A Copula-Based Partition Markov Procedure

Jesús E. García[1]

III LACSC
San José, Costa Rica

[1]IMECC-UNICAMP e-mail: jg@ime.unicamp.br

The number of parameters needed to specify a discrete multivariate Markov model grows exponentially with the order and dimension of the chain. For a Markov chain with fixed order o and multivariate alphabet $A = B^k$, the number of parameters needed is,

$$(|B|^k - 1)|B|^{ko}.$$

Example

Consider $B = \{0, 1\}$,

- for $k = 2$ we have $A = \{0, 1\}^2 = \{(0, 0); (0, 1); (1, 0); (1, 1)\}$.
  - For $o = 1$ the number of parameters needed to specify the model will be $(4 - 1) * 2^{2*1} = 12$.
  - For $o = 2$, the number of parameters needed to specify the model will be $(4 - 1) * 2^{2*2} = 48$
- For $k = 3$ and $o = 3$ the number of parameters needed to specify the model will be $(8 - 1) * 2^{3*3} = 3584$

The amount of data could be enough to produce good marginal models and not big enough to produce multivariate models.

- We show how a new estimation strategy can help to extend the memory (length of the past) to be considered in those series.
- The strategy consists on choosing a model for each marginal, a model for the joint process and then joining those models using the copula.
- In this paper we present the situation on the scope of Partition Markov Models (PMM).

Notation

- Let (X_t) a discrete time, order M Markov chain on a finite alphabet $A = B^k$.
- $X_t = (X(1)_t, ..., X(k)_t)$,
- $X(i)_t$ is the state of the marginal i at time t.
- $X(i)_t \in B$ and
- $X_t \in A = B^k$.
- Denote the string $a_m a_{m+1} \dots a_n$ by a_m^n, where $a_i \in A$, $m \leq i \leq n$.
- x_1^n will be a size n realization of X_t.

Notation

For each $s \in S = A^M$, $a \in A$,

- $$N(s) = \sum_{i=M+1}^{n} 1_{\{x_{i-M}^{i-1}=s\}},$$
- $$N(s,a) = \sum_{i=M+1}^{n} 1_{\{x_{i-M}^{i-1}=s, x_i=a\}},$$
- we denote the conditional joint probability of the process by,

$$P^J(a|s) = \text{Prob}(X_t = a | X_{t-M}^{t-1} = s).$$

For each $b \in B$ and $1 \leq i \leq k$, the conditional marginal probability of the marginal i is,

$$P_i(b|s) = \text{Prob}(X(i)_t = b | X_{t-M}^{t-1} = s).$$

Equivalence

Definition

(Equivalence relationship based on the joint distribution) For each $s, r \in S$, $s \sim r$ if $P(a|s) = P(a|r)\ \forall a \in A$.

Definition

(Equivalence relationship based on the marginal distributions) For each $i \in \{1, 2, ...k\}$ and $s, r \in S$,
$s \sim_i r$ if $P_i(b|s) = P_i(b|r)\ \forall b \in \{0, 1\}$.

Remark

For each $s, r \in S$, $s \sim r \ \Rightarrow\ s \sim_i r\ \forall i \in \{1, 2, ...k\}$.

Remark

If all the sources are independent all the time then, for each $s, r \in S$,

$$s \sim r \iff s \sim_i r \ \ \forall i \in \{1, 2, ...k\},$$

Example

Bi-variate case, order 2.

- $k = 2$,
- $M = 2$,
- $A = \{0, 1\}^2$ and $S = A^2$.
- For any $s, r \in S$, $s \sim r$ if and only if,
  1) $P_1(0|s) = P_1(0|r)$,
  2) $P_2(0|s) = P_2(0|r)$ and
  3) $P((0,0)|s) = P((0,0)|r)$.

Partition Markov Models

Let (X_t) be a discrete time order M Markov chain on a finite alphabet A. Let us call $\mathcal{S} = A^M$ the state space. Denote the string $a_m a_{m+1} \dots a_n$ by a_m^n, where $a_i \in A$, $m \leq i \leq n$. For each $a \in A$ and $s \in \mathcal{S}$, $P(a|s) = \text{Prob}(X_t = a | X_{t-M}^{t-1} = s)$. Let $\mathcal{L} = \{L_1, L_2, \dots, L_K\}$ be a partition of $\mathcal{S}$, for $a \in A$, $L \in \mathcal{L}$,

$$P(L, a) = \sum_{s \in L} \text{Prob}(X_{t-M}^{t-1} = s, X_t = a),$$

$$P(L) = \sum_{s \in L} \text{Prob}(X_{t-M}^{t-1} = s) \text{ and } P(a|L) = \frac{P(L, a)}{P(L)}.$$

Definition

Let (X_t) be a discrete time order M Markov chain on a finite alphabet A. We will say that $s, r \in \mathcal{S}$ are equivalent (denoted by $s \sim_p r$) if $P(a|s) = P(a|r)\ \forall a \in A$. For any $s \in \mathcal{S}$, the equivalence class of s is given by $[s] = \{r \in \mathcal{S} | r \sim_p s\}$.

The previous definition allows to define a Markov chain with "minimal partition", that is the one which respects the equivalence relationship.

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 the equivalence relationship $\sim_p$ introduced by the previous definition.

Estimation of the minimal partition

- There are several algorithms for the estimation of the minimal patition (García and González-López (2011)[1] and García and González-López (2017)[2]).
- Those algorithms are based on the BIC criterion or on a distance defined on the state space.
- Those algorithms are consistent in the sense that the estimated partition $\hat{\mathcal{L}}_n$ converges eventually almost surely to $\mathcal{L}$, where $\mathcal{L}$ is the partition of $\mathcal{S}$ defined by the equivalence relationship $\sim_p$.

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

- $N_n(s,a) = \left|\{t : M < t \leq n, x_{t-M}^{t-1} = s, x_t = a\}\right|$,
  $N_n(s) = \left|\{t : M < t \leq n, x_{t-M}^{t-1} = s\}\right|$.
- The estimator of $P(a|s)$ is defined by

$$\hat{P}^J(a|s) = \frac{\sum_{r \in \hat{L}} N_n(r,a)}{\sum_{r \in \hat{L}} N_n(r)}.$$

such that $s \in \hat{L}$ and $\hat{L}$ is a part of $\hat{\mathcal{L}}_n$. J: indicates its dependence of the joint process.

Notation

- $X_t = (X(1)_t, ..., X(k)_t)$, where $X(i)_t \in B$ and it is the state of the i-source at time t for $i = 1, ..., k$,
- $X_t \in A$, where $A = B^k$ and B is the finite alphabet for the one-dimensional marginal processes.
- We will assume that for $1 \leq i \leq k$, $X(i)_t$ is an order o_M Markov chain, with $o_M < \infty$. The marginal state space is B^{o_M}.
- For each $s \in B^{o_M}$ and $b \in B$, we denote
  - the marginal conditional probability for the i coordinate by $P_i(b|s) = \text{Prob}(X(i)_t = b | X(i)_{t-o_M}^{t-1} = s)$.
  - the marginal cumulative function by $F_i(\cdot|s), i = 1, \ldots, k$.
- Given a sample x_1^n of X_t, we assume $o = \lfloor \log_{|A|}(n) \rfloor - 1$ and $o_M = \lfloor \log_{|B|}(n) \rfloor - 1$, as a consequence $(o_M > o)$.

Order

In practical situations, given a data set, a way to estimate the maximal order for the posible model is to use the following rule, Suppose that the alphabet is a finite set D, with cardinal $|D|$ and suppose also that the data set is of size N, the relation that will be followed is

$$|D|^{l}(|D|-1)\log(N)=N,$$

where l represents the maximum order possible to be used in the estimation.

The relation means that the number of parameters is $N/\log(N)$.
Then,

$$l < \log_{|D|}(N) - 1,$$

for N large.
Because $l = \frac{\log(N)}{\log(|D|)} - \frac{\log(|D|-1)}{\log(|D|)} - \frac{\log(\log(N))}{\log(|D|)}$.
According to our assumptions,
Joint Case:

$$l_{joint} = o < \frac{\log(n)}{k \log(|B|)} - 1.$$

Marginal Case:

$$l_{marg} = o_M < \frac{\log(n)}{\log(|B|)} - 1.$$

The Procedure

- Fit a PMM to the process X_t with a maximum order equal to o on A^o, obtaining $\mathcal{L}^o = \{L_1^o, ..., L_{m_o}^o\}$, then define a partition on A^{o_M},

$$\mathcal{P}_\alpha = \{L_1^\alpha, ..., L_{m_o}^\alpha\}, \text{ with } L_j^\alpha = \cup_{s \in L_j^o} \{w.s : w \in A^{(o_M - o)}\}$$

$$1 \leq j \leq m_o.$$

- Fit a PMM to each marginal process. Call $\mathcal{L}^i = \{L_1^i, ..., L_{m_i}^i\}$ the partition of B^{o_M} corresponding to the model fitted to the marginal process $X(i)_t, i = 1, \ldots, k$.
- Define the following partition of A^{o_M}.

$$\mathcal{P}_\beta = \{L_{j_1}^1 \times ... \times L_{j_k}^k : 1 \leq j_1 \leq m_1, ..., 1 \leq j_k \leq m_k\}.$$

Where $L_{j_1}^1 \times ... \times L_{j_k}^k = \{s(1) \times \ldots \times s(k) : s(i) \in L_{j_i}^i\}$, with $s(1) \times \ldots \times s(k) = \{s(1)_i, \ldots, s(k)_i\}_{i=1}^{o_M}$.

Definition

$s, r \in \mathcal{S}$ are equivalent, denoted by $s \sim r$ if there exist parts $L \in \mathcal{P}_\alpha$ and $L' \in \mathcal{P}_\beta$ such that $s, r \in L \cap L'$. The partition of $\mathcal{S}$ given by the relation "$\sim$" is denoted by $\mathcal{P}$.

This means that two states s and r belong to the same part of $\mathcal{P}$ if and only if, they belong to the same part of both $\mathcal{P}_\alpha$ and $\mathcal{P}_\beta$.
As we see, the longest memory possible in this context is o_M.
Now we introduce how to compute the transition probability from each string in A^{o_M} to $a \in A$.

Given $s \in A^{o_M}$ and $a \in A$:

- Let w be the size o suffix of s, that means $s = q.w$ for an appropriated string q. Consider the estimator $\hat{P}^J(a|w) = \frac{\sum_{r \in \hat{L}} N(r,a)}{\sum_{r \in \hat{L}} N(r)}$, for $w \in \hat{L}$ such that $\hat{L}$ is a part of $\mathcal{L}^o$.
- For $1 \leq i \leq k$, let $s(i)$ be the sequence in B^{o_M}, that is the sequence consisting of the concatenation of elements of s in the coordinate i.
  Denote by
  - $\hat{P}_i(a(i)|s(i))$ the estimate of the marginal probability,
  - $\hat{F}_i(a(i)|s(i))$ the estimate of the cumulative function,

  from the i process of order o_M, where $a(i)$ is the i-coordinate of a.

Incorporating Copula Function

The two set of probabilities are combined in the following way.

- We define a k-dimensional copula distribution $\hat{C}\left(u_1, \ldots u_k | w\right)$ from the joint probabilities $\hat{P}^J(a|w)$.
- With $u_i \in [0,1], 1 \leq i \leq k$. The copula distribution is evaluated on the marginal distributions, and the estimator is given by

$$\hat{P}(X_t \leq a|s) = \hat{C}\Big(\hat{F}_1(a(1)|s(1)), \ldots, \hat{F}_k(a(k)|s(k))|w\Big).$$

Theorem

For any $L \in \mathcal{P}$ if $s, r \in L$ then $\hat{P}(X_t \leq a|s) = \hat{P}(X_t \leq a|r)$, $\forall a \in A$.

For details about the properties and behavior of this procedure see Fernández, García & González-López (2017) [3].

General Form of a K-Discrete Copula

- $(Z_1, \ldots, Z_K)$ is a vector of discrete random variables, Z_k takes values in the domain $D_k = \{z_{k1}, z_{k2}, \ldots, z_{km_k}\}$, for $k = 1, \ldots, K$,
- $\text{Prob}(Z_1 = z_{1i_1}, \ldots, Z_K = z_{Ki_K}) = p_{z_{1i_1} \ldots z_{Ki_K}}$,
- The univariate marginal distribution of Z_1 : $p_{z_{1i_1} \bullet \ldots \bullet} = \sum_{i_2} \cdots \sum_{i_K} p_{z_{1i_1} z_{2i_2} \ldots z_{Ki_K}}$, for each $z_{1i_1} \in D_1$.
- The cumulative marginal distribution F_1, applied to an arbitrary point z: $F_1(z) = \sum_{z_{1i_1} \leq z} p_{z_{1i_1} \bullet \ldots \bullet}$.

The k-variate copula density is given by

$$c(u_1,\ldots,u_K)=\begin{cases}\dfrac{p_{z_{1i_1}\ldots z_{Ki_K}}}{p_{z_{1i_1}\bullet\ldots\bullet}\times\ldots\times p_{\bullet\ldots\bullet z_{Ki_K}}}, \\ \qquad\qquad \text{if}(u_1,\ldots,u_K)\in\otimes_{k=1}^{K}[F_k(z_{ki_k-1}),F_k(z_{ki_k}))\\ 0, \qquad\qquad \text{otherwise,}\end{cases}$$

with $F_k(z_{k1-1})=F_k(z_{k0})=0$, for $k=1,\ldots,K$.
The function $c(u_1,\ldots,u_K)$ satisfies the following characteristics

- (i) it is a probability mass function, displayed in $[0,1]^K$,
- (ii) the univariate marginal distributions are $U(0,1)$ and
- (iii) the cumulative distribution C of c verifies

$$\text{Prob}(Z_1\le z_{1i_1},\ldots,Z_K\le z_{Ki_K})=C\big(F_1(z_{1i_1}),\ldots,F_K(z_{Ki_K})\big),$$

for all $(z_{1i_1},\ldots,z_{Ki_K})\in\otimes_{k=1}^{K}D_k$.

see Fernández, García & González-López (2015) [4].

Conjecture and Evidences

- According to the linguistic conjecture (Lloyd (1940) [5], Pike (1945) [6] and Abercrombie (1967) [7]), the languages are divided into three classes by their rhythmic properties, those are: *stress-timed*, *syllable-timed* and *mora-timed*.
- Examples in each class are, English and Dutch (stress-timed), French, Spanish and Italian (syllable-timed), Japanese (mora-timed).
- Was reported the existence of intermediate languages, Catalan and Polish are examples. Polish shows a high syllable complexity but without the expected vowel reduction for a stress-timed language. Catalan has the same syllabic system as Spanish but it has vowel reduction.
- Ramus et al. (1999) [8], García, González-López & Viola (2012) [1], García & González-López (2014) [2].

Linguistic Data

The data set consists of 576 recorded sentences belonging to 3 languages and it is described in the next table,

Table: Sentences from English (EN), Japanese (JA), Spanish (SP). From a corpus belonging to the *Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS/CNRS)*.

Language	EN	JA	SP
Number of sentences	152	212	212

The sentences have lengths going from 2 to 3.5 seconds, digitalized at 16.000 samples per second, i.e. sample rate of 16 kHz.

Fixed a language l we consider the sentence j of length $T_{l,j}$. Given a frequency f we denote by $\vartheta_t^{l,j}(f)$ the power spectral density at time t for that sentence j and language l where $t = 1, \cdots, T_{l,j}$. For each time t we consider the stochastic processes (energies)

$$\chi_1^{l,j}(t) = \sum_{f=80,100,\ldots,800} \vartheta_t^{l,j}(f),$$

$$\chi_2^{l,j}(t) = \sum_{f=820,1520,\ldots,1480} \vartheta_t^{l,j}(f),$$

$$\chi_3^{l,j}(t) = \sum_{f=1500,1520,\ldots,5000} \vartheta_t^{l,j}(f).$$

The definition of the energies bands, including the frequencies for the bands, where chosen based on previous works about automatic segmentations of speech signal in vowels and consonants, see for example García et al. (2002) [3].

For each sentence j from language l and an energy band k, where $k = 1$ represents the inferior band of energy $\chi_1^{l,j}(t)$, $k = 2$ represents the midband of energy $\chi_2^{l,j}(t)$ and $k = 3$ represents the superior band of energy $\chi_3^{l,j}(t)$ we define $Y_t^{l,j,k} = 1$ if $\chi_k^{l,j}(t+1) > \chi_k^{l,j}(t)$, and $Y_t^{l,j,k} = 0$ otherwise.
From the above description the size of the data set for each language is given by the next table.

Table: Sample size available, for each language.

Language	EN	JA	SP
Size	17010	22035	26528

According to the rules established to define o and o_M, we have

Table: Orders for each language.

Language	EN	JA	SP
o_M	13	13	13
o	3	3	3

The next table shows a brief description of the Partition Markov Models: marginals and joint, estimated for each language.

Table: Cardinal of partitions. Marginals: $\mathcal{L}^i$, $i = 1$ - inferior band, $i = 2$ - midband, $i = 3$ - superior band.

Language	$\lvert\mathcal{L}^1\rvert$	$\lvert\mathcal{L}^2\rvert$	$\lvert\mathcal{L}^3\rvert$	$\lvert\mathcal{L}^o\rvert$ (Joint)
EN	7	8	7	11
JAP	8	8	8	14
SP	7	9	8	12

Profiles

We say that the profile of the language has been well charted if comparing the predictive ability of this model (PMM's+Copula) to others, it brings an improvement. For this, we use to compare two settings (1) the joint PMM model of order o and (2) the model of independence, between the marginal processes.

Table: Predictive ability: proportion of correct predictions.

Language	Independence	joint PMM (order o)	PMM's+Copula
EN	0.4120	0.3454	**0.4564**
JAP	0.3795	0.3743	**0.4422**
SP	0.3618	0.3909	**0.4455**

Applying equation (**??**) and from $\hat{P}^J$ with memory $o = 2$ we obtain
$\hat{P}^J\Big(x_{t-1}x_t = (0,0)(2,0)|x_{t-3}^{t-2} = (1,0)(2,0)\Big)$

$$= \hat{P}^J\Big(x_{t-1} = (0,0)|x_{t-3}^{t-2}\Big)\hat{P}\Big(x_t = (2,0)|x_{t-2}^{t-1}\Big) = 0.00083.$$

But if we compare with the estimation made with $\hat{P}$
$\hat{P}\Big(x_{t-1}x_t = (0,0)(2,0)|x_{t-7}^{t-2}\Big)$

$$= \hat{P}\Big(x_{t-1} = (0,0)|x_{t-7}^{t-2}\Big)\hat{P}\Big(x_t = (2,0)|x_{t-6}^{t-1}\Big) = 0.23862.$$

We obtain a probability (0.23862) which could be meaningful for a financial decision.

Final Remarks

- In this paper we show how it can be relevant to consider a longer past, to make predictions of events of interest.
- We show how this issue is a challenge in the Markov models and for this reason it makes sense to use models built from an economic conception, as in the case of Partition Markov Models.
- The incorporation of the concept of copula brings great benefit to those models, allowing to extend the memory to be considered in the statistical estimation of the process.
- With the procedure described in this article, we see that it is possible to produce an improvement in the predictive power of the model, and this is because we incorporate through a copula the marginal estimates, which will require a smaller sample size than required by a traditional joint estimate.

J.E. García and V.A. González-López. Minimal Markov Models. In *Proceedings of the Fourth Workshop on Information Theoretic Methods in Science and Engineering. Helsinki.* v.1. p.25 - 28, 2011.

J.E. García and V.A. González-López (2017) Consistent Estimation of Partition Markov Models, *Entropy*, 19(4), 160.

M. Fernández, J.E. García and V.A. González-López (2017) A copula-based partition Markov procedure, *Communications in Statistics-Theory and Methods* (`https://doi.org/10.1080/03610926.2017.1359291`)

M. Fernández, J.E. García and V.A. González-López (2015). Multivariate Markov chain predictions adjusted with copula models. In *New Trends in Stochastic Modeling and Data Analysis* (chapter 8, page 389).

J. Lloyd, Speech signal in telephony, (Sir I. Pitman & sons, London, 1940).

K. L. Pike, The intonation of American English, (Ann Arbor: University of Michigan Press, 1945).

D. Abercrombie, Elements of general phonetics (Vol. 203). (Edinburgh University Press, Edinburgh, 1967).

F. Ramus, M. Nespor and J. Mehler, Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**(3), 265-292, 1999.

J. E. García, V. A. González-López and M. L. L. Viola, Robust model selection and the statistical classification of languages. *AIP Conference Proceedings*, **1490**, 160-170, 2012. Doi: 10.1063/1.4759600

J. E. García and V. A. González-López, Modeling of acoustic signal energies with a generalized Frank copula. A linguistic conjecture is reviewed. *Communications in Statistics-Theory and Methods*, **43**(10-12), 2034-2044, 2014. Doi: 10.1080/03610926.2013.866248

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

J.E. García and M. Fernández. Copula based model correction for bivariate Bernoulli financial series. In *11TH INTERNATIONAL CONFERENCE OF NUMERICAL ANALYSIS AND APPLIED MATHEMATICS 2013: ICNAAM 2013,* vol. 1558, no. 1, pp. 1487-1490. AIP Publishing, 2013.

Thanks for your attention!,
Jesús