathcal{M}^*_{|K}$ and the corresponding F satisfies condition (3)$\}$. Then, for $n > \max\{L_{\epsilon',\kappa}, \frac{2\lceil \log_2 Q_\epsilon \rceil}{\epsilon'}\}$, there exists a family of causal subsequence selection rules Ψ_U, depending only on α_0, α, κ and ϵ, such that $\forall \mathcal{M}_{|K}$, and $\forall P \in \mathcal{M}^*_{|K,R}$:

$$P(\bigcap_{y \in \mathcal{X}^K} \{X^n : [ch(\mathcal{M}_{|K}(y))]^{4\epsilon} \supset [\hat{\mathcal{M}}^{\Psi_U,y}_{|K}]^{3\epsilon} \supset \mathcal{M}^f_{|K}(y)\}) \geq 1 - \delta,$$

where $\gamma = \frac{\alpha_0^2\alpha^2\epsilon^2}{2} - \beta$ *and* $\delta = 2\xi^{K+1}n^{\kappa}e^{\alpha_0^2\alpha^2\epsilon^2 K}e^{-\gamma n}$.

Remark 4.2: Note that if $\lambda_{y,n} < m_0$, then by definition we have $\hat{\mathcal{M}}_{|K}^{\Psi_U,y} = \mathcal{M}_{|K}^{f}(y) = \emptyset$. Thus, we fail to estimate $\mathcal{M}_{|K}(y)$ in this case. But the fraction of times a string of outcomes $y \in \mathcal{X}^K$ such that $\lambda_{y,n} < m_0$ appears in a sequence X^n is bounded from above by $(1/f_0)$. Therefore, for f_0 sufficiently large it is reasonable to expect that such measures may not be estimated.

Remark 4.3: Note also that if $\lambda_{y,n} \geq m_0$, then the fraction of times a measure in $\mathcal{M}_{|K}(y) \setminus \mathcal{M}_{|K}^{f}(y)$ is used to generate an outcome in a sequence X^n is bounded from above by $(1/f)$. Therefore, for f sufficiently large it is reasonable to expect that such measures may not be estimated.

Proof: Define a family of selection functions, Ψ_G, that corresponds to $F_{\epsilon,n}$ as follows: $\Psi_G = \{\varphi_i^G$, for $1 \leq i \leq N_\epsilon\}$, where, for $0 \leq j \leq n-1$:

$$\varphi_i^G(x^j) \doteq \begin{cases} 1 & \text{if } F_{\epsilon,n}(x^j) = \mu_i \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

As each φ_i^G is a function of $F_{\epsilon,n}$ and μ_i, and $\lceil \log_2 Q_\epsilon \rceil$ is an upper bound on the number of bits necessary to specify the index i of the particular measure μ_i, the Kolmogorov complexity, $K(\varphi_i^G)$, of φ_i^G satisfies:

$$\max_i K(\varphi_i^G) \leq K(F_{\epsilon,n}) + \lceil \log_2 Q_\epsilon \rceil \tag{5}$$

It then follows, from our hypothesis, that for $1 \leq i \leq N_\epsilon$ and $\forall n \geq L_{\epsilon',\kappa}$, $K(\varphi_i^G)$ satisfies the following condition:

$$\frac{K(\varphi_i^G)}{n} < \beta \log_2 e + \frac{\kappa \log_2 n}{n} - \epsilon' + \frac{\lceil \log_2 Q_\epsilon \rceil}{n}$$

Therefore, for $n > \max(L_{\epsilon',\kappa}, \frac{2\lceil \log_2 Q_\epsilon \rceil}{\epsilon'})$:

$$\frac{K(\varphi_i^G)}{n} < \beta \log_2 e + \frac{\kappa \log_2 n}{n} - \frac{\epsilon'}{2}$$

Let Ψ_U consist of all rules of Kolmogorov complexity less than or equal to $\beta n \log_2 e + \kappa \log_2 n - 1$. Note that since for $n > \max\{L_{\epsilon',\kappa}, \frac{2\lceil \log_2 Q_\epsilon \rceil}{\epsilon'}\}$, $\frac{n\epsilon'}{2} > 1$, so Ψ_U includes Ψ_G for n large enough.

As $||\Psi_U|| \leq 2^{n\beta \log_2 e + \kappa \log_2 n} = n^\kappa e^{\beta n}$, $m_0 = \alpha_0(n-K)$, $m = \alpha\lambda_{y,n}$ and $\gamma = \frac{\alpha_0^2\alpha^2\epsilon^2}{2} - \beta > 0$, by the causal faithfulness theorem, for any $P \in \mathcal{M}_{|K}^*$,

$$\begin{aligned}
&P(X^n : \max_{y\in\mathcal{X}^K} \max_{\varphi\in\Psi_U} \{d(\overline{\mu}_{\varphi,n,y}, \overline{\nu}_{\varphi,n,y}) : \\
&\lambda_{y,n} \geq m_0, \lambda_{\varphi,n,y} \geq m\} \geq \epsilon) = \\
&= P(X^n : \max_{y\in\mathcal{X}^K} \max_{\varphi\in\Psi_U} \{d(\overline{\mu}_{\varphi,n,y}, \overline{\nu}_{\varphi,n,y}) : \\
&\lambda_{y,n} \geq m_0, \lambda_{\varphi,n,y} \geq \alpha\lambda_{y,n}\} \geq \epsilon) \leq \\
&\leq P(X^n : \max_{y\in\mathcal{X}^K} \max_{\varphi\in\Psi_U} \{d(\overline{\mu}_{\varphi,n,y}, \overline{\nu}_{\varphi,n,y}) : \\
&\lambda_{y,n} \geq m_0, \lambda_{\varphi,n,y} \geq \alpha m_0\} \geq \epsilon) \leq \\
&\leq P(X^n : \max_{y\in\mathcal{X}^K} \max_{\varphi\in\Psi_U} \{d(\overline{\mu}_{\varphi,n,y}, \overline{\nu}_{\varphi,n,y}) : \\
&\lambda_{\varphi,n,y} \geq \alpha_0\alpha(n-K)\} \geq \epsilon) \leq \\
&\leq 2\xi^{K+1}n^\kappa e^{\frac{\alpha_0^2\alpha^2\epsilon^2 K}{2}} e^{-\gamma n}
\end{aligned}$$

Note that, since for $\alpha_0 = (f_0\xi^K)^{-1}$, for all X^n, there exists y such that $\lambda_{y,n} \geq m_0$. And for $\alpha = (fN_\epsilon^{-1})$, we know that for all X^n and for all y, there exists i such that $\lambda_{\varphi_i^G,n,y} \geq m$, as $\Psi_G \subset \Psi_U$, we have that for all X^n the maximum above is taken over a non-empty set.

To prove the theorem, let φ_i^G be as defined in Equation (4), then for a fixed X^n, by definition of $\mathcal{M}_{|K}^{f}(y)$, $\forall \nu \in \mathcal{M}_{|K}^{f}(y)$, $\exists \mu_i \in \mathcal{M}_\epsilon(y)$ such that $d(\nu,\mu_i) < \epsilon$, $\lambda_{\varphi_i^G,n,y} \geq m$ and $\lambda_{y,n} \geq m_0$ (Note the index i depends on X^n). Then, using the triangle inequality property:

$$\begin{aligned}
&\max_{y\in\mathcal{X}^K} \sup_{\nu\in\mathcal{M}_{|K}^{f}(y)} d(\nu, \overline{\mu}_{\varphi_i^G,n,y}) \leq \\
&\max_{y\in\mathcal{X}^K} \sup_{\nu\in\mathcal{M}_{|K}^{f}(y)} d(\overline{\mu}_{\varphi_i^G,n,y}, \overline{\nu}_{\varphi_i^G,n,y}) + \\
&\max_{y\in\mathcal{X}^K} \sup_{\nu\in\mathcal{M}_{|K}^{f}(y)} d(\overline{\nu}_{\varphi_i^G,n,y}, \mu_i) + \\
&\max_{y\in\mathcal{X}^K} \sup_{\nu\in\mathcal{M}_{|K}^{f}(y)} d(\mu_i, \nu)
\end{aligned}$$

and since $\overline{\nu}_{\varphi_i^G,n,y}$ is the time average of the actual measures selected by F in the ball $B(\epsilon,\mu_i)$, $d(\overline{\nu}_{\varphi_i^G,n,y}, \mu_i) < \epsilon$, and as $\Psi_G \subset \Psi_U$, the following holds,

$$\begin{aligned}
&\{X^n : \max_{y\in\mathcal{X}^K} \max_{\varphi\in\Psi_U} \{d(\overline{\mu}_{\varphi,n,y}, \overline{\nu}_{\varphi,n,y}) : \\
&\lambda_{y,n} \geq m_0, \lambda_{\varphi,n,y} \geq m\} < \epsilon\} \subset \\
&\{X^n : \max_{y\in\mathcal{X}^K} \sup_{\nu\in\mathcal{M}_{|K}^{f}(y)} \min_{\varphi\in\Psi_U} \{d(\nu, \overline{\mu}_{\varphi,n,y}) : \\
&\lambda_{y,n} \geq m_0, \lambda_{\varphi,n,y} \geq m\} < 3\epsilon\}.
\end{aligned} \tag{6}$$

Equation 6 implies,

$$\begin{aligned}
&\{X^n : \max_{y\in\mathcal{X}^K} \max_{\varphi\in\Psi_U} \{d(\overline{\mu}_{\varphi,n,y}, \overline{\nu}_{\varphi,n,y}) : \\
&\lambda_{\varphi,n,y} \geq m\} < \epsilon\} \subset
\end{aligned}$$

$$\{\bigcap_{y\in\mathcal{X}^K}\{X^n:$$
$$[ch(\mathcal{M}_{|K}(y))]^{4\epsilon}\supset[\hat{\mathcal{M}}_{|K}^{\Psi_U,y}]^{3\epsilon}\supset\mathcal{M}_{|K}^{f}(y)\}\} \quad (7)$$

Theorem 4.1 follows from the causal faithfulness Theorem 3.2. ∎

The problem with the sort of estimator provided by the above theorem is that on one hand it is able to conditionally expose all measures in $\mathcal{M}_{|K}(y)$ that appeared frequently enough in the process, if y also appeared frequently enough in the outcomes. On the other hand, we have that for each $y\in\mathcal{X}^K$ that appeared frequently enough, the estimator is only guaranteed to be included in an enlarged neighborhood of $\mathcal{M}_{|K}(y)$'s convex hull and in some cases this can be rather larger than $\mathcal{M}_{|K}(y)$.

The following section proves a theorem that given x^n provides a methodology for finding a universal family of subsequences, $\Psi(x^n)$, that is both able to conditionally expose all measures in $\mathcal{M}_{|K}(y)$ that appeared frequently enough in the process, if y appeared frequently enough, and contains only these subsequences whose empirical time averages are close enough to $\mathcal{M}_{|K}(y)$ with high probability. We will call this family to be **conditionally strictly faithful**.

5 Conditionally Strictly Faithful Family of Subsequences

In this section, we propose a methodology for finding a conditionally strictly faithful family of subsequences that can both conditionally expose all measures in $\mathcal{M}_{|K}(y)$ that appear frequently enough in the process, if y appears frequently enough; and contains only these subsequences whose empirical time averages are close enough to $\mathcal{M}_{|K}(y)$ with high probability.

The problem with the set of rules Ψ_U is that it may contain rules that are not conditionally homogeneous, i.e., rules that given that the previous outcomes are equal to y select subsequences generated by mixtures of measures μ_i's. In our proposed methodology in this section, we will analyze each rule $\varphi\in\Psi_U$ with a universal family Ψ_U^{φ} (see definition below) and include φ in $\Psi(x^n)$ only if it is conditionally homogeneous. As Ψ_U^{φ} is universal for the subsequence selected by φ, it will be able to identify if it is or not conditionally homogeneous with high probability. Thus, our family of sequences $\Psi(x^n)$ is constructed in a two-stage process: first we consider the family of selection rules Ψ_U which consists of all rules of at most a certain complexity value which is able to make $\mathcal{M}_{|K}$ conditionally visible; then we filter the rules contained in Ψ_U so that it contains only conditionally homogenous subsequences whose relative frequencies are close enough to a measure in $\cup_{y\in\mathcal{X}^K}\mathcal{M}_{|K}(y)$.

The following theorem proves the desired result, i.e., if the Kolmogorov complexity of the conditional chaotic measure selection function is not too high and we have a long enough data sequence, with high probability we can conditionally make visible all and only measures that were used frequently enough in the sequence.

Theorem 5.1: *Choose $f_0, f\geq 1$, $\alpha_0=(f_0\xi^K)^{-1}$, $\alpha_1=(fN_\epsilon)^{-1}$, $\alpha_2=N_\epsilon^{-1}$ and let $m_0=\alpha_0(n-K)$, $m=\alpha_1\lambda_{y,n}$. Define $\mathcal{M}_{|K}^{f}(y)\doteq\{\nu:\nu\in\mathcal{M}_{|K}(y)$ and $\exists\mu_i\in\mathcal{M}_\epsilon(y)$ such that $d(\nu,\mu_i)<\epsilon$ and μ_i is selected at least m times by $F_{\epsilon,n}$ when the previous K outcomes were equal to y and $\lambda_{y,n}\geq m_0\}$ and define $\mathcal{M}_{\epsilon}^{f}(y)\doteq\{\mu_i:\mu_i\in\mathcal{M}_\epsilon(y)$ and μ_i is selected at least $\alpha_2 m$ times by $F_{\epsilon,n}$ when the previous K outcomes were equal to y and $\lambda_{y,n}\geq m_0\}$. Given β smaller than $\frac{\alpha_0^2\alpha_1^2\alpha_2^2\epsilon^2}{2}$, choose $\epsilon'\in(0,\beta\log_2 e)$ and assume the Kolmogorov complexity, $K(F_{\epsilon,n})$, of $F_{\epsilon,n}$ satisfies the same condition (3), i.e.,:*

$$\exists\kappa\geq 0,\ \exists L_{\epsilon',\kappa}\ \textit{such that}\ \forall n\geq L_{\epsilon',\kappa},$$
$$\frac{K(F_{\epsilon,n})}{n}<\beta\log_2 e+\frac{\kappa\log_2 n}{n}-\epsilon'$$

Define $\mathcal{M}_{|K,R}^{*}\doteq\{P:P\in\mathcal{M}_{|K}^{*}$ and the corresponding F satisfies condition (3)$\}$. Then, for $n>\max\{L_{\epsilon',\kappa},\frac{2\lceil\log_2 Q_\epsilon\rceil}{\epsilon'}\}$, for each x^n, there exists a family of subsequences $\Psi(x^n)$, depending only on α_0, α_1, α_2, κ and ϵ, such that $\forall\mathcal{M}_{|K}$ and $\forall P\in\mathcal{M}_{|K,R}^{*}$:[3]

$$P(\{X^n:\max_{y\in\mathcal{X}^K}\sup_{\mu\in\mathcal{M}_{|K}^{f}(y)}\min_{\nu\in\hat{\mathcal{M}}_{|K}^{\Psi(X^n),y}}d(\mu,\nu)<3\epsilon\}$$
$$\cap\{X^n:\max_{y\in\mathcal{X}^K}\max_{\nu\in\hat{\mathcal{M}}_{|K}^{\Psi(X^n),y}}\min_{\mu\in\mathcal{M}_{\epsilon}^{f}(y)}d(\mu,\nu)<6\epsilon\})$$
$$\geq 1-\delta_1$$

where $\gamma_1=\frac{\alpha_0^2\alpha_1^2\alpha_2^2\epsilon^2}{2}-\beta$, $S_\epsilon\doteq\min\{Q_\epsilon,n^\kappa e^{\beta n}\}$ and $\delta_1=4\xi^{K+1}S_\epsilon n^\kappa e^{\frac{\alpha_0^2\alpha_1^2\alpha_2^2\epsilon^2 K}{2}}e^{-\gamma_1 n}$.

Proof: It follows closely the proof of Theorem 3 contained in the appendix of [12]; we omit details here. ∎

[3] If $\Psi(X^n)=\emptyset$, we adopt the following convention:

$$\max_{y\in\mathcal{X}^K}\sup_{\mu\in\mathcal{M}_{|K}^{f}(y)}\min_{\nu\in\hat{\mathcal{M}}_{|K}^{\emptyset,y}}d(\mu,\nu)=\infty$$

and

$$\max_{y\in\mathcal{X}^K}\max_{\nu\in\hat{\mathcal{M}}_{|K}^{\emptyset,y}}\min_{\mu\in\mathcal{M}_{\epsilon}^{f}(y)}d(\mu,\nu)=0.$$

6 Interpretation as Generalized Markov Chain

The conditional chaotic probability model studied in this paper can be given the interpretation of a generalized Markov chain (GMC). The difference from the standard Markov chain is that the transition probabilities are given by sets of probability measures instead of single probabilities. Therefore, consider the following definitions of the parameters of the GMC:

- **States:** There are ξ^K states, one state for each $y \in \mathcal{X}^K$.
- **Initial Probabilities:** They are given by the initial probability of the first K symbols of the sequence, $\mu_F \in \mathcal{P}^K$.
- **Transition Set of Probabilities:**

$$\mathcal{M}_{|K}(y_{i+1}|y_i) \doteq \begin{cases} \{\nu(y_{i+1}(K)) : \nu \in \mathcal{M}_{|K}(y_i)\}, \\ \qquad \text{if } y_i(l+1) = y_{i+1}(l), \\ \qquad \text{for } 1 \leq l \leq K-1 \\ \{0\}, \text{ otherwise.} \end{cases}$$

  where $y_i(l)$ is the l-th position of the i-th state of the GMC.

Although this GMC looks like a partially specified Markov chain, they differ in the fact that in the GMC there is no single underlying "true" transition probability as in the partially specified Markov chain.

As pointed out by an anonymous referee, we must take care with the interpretation of the conditional chaotic probability model as a GMC. On one hand, usually a Markov chain describes a random phenomenon without memory. On the other hand, the instrumental interpretation of the conditional chaotic probability model proposed is sensible to the initial conditions of the realization of the random experiment; each initial condition determines a unique process P as defined in (2). As argued in Section 2.1, the issue is that P does not reflect the reality of the underlying phenomena. In a chaotic probability model, all that can be learnt and used to predict the next outcome in the sequence is $\mathcal{M}_{|K}(y)$ for each $y \in \mathcal{X}^K$, i.e., the transition set of probabilities of the GMC. Thus, a GMC is memoryless in the sense that once one knows the transition set of probabilities of the GMC all that we can learn and use to predict about the distribution of the next outcome in the sequence is given by the present state $y \in \mathcal{X}^K$ of the GMC, and it is chaotic in the sense that given that the present state is $y \in \mathcal{X}^K$ which measure will actually produce the next outcome varies unpredictably while remaining in $\mathcal{M}_{|K}(y)$.

7 Conclusions and Future Work

For ease of exposition, in this paper we focused on the case of conditioning on the previous K outcomes. It is easy to see that the results presented can be easily generalized to conditioning on a family of selection rules Φ such that $\exists L < n-1$ such that the following two conditions hold:

1. $\forall \phi_1, \phi_2 \in \Phi,\ \phi_1 \neq \phi_2$ implies $\phi_1(x^i) \cdot \phi_2(x^i) = 0,\ L < i \leq n-1$
2. $\sum_{\phi \in \Phi} \phi(x^i) = 1,\ L < i < n$

The development of chaotic probability theory is an important conceptual achievement, since it will provide us with a more powerful and general tool for analyzing time series. With the increasing size and number of data sets available nowadays, a different way of looking at them, provided by this theory, can have a huge impact in our world.

Although we do not have analyzed any practical real world data supporting the model, the main mathematical tools that enhance our capability of recognizing such phenomena (since we believe that we are only likely to find what we expect to see) have been presented. Therefore, new concepts of probability are likely to open our perception and understanding of chance phenomena.

To further develop the chaotic probability theory, a method to evaluate self-consistency of simulation and estimation needs to be studied (for details, see Fierens and Fine 2003 [2] [4]). Also, implications of this theory for inference and decision making problems have to be investigated.

In a broader perspective, the possibility of modeling physical chance phenomena with a set of measures, raises the question about the existence of other physical quantities that have properties that cannot be quantified by a single real number, but only as a set of them.

Acknowledgements

We want to specially thanks Terry Fine and Pablo Fierens for useful talks about Chaotic Probabilities Models. We also would like to thank Terry Fine for important suggestions and comments in early drafts of this work. Last but not least, we thank anonymous referees that made useful comments about this work.

References

[1] F. Cozman and L. Chrisman, *Learning convex sets of probability from data*, Tech. Report CMU-RI-TR-97-25, Robotics Institute, Carnegie Mellon University, 1997.

[2] P. I. Fierens, *Towards a Chaotic Probability Model for Frequentist Probability*, PhD thesis, Cornell University, 2003.

[3] P. I. Fierens, *An extension of chaotic probability models to real-valued variables*, in ISIPTA'07 Proceedings, July 2007.

[4] P. I. Fierens and T. L. Fine, *Toward a Chaotic Probability Model for Frequentist Probability: The Univariate Case.*, July 2003, pp. 245–259.

[5] M. Gell-Mann, *The Quark and The Jaguar*, W. H. Freeman and Company, 1994.

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

[7] A. N. Kolmogorov, *On logical foundations of probability theory*, vol. 1021 of Lecture Notes in Mathematics, Springer-Verlag, 1983.

[8] A. N. Kolmogorv, *On tables of random numbers*, Sankhya: The Indian Journal of Statistics, (1963), p. 369.

[9] A. Kumar and T. L. Fine, *Stationary lower probabilities and unstable averages*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 69 (1985), pp. 1–17.

[10] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Graduate Texts in Computer Science, Springer, second ed., 1997.

[11] A. Papamarcou and T. L. Fine, *A note on undominated lower probabilities*, Annals of Probability, 14 (1986), pp. 710–723.

[12] L. C. Rêgo and T. L. Fine, *Estimation of chaotic probabilities*, in ISIPTA'05 Proceedings, 2005, pp. 297–305.

[13] A. Sadrolhefazi and T. L. Fine, *Finite-dimensional distribution and tail behavior in stationary interval-valued probability models*, Annals of Statistics, 22 (1994), pp. 1840–1870.

[14] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall Pubs., 1991.