Estimation and model selection for mixing graphical models

63rd ISI World Statistics Congress 2021

Magno T.F. Severino*† and Florencia Leonardi †

March 22, 2021

Abstract

In this short paper, we propose a model selection criterion to estimate the graph of conditional dependencies of a random vector on a finite alphabet, whose distribution corresponds to the stationary distribution of a mixing stochastic process. The method is based on the regularization of the maximum conditional likelihood of each node, given the remaining nodes in the graph. We evaluate the convergence of the estimator on simulated data and discuss an application to real data.

Keywords: mixing process, regularized maximum likelihood, structure estimation, dimensionality reduction.

1 Introduction

In this short paper, we propose a method to estimate the graph of conditional dependencies of a vector on a finite alphabet whose distribution corresponds to the stationary distribution of a mixing stochastic process. This method extends the previous works Leonardi et al. (2019) and Leonardi et al. (2021), which considered model selection for independent samples of graphical models and detection of independent blocks in a vector for mixing processes, respectively.

In the following section, we define the estimator of the graph of conditional dependencies of the stationary measure of the mixing process. In Section 3 we show the results of an evaluation of the estimator on simulated data, and in Section 4 we apply the method to real data of water flow measurements in São Francisco river, Brazil, also considered in Leonardi et al. (2021). We show that the results obtained by our method are consistent with the conclusions of the previous analysis.

2 Graph estimator on mixing processes

Let $\{X^{(i)} \colon i \in \mathbb{N}\}$ be a multivariate stochastic process taking values in A^d with transition matrix Q. We assume the process is in a stationary regime, with invariant distribution denoted by π. The indices of the random vector $X^{(i)}$ belonging to $V = \{1, \dots, d\}$ will be called vertices, and then $X_v^{(i)}$ is the random variable observed at time i on vertex $v \in V$. As $X^{(i)}$ has distribution π for all $i \in \mathbb{N}$, then when not necessary, the time index will be dropped.

Any subset of vertices $W \subset V$, with $v \notin W$ is referred to as a *neighborhood* of $v \in V$. The cardinal of the set W is denoted by $|W|$. A Markov neighborhood of v is a neighborhood W satisfying $\pi(X_v|X_W) = \pi(X_v|X_{-v})$ for all $x \in A^V$ with $\pi(X_{-v}) > 0$. Here, X_{-v} refers to the vector $(X_w \colon w \in V \setminus \{v\})$. The smallest Markov neighborhood of v, called "basic" neighborhood of v, will be denoted by $\text{ne}(v)$.

*Corresponding author: magno@ime.usp.br

†Universidade de São Paulo

Based on the basic neighborhoods of each node $v \in V$, define the graph $G^* = (V, E)$ by

$$(v, w) \in E \text{ if and only if } w \in \mathrm{ne}(v). \tag{1}$$

It is easy to see that the graph G^* defined in (1) is undirected, see Leonardi et al. (2019) for details.

Assume we observe a sample of size n of the process, denoted by $x^{(1)}, \ldots, x^{(n)}$. Given a vertex $v \in V$ and a neighborhood W, the maximal conditional likelihood function of X_v given $X_W = x_W$ computed on the sample is given by

$$\widehat{\mathbb{P}}(x_v^{(1:n)} | x_W^{(1:n)}) = \prod_{a_w \in A^W} \prod_{a \in A} \hat{\pi}(X_v = a | X_W = a_W)^{N(a, a_W)},$$

with the product being over all $a_W \in A^W$ for which $N(a_W) > 0$. Here, $N(a, a_W)$ is the number of times the sub-vector (a, a_W) appears in the sample $\{x_w^{(1:n)} \colon w \in \{v\} \cup W\}$ and

$$\hat{\pi}(X_v = a | X_W = a_W) \;=\; \frac{N(a, a_W)}{\sum_b N(b, a_W)}\,.$$

Given a constant $\lambda > 0$ and a vertex $v \in V$ we define the neighborhood estimator for the set $\mathrm{ne}(v)$ as

$$\widehat{\mathrm{ne}}(v) \;=\; \operatorname*{arg\,max}_{W \subset V \setminus \{v\}} \{\log \hat{\mathbb{P}}(x_v^{(1:n)} | x_W^{(1:n)}) + \lambda |A|^{|W|} \log n\}. \tag{2}$$

Since we are interested in estimating the graph G^*, we can estimate the neighborhood of each node and reconstruct the graph based on the set of estimated neighborhoods, considering the intersection of the estimated edges (conservative approach) or the union of the estimated edges (non-conservative approach).

We say the process $\{X^{(i)} \colon i \in \mathbb{N}\}$ satisfies a mixing condition with rate $\{\psi(\ell)\} \downarrow 0$ as $\ell \to \infty$ if for each $k, m \in \mathbb{N}$ and each $x^{(1:k)} \in (A^d)^k$, $x^{(1:m)} \in (A^d)^m$ with $\mathbb{P}(X^{(1:m)} = x^{(1:m)}) > 0$ we have that

$$\left|\mathbb{P}(X^{((n+1):(n+k))} = x^{(1:k)} \,|\, X^{(1:m)} = x^{(1:m)}) - \mathbb{P}(X^{(1:k)} = x^{(1:k)})\right| \leq \psi(n - m)\mathbb{P}(X^{(1:k)} = x^{(1:k)}), \tag{3}$$

for $n \geq m$. Processes satisfying a mixing condition as (3) were considered in Leonardi et al. (2021) to prove the consistency of the estimator of the points of independence of a random vector. This estimator produces a decomposition of the distribution function of the vector into independent blocks. The proof is based on concentration inequalities for the empirical probabilities obtained by Csiszár (2002), and was obtained for a process satisfying (3) with $\psi(n - m) = \delta^{n-m}$ for some $0 < \delta < 1$.

We conjecture that the estimator of the basic neighborhood of a node $v \in V$, given by (2), is consistent when $n \to \infty$. The particular case for $\{X^{(i)} \colon i \in \mathbb{N}\}$ independent and identically distributed random vectors with distribution π was recently proved in Leonardi et al. (2019). Our goal is to generalize these findings for the non-i.i.d. mixing case.

In the next sections, we present some empirical results on synthetic and real data that corroborate the ideas presented above.

3 Simulation

Let $G^* = (V, E)$ be the graph with $V = \{1, 2, 3, 4, 5\}$ and $E \subset V \times V$ corresponding to the graph shown in Figure 1(a). Let $X^{(i)} = (X_1^{(i)}, \ldots, X_5^{(i)}) \in \{0, 1\}^5$ be a stochastic process with stationary distribution factorizing as

$$\pi(X^{(i)}) = \pi(X_3^{(i)})\pi(X_1^{(i)}|X_3^{(i)})\pi(X_2^{(i)}|X_1^{(i)}, X_3^{(i)})\pi(X_4^{(i)}|X_3^{(i)})\pi(X_5^{(i)}|X_3^{(i)}). \tag{4}$$

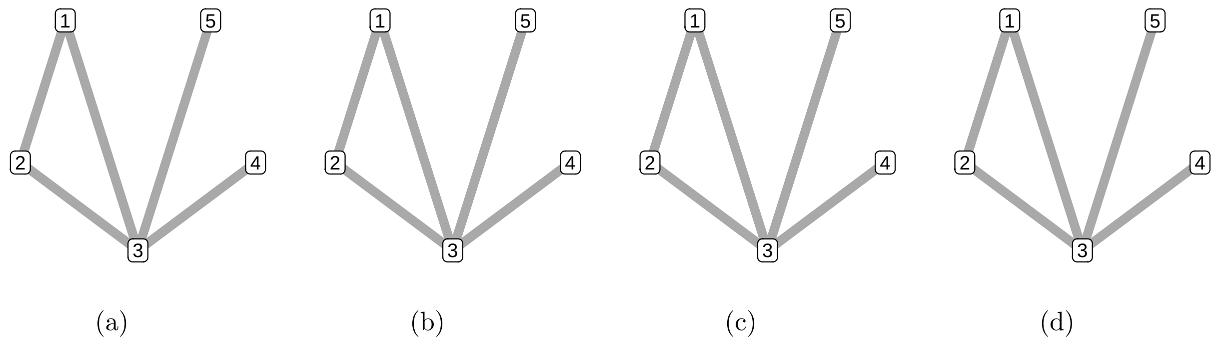


Figure 1: (a) target graph, the remaining are estimated graphs with different penalizing constant λ, namely (b) $\lambda = 0.1$, (c) $\lambda = 0.5$, and (d) $\lambda = 1$.

Based on the distribution (4) and the full conditionals $\pi(X_i^{(i)}|X_{-i}^{(i)})$ for each node $i = 1, 2, 3, 4, 5$ we used a Gibbs sampler algorithm (Geman and Geman, 1984) to obtain a sample of size 10,000 of the process $X^{(1)}, \ldots, X^{(n)}$, with the first 5,000 being discarded, to yield a final sample of size 5,000. To estimate the graph of conditional dependencies defined by (2), implemented in the R package `mrfse` (Carvalho and Leonardi, 2020), considering different values for the penalization term, $\lambda = 0.1, 0.5, 1$. The estimated graphs with these different penalizing constants can be seen in Figure 1 (b), (c), and (d), respectively. Note that in all the scenarios, the algorithm correctly estimated G^*.

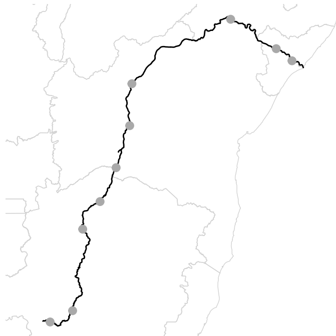
4 São Francisco river data

We applied the estimator proposed in this work to the volumetric discharge dataset in the São Francisco River in Brazil, previously analyzed in Leonardi et al. (2021). We considered measurements at 10 gauges located along the course of the river (they are numbered according to the order in which they appear on the river), see Figure 2 (a). The Brazilian National Water Agency (see Sistema Nacional de Informações sobre Recursos Hídricos, 2019) provides the data and it is publicly available. We considered 360 monthly averaged data registered between January 1977 and January 2016 in these 10 stations, and this data was discretized into 5 different levels, the size of the alphabet considered. This measurements form the $X^{(1)}, \ldots, X^{(n)}$ multivariate stochastic process described above. This study aims to estimate the graph of conditional dependencies among the stream gauges and compare it with the results in Leonardi et al. (2021), where only consecutive independent blocks were considered. By the natural spatial configuration of the river, the blocked structure seems adequate for this problem.

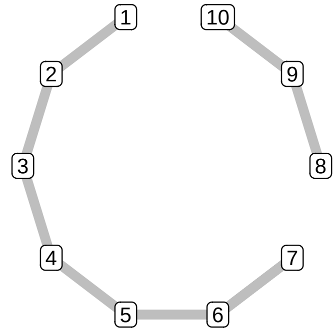
The resulting graph of conditional probability distributions is shown in Figure 2 (b). It consists of two disconnected sub-graphs with nodes $\{1, 2, 3, 4, 5, 6, 7\}$ and $\{8, 9, 10\}$, respectively, with each node being linked with the previous and subsequent stations in each component, as expected for a dataset with a Markovian spatial dependence. As discussed with more detail in Leonardi et al. (2021), between stations 7 and 8 is located the biggest hydroelectric dam along the São Francisco river, the Sobradinho dam. Therefore, the regime of measurements taken before and after Sobradinho is considerably different. This explains the break in the conditional dependencies found by the algorithm.

5 Discussion and future work

In this work, we proposed a model selection criterion to estimate the graph of conditional dependencies of a random vector on a finite alphabet, whose distribution corresponds to the stationary distribution



(a)



(b)

Figure 2: (a) The São Francisco River, in northeastern Brazil. Gray points represent the ten streamflow gauges considered in our analysis, numbered in increasing order from bottom to top; (b) Estimated graph of conditional dependencies between the gauges.

of a mixing stochastic process. We conjecture this estimator is consistent, as shown in the previous work Leonardi et al. (2019) for i.i.d data and suggested by the simulations in Section 3. The authors are working on the formal proof of this conjecture.

References

Carvalho, R. and Leonardi, F. (2020). *mrfse: Markov Random Field Structure Estimator*. R package version 0.2.

Csiszár, I. (2002). Large-scale typicality of Markov sample paths and consistency of MDL order estimators. *IEEE Trans. Inform. Theory*, 48(6):1616–1628. Special issue on Shannon theory: perspective, trends, and applications.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.

Leonardi, F., Carvalho, R. R., and Frondana, I. M. (2019). Strong structure recovery for partially observed discrete markov random fields on graphs. *arXiv preprint arXiv:1911.12198*.

Leonardi, F., Lopez-Rosenfeld, M., Rodriguez, D., Severino, M. T., and Sued, M. (2021). Independent block identification in multivariate time series. *Journal of Time Series Analysis*, 42(1):19–33.

Sistema Nacional de Informações sobre Recursos Hídricos (2019). Portal HidroWeb `http://www.snirh.gov.br/hidroweb/`. Accessed: 2019-09-03.