Stochastic Processes and Temporal Data Mining

Paul Cotofrei
Information Management Institute
University of Neuchâtel
Neuchâtel, Switzerland
paul.cotofrei@unine.ch

Kilian Stoffel
Information Management Institute
University of Neuchâtel
Neuchâtel, Switzerland
kilian.stoffel@unine.ch

ABSTRACT

This article tries to give an answer to a fundamental question in temporal data mining: "Under what conditions a temporal rule extracted from up-to-date temporal data keeps its confidence/support for future data". A possible solution is given by using, on the one hand, a temporal logic formalism which allows the definition of the main notions (*event, temporal rule, support, confidence*) in a formal way and, on the other hand, the stochastic limit theory. Under this probabilistic temporal framework, the equivalence between the existence of the support of a temporal rule and the law of large numbers is systematically analyzed.

Categories and Subject Descriptors

H.2.8 [**DATABASE MANAGEMENT**]: Database Applications—*data mining*; G.3 [**Mathematics of Computing**]: PROBABILITY AND STATISTICS—*Stochastic processes*; F.4.1 [**MATHEMATICAL LOGIC AND FORMAL LANGUAGES**]: Mathematical Logic—*temporal logic*

General Terms

THEORY

Keywords

Consistency of temporal rules, stochastic limit theory, stochastic processes, temporal data mining, temporal logic formalism

1. INTRODUCTION

The domain of temporal data mining focuses on the discovery of causal relationships among events that are ordered in time and may be causally related. The contributions in this domain encompass the discovery of temporal rule, of sequences and of patterns. However, in many respects this is just a terminological heterogeneity among researchers that are, nevertheless, addressing the same problem, albeit from different starting points and domains.

For the temporal data mining task which consists in extracting knowledge represented as temporal rules (expressing the intrinsic dependence between successive events in time), one of the most important goals is to guarantee that a rule learned from a local data subset keeps its "correctness" (expressed by the confidence measure) when applied on future data subsets. In a simplistic approach we could affirm that this guarantee is made if the data model does not change over time. In fact, if temporal data is modelled by a stochastic process, the model is characterized not only by the marginal distribution of the coordinates of the process, but also by the amount of dependence between these coordinates. And if these coordinates represent real events occurring at time moment i, it is possible to have such a great amount of dependence between all events over time that all rules learned from a local set of events (those occurring in a given period of time) are effective only for this events set. This effect, which is similar to the overfitting effect for the classification trees, can be avoided only if different local sets of events are "almost" independent. It is obvious the trade-off we must assure between the necessity of dependency between events nearby in time (which makes the rules meaningful) and the necessity of independence between events faraway in time (which makes the rules effective for all data).

Before defining a stochastic model for a temporal data mining task, we need to give a formal definition for the basic notions such as *event, temporal rule* or *confidence*. Although there is a rich bibliography concerning formalism for temporal databases, there are very few articles on this topic for temporal data mining. In [1, 2, 16] general frameworks for temporal mining are proposed, but usually the research on causal and temporal rules is more concentrated on the methodological/algorithmic aspect, and less on the theoretical aspect. Based on a methodology for temporal rule extraction, described in [4], we proposed in [5, 6] an innovative formalism based on first-order temporal logic, which permits an abstract view on temporal rules. The formalism is developed around a time model for which the events are those that describe system evolution (event-based temporal logics).

If first-order logic is widely recognized as being a fundamental building block in knowledge representation, it does not have, however, the expressive power to deal with many situations of interest, especially those related to uncertainty [15]. And if the uncertainty is a fundamental and irreducible aspect of our knowledge about the world, the probability is the most well-understood and widely applied logic for computational scientific reasoning under uncertainty. By attaching a probabilistic model (more precise, a stochastic process ψ) to our formalism we obtain a probabilistic temporal framework. An important concept defined in this formalism is the property of *consistency*, which guarantees the preservation over time of the confidence/support of a temporal rule. Even if the condition of independence for the stochastic process is sufficient to induce the property of consistency for a temporal structure gen-

erated by ψ, this condition is not suitable for modelling temporal data mining tasks. By using advanced theorems from the stochastic limit theory, we succeeded improving that a certain amount of dependence of the stochastic process (called near-epoch dependence) is the highest degree of dependence which is sufficient to induce the property of consistency.

The rest of the paper is structured as follows. In the next section, the first-order temporal logic formalism is briefly described (definitions of the main terms – *event, temporal rules, confidence* – and concepts – *consistent linear time structure, general interpretation*). The definitions and theorems concerning the extension of the formalism towards a stochastic temporal logic and a deep analysis of the influence of different types of stochastic dependencies on the model consistency are presented in Section 3. Finally, the last section summarizes our work.

2. FORMALISM OF TEMPORAL RULES

Time is ubiquitous in information systems, but the mode of representation/perception varies in function of the purpose of the analysis [3]-[9]. Firstly, there is a choice of a *temporal ontology*, which can be based either on *time points* (instants) or on *intervals* (periods). Secondly, time may have a *discrete* or a *continuous* structure. Finally, there is a choice of *linear* vs. *nonlinear* time (e.g., acyclic graph). Our choice is a temporal domain represented by linearly ordered discrete instants.

For the purpose of our approach we consider a restricted first-order temporal language L which contains only constant symbols $\{c, d, ..\}$, n-ary ($n \geq 1$) function symbols $\{f, g, ..\}$, variable symbols $\{y_1, y_2, ...\}$, n-ary predicate symbols ($n \geq 1$, so no proposition symbols), the set of relational symbols $\{=, <, \leq, >, \geq\}$, the logical connective **AND** with two graphical forms ($\wedge$ and $\mapsto$) and a temporal connective of the form ∇_k, $k \in \mathbb{Z}$, where k strictly positive means *after k time instants*, k strictly negative means *before k time instant* and $k = 0$ means *now*.

Definition 1. An event (or temporal atom) is an atom formed by the predicate symbol E followed by a bracketed n-tuple of terms ($n \geq 1$) $E(t_1, t_2, \ldots, t_n)$. The first term of the tuple, t_1, is a constant symbol representing the name of the event and all other terms are expressed according to the rule $t_i = f(t_{i1}, \ldots, t_{ik_i})$. A short temporal atom (or the event's head) is the atom $E(t_1)$.

Definition 2. A constraint formula for the event $E(t_1, t_2, \ldots t_n)$ is a conjunctive compound formula, $E(t_1, t_2, \ldots t_n) \wedge C_1 \wedge C_2 \wedge \cdots \wedge C_k$. Each C_j is a relational atom $t \rho c$, where the first term t is one of the terms t_i, $i = 1 \ldots n$, ρ is a relational symbol and the second term is a constant symbol.

For a short temporal atom $E(t_1)$, the only constraint formula that is permitted is $E(t_1) \wedge (t_1 = c)$. We denote such constraint formula as *short constraint formula*.

Definition 3. A temporal rule is a formula of the form $H_1 \wedge \cdots \wedge H_m \mapsto H_{m+1}$, where H_{m+1} is a short constraint formula and $H_i, i = 1..m$ are constraint formulae, prefixed by the temporal connectives ∇_{-k}, $k \geq 0$. The maximum value of the index k is called the time window of the temporal rule.

If we change in Definition 1 the conditions imposed on the terms $t_i, i = 1...n$ to "each term t_i is a variable symbol", we obtain the definition of a temporal atom template. We denote such a template as $E(y_1, \ldots, y_n)$. Following the same rationale, a constraint formula template for $E(y_1, \ldots, y_n)$ is a conjunctive compound formula, $C_1 \wedge C_2 \wedge \cdots \wedge C_k$, where the first term of each relational

atom C_j is one of the variables y_i, $i = 1 \ldots n$. Finally, by replacing in Definition 3 the notion "constraint formula" with "constraint formula template" we obtain the definition of a temporal rule template.

The semantics of L is provided by an interpretation I over a domain D (in our formalism, D is always a linearly ordered domain). The interpretation assigns an appropriate meaning over D to the (non-logical) symbols of L. Usually, the domain D is imposed during the discretisation phase, which is a pre-processing phase used in almost all knowledge extraction methodologies. Based on Definition 1, an event can be seen as a labeled (constant symbol t_1) sequence of points extracted from raw data and characterized by a finite set of features (terms $t_2, \cdots, t_n$). Consequently, the domain D is the union $D_e \cup D_f$, where the set D_e contains all the strings used as event names and the set D_f represents the union of all domains corresponding to chosen features.

Example 1. Consider a temporal database containing the results of a series of experiments, each experiment being characterized by a *decision* ($\{Run, Stop\}$) and by a real positive parameter h (the average of two measurements). In the frame of our formalism, the language L will contain a 2-ary predicate symbol E, two variable symbols y_1, y_2, a 2-ary function symbol h, two sets of constant symbols – $\{d_1, d_2\}$ and $\{c_1, \ldots, c_n\}$ – and the usual set of relational symbols and logical(temporal) connectives. According to the syntactic rules of L, an event is defined as $E(d_i, h(c_j, c_k))$, an event template as $E(y_1, y_2)$, whereas $\nabla_{-2}(y_1 = d_1) \wedge \nabla_{-1}(y_1 = d_2 \wedge y_2 < c_j) \mapsto (y_1 = d_1)$ is an example of a temporal rule template with a time window of 2. Concerning the semantics of L, the domain D is defined as the union of $D_e = \{Run, Stop\}$ and $D_f = \Re^+$.

To define a first-order linear temporal logic based on L, we need a structure having a temporal dimension and capable of capturing the relationship between a time moment and the interpretation **I** at this moment.

Definition 4. Given L and a domain D, a (first order) linear time structure is a triple $M = (S, x, \mathbf{I})$, where S is a set of states, $x : \mathbb{N} \to S$ is an infinite sequence of states $(s_{(1)}, s_{(2)}, \ldots, s_{(n)}, \ldots)$ and **I** is a function that associates with each state s an interpretation $\mathbf{I}_s$ of all symbols from L.

In the framework of linear temporal logic, the set of symbols is divided into two classes, the class of global symbols and the class of local symbols. The formalism of temporal rules assumes that all function symbols (including constants) and all relational symbols are global, whereas the predicate symbols and variable symbols are local. Consequently, the meaning of truth for temporal atoms, constraint formulae, temporal rules and the corresponding templates depends on the state at which they are evaluated. Given a first order time structure M and a formula p, we denote the instant i (or equivalently, the state $s_{(i)}$) for which $\mathbf{I}_{s_{(i)}}(p) = true$ by $i \models p$ (or $(M, i) \models p$), i.e. at time instant i the formula p is true. Therefore, $i \models E(t_1, \ldots, t_n)$ means that at time i an event with the name $\mathbf{I}(t_1)$ and characterized by the global features $\mathbf{I}(t_2), \ldots, \mathbf{I}(t_n)$ occurs.

The connection between the restricted first-order temporal logic and the temporal data mining task this logic tries to formalize (temporal rules extraction) is made by the following assumptions:

A. For each formula p in L, there is an algorithm that calculates the value of the interpretation $\mathbf{I}_s(p)$, for each state s, in a finite number of steps.

B. There are states (called incomplete states) that do not contain enough information to calculate the interpretation for all formulae defined at these states.

C. It is possible to establish a measure, (called *general interpretation*) about the degree of truth of a compound formula along the entire sequence of states $(s_{(1)}, s_{(2)}, \ldots, s_{(n)}, \ldots)$.

The first assumption expresses the calculability of the interpretation $\mathbf{I}$. The second assumption expresses the situation when only the body of a temporal rule can be evaluated at a time moment i, but not the head of the rule. Therefore, for the state $s_{(i)}$, we cannot calculate the interpretation of the temporal rule, and the only solution is to estimate it using a general interpretation. This solution is expressed by the third assumption. (*Remark:* The second assumption violates the condition about the existence of an interpretation in each state $s_{(i)}$, as defined in Definition 4. But it is well known that in data mining sometimes data is incomplete or is missing. Therefore, we must modify this condition as "*$\mathbf{I}$ is a function that associates with* `almost` *each state s an interpretation $\mathbf{I}_s$ of all symbols from L*").

Example 2. (cont.) The database of events contains tuples with two values, (v_1, v_2). For a tuple with a recording index i, the first value expresses the name of the event – *Run*, *Stop* – which occurs at time moment i and the second is the value of the parameter h. Therefore, to specify a linear time structure $M = (S, x, \mathbf{I})$ we define the state s as a tuple $(decision_i, h_i)$ (see Table 1), the set S as the set of all tuples from database and the sequence x as the ordered sequence of tuples in the database. At this

Table 1: The first ten states of the linear time structure M

event	Run	Run	Stop	Run	Run	Run	Stop	Run	Run	Run
h	30	34	32	30	32	35	29	33	32	34

stage the interpretation of all symbols (global and local) can be defined. For the global symbols, the interpretation is quite intuitive: the meaning $\mathbf{I}(d_1)$ is *Run*, $\mathbf{I}(d_2)$ is *Stop* and $\mathbf{I}(h)$ is the function $f : R^2 \mapsto R, f(x_1, x_2) = (x_1 + x_2)/2$. For the local symbols (the predicate symbol E and the variables y_1, y_2), the interpretation depends on the state s_i: if $s_1 = (Run, 30)$, as example, then $\mathbf{I}_{s_1}(y_1) = Run$, $\mathbf{I}_{s_1}(y_2) = 30$, $\mathbf{I}_{s_1}(E(d_2, h(10, 30))) = false$ and $\mathbf{I}_{s_i}(E(d_1, h(20, 40))) = true$. Having defined the language L, the syntax and the semantics of L, as well as the linear time structure M, we can construct the temporal atoms evaluated as *true* at the time moment i (see Table 2).

Table 2: The temporal atoms (templates) evaluated as *true*

State	Temporal atoms (templates)
s_1	$E(y_1, y_2), E(d_1, h(c_1, c_2))$ where $(c_1 + c_2)/2 = 30$
s_2	$E(y_1, y_2), E(d_1, h(c_1, c_2))$ where $(c_1 + c_2)/2 = 34$
s_3	$E(y_1, y_2), E(d_2, h(c_1, c_2))$ where $(c_1 + c_2)/2 = 32$
$\ldots$	$\ldots$
s_i	$E(y_1, y_2), \ldots$

2.1 Consistency

To ensure that the concept of *general interpretation* (introduced by assumption C) is well defined, the linear time structure must present some property of consistency. In practical terms, this means that if we take any sufficiently large subset of time instants, the conclusions (rules) we may infer from this subset are sufficiently close to those inferred from the entire set of time instants. Therefore,

Definition 5. Given L and a linear time structure M, we say that M is a consistent time structure for L if, for every formula p, the limit $supp(p) = \lim_{n \to \infty} n^{-1} \#A$ exists, where $\#$ means "cardinality" and $A = \{i = 1..n \mid i \models p\}$. The notation $supp(p)$ denotes the support (of truth) of p.

Now we define the general interpretation for an n-ary predicate symbol P as:

Definition 6. Given L and a consistent linear time structure M for L, the general interpretation I_G for an n-ary predicate P is a function $D^n \to [0, 1]$, such that, for each n-tuple of terms $\{t_1, .., t_n\}$, $I_G(P(t_1, \ldots, t_n)) = supp(P(t_1, \ldots, t_n))$.

The general interpretation is naturally extended to constraint formulae, temporal rules and the corresponding templates. There is another useful measure, called *confidence*, but available only for temporal rules (templates). This measure is calculated (as we expect) as a limit ratio between the number of certain applications (time instants where both the body and the head of the rule are true) and the number of potential applications (time instants where only the body of the rule is true).

Definition 7. The confidence of a temporal rule (template) $H_1 \wedge \cdots \wedge H_m \mapsto H_{m+1}$ is the limit $\lim_{n \to \infty} (\#B)^{-1} \#A$, where $A = \{i = 1..n \mid i \models H_1 \wedge \cdots \wedge H_m \wedge H_{m+1}\}$ and $B = \{i = 1..n \mid i \models H_1 \wedge \cdots \wedge H_m\}$.

The relation between the property of consistency and the existence of the confidence for a temporal rule is expressed in the following lemma.

Lemma 1 *If M is a consistent linear time structure for L then every temporal rule (template) $H_1 \wedge \cdots \wedge H_m \mapsto H_{m+1}$ for which $supp(H_1 \wedge \cdots \wedge H_m) \neq 0$ has a well-defined confidence.*

For different reasons, (the user has no access to the entire sequence of states, or the states he has access to are incomplete), the general interpretation cannot be calculated. A solution is to estimate I_G using a finite linear time structure, i.e. a model.

Definition 8. Given L and a consistent time structure $M = (S, x, \mathbf{I})$, a model for M is a structure $\tilde{M} = (\tilde{T}, \tilde{x})$ where $\tilde{T}$ is a finite temporal domain $\{i_1, .., i_n\}$, $\tilde{x}$ is the subsequence of states $\{s_{(i_1)}, .., s_{(i_n)}\}$ (the restriction of x to the temporal domain $\tilde{T}$) and for each $i_j, j = 1, \ldots, n$, the state $s_{(i_j)}$ is a complete state.

For a given model $\tilde{M}$, the local support for a formula p (denoted $supp(p, \tilde{M})$) is defined as the ratio $\frac{\#A}{\#\tilde{T}}$, where $A = \{i \in \tilde{T} \mid i \models p\}$. Similarly, the estimation of the confidence for a temporal rule (template) is defined as:

Definition 9. Given a model $\tilde{M} = (\tilde{T}, \tilde{x})$ for M, the estimation of the confidence for the temporal rule (template) $H_1 \wedge \cdots \wedge H_m \mapsto H_{m+1}$ is the ratio $(\#B)^{-1} \#A$, where $A = \{i \in \tilde{T} \mid i \models H_1 \wedge \cdots \wedge H_m \wedge H_{m+1}\}$ and $B = \{i \in \tilde{T} \mid i \models H_1 \wedge \cdots \wedge H_m\}$.

Example 3. (cont.) Consider the following temporal rule (denoted H):

$$(y_1 = d_1) \wedge (y_2 > 28) \wedge \nabla_1(y_1 = d_1) \wedge \nabla_1(y_2 \geq 32) \mapsto \nabla_2(y_1 = d_2)$$

"translated" in a natural language as:

"IF at time t decision=Run AND $h > 28$ AND at time $t + 1$ decision=Run AND $h \geq 32$ THEN at time $t + 2$ decision=Stop"

If M is a consistent time structure, then the model $\tilde{M}$ formed by the first ten states s_i (Table 1) can be used to estimate the confidence of the temporal rule. And, due to the consistency property, this estimation (0.4) represents reliable[1] information about the success rate for this rule when applied to future data.

3. FIRST ORDER STOCHASTIC TEMPORAL LOGIC

The concept of consistency has deep consequences for any methodology of temporal rule extraction from high-dimensional data. Since it is (almost) impossible to use all data, the process of knowledge extraction is applied on subsets of data. For an end-user, the main question is if a temporal rule extracted from such a subset may be applied, with the same confidence, to any part of data (current or future). And the answer is positive if the linear time structure which models the process and data is consistent.

According to Definition 5, to verify the consistency of a time structure involves verifying the existence of the support for each well-defined formula. Because this approach is hard to apply in practice, we decided to concentrate on particular behaviors for the sequence of states x, behaviors which (as we will prove) are sufficient to ensure the existence of the consistency property. And the key to this approach is given by the observation that the sequence x may be considered as a particular realization of a stochastic process.

To extend our formalism with a probabilistic model we start by adding probabilities to a first order time structure $M = (S, x, \mathbf{I})$. If $S = \{s_0, s_1, \ldots\}$ is a countable set of states, consider $\sigma(S)$ the σ−algebra generated by S. The probability measure P on $\sigma(S)$ is defined such that $P(s_i) = p_i > 0, \forall i \in \mathbb{N}$. Consider now a random variable $\mathbb{X} : S \to \mathbb{R}$ such that the probability $P(\mathbb{X} = s_i) = p_i$ for all $i \in \mathbb{N}$ – this condition assures that the probability systems $(S, \sigma(S), P)$ and $(\mathbb{R}, \mathcal{B}, P_{\mathbb{X}})$ model the same experiment. If $S^{\mathbb{N}} = \{\omega \,|\, \omega = (\omega_1, \omega_2, \ldots, \omega_t, \ldots), \omega_t \in S, t \in \mathbb{N}\}$, then the variable $\mathbb{X}$ induces the stochastic sequence $\psi : S^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, where $\psi(\omega) = \{\mathbb{X}_t(\omega), t \in \mathbb{N}\}$ and $\mathbb{X}_t(\omega) = \mathbb{X}(\omega_t)$ for all $t \in \mathbb{N}$. The fact that each $\omega \in S^{\mathbb{N}}$ may be uniquely identified with a function $x : \mathbb{N} \to S$ and that $\mathbb{X}$ is a bijection between S and $\mathbb{X}(S)$ allows us to uniquely identify the function x with a single realization of the stochastic sequence. In other words, the sequence $x = (s_{(1)}, s_{(2)}, \ldots, s_{(i)}, \ldots)$ from the structure M can be seen as one of the outcomes of an infinite sequence of experiments, each experiment being modelled by the probabilistic system $(S, \sigma(S), P)$. To each such sequence corresponds a single realization of the stochastic sequence, $\psi(x) = (\mathbb{X}(s_{(1)}), \mathbb{X}(s_{(2)}), \ldots, \mathbb{X}(s_{(i)}), \ldots)$.

Definition 10. Given L and a domain D, a stochastic (first order) linear time structure is a quintuple $\mathbb{M} = (S, P, \mathbb{X}, \psi, \mathbf{I})$, where

- $S = \{s_1, s_2, \ldots\}$ is a (countable) set of states,

- P is a probability measure on the σ−algebra $\sigma(S)$ such that $P(s_i) = p_i > 0, i \in \mathbb{N}$

- $\mathbb{X}$ is a random variable such that $P(\mathbb{X} = s_i) = p_i$,

- ψ is a random sequence, $\psi(\omega) = \{\mathbb{X}(\omega_i)\}_1^{\infty}$ where $\omega \in S^{\mathbb{N}}$,

- $\mathbf{I}$ is a function that associates with each state s an interpretation $\mathbf{I}_s$ for all symbols from L.

To each realization of the stochastic sequence ψ, obtained by random drawing of a point in $\mathbb{R}^{\infty}$ (or equivalently, of a point ω in $S^{\mathbb{N}}$), corresponds a realization of the stochastic structure $\mathbb{M}$. This realization (in the following called "world") is given by the (ordinary) linear time structure $M_{\omega} = (S, \omega, \mathbf{I})$, which implies that the semantics attached to the symbols of L, described in Section 2, are totally effective.

3.1 Dependence and the Law of Large Numbers.

Much of the largest part of stochastic process theory has to do with the joint distribution of sets of coordinates, under the general heading of dependence. The degree to which random variations of sequence coordinates are related to those of their neighbors, in the time ordering, is sometimes called the memory of a sequence; in the context of time-ordered observations, one may think in terms of amount of information contained in the current state of the sequence about its previous states. A sequence with no memory is a rather special kind of object, because the ordering ceases to have significance. It is like the outcome of a collection of independent random experiments conducted in parallel, and indexed arbitrarily. Indeed, independence and stationarity are the best-known restrictions on the behavior of a sequence. But while the emphasis in our framework will mainly be on finding ways to relax these conditions, they remain important because of the many classic theorems in probability and limit theory which are founded on them.

The amount of dependence in a sequence is the chief factor determining how informative a realization of given length can be about the distribution that generated it. At one extreme, the i.i.d. sequence is equivalent to a true random sample. The classical theorems of statistics can be applied to this type of distribution. At the other extreme, it is easy to specify sequences for which a single realization can never reveal the parameters of the distribution, even in the limit as its length tends to infinity. This last possibility is what concerns us most, since we want to know whether averaging operations applied to sequences have useful limiting properties.

Let $\{\mathbb{X}_t\}_1^{\infty}$ be a stochastic sequence and define $\overline{\mathbb{X}}_n = n^{-1} \sum_1^n \mathbb{X}_t$. Suppose that $E(\mathbb{X}_t) = \mu_t$ and $n^{-1} \sum_1^n \mu_t$ converges to μ, with $|\mu| < \infty$. In this simple setting, the sequence is said to obey the weak law of large numbers (WLLN) when $\overline{\mathbb{X}}_n$ converges in probability to μ, and the strong law of large numbers (SLLN) when $\overline{\mathbb{X}}_n$ converges almost sure to μ. To obey the law of large numbers, a sequence must satisfy regularity conditions relating to two distinct factors: the probability of extreme values (limited by bounding absolute moments) and the degree of dependence between coordinates. The necessity of a set of regularity conditions is usually hard to prove (except if the sequences are independent), but various configurations of dependency and boundedness conditions can be shown to be sufficient. These results usually exhibit a trade-off between the two dimensions of regularity; the stronger the moment restrictions are, the weaker the dependence restrictions can be, and vice-versa.

Consider now the sequence of the indicator function for an event[2] A (i.e. $\mathbb{X}_t = \mathbf{1}_A$ for all t). In this case, $\mu_t = \mu = P(A)$ and $\overline{\mathbb{X}}_n(\omega) = n^{-1} \sum_{t=1}^n \mathbf{1}_A(\omega) = n^{-1} \#\{i \in 1..n \,|\, \mathbb{X}_i(\omega) = 1\}$. If M_{ω} is a realization of a stochastic structure, p a formula defined on language L and A the event "the interpretation of the formula p is

---

[1]In practice, the size of the model $\tilde{M}$ must be at least of order of hundreds states to be able to consider the confidence estimate as *truly* reliable information

[2]In this context, an event is a set of possible outcomes of a random experiment

true", then the expression for $\overline{\mathbb{X}}_n(\omega)$ is equivalent (under some conditions) with the expression which gives, at the limit, the support of p. Indeed,

$$\overline{\mathbb{X}}_n(\omega) = \frac{\sum_1^n \mathbf{1}_{A_p}(\omega_i)}{n} = \frac{\#\{i \le n \mid \mathbf{1}_{A_p}(\omega_i) = 1\}}{n} =$$

$$= \frac{\#\{i \le n \mid \mathbf{I}_{s_{(i)}}(p) = true\}}{n} = \frac{\#\{i \le n \mid (M_\omega, i) \models p\}}{n}$$

Consequently, *supp(p) exists (almost sure) if the stochastic sequence $\{\mathbf{1}_A\}_1^\infty$ satisfies the strong law of large numbers*.

Given a stochastic linear time structure $\mathbb{M} = (S, P, \mathbb{X}, \psi, \mathbf{I})$, the sequence $\{\mathbf{1}_A\}_1^\infty$ is obtained – as we will prove in the following – by applying a particular transformation to the random sequence ψ. Therefore, the sufficient conditions for $\{\mathbf{1}_A\}_1^\infty$ to obey SLLN are inherited from the regularity conditions the "basic" stochastic process ψ must satisfy. And because all absolute moments of $\mathbf{1}_A$ are finite (bounded by 0 and 1), the only regularity condition we may modify concerns the degree of dependence for ψ.

Example 4. If the first ten realizations of the stochastic process ψ are those for Table 1 and p is the temporal rule H, then the corresponding sequence $\{\mathbf{1}_A\}_1^\infty$ has at the first eight positions the values $1, 0, 0, 0, 1, 0, 0, 0$.

3.2 The Independence Case.

In this case the stochastic process $\psi = \{\mathbb{X}(\omega_i)\}_1^\infty$ is i.i.d. Firstly, let p be a temporal free formula (e.g. a temporal atom). On the probabilistic system $(S, \sigma(S), P)$ one defines the event $A_p = \{s \in S \mid \mathbf{I}_s(p) = true\}$.

Lemma 2 *If $\{\mathbb{X}(\omega_i)\}_1^\infty$ is i.i.d. then $\{(\mathbf{I}_{A_p})_i\}_1^\infty = \{\mathbf{I}_{A_p}(\omega_i)\}_1^\infty$ is also i.i.d.*

The proof is elementary and is based on the fact that if the random variables $\mathbb{X}_i(\omega) = \mathbb{X}(\omega_i)$ and $\mathbb{X}_j(\omega) = \mathbb{X}(\omega_j)$ are independent then the random variables $\mathbf{1}_{A_p}(\omega_i)$ and $\mathbf{1}_{A_p}(\omega_j)$ are also independent (see [18]).

As we mentioned, the regularity conditions for SLLN concern the dependence restrictions and the moment restrictions. For the independence case, the Kolmogorov classical version of the SLLN may be applied to the sequence $\{\mathbf{1}_{A_p}(\omega_i)\}_1^\infty$. Therefore, we may conclude that

Corollary 1. If the random process ψ from the stochastic first-order linear time structure $\mathbb{M}$ is i.i.d., then for almost all worlds M_ω the support of p, $supp(p)$, where p is a temporal free formula in L, exists and it is equal to $P(A_p)$.

Consider now the temporal formula $\nabla_k p$, $k > 0$. For a fixed world M_ω, we have $(M_\omega, i) \models \nabla_k p$ iff $(M_\omega, i + k) \models p$. Therefore, the stochastic sequence corresponding to $\nabla_k p$ is given by $\{(\mathbf{1}_{A_{\nabla_k p}})_i\}_1^\infty = \{\mathbf{1}_{A_p}(\omega_{i+k})\}_1^\infty = \{(\mathbf{1}_{A_p})_i\}_{k+1}^\infty$, the last sequence being the one corresponding to the formula p, but without the first k coordinates. Because the approach for $k < 0$ is similar, we may conclude that:

Corollary 2. If the random process ψ from the stochastic first-order linear time structure $\mathbb{M}$ is i.i.d., then for almost all worlds M_ω the support of $\nabla_k p$, where p is a temporal free formula in L and $k \in \mathbb{N}$, exists and it is equal to $P(A_p)$.

The last type of formula we consider is $\nabla_{k_0} p_0 \wedge \nabla_{k_1} p_1 \wedge \ldots \wedge \nabla_{k_n} p_n$, where $0 = k_0 \le k_1 \le \cdots \le k_n$ (e.g. a temporal rule). If Tp is an abbreviation for this formula and M_ω is a fixed world,

we have $(M_\omega, i) \models$ Tp if and only if $(M_\omega, i + k_j) \models p_j$ for all $j = 0..n$. To construct the stochastic sequence corresponding to Tp we first introduce the following Borel transformation $g_p^k(x)$:

$$g_p^k(\mathbb{X}_i(\omega)) = (g_p^k \circ \mathbb{X}_i)(\omega) = \begin{cases} 1 & \text{if } \omega_{i+k} \in A_p, \\ 0 & \text{if not} \end{cases} \quad (1)$$

Therefore, the stochastic sequence for the formula p was obtained by applying to $\{\mathbb{X}_i\}$ the transformation g_p^0, whereas for the formula $\nabla_k p$ one applied the transformation g_p^k. Given the formula Tp, consider the stochastic sequences

$$\{_0\mathbb{G}_i\}_1^\infty = \{g_{p_0}^{k_0}(\mathbb{X}_i)\}_1^\infty, \ldots, \{_n\mathbb{G}_i\}_1^\infty = \{g_{p_n}^{k_n}(\mathbb{X}_i)\}_1^\infty,$$

corresponding to the formulae $\nabla_{k_0} p_0, \nabla_{k_1} p_1, \ldots, \nabla_{k_n} p_n$. From these sequences we define the stochastic sequence $\{\mathbb{G}_i\}_1^\infty$, $\mathbb{G}_i(\omega) = \prod_{j=0}^n {}_j\mathbb{G}_i(\omega)$. According to the following lemma, $\{\mathbb{G}_i\}$ is the sequence corresponding to the formula Tp.

Lemma 3 $\mathbb{G}_i(\omega) = 1$ *if and only if* $(M_\omega, i) \models Tp$.

Because $g_{p_j}^{k_j}(\mathbb{X}_i) = g_{p_j}^0(\mathbb{X}_{i+k_j})$, the random variable $\mathbb{G}_i$ can be expressed as $h(\mathbb{X}_i, \ldots, \mathbb{X}_{i+k_n})$, where h is a Borel function (a composition between the product function and the $g_{p_j}^{k_j}$ functions). The sequence $\{\mathbb{G}_i\}$ is identically distributed (condition inherited from the sequence $\{\mathbb{X}_i\}$ by applying the function h), but it is not independent (the events "Tp true at i" and "Tp true at $i+1$" are not independent). In exchange we may prove the following result:

Lemma 4 *For all $i \in \mathbb{N}$ and all $m \in \mathbb{N}, m \ge k_n + 1$, the random variables $\mathbb{G}_i$ and $\mathbb{G}_{i+m}$ are independent.*

This Lemma affirms that the sequence $\{\mathbb{G}_i\}$ is what is called in stochastic process theory a k_n-dependent sequence, which is a particular case of a mixing sequence. In a brief description, we say that a sequence is α-mixing (or strong mixing) if the supremum of the strong mixing coefficient[3] α_m, which is a measure of the dependence between coordinates situated at a distance m, converges to zero for $m \to \infty$. A consequence of Lemma 4 is that α_m is zero for $m \ge k_n + 1$, and evidently $\{\mathbb{G}_i\}$ is a strong mixing sequence. The importance of this result lies in the fact that this kind of dependence is sufficient, under certain conditions, that $\{\mathbb{G}_i\}$ obeys SLLN.

Theorem 1 *([11], pp. 40) Let $\{\mathbb{X}_t\}_1^\infty$ be a α-mixing sequence such that $E(\mathbb{X}_t) = \mu$ and $E(\mathbb{X}_t^2) < \infty, t \ge 1$. Suppose that*

$$\sum_{t=1}^\infty b_t^{-2} Var(\mathbb{X}_t) < \infty \text{ and } \sup_n b_n^{-1} \sum_{t=1}^n E(|\mathbb{X}_t|) < \infty \quad (2)$$

where $\{b_t\}$ is a sequence of positive constants increasing to ∞. Then

$$b_n^{-1} \sum_{t=1}^n \mathbb{X}_t \xrightarrow{a.s.} \mu$$

For the particular case $b_n = n$, the conclusion of the theorem becomes $\overline{\mathbb{X}}_n \xrightarrow{a.s.} \mu$. It is not difficult to prove that the sequence $\{\mathbb{G}_i\}$ verifies the hypothesis of Theorem 1. In conclusion

---

[3] $\alpha_m = \sup_t \alpha(\mathcal{X}_{-\infty}^t, \mathcal{X}_{t+m}^\infty)$, where $\alpha(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}} |P(G \cap H) - P(G)P(H)|$ and $\mathcal{X}_s^t = \sigma(\mathbb{X}_s, \ldots, \mathbb{X}_t)$

Corollary 3. If the random process ψ from the stochastic first-order linear time structure $\mathbb{M}$ is i.i.d., then for almost all worlds M_ω the support of Tp, where Tp is a temporal formula $\nabla_{k_0} p_0 \wedge \nabla_{k_1} p_1 \wedge \ldots \wedge \nabla_{k_n} p_n$, exists and it is equal to $\prod_{j=0}^{n} P(A_{p_j})$.

Finally, based on Corollaries 1-3, we can prove the following fundamental theorem:

Theorem 2 *(Independence and Consistency)* *If the random process ψ from the stochastic first-order linear time structure $\mathbb{M} = (S, P, \mathbb{X}, \psi, \boldsymbol{I})$ is i.i.d., then almost all worlds $M_\omega = (S, \omega, \boldsymbol{I}_s)$ are consistent linear time structures.*

But the independence restriction for the random process ψ, even if it is a sufficient condition for the property of consistency for linear time structures M_ω, represents a serious drawback for a temporal data mining methodology. Indeed, what we try to discover are temporal rules expressing a dependence between the event occurred at time t and the events occurred before time t. It's obvious that the independence implies a null correlation between the body and the head of the temporal rule – in other words the rule is not meaningful. The question is how much do we have to relax the independence condition for ψ to still conserve the property of consistency.

3.3 The Mixing Case.

Since mixing is not so much a property of the sequence $\{\mathbb{X}_i\}$ as of the sequence of σ-fields generated by $\{\mathbb{X}_i\}$, it holds for any random variables measurable on those σ-fields. More generally, we have the following important implication:

Theorem 3 *([7]) Let $\mathbb{Y}_i = g(\mathbb{X}_i, \mathbb{X}_{i-1}, .., \mathbb{X}_{i-k})$ be a Borel function, for finite k. If $\mathbb{X}_i$ is α-mixing, then $\mathbb{Y}_i$ is too.*

This theorem is the key to proving that ψ α-mixing is a sufficient condition for consistency. Indeed, the previously defined functions $g_{p_j}^{k_j}$ and $h = \prod_{j=0}^{n} g_{p_j}^{k_j}$ are Borel transformations. Consequently, the sequence $\{g_p^0(\mathbb{X}_t)\}$ (corresponding to temporal free formula p), the sequence $\{g_p^k(\mathbb{X}_t)\}$ (corresponding to temporal formula $\nabla_k p$) and the sequence $\{h(\mathbb{X}_t)\}$ (corresponding to temporal formula Tp) are also α-mixing. Because all these sequences fulfil the conditions of Theorem 1 we can conclude that for any formula p in L the support of p exists (but, unlike in the independent case, we can not give an exact expression for the support of a temporal formula like Tp). This result is formalized in the following theorem.

Theorem 4 *(Mixing and Consistency)* *If the random process ψ from the stochastic first-order linear time structure $\mathbb{M} = (S, P, \mathbb{X}, \psi, \boldsymbol{I})$ is α-mixing, then almost all worlds $M_\omega = (S, \omega, \boldsymbol{I}_s)$ are consistent linear time structures.*

Remark If ψ is i.i.d., a consequence of Corollary 3 is that the confidence of the rule Tp is $P(A_{p_n})$. If ψ is α-mixing, we can obtain only an upper bound for the confidence of the temporal rule. By denoting $\mathbb{A}$ the event "the implicated clause of the rule is satisfied" and $\mathbb{B}$ the event "implication clauses of the rule are satisfied" then the following Lemma holds.

Lemma 5 *If ψ is α-mixing, the confidence of the temporal rule (template) Tp satisfies the relation*

$$conf(Tp) \leq \frac{\alpha_1}{P(\mathbb{B})} + P(\mathbb{A}).$$

3.4 The Near Epoch Dependence Case.

The mixing concept has a serious drawback from the viewpoint of applications in stochastic limit theory, in that a function of a mixing sequence (even of an independent sequence) that depends on an infinite number of coordinates of the sequence is not generally mixing. Let $\mathbb{X}_i = g(\ldots, \mathbb{V}_{i-1}, \mathbb{V}_i, \mathbb{V}_{i+1}, \ldots)$, where $\mathbb{V}_i$ is a vector of mixing processes. The idea is that although $\mathbb{X}_i$ may not be mixing, if it depends almost entirely on the "near epoch" of $\{\mathbb{V}_i\}$ it will often have properties permitting the application of limit theorems, including SLLN. Near-epoch dependence[4] is not an alternative to a mixing assumption; it is a property of the mapping from $\{\mathbb{V}_i\}$ to $\{\mathbb{X}_i\}$, not of the random variables themselves.

The main approach we applied in the previous cases is the property of a Borel transformation g to inherit the type of dependence (the independence or the mixing dependence) from the initial sequence. For the near-epoch dependence this property is achieved only if the function g satisfies additional conditions and only for particular L_q-NED sequences. Concretely, let $g(x) : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$ be a Borel function and consider the metric on $\mathbb{R}^n$, $\rho(\boldsymbol{x}^1, \boldsymbol{x}^2) = \sum_1^n |x_i^1 - x_i^2|$ for measuring the distance between points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$. If g satisfies

   i. g continuous and

   ii. $|g(\boldsymbol{X}^1) - g(\boldsymbol{X}^2)| \leq M\rho(\boldsymbol{X}^1, \boldsymbol{X}^2)$ a.s., where $\boldsymbol{X}^i \in \mathbb{R}^n$,

then the following theorem holds:

Theorem 5 *([7], pp. 269) Let $\mathbb{X}_{ji}$ be L_2-NED of size $-a$ on$\{\mathbb{V}_i\}$ for $j = 1..n$, with constants d_{ji}. If g satisfies the conditions (i)-(ii), then $\{g(\mathbb{X}_{1i}, \ldots, \mathbb{X}_{ni})\}$ is also L_2-NED on $\{\mathbb{V}_i\}$ of size $-a$, with constants a finite multiple of $\max_i\{d_{ji}\}$.*

Suppose the process $\psi = \{\mathbb{X}_i\}$ is L_2-NED of size $-a$ on $\{\mathbb{V}_i\}$. As we have already seen in the previous cases, for p a temporal free formula, the corresponding sequence is $\{g_p^0(\mathbb{X}_i)\}$. The function $g_p^0(\cdot)$, as defined in (1), doesn't satisfies the condition (i). But it is possible to define a function $\tilde{g}_p$ which takes the value one for the arguments $x \in \mathbb{X}(A_p) = \{\mathbb{X}(s) : s \in A_p\}$, the value zero for the arguments $x \in \{\mathbb{X}(s) : s \in S - A_p\}$ and to be continuous for $x \in \mathbb{R}$. Because $g_p^0(\mathbb{X}_i(\omega)) = \tilde{g}_p(\mathbb{X}_i(\omega)) \in \{0, 1\}$, it is possible (the support of $\mathbb{X}$ being a discrete set) to choose the constant M_p such that $|\tilde{g}_p(x) - \tilde{g}_p(y)| \leq M_p|x - y|$, for any $x, y \in \mathbb{X}(S)$. Therefore, the conditions of Theorem 5 are verified and so $\{\tilde{g}_p(\mathbb{X}_i)\} = \{\mathbf{1}_{A_p}\}$ is also L_2-NED of size $-a$ on $\{\mathbb{V}_i\}$.

For the temporal formula $\nabla_k p$, the corresponding sequence is $\{g_p^k(\mathbb{X}_i)\} = \{g_p^0(\mathbb{X}_{i+k})\}$. Because $\mathbb{X}_{i+k}$ is L_2-NED, the same argument as in the previous paragraph proves that $\{\tilde{g}_p(\mathbb{X}_{i+k})\}$ is L_2-NED. Finally, consider the temporal formula $\nabla_{k_0} p_0 \wedge \ldots \wedge \nabla_{k_n} p_n$. The corresponding sequence is

$$\left\{\prod_{j=0}^{n} g_{p_j}^{k_j}(\mathbb{X}_i)\right\} = \left\{\prod_{j=0}^{n} g_{p_j}^0(\mathbb{X}_{i+k_j})\right\} = \left\{\prod_{j=0}^{n} \tilde{g}_{p_j}(\mathbb{X}_{i+k_j})\right\}$$

$$= \{\tilde{h}(\mathbb{X}_i, \ldots, \mathbb{X}_{i+k_n})\} = \{\tilde{h}(\mathbb{X}'_{i0}, \ldots, \mathbb{X}'_{in})\},$$

where $\mathbb{X}'_{ij} = \mathbb{X}_{i+k_j}$. Concerning the transformation $\tilde{h}$, it satisfies (i) as being a product of continuous functions and satisfies (ii) because, by denoting $\boldsymbol{X}_i = (\mathbb{X}_i, \ldots, \mathbb{X}_{i+k_n})$,

---

[4] Let $\mathcal{X}_{t-m}^{t+m} = \sigma(V_{t-m}, .., V_{t+m})$, for a stochastic sequence $\{V_t\}_{-\infty}^{\infty}$. If, for $q > 0$, a sequence of integrable r.v.s $\{X_t\}_{-\infty}^{\infty}$ satisfies $\| X_t - E(X_t | \mathcal{X}_{t-m}^{t+m}) \|_q \leq d_t \nu_m$, where $\nu_m \rightarrow 0$, and $\{d_t\}_{-\infty}^{\infty}$ is a sequence of positive constants, X_t is said to be `near-epoch dependent in` L_q `norm` $(L_q$-NED$)$ on $\{V_t\}$.

$$\left| \prod_{i=0}^{n} \tilde{g}_{p_j}(\mathbb{X}_{i+k_j}^1) - \prod_{i=0}^{n} \tilde{g}_{p_j}(\mathbb{X}_{i+k_j}^2) \right| \leq$$

$$\leq \sum_{j=0}^{n} \left| \tilde{g}_{p_j}(\mathbb{X}_{i+k_j}^1) - \tilde{g}_{p_j}(\mathbb{X}_{i+k_j}^2) \right| \leq$$

$$\leq \sum_{j=0}^{n} M\rho(\mathbb{X}_{i+k_j}^1, \mathbb{X}_{i+k_j}^2) \leq M\rho(\boldsymbol{X}_i^1, \boldsymbol{X}_i^2).$$

The first inequality comes from the fact that $|\prod_i x_i - \prod_i y_i| \leq \sum_i |x_i - y_i|$ if $x_i, y_i \in \{0, 1\}$ and the second inequality is the condition (ii) for the transformations $\tilde{g}_{p_j}$. Therefore, Theorem 5 holds and so the sequence corresponding to the temporal formula Tp is L_2-NED. In conclusion,

Corollary 4. If ψ is L_2-NED then for any formula in L the corresponding sequence is also L_2-NED.

The following step is to establish the sufficient condition for the application of SLLN to an L_q-NED sequence. In [8] are summarized the up-to-date strong laws for dependent heterogeneous processes, including NED sequences. We consider the following form for the limit theorem, which includes the case $q = 2$.

Theorem 6 *([8]) Let a sequence $\{\mathbb{X}_i\}$ with means $\{\mu_i\}$ be L_q-NED, $q \in [1, 2]$, of size $-b$, on a sequence $\{\mathbb{V}_i\}$ which is α-mixing of size $-a$. If $a_n/\sqrt{n} \uparrow \infty$ as $n \to \infty$, and*

$$a_n^{-1} \parallel \mathbb{X}_n - \mu_i \parallel_q^{2-q/2} = O(n^\epsilon), \tag{3}$$

where

$$\epsilon < 1/2 - 1/q + \min\{-1/2, \min\{bq/2, a/2\} - 1\} \tag{4}$$

then $a_n^{-1} \sum_{i=1}^{n} (\mathbb{X}_i - \mu_i) \to 0$, a.s.

It is not difficult to verify that an L_2-NED sequence bounded by 0 and 1 fulfils the hypotheses of this theorem and so obeys the SLLN. Therefore, as in the previous cases, we may conclude that

Theorem 7 *(Near-Epoch Dependence and Consistency) If the random process ψ from the stochastic first-order linear time structure $\mathbb{M} = (S, P, \mathbb{X}, \psi, \boldsymbol{I})$ is L_2-NED on an α-mixing sequence, then almost all worlds $M_\omega = (S, \omega, \boldsymbol{I}_s)$ are consistent linear time structures.*

The near-epoch dependence is, according to the stochastic limit theory, the highest degree of dependence for which theorems concerning SLLN still hold.

3.5 From Theory to Practice

From a theoretical viewpoint, proving that a certain amount of dependence for the stochastic process ψ is a sufficient condition for the consistency of the linear time structure M_ω it is a useful achievement. But from a practical viewpoint, checking that a sample series (two series in our example, the series of h values and the series of *decisions*) holds a given amount of dependence is a difficult task [17]. Serial dependency is often characterized by the standardized spectral density function

$$\phi(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho(j)e^{-ij\omega}, \quad \omega \in [-\pi, \pi]$$

where $i = \sqrt{-1}$ and ρ is the autocorrelation function. If ϕ is constant then the series is independent, if ϕ is bounded above and

below then the series has a short memory (or is n-dependent) and if $\phi(0)$ is infinity then the series has a long memory (is α-mixing or L_q-NED). Nonparametric tests for serial independence were constructed based on spectral density or on higher-order spectra, but the need of Gaussian assumption and of restrictive moment conditions when testing for a specific type of dependence makes these tests unappropriate in our case [10]. A possible solution is given by the generalized spectral density [12, 13, 14], which needs no moment condition:

$$f(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u,v)e^{-ij\omega}, \quad \omega \in [-\pi, \pi],$$

where $\sigma_j(u,v) \equiv cov(e^{iuX_t}, e^{ivX_{t-|j|}})$. It applies to time series generated from either discrete or continuous distribution with possibly infinite moments, as is often encountered in high-frequency economic and financial data. This is appropriate for temporal data mining, because data for which temporal rules are extracted contain at least a series of discrete events (as in our example, the series of possible *decisions*, $\{Run, Stop\}$). The estimates of the generalized spectral function and of its derivatives can be used to test generic serial dependence and hypotheses of various specific aspects of serial dependence (serial uncorrelatedness, martingale, conditional homoscedasticity, conditional symmetry, etc.). The computational complexity of these estimates is high, involving the selection of a data-dependent asymptotically optimal bandwidth (or lag order) for the kernel and a four-dimensional integration, but remains manageable for the present generation of computers.

4. CONCLUSIONS

To give an answer to the practical question "How we can be sure that a temporal rule learned from a data subset can be applied with the same confidence on future data?" we developed a probabilistic temporal logic framework by combining stochastic theory with first-order temporal logic. The connection between a practical temporal data mining methodology and the abstract framework is represented by the sequence of states ω from the linear time structure $M_\omega = (S, \omega, \boldsymbol{I})$. This sequence is constructed based on raw data, and by modeling it as a realization of a stochastic process ψ we achieve two goals: *i*) we express the intrinsic dependency of temporal raw data and *ii*) we get a certain "independence" for the analysis of the confidence/support preservation regardless of the original raw data.

According to the proposed formalism, a temporal rule preserves its confidence over future data sets if the model satisfies the property of consistency. The model is consistent if each formula p has a support or, as we proved, if a particular stochastic sequence (depending on p) obeys the strong law of large numbers. And because the sequence corresponding to formula p is constructed from the stochastic sequence ψ, using appropriate transformations, we studied the necessary conditions for ψ which assures the applicability of SLLN.

The independence of ψ is a sufficient condition for the consistency, but it is not useful for temporal rule extraction. For other two type of dependence, the α-mixing (the degree of dependence converges to zero if the distance between coordinates converges to ∞) and the near-epoch dependence (a function of a mixing sequence with an infinite number of parameters) we could prove that the linear time structure $M_\omega = (S, \omega, \boldsymbol{I})$ is consistent. Only a degree of dependence greater than L_2-NED makes any temporal rule extracted (anyhow) from a local data set inappropriate for forecasting use. In our opinion, these results also imply that future research must concentrate on the connection between the degree of depen-

dence of raw data and the quality (and efficiency) of temporal rules extraction algorithms.

5. REFERENCES

[1] S. Al-Naemi. A theoretical framework for temporal knowledge discovery. In *Proc. of Int. Workshop on Spatio-Temporal Databases*, pages 23–33, Spain, 1994.

[2] X. Chen and I. Petrounias. A Framework for Temporal Data Mining. *Lecture Notes in Computer Science*, 1460:796–805, 1998.

[3] J. Chomicki and D. Toman. Temporal Logic in Information Systems. *BRICS Lecture Series*, LS-97-1:1–42, 1997.

[4] P. Cotofrei and K. Stoffel. Classification Rules + Time = Temporal Rules. In *Lecture Notes in Computer Science*, volume 2329, pages 572–581. Springer Verlang, 2002.

[5] P. Cotofrei and K. Stoffel. From temporal rules to temporal meta-rules. In *Proc. of 6th Int. Conf. DaWaK 2004*, Lecture Notes in Computer Science, vol. 3181, pages 169–178, Zaragoza, Spain, 2004.

[6] P. Cotofrei. *Methodology for Mining Meta Rules from Sequential Data*. PhD. Thesis. University of Neuchtel, 2005.

[7] J. Davidson. *Stochastic Limit Theory*. Oxford University Press, 1994.

[8] J. Davidson and R. de Jong. Strong laws of large numbers for dependent and heterogeneous processes: a synthesis of new and recent results. *Econometric Reviews*, 16(3):251–79, 1997.

[9] E. A. Emerson. Temporal and Modal Logic. *Handbook of Theoretical Computer Science*, pages 995–1072, 1990.

[10] C. Granger and T. Terasvirta. *Modelling Nonlinear Economic Relationships*. Oxford University Press, New York, 1993.

[11] P. Hall and C. Heyde. *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, 1980.

[12] Y. Hong. Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach. *JASA*, 94(448):1201–1220, 1999.

[13] Y. Hong and T. H. Lee. Diagnostic checking for adequacy of nonlinear time series models. *Econometric Theory*, 19:10651121, 2003.

[14] Y. Hong and T. H. Lee. Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Review of Economic Studies*, 72:499541, 2005.

[15] D. Koller and J. Y. Halpern. Irrelevance and conditioning in first-order probabilistic logic. In *AAAI/IAAI, Vol. 1*, pages 569–576, 1996.

[16] D. Malerba, F. Esposito, and F. Lisi. A logical framework for frequent pattern discovery in spatial data. In *FLAIRS Conference*, 2001.

[17] P. Nze and P. Doukhan. Weak dependence: models and applications to econometrics. *Econometric Theory*, 20(6):995–1045, 2004.

[18] P. Pfeiffer. *Probability for Applications*. Springer Texts in Statistics. Springer-Verlag, 1989.