Structure recovery for partially observed discrete Markov random fields on graphs[1]

Florencia Leonardi

October 18, 2021

Probability Webinar - IM-UFRJ

[1] This talk is based on joint work with Iara Frondana, Rodrigo Carvalho and Magno Severino

Markov random fields on graphs

Let $\{X_v : v \in V\}$ be a family of random variables taking values on A (finite), with joint distribution $\mathbb{P}$

Markov random fields on graphs

Let $\{X_v : v \in V\}$ be a family of random variables taking values on A (finite), with joint distribution $\mathbb{P}$

V is the set of vertices or nodes and can be finite or countable infinite

Markov random fields on graphs

Let $\{X_v : v \in V\}$ be a family of random variables taking values on A (finite), with joint distribution $\mathbb{P}$

V is the set of vertices or nodes and can be finite or countable infinite

For each $\Delta \subset V$ (finite) we write

$$p(a_\Delta) = \mathbb{P}(X_\Delta = a_\Delta) \text{ with } a_\Delta \in A^\Delta$$

and if $p(a_\Delta) > 0$ we denote by

$$p(a_\Phi | a_\Delta) = \mathbb{P}(X_\Phi = a_\Phi | X_\Delta = a_\Delta) \text{ for } a_\Phi \in A^\Phi,\, a_\Delta \in A^\Delta$$

the corresponding conditional probabilities.

Markov random fields on graphs

Given $v \in V$, a neighborhood W of v is any finite set of vertices such that $v \notin W$.

Markov random fields on graphs

Given $v \in V$, a neighborhood W of v is any finite set of vertices such that $v \notin W$.

A Markov neighborhood is a neighborhood W that satisfies

$$p(a_v | a_W) \; = \; p(a_v | a_\Delta)$$

for all finite set $\Delta \supset W$, $v \notin \Delta$ and all $a_v \in A, a_\Delta \in A^\Delta$ with $p(a_\Delta) > 0$.

Markov random fields on graphs

Given $v \in V$, a neighborhood W of v is any finite set of vertices such that $v \notin W$.

A Markov neighborhood is a neighborhood W that satisfies

$$p(a_v|a_W) \;=\; p(a_v|a_\Delta)$$

for all finite set $\Delta \supset W$, $v \notin \Delta$ and all $a_v \in A, a_\Delta \in A^\Delta$ with $p(a_\Delta) > 0$.

Note! if W is a Markov neighborhood of $v \in V$, then any finite set $\Delta \supset W$ is also a Markov neighborhood of v. On the other hand, if W_1 and W_2 are Markov neighborhoods of v, it is not always true that $W_1 \cap W_2$ is a Markov neighborhood of v.

Hypotheses

(I) Positivity condition: for all finite set $W \subset V$ and all $a_W \in A^W$ we have that $p(a_W) > 0$.

(II) Markov intersection property: for all $v \in V$ and all W_1 and W_2 Markov neighborhoods of v, the set $W_1 \cap W_2$ is also a Markov neighborhood of v.

Hypotheses

(I) Positivity condition: for all finite set $W \subset V$ and all $a_W \in A^W$ we have that $p(a_W) > 0$.

(II) Markov intersection property: for all $v \in V$ and all W_1 and W_2 Markov neighborhoods of v, the set $W_1 \cap W_2$ is also a Markov neighborhood of v.

(I) $\Rightarrow$ **(II)** (Lauritzen, 1996)

Hypotheses

(I) Positivity condition: for all finite set $W \subset V$ and all $a_W \in A^W$ we have that $p(a_W) > 0$.

(II) Markov intersection property: for all $v \in V$ and all W_1 and W_2 Markov neighborhoods of v, the set $W_1 \cap W_2$ is also a Markov neighborhood of v.

(I) $\Rightarrow$ **(II)** (Lauritzen, 1996)

(II) $\not\Leftarrow$ **(I)** Markov chain with transition matrix

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}$$

Basic neighborhood

If $\mathbb{P}$ satisfies the Markov intersection property we can define the basic neighborhood of a vertex $v \in V$ as

$$\text{ne}(v) = \bigcap_{W \in \Theta(v)} W.$$

where $\Theta(v) = \{W : W \text{ is a Markov neighborhood of } v\}$.

Basic neighborhood

If $\mathbb{P}$ satisfies the Markov intersection property we can define the basic neighborhood of a vertex $v \in V$ as

$$\mathrm{ne}(v) = \bigcap_{W \in \Theta(v)} W.$$

where $\Theta(v) = \{W : W \text{ is a Markov neighborhood of } v\}$.

$\mathrm{ne}(v)$ is the smallest Markov neighborhood of $v \in V$.

Basic neighborhood

If $\mathbb{P}$ satisfies the Markov intersection property we can define the basic neighborhood of a vertex $v \in V$ as

$$\text{ne}(v) = \bigcap_{W \in \Theta(v)} W.$$

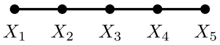
where $\Theta(v) = \{W : W \text{ is a Markov neighborhood of } v\}$.

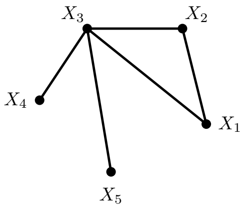
$\text{ne}(v)$ is the smallest Markov neighborhood of $v \in V$.

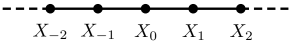
The graph of intercation $G = (V, E)$ for the measure $\mathbb{P}$ is defined by

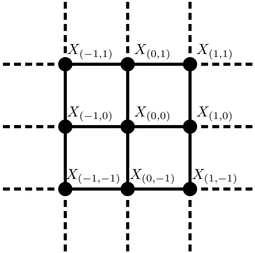
$$(v, w) \in E \text{ if and only if } w \in \text{ne}(v).$$

Examples


X_1 X_2 X_3 X_4 X_5


X_3
X_2
X_4
X_1
X_5


X_{-2} X_{-1} X_0 X_1 X_2


$X_{(-1,1)}$ $X_{(0,1)}$ $X_{(1,1)}$
$X_{(-1,0)}$ $X_{(0,0)}$ $X_{(1,0)}$
$X_{(-1,-1)}$ $X_{(0,-1)}$ $X_{(1,-1)}$

Neighborhood estimator

Let $V_n \nearrow V$ and let $\{x_{V_n}^{(i)}, i = 1, \dots, n\}$ an i.i.d (partial) sample of the distribution $\mathbb{P}$.

Neighborhood estimator

Let $V_n \nearrow V$ and let $\{x^{(i)}_{V_n}, i = 1, \ldots, n\}$ an i.i.d (partial) sample of the distribution $\mathbb{P}$.

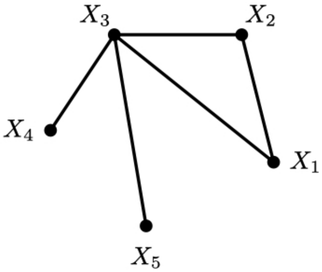
The maximum pseudo-likelihood for the vertex $v \in V$ given the neighborhood W is defined by

$$\widehat{\mathbb{P}}(x_v^{(1:n)}|x_W^{(1:n)}) = \prod_{a_W \in A^W} \prod_{a_v \in A} \hat{p}(a_v|a_W)^{N(a_v,a_W)} \,,$$

where the product is computed over all $a_W \in A^W$ with $\hat{p}(a_W) > 0$ and all $a_v \in A$ with $\hat{p}(a_v|a_W) > 0$.

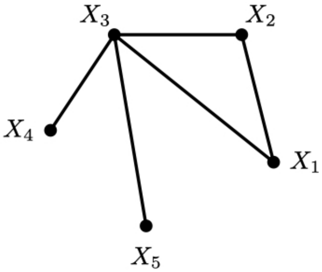
Here, $N(a_v, a_W)$ counts the number of occurrencies of the "pattern" (a_v, a_W) in the sample and $\hat{p}(a_v|a_W)$ are the usual empirical probabilities deriver from these counters.

?



X_1	X_2	X_3	X_4	X_5
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
		⋮		
x_{n1}	x_{n2}	x_{n3}	x_{n4}	x_{n5}

?

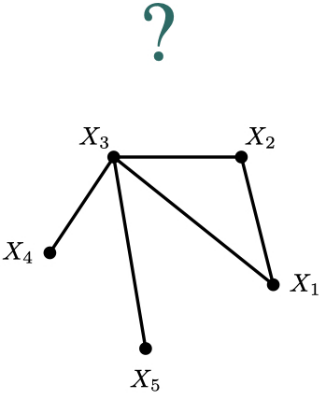


X_1	X_2	X_3	X_4	X_5
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
		$\vdots$		
x_{n1}	x_{n2}	x_{n3}	x_{n4}	x_{n5}

$$v = 3\,, \quad W = \{4{,}5\}$$

$$a_v = 0\,, \quad a_W = \{0{,}1\}$$

$$N(a_v, a_W) = \sum_{i=1}^{n} \mathbf{1}\{x_{i3} = 0, x_{i4} = 0, x_{i5} = 1\}$$



X_1	X_2	X_3	X_4	X_5
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
		$\vdots$		
x_{n1}	x_{n2}	x_{n3}	x_{n4}	x_{n5}

$$v = 3\,, \quad W = \{4,5\}$$

$$a_v = 0\,, \quad a_W = \{0,1\}$$

$$\hat{p}(0_3 \,|\, 0_4, 1_5) = \frac{N(0_3, 0_4, 1_5)}{N(0_4, 1_5)}$$

Neighborhood estimator

Given the sample $\{x_{V_n}^{(i)}, i = 1, \dots, n\}$ and a constant $c > 0$, the empirical neighborhood of $v \in V_n$ is the set of vertices $\widehat{\mathsf{ne}}(v)$ defined by

$$\widehat{\mathsf{ne}}(v) \;=\; \underset{W \subset V_n \setminus \{v\}}{\arg\max} \left\{ \log \hat{\mathbb{P}}(x_v^{(1:n)} | x_W^{(1:n)}) - c\,|A|^{|W|} \log n \right\}.$$

Consistency of the neighborhood estimator

Theorem (L, Carvalho & Frondana, 2021)

Let $v \in V_n$ for all $n \geq n_0$ and assume $|V_n| = o(\log n)$. Then for all $c > 0$, the neighborhood estimator satisfies $\widehat{\text{ne}}(v) = \text{ne}(v)$ with probability converging to 1 when $n \to \infty$. Moreover, if $c > |A|^2[p_{\min}(|A|-1)]^{-1}$ then $\widehat{\text{ne}}(v) = \text{ne}(v)$ eventually almost surely as $n \to \infty$.

Consistency of the neighborhood estimator

Theorem (L, Carvalho & Frondana, 2021)

Let $v \in V_n$ for all $n \geq n_0$ and assume $|V_n| = o(\log n)$. Then for all $c > 0$, the neighborhood estimator satisfies $\widehat{\text{ne}}(v) = \text{ne}(v)$ with probability converging to 1 when $n \to \infty$. Moreover, if $c > |A|^2[p_{\min}(|A|-1)]^{-1}$ then $\widehat{\text{ne}}(v) = \text{ne}(v)$ eventually almost surely as $n \to \infty$.

$$p_{\min}(v) = \min_{W \subseteq \text{ne}(v)} \min_{a_v, a_W} \{\, p(a_v|a_W) \colon p(a_v|a_W) > 0 \,\}$$

Graph estimator

Given $V' \subset V$ finite, we define $G_{V'}$ as the sub-graph induced by G; i.e the pair (V', E') where

$$E' = \{(v, w) \in E \colon v, w \in V'\}$$

Graph estimator

Given $V' \subset V$ finite, we define $G_{V'}$ as the sub-graph induced by G; i.e the pair (V', E') where

$$E' = \{(v, w) \in E \colon v, w \in V'\}$$

To estimate G' based on $\widehat{\text{ne}}(v)$, we have two possibilities:

$$\widehat{E}'_{\wedge} = \{(v, w) \in V' \times V' \colon v \in \widehat{\text{ne}}(w) \text{ and } w \in \widehat{\text{ne}}(v)\} \quad \rightarrow \quad \text{conservative}$$

Graph estimator

Given $V' \subset V$ finite, we define $G_{V'}$ as the sub-graph induced by G; i.e the pair (V', E') where

$$E' = \{(v, w) \in E \colon v, w \in V'\}$$

To estimate G' based on $\widehat{\text{ne}}(v)$, we have two possibilities:

$$\widehat{E}'_{\wedge} = \{(v, w) \in V' \times V' \colon v \in \widehat{\text{ne}}(w) \text{ and } w \in \widehat{\text{ne}}(v)\} \quad \rightarrow \quad \text{conservative}$$

$$\widehat{E}'_{\vee} = \{(v, w) \in V' \times V' \colon v \in \widehat{\text{ne}}(w) \text{ or } w \in \widehat{\text{ne}}(v)\} \quad \rightarrow \quad \text{non-conservative}$$

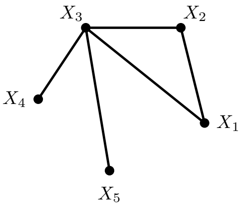
Consistency of the sub-graph estimator

Corollary (L, Carvalho & Frondana, 2021)

Let $G' = (V', E')$ be the sub-graph induced by G with a finite set of vertices V'. Then for all $c > 0$, if $|V_n| = o(\log n)$ we have that $\widehat{E}'_{\wedge} = \widehat{E}'_{\vee} = E'$ with probability converging to 1 when $n \to \infty$. If $c > |A|^2[p_{\min}(V')(|A|-1)]^{-1}$ then $\widehat{E}'_{\wedge} = \widehat{E}'_{\vee} = E'$ eventually almost surely as $n \to \infty$.

Simulations

Graph of the generated distribution



$$A = \{0, 1, 2\}$$

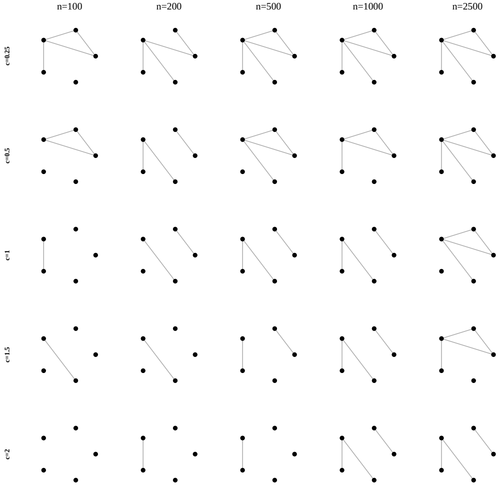
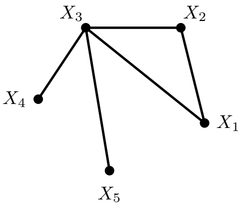


Figura: Estimated graph with the conservative approach

Simulations

Graph of the generated distribution



$$A = \{0, 1, 2\}$$

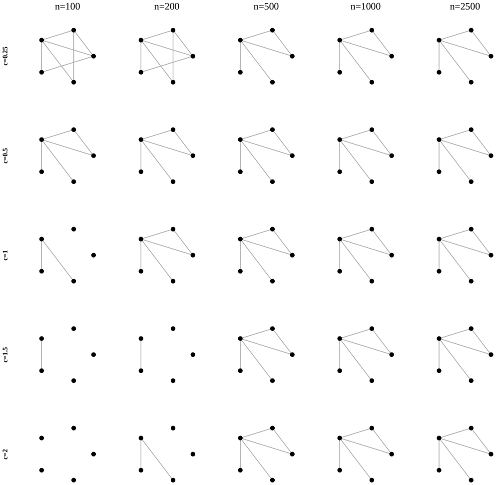


Figura: Estimated graph with the non-conservative approach

Simulations

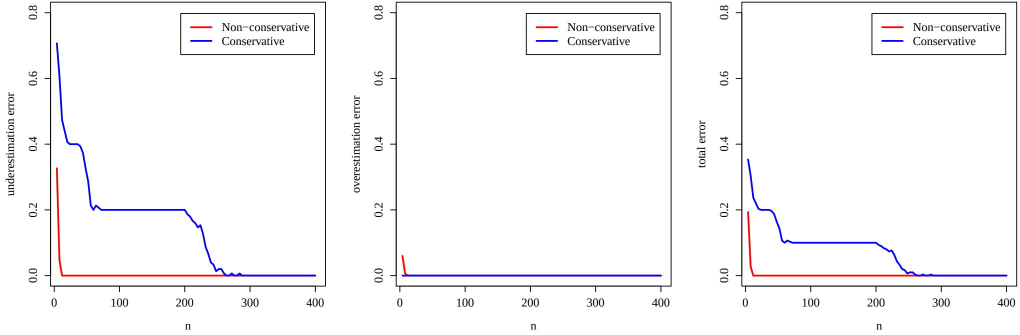


Figura: Mean erros of overestimation, underestimation and total error computed on 30 runs of the simulations, for the conservative and non-conservative estimators

World wide stock indexes

- To illustrate the application of the graph estimator we analyzed stock indexed from different countries extracted from `https://br.investing.com/indices/world-indices`
- The sample consists of $n = 530$ observation in time, where each entry corresponds to the indicator function of a positive change from the previous day, for 15 countries
- In order to reduce dependency, we considered an interval of 4 days between observations

World wide stock indexes

Figura: Grafo estimado para os dados dos índices de acões de diferentes países, com o estimador “não conservador”

World wide stock indexes

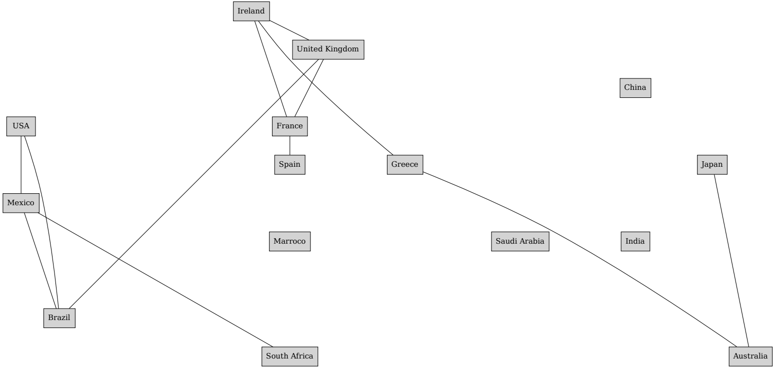
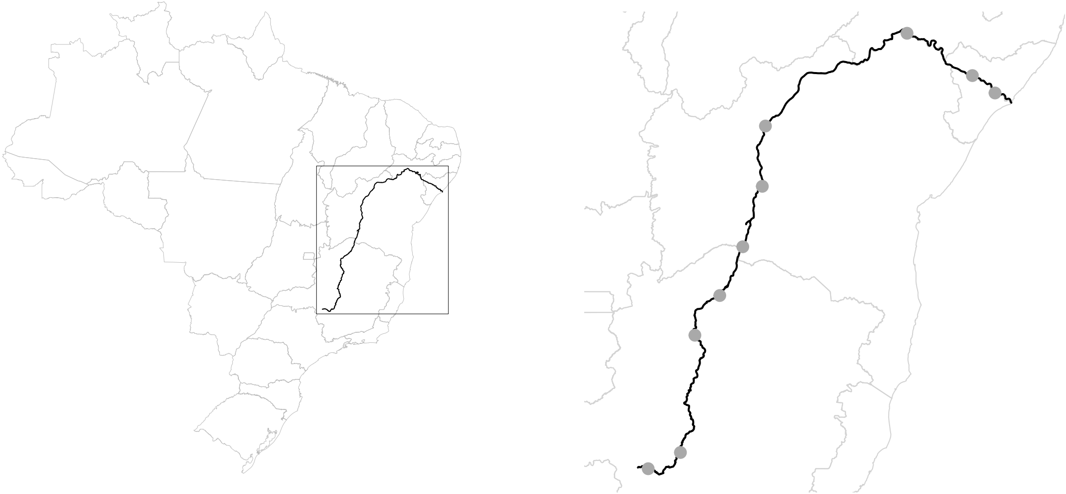


Figura: Grafo estimado para os dados dos índices de acões de diferentes países, com o estimador "conservador"

São Francisco River



São Francisco River

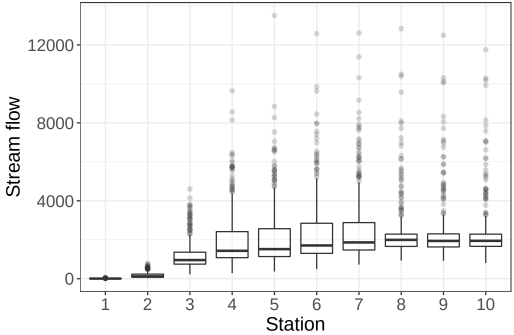
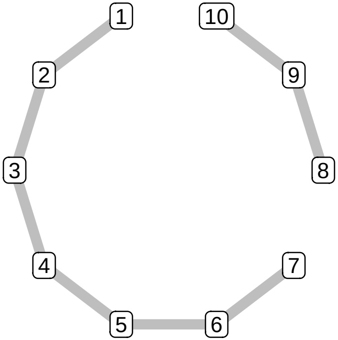


Figure from Leonardi et al. (2021). *Independent block identification in multivariate time series*, Journal of Time Series Analysis, 2021.



Estimated graph

Discussion and future work

- The hypothesis of independence in the observations is strong, but the estimator seems to be consistent also on dependent data

Discussion and future work

- The hypothesis of independence in the observations is strong, but the estimator seems to be consistent also on dependent data
- We are working on consistency for *mixing* processes (Magno T.F. Severino' PhD project)

Discussion and future work

- The hypothesis of independence in the observations is strong, but the estimator seems to be consistent also on dependent data
- We are working on consistency for *mixing* processes (Magno T.F. Severino' PhD project)
- Computing the estimator is a problem... the exhaustive search is prohibited even for medium size graphs... So we are looking also for efficient algorithms

Discussion and future work

- The hypothesis of independence in the observations is strong, but the estimator seems to be consistent also on dependent data
- We are working on consistency for *mixing* processes (Magno T.F. Severino' PhD project)
- Computing the estimator is a problem... the exhaustive search is prohibited even for medium size graphs... So we are looking also for efficient algorithms
- An R package, called `mrfse` , was developed with the implementation of the algorithm by Rodrigo R.S. Carvalho, who is also working now on other algorithmic approaches

Thank you!

florencia@ime.usp.br

www.ime.usp.br/~leonardi

This work has the support of Brazilian agencies FAPESP and CNPq.