Model correction for bivariate Bernoulli Markov procecesses based on copula theory

Jesús E. García   M. Fernández

September 18, 2013

Introduction

- In this paper we address the problem of simultaneously modeling two binary time series.
- We describe jointly both series using a new family of Markov models called Partition Markov Models.
- This family has the characteristic of finding a Markov model with a minimal number of parameters.
- However, sometimes the sample size does not allow to extract enough information about the model.
- We develop a copula based methodology to correct the joint probabilities by using the marginal models obtained by the Partition Markov Models.

Dataset

- Our methodology is applied to series given by the direction of the daily variation of two asset and denoted by $X(1)_t$ and $X(2)_t$.
- the first series, $\{b_t\}_{t=0}^{T}$, corresponds to the main Brazilian stock exchange index (Ibovespa),
- the second series, $\{d_t\}_{t=0}^{T}$, corresponds to the exchange rate of United States Dollar to Brazilian Real
- the two series where observed from 2003 to 2012.
- $X(1)_t = 0$ if $d_t \leq d_{t-1}$ and $X(1)_t = 1$ otherwise,
- $X(2)_t = 0$ if $b_t \leq b_{t-1}$ and $X(2)_t = 1$ otherwise.
- We will analyze the bivariate series $X_t = (X(1)_t, X(2)_t)$.
- the joint series X_t has alphabet $A = \{0, 1\}^2$.

Problem

- For an alphabet of size 4, the number of parameters for an order k Markov chain model will be $4^k \times 3$.
- This makes the problem unfeasible even for not too large k.
- We will use a very economic family of Markov models called Partition Markov Models (PMM).
- Even using PMM, our dataset with $T = 2356$ is not large enough to find a good model of X_t.
- T is large enough to model the margins $X(1)_t$ and $X(2)_t$ (with PMM).
- This motivates the new methodology proposed here, which improves the estimated probabilities for X_t by applying a correction through the marginals models.

Partition Markov models
Notation

Let (X_t) be a discrete time order M Markov chain

- A the finite alphabet;
- $\mathcal{S} = A^M$ the state space;
- $P(a|s) = \text{Prob}(X_t = a | X_{t-M}^{t-1} = s),\ a \in A,\ s \in \mathcal{S}$ the transition probabilities.

Equivalence relationship on $\mathcal{S}$

Definition

for $s, r \in \mathcal{S}$; $s \sim_p r \iff P(a|s) = P(a|r)\ \forall a \in A.$

- For any $s \in \mathcal{S}$, the equivalence class of s is given by $[s] = \{r \in \mathcal{S} | r \sim_p s\}$.
- The classes defined by $\sim_p$ are the subsets of $\mathcal{S}$ with the same transition probabilities.

Equivalence relationship on $\mathcal{S}$

- The equivalence relation defines a partition $\mathcal{L}$ of $\mathcal{S}$.
- We have $(|A| - 1)$ transition probabilities for each "part" (element of $\mathcal{L}$), obtaining a model with $(|A| - 1)|\mathcal{L}|$ parameters.
- The elements of $\mathcal{S}$ on the same equivalence class activate the same random mechanism to choose the next element in the Markov chain.

Markov chain with partition $\mathcal{L}$

Definition

let (X_t) be a discrete time, order M Markov chain on A and let $\mathcal{L} = \{L_1, L_2, \ldots, L_K\}$ be a partition of $\mathcal{S}$. We will say that (X_t) is a Markov chain with partition $\mathcal{L}$ if this partition is the one defined by $\sim_p$.

Example

- $A = \{0, 1\}$, $M = 2$, $\quad \mathcal{S}(= A^M) = \{00, 01, 10, 11\}$,
  $P(0|00) = P(0|01) = 0.4 \;\; P(0|10) = P(0|11) = 0.2$;
- $P(1|s) = 1 - P(0|s) \;\; \forall s \in \mathcal{S}$;
- the partition for this Markov chain is $\mathcal{L} = \{\{00, 01\}, \{10, 11\}\}$ with parts $L_1 = \{00, 01\}$ and $L_2 = \{10, 11\}$;
- the parameters of the Markov chain with partition $\mathcal{L}$ are $P(0|L_1) = 0.4$ and $P(0|L_2) = 0.2$.

Model selection problem

Given a sample generated by a finite memory stationary process, how to choose a partition defining a good Markov model for the source?

Notation

Let x_1^n be a sample of the process $\left(X_t\right)$, $s \in \mathcal{S}$, $a \in A$ and $n > M$.

$$N_n(s, a) = \left|\left\{t : M < t \leq n, x_{t-M}^{t-1} = s, x_t = a\right\}\right|, \tag{1}$$

$$N_n(s) = \left|\left\{t : M < t \leq n, x_{t-M}^{t-1} = s\right\}\right|. \tag{2}$$

To simplify the notation we will omit the $_n$ on N_n.

A distance in S

Definition

We define the distance d in S,

$$\begin{aligned} d(s,r) &= \frac{2}{(|A|-1)\ln(n)} \sum_{a \in A} \left\{ N(s,a) \ln\left(\frac{N(s,a)}{N(s)}\right) \right. \\ &+ N(r,a) \ln\left(\frac{N(r,a)}{N(r)}\right) \\ &- \left. (N(s,a)+N(r,a)) \ln\left(\frac{N(s,a)+N(r,a)}{N(s)+N(r)}\right)\right) \end{aligned}$$

for any $s, r \in S$.

A distance in S

Proposition

For any $s, r, t \in S$,

i. $d(r,s) \geq 0$ *with equality if and only if* $\frac{N(s,a)}{N(s)} = \frac{N(r,a)}{N(r)} \;\; \forall a \in A$,

ii. $d(r,s) = d(s,r)$,

iii. $d(r,t) \leq d(r,s) + d(s,t)$.

Remark

d can be generalized to subsets (see García, J. and González-López, V. A. (2010)).

Consistence in the case of a Markov source

Theorem

Let (X_t) be a discrete time, order M Markov chain on a finite alphabet A. Let x_1^n be a sample of the process and $0 < \alpha < \infty$, then for n large enough and for each $s, r \in S$, $d_n(r, s) < \alpha$ iff s and r belong to the same class.

Algorithm

Input: $d(s, r) \forall s \neq r \in S$; **Output**: $\hat{\mathcal{L}}_n$.
$B = S$; $\hat{\mathcal{L}}_n = \emptyset$
while $B \neq \emptyset$

    select $s \in B$

    define $L_s = \{s\}$

    $B = B \setminus \{s\}$

    for each $r \in B, r \neq s$

        if $d(s, r) < \alpha$

            $L_s = L_s \cup \{r\}$
            $B = B \setminus \{r\}$

    $\hat{\mathcal{L}}_n = \hat{\mathcal{L}}_n \cup \{L_s\}$

Return: $\hat{\mathcal{L}}_n = \{L_1, L_2, \ldots, L_K\}$

Notation for the bivariate process

- $A = \{0, 1\}^2$.
- X_t will be the state of the two sources at time t.
- $X_t = (X(1)_t, X(2)_t)$,
- $X(i)_t$ is the state of the source number i at time t.
- $X(i)_t \in \{0, 1\}$ and
- $X_t \in A = \{0, 1\}^2$.

Correction through the marginal processes

- Here we introduce a methodology to improve the probabilities estimated using a PMM fitted on a multivariate time series with a dataset of insufficient size.
- To enhance the multivariate estimations we use models fitted to the marginals univariate time series.
- Let $\mathcal{L}_X$, $\mathcal{L}_{X(1)}$ and $\mathcal{L}_{X(2)}$ be the fitted PMM partitions for the datasets X, $X(1)$ and $X(2)$, with memory lengths M_X, $M_{X(1)}$ and $M_{X(2)}$.
- $X_t \in \{0,1\}^2 = A,\ X(1)_t, X(2)_t \in \{0,1\} = B,$
- To simplify the notation, we will assume that $M_{X(1)} = M_{X(2)} = M$ and $M_X \leq M$.

Correction through the marginal processes

- Consider t such that $M < t \leq T$.
- $s = x_{t-M}^{t-1}, \;\; s \in A^M$,
- $v = x(1)_{t-M}^{t-1}, \;\; v \in \{0,1\}^M$ and
- $w = x(2)_{t-M}^{t-1}, \;\; w \in \{0,1\}^M$,

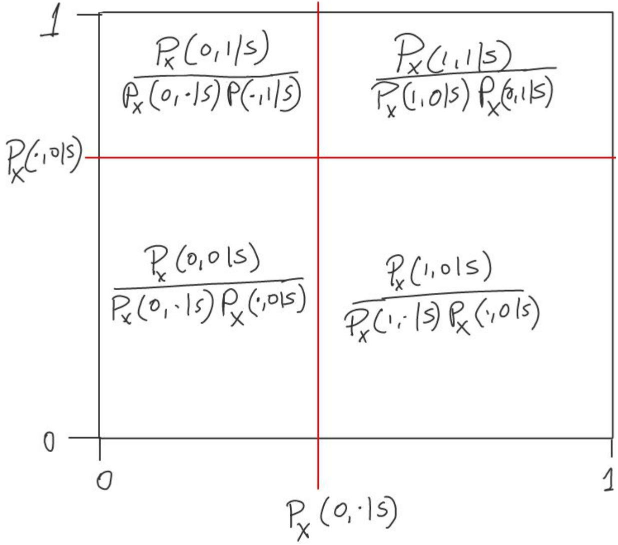
Correction through the marginal processes

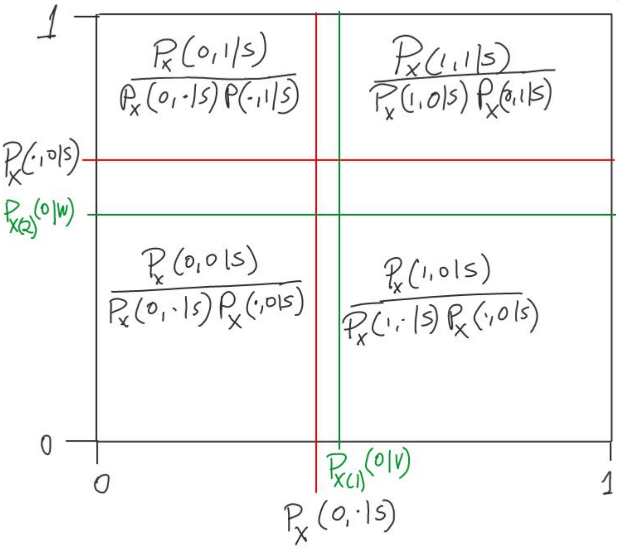
- $P_X(.|s)$, $P_{X(1)}(.|v)$ and $P_{X(2)}(.|w)$ will be the estimated probabilities using the PMM with the corresponding partition $\mathcal{L}_X$, $\mathcal{L}_{X(1)}$ and $\mathcal{L}_{X(2)}$.
- **Objective:** to find an estimative of $P(.|s)$ better than $P_X(.|s)$ using $P_{X(1)}(.|v)$ and $P_{X(2)}(.|w)$.
- **Notation:** for $i, j \in \{0, 1\}$,
  $P_X(i, j|s) = P_X(X(1) = i, X(2) = j|s)$,
  $P_X(i.|s) = P_X(X(1) = i|s)$ and
  $P_X(.j|s) = P_X(X(2) = j|s)$.

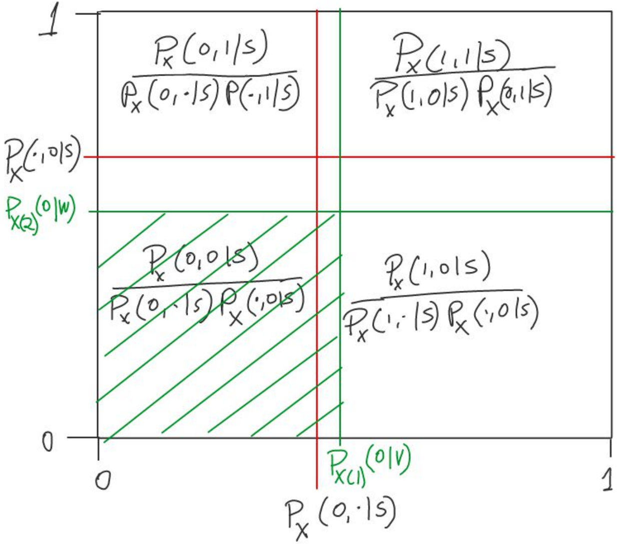
Copula density

A copula density for $P_X(X(1), X(2)|s)$ can be written as follows (see [10]),

$$c_X(u,v) = \begin{cases} \frac{P_X(0,0|s)}{P_X(0.|s)P_X(.0|s)}, & \text{if} \quad 0 \leq a < P_X(0.|s) \text{ and } 0 \leq b < P_X(.0|s), \\ \frac{P_X(0,1|s)}{P_X(0.|s)P_X(.1|s)}, & \text{if} \quad 0 \leq a < P_X(0.|s) \text{ and } P_X(.0|s) \leq b < 1, \\ \frac{P_X(1,0|s)}{P_X(1.|s)P_X(.0|s)}, & \text{if} \quad P_X(0.|s) \leq a < 1 \text{ and } 0 \leq b < P_X(.0|s), \\ \frac{P_X(1,1|s)}{P_X(0.|s)P_X(.0|s)}, & \text{if} \quad P_X(0.|s) \leq a < 1 \text{ and } P_X(.0|s) \leq b < 1. \end{cases}$$


1
$\frac{P_X(0,1|s)}{P_X(0,\cdot|s)P(\cdot,1|s)}$
$\frac{P_X(1,1|s)}{P_X(1,0|s)P_X(0,1|s)}$
$P_X(\cdot,0|s)$
$\frac{P_X(0,0|s)}{P_X(0,\cdot|s)P_X(\cdot,0|s)}$
$\frac{P_X(1,0|s)}{P_X(1,\cdot|s)P_X(\cdot,0|s)}$
0
0
1
$P_X(0,\cdot|s)$


$\frac{P_X(0,1|s)}{P_X(0,\cdot|s)P(\cdot,1|s)}$
$\frac{P_X(1,1|s)}{P_X(1,0|s)P_X(0,1|s)}$
$P_X(\cdot,0|s)$
$P_{X(2)}(0|w)$
$\frac{P_X(0,0|s)}{P_X(0,\cdot|s)P_X(\cdot,0|s)}$
$\frac{P_X(1,0|s)}{P_X(1,\cdot|s)P_X(\cdot,0|s)}$
1
0
0
1
$P_{X(1)}(0|v)$
$P_X(0,\cdot|s)$


1
$\frac{P_X(0,1|s)}{P_X(0,\cdot|s)P(\cdot,1|s)}$
$\frac{P_X(1,1|s)}{P_X(1,0|s)P_X(0,1|s)}$
$P_X(\cdot,0|s)$
$P_{X(2)}(0|w)$
$\frac{P_X(0,0|s)}{P_X(0,\cdot|s)P_X(\cdot,0|s)}$
$\frac{P_X(1,0|s)}{P_X(1,\cdot|s)P_X(\cdot,0|s)}$
0
0
$P_{X(1)}(0|v)$
1
$P_X(0,\cdot|s)$

To complete the correction,

- We use the margin $P_{X(1)}(X = i|v)$ instead of $P_X(i.|s)$ and
- $P_{X(2)}(i|w)$ will substitute $P_X(.i|s)$
- Having the new marginals and the copula density, we produce the new corrected joint probability obtained by using c_X and the computed $P_{X(1)}$ and $P_{X(2)}$ margins.
  More precisely the corrected probability of $X_t = (0,0)$ given $s = X_{t-M}^{t-1}$ is,

$$P_c(0,0|s) = \int_0^{P_{X(1)}(0|v)} \int_0^{P_{X(2)}(0|w)} c_X(a,b|s)dbda$$

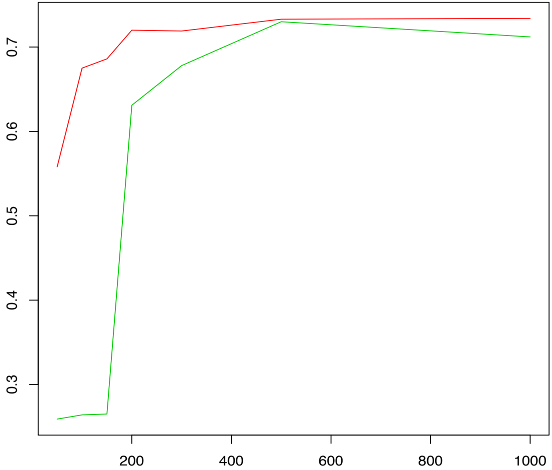
Our methodology can be summarized as follows,

1) fit a PMM model to the joint time series X_t,
2) compute the copula density $c_X(u, v|L)$ of $P_X(.|L)$ for each $L \in \mathcal{L}_X$,
3) fit PMM models to the margins $X(1)_t$ and $X(2)_t$,
4) compute the corrected probabilities for X_t by using the margins fitted in 3) and the expression of $c_X(u, v)$.

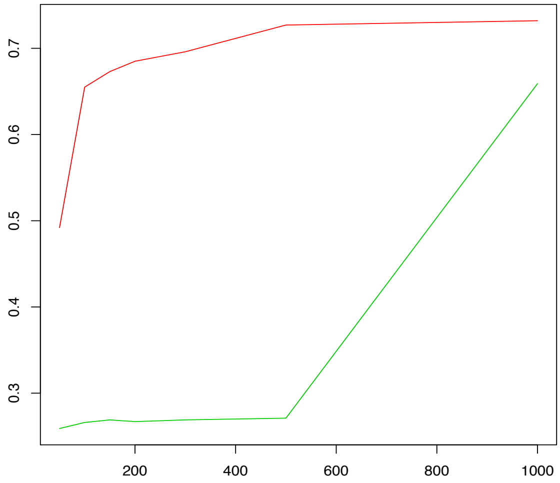
Simulations

- We simulated data corresponding to 5 PMM.
- For each model we used the sample sizes: 50, 100, 150, 200, 300, 500 and 1000.
- For each model M and sample size T, we simulated 100 datasets consisting of two parts,
- part one is a size T sample and part two is a size 10000 sample.
- for each dataset, first, the model selection and correction methodology was applied to the size T sample. Second, the corrected probabilities obtained from the size T sample are used on the size 10000 sample to predict the successive values in the sample. Obtaining the rate of correct predictions on the size 10000 sample.

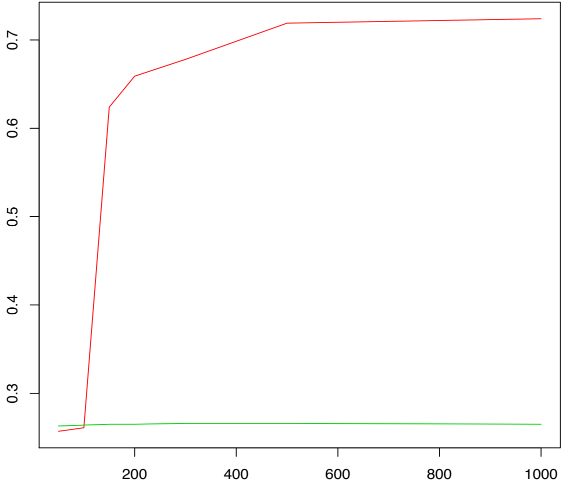
Simulation model 1



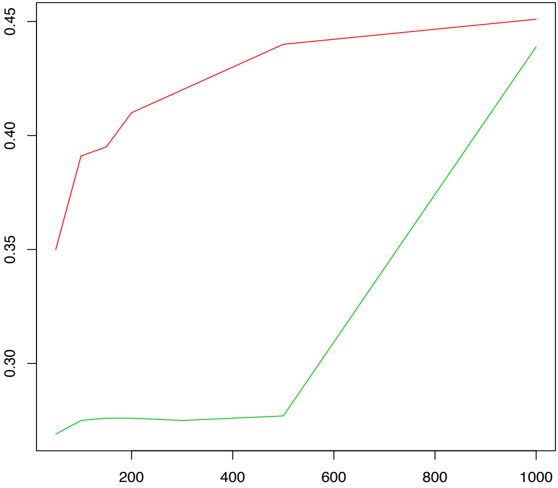
Simulation model 2



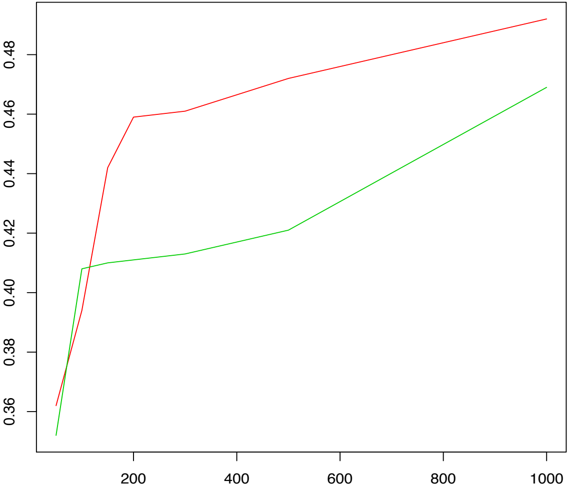
Simulation model 3



Simulation model 4



Simulation model 5

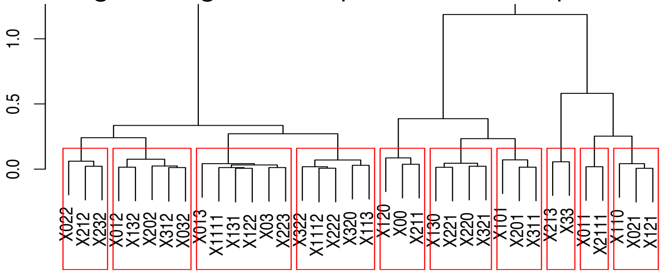


Application

- **Dataset:** directions of the daily variation of two asset and denoted by $X(1)_t$ and $X(2)_t$.
- the first series, $\{b_t\}_{t=0}^{T}$, corresponds to the main Brazilian stock exchange index (Ibovespa),
- the second series, $\{d_t\}_{t=0}^{T}$, corresponds to the exchange rate of United States Dollar to Brazilian Real,
- the two series where observed from 2003 to 2012.
- $X(1)_t = 0$ if $d_t \leq d_{t-1}$ and $X(1)_t = 1$ otherwise,
- $X(2)_t = 0$ if $b_t \leq b_{t-1}$ and $X(2)_t = 1$ otherwise.
- We will analyze the bivariate series $X_t = (X(1)_t, X(2)_t)$.

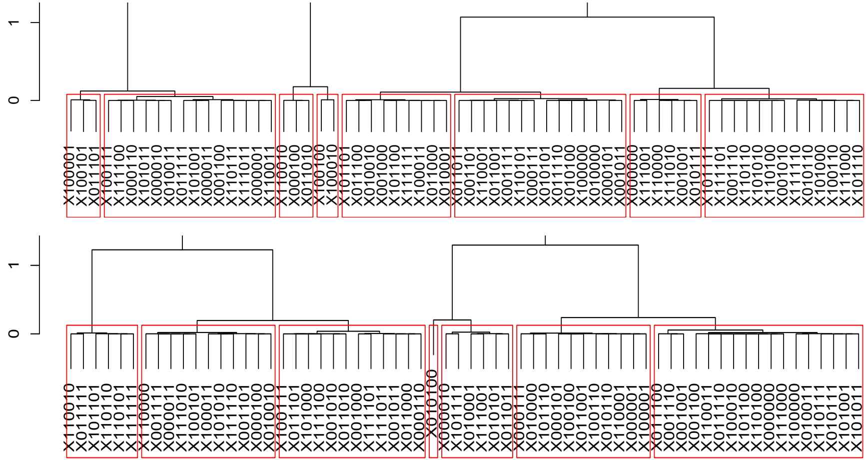
Application

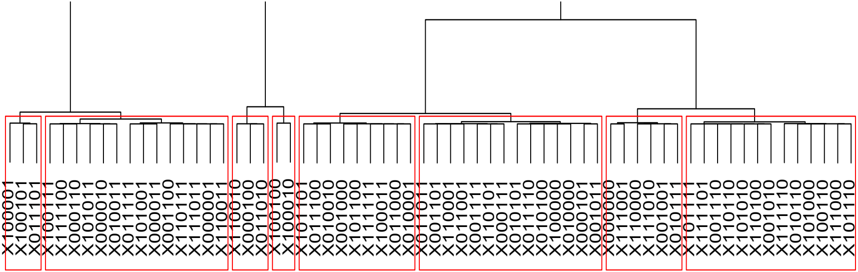
First we fit a PMM to the bivariate series X_1^T and obtain the following dendrogram. The partition have 10 parts.



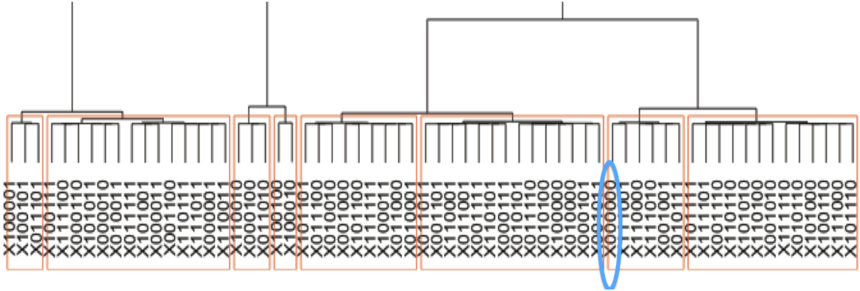
The longest string, for which we have enough repetitions to estimate probabilities, has size 4.

We fit PMM for the marginals $X(1)_1^T$ and $X(2)_1^T$,

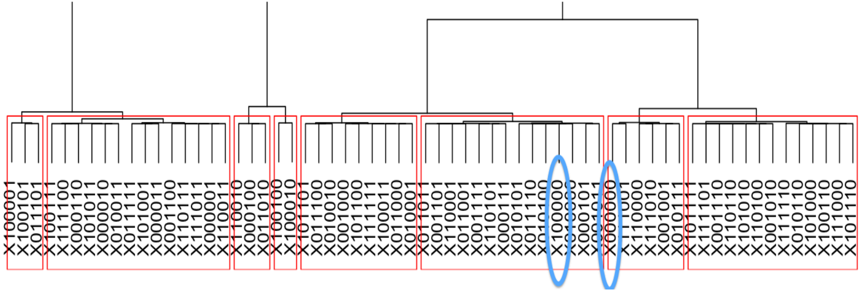




- For the marginal series $X(1)_1^T$ and $X(2)_1^T$, all strings of size 6 are included in the parts of the estimated partitions.
- From the partition is easy to see that the memory can not be reduced,
- for example for X_t, the strings 000000 and 100000 are in very different parts.



- For the marginal series $X(1)_1^T$ and $X(2)_1^T$, all strings of size 6 are included in the parts of the estimated partitions.
- From the partition is easy to see that the memory can not be reduced,
- for example for X_t, the strings 000000 and 100000 are in very different parts.



- For the marginal series $X(1)_1^T$ and $X(2)_1^T$, all strings of size 6 are included in the parts of the estimated partitions.
- From the partition is easy to see that the memory can not be reduced,
- for example for X_t, the strings 000000 and 100000 are in very different parts.

- The model fitted for X_t has smaller memory than the model for the marginals.
- We applied the uncorrected model to the dataset and it successfully predicted the next value for the X_t series 32% of the time.
- Using our corrected model we improved the assertion rate to 35%.

Conclusions

- The present paper shows a new methodology for the problem of insufficient data for modeling jointly two Bernoulli time series when the marginals are well modeled.
- In the application to real finance data, our methodology produce an increment from 32% to 35% in the percentage of good predictions for the next symbol in the sequence.
- The methodology can be used with others families of Markov models, such as the variable length ones, for which there exists several model selection methods (see [11], [2] and [3]).

Acknowledgments

The authors acknowledge the support provided by USP project "Mathematics, computation, language and the brain. FAPESP projects "Portuguese in time and space: linguistic contact, grammars in competition and parametric change", 2012/06078-9 and "Research, Innovation and Dissemination Center for Neuromathematics - NeuroMat", 2013/07699-0. Additionally, the authors acknowledge the financial support for this research provided by FAPESP Post-doctoral Grant 2011/18285-6 to M. Fernández.

Cassandro, M., Collet, P., Duarte, D., Galves, ., and Garcia, J. "A Stochastic model for the speech sonority." *Mathématiques et sciences humaines. Mathematics and social sciences* **180**, 43–55, 2007.

Csiszár, I. and Talata, Z., "Context tree estimation for not necessarily finite memory processes, via BIC and MDL", *IEEE Trans. Inform. Theory* **52**, 1007–1016, 2006.

Galves, A., Galves, C., Garcia, J. E., Garcia, N. L. and Leonardi, F., "Context tree selection and linguistic rhythm retrieval from written texts", *Annals of Applied Statistics*, **6** 1, 186 – 209, 2012.

Garcia, J. and Gonzalez-Lopez, V. A., "Minimal Markov Models", *arXiv:1002.0729v1*, 2010.

Garcia, J. E., Gonzalez-Lopez, V. A. and Nelsen, R.B., "A new index to measure positive dependence in trivariate distributions". *Journal of Multivariate Analysis* **115**, 481–495, 2013.

Garcia, J. E., Gonzalez-Lopez, V. A. and Viola, M. L. L., "Robust model selection and the statistical classification of languages", *AIP Conference Proceedings* **1490**, 160 – 170, 2012.

Garcia, J. E., Gonzalez-Lopez, V. A. and Viola, M. L. L., "Robust model selection for stochastic processes", *In fifth Workshop on information theoretic methods in science and engineering,* p79, 2012.

Garcia, J. E., Gonzalez-Lopez, V. A. and Viola, M. L. L., "Model selection for multivariate stochastic processes", *In fifth Workshop on information theoretic methods in science and engineering,* p68, 2012.

Garcia, J. E., Gut, U., and Galves A. "Vocale-a semi automatic annotation tool for prosodic research." *In Speech Prosody 2002, Internationa Conference.* 2002.

Joe, H., *Multivariate models and dependence concepts* (Vol. 73), CRC Press, 1997.

Rissanen J., "A universal data compression system", *IEEE Trans. Inform. Theory* **29**(5), 656 – 664, 1983.