

Sumário

1	Introdução	2
2	Granger-Causalidade	2
3	LiNGAM para séries temporais	4
4	Causalidade TiMINo	6
5	Experimento: Influência de Confundidores	8
6	Conclusão	11
A	Códigos	12
	Referências	14

1 Introdução

Inferência causal foi sem dúvida um dos assuntos de grande importância do curso. Vimos como diferentes estruturas causais intrínsecas do processo gerador de dados podem impactar no resultado de modelos estatísticos, podendo inclusive gerar falsas conclusões.

Os DAG's (*Directed Acyclical Graphs*) são introduzidos de forma a esquematizar tais estruturas causais, e a partir deles decidimos quais variáveis incluir ou não em nossos modelos para que a saída represente de fato o efeito causal de interesse.

Este trabalho é um estudo sobre inferência causal no contexto de séries temporais. A ideia principal é a partir de dados temporais, conseguir recuperar o DAG que represente a estrutura causal do problema. Nos basearemos no artigo[5], que apresenta uma classe específica de modelos de equações estruturais (SEM¹) que os autores batizam de TiMINo (*Time Series Models with Independent Noise*). Os autores apresentam um algoritmo para recuperar a estrutura do grafo baseado no modelo desenvolvido, e compararam a capacidade de recuperação diversos outros modelos, entre eles Granger-causalidade[6] e LiNGAM (*Linear Non-Gaussian Acyclic Model*)[2][1]. O objetivo deste texto é apresentar com um pouco mais detalhadamente esses dois últimos métodos, e fazer um experimento simulado no final testando a influência de variáveis confundidoras nos algoritmos. Vamos agora introduzir a notação e conceitos iniciais deste texto.

Seja V um conjunto finito de índices de tamanho v e $i \in V$; nossos objetos de estudos serão séries temporais $(X_t^i)_{t \in \mathbb{N}}$. Denotaremos \mathbf{X}_t como o vetor que armazena as realizações das v séries temporais no tempo t . O **grafo temporal completo** é o grafo que contém cada observação de cada série como um nó, isto é, cada X_t^i será um nó.

Em um contexto com muitas séries e/ou muitas observações em cada uma delas, esse tipo de visualização fica inviável. Com isso definimos o **grafo temporal reduzido** que contém v nós representando as séries $(X_t^i)_{t \in \mathbb{N}}$. Um aresta direcionada de X^i para X^j ($i \neq j$) é criada caso, para algum k exista uma aresta de X_{t-k}^i para X_t^j . Tais arestas direcionadas representariam causalidade.

O objetivo dos métodos que apresentaremos é de recuperar o grafo temporal reduzido a partir de observações $(\mathbf{X}_1, \dots, \mathbf{X}_T)$ de uma série temporal multivariada.

2 Granger-Causalidade

Vamos introduzir agora o conceito de **Granger-Causalidade** (ou **G-causalidade**). Sejam X^i e X^j duas séries temporais. Denote por $\mathcal{U}_t = (U_{t-1}, U_{t-2}, \dots)$ toda a informação do universo até o tempo t , e $X_{<t}^i = (X_{t-1}^i, X_{t-2}^i, \dots)$ toda a informação de X^i até o tempo t .

Seja $\sigma^2(X_t^j | \mathcal{U}_t)$ a variância dos resíduos na predição de X_t^j usando \mathcal{U}_t . Similarmente, denotemos por $\sigma^2(X_t^j | \mathcal{U}_t \setminus X_{<t}^i)$ a variância dos resíduos na predição de X_t^j usando toda a informação

¹Structural Equation Models

\mathcal{U}_t exceto $X_{<t}^i$.

Definição 1. (*G-causalidade*) Se $\sigma^2(X_t^j|\mathcal{U}_t) < \sigma^2(X_t^j|\mathcal{U}_t \setminus X_{<t}^i)$ dizemos que X^i **Granger-causa** (ou **G-causa**) X^j . Em outras palavras, X^i G-causa X^j se, dada todas as informações do universo, a remoção do passado de X^i piora a predição de X^j .

Na prática, é impossível ter acesso ao conjunto \mathcal{U}_t , então substituímos ele por $X_{<t}^V = (X_{<t}^1, X_{<t}^2, \dots, X_{<t}^v)$, o conjunto de todas as séries disponíveis no estudo até o tempo t (exclusivo). Formalmente, dizemos que X^i G-causa X^j com respeito a $X_{<t}^V$, porém omitiremos tal sufixo neste texto e quando nos referirmos a Granger-causalidade $X^i \rightarrow X^j$ estará implícito que nosso universo é X^V . Perceba que estamos incluindo sempre o passado das variáveis causadoras, excluindo o próprio t do sistema de causas, porém é possível fazer essa inclusão, o que gera a seguinte definição:

Definição 2. (*Causalidade instantânea*) Dizemos que duas séries X^i e X^j possuem causalidade instantânea, se para algum $t \in \mathbb{N}$, incluir X_t^i no conjunto de informações preditoras, ajuda na predição de X_t^j . Usando a notação anterior, isto quer dizer que:

$$\sigma^2(X^j|\mathcal{U}_t \cup \{X_t^i\}) < \sigma^2(X^j|\mathcal{U}_t)$$

Note que quando definimos \mathcal{U}_t excluimos as observações do universo no próprio tempo t , logo a definição é consistente. E novamente, em experimentos práticos \mathcal{U}_t será substituído por $X_{<t}^V$.

Vetor Autorregressivo e G-causalidade linear

O modelo de Vetor Autorregressivo é um dos mais conhecidos em séries temporais e é abreviado por VAR (*Vector Autoregression*). Se baseia na hipótese de linearidade entre \mathbf{X}_t e as realizações passadas das séries $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots, \mathbf{X}_{t-p}$, onde p é o lag máximo considerado. A G-causalidade linear, toma então como base o modelo VAR:

$$\mathbf{X}_t = \sum_{\tau=1}^p \mathbf{A}(\tau) \mathbf{X}_{t-\tau} + \mathbf{N}_t, \quad (1)$$

Onde $\mathbf{X}_t, \mathbf{N}_t$ são vetores colunas representando as séries e o ruído (*noise*) respectivamente, e $\mathbf{A}(\tau)$ é uma matriz quadrada de coeficientes com o mesmo número de linhas do vetor das séries. Algumas condições sobre o termo de ruído são necessárias como hipótese de modelo como média 0 e descorrelação entre \mathbf{N}_t e \mathbf{N}_{t-k} para todo k inteiro positivo não nulo. A estimação das matrizes de coeficientes se dá por minimização quadrática multivariada.

Para testar se X^i G-causa X^j nesse contexto, estima-se um modelo VAR completo usando todas as séries, e depois um modelo restrito que elimina X^i da predição de X_t^i dentro de \mathbf{X}_t isso pode ser feito forçando a i -ésima coluna de $\mathbf{A}(\tau) = 0$ para todo τ tal que $1 \leq \tau \leq p$, com isso, as multiplicações $\mathbf{A}(\tau) \mathbf{X}_{t-\tau}$ anularão a i -ésima entrada. Sendo RSS_{full} a soma quadrática dos resíduos do modelo completo, e RSS_{restr} a mesma medida para o modelo

restrito sem X^i , utiliza-se a estatística abaixo para testar se a redução nos resíduos foi significativa.

$$\frac{(RSS_{restr} - RSS_{full})/(p_{full} - p_{restr})}{RSS_{full}/(T - p_{full})},$$

onde T é o tamanho das séries da amostra, p_{full} o número de parâmetros do modelo completo e p_{restr} do modelo restrito. Tal estatística segue distribuição $F(p_{full} - p_{restr}, T - p_{full})$, daí o teste depende do estabelecimento de um nível de significância para se comparar o quantil teórico da distribuição F com a estatística apresentada acima, nos testes utilizaremos $\alpha = 0.05$.

Com isso é possível codificar um algoritmo que recupera a estrutura do DAG a partir da amostra de séries temporais. O autor do artigo[5] disponibiliza seus scripts em R usados nos experimentos, No Apêndice A apresentamos duas das funções que usaremos aqui. A função `granger_partial` testa se uma série G-cause outra, dada z nosso universo. Já a função `granger_dag_partial` recebe uma matriz de realizações de séries e aplica iterativamente a função `granger_partial` em cada par possível para retornar o DAG final. A quantidade de lags (p) para o modelo VAR é escolhida via critério de informação AIC dado o máximo pelo parâmetro `max_lag`.

Nosso artigo de referência apresenta ainda, extensões não-lineares que utilizam modelos mais complexos de séries temporais. Por aqui, trabalharemos apenas com o caso linear por simplicidade.

3 LiNGAM para séries temporais

O modelo LiNGAM (*Linear Non-Gaussian Acyclic Model*) foi introduzido em [1] para variáveis gerais, fora do contexto de séries temporais. Considere que variáveis observadas x_i , $i \in \{1, \dots, m\}$ possam ser reordenadas em uma ordem causal, isto é, que nenhum variável cause outra variável que a anteceda. Dessa forma é possível representar a estrutura causal com um DAG. Note que a possibilidade de tal ordenação exclui a possibilidade de presença de variáveis confundidoras. Seja então $k(i)$ a ordem causal da variável x_i . A modelagem proposta assume que x_i é função linear das variáveis que a antecedem na ordem causal mais um erro e_i .

$$x_i = \sum_{k(j) < k(i)} b_{ij}x_j + e_i + c_i,$$

onde c_i é uma constante, e os erros e_i 's são variáveis aleatórias independentes entre si, com distribuição não-gaussiana e variância não-nula. Tais propriedades inspiram o nome do modelo, que é linear, acíclico e não-gaussiano.

As referências [2][4] fazem as adaptações desse modelo no caso de séries temporais. Considerando nosso conjunto observacional de v séries temporais X^1, X^2, \dots, X^v cada uma delas com tamanho T . Seja $\mathbf{A}(\tau)$ a matriz $v \times v$ dos efeitos causais entre as séries e o τ -ésimo lag, novamente restringindo $1 \leq \tau \leq p$. Sendo \mathbf{X}_t o vetor coluna de observações das v séries no

tempo t , a estrutura causal seguirá o modelo:

$$\mathbf{X}_t = \sum_{\tau=0}^p \mathbf{A}(\tau) \mathbf{X}_{t-\tau} + \mathbf{N}_t \quad (2)$$

Tau modelo difere de 1 pela inclusão de $\tau = 0$ na soma, permitindo a modelagem de causalidade instantânea segundo a Definição 2. A matriz $\mathbf{A}(0)$ será a matriz de efeitos instantâneos. Os ruídos \mathbf{N}_t além de serem mutualmente independentes e temporalmente descorrelacionados, agora tem distribuição não-gaussiana, premissa que distingue tal modelo dos convencionais. Além disso, supondo que as entradas do vetor \mathbf{X}_t estejam ordenadas por ordem causal, as matrizes $\mathbf{A}(\tau)$ serão triangulares inferiores. Por exemplo, considere o caso de três séries X^1, X^2, X^3 onde a ordem causal é justamente $X^1 \rightarrow X^2 \rightarrow X^3$, sem efeitos instantâneos ($\mathbf{A}(0) \equiv 0$) e $p = 1$; Considere então uma estrutura temporal na qual $X_{t-1}^1 \rightarrow X_t^2$ e $X_{t-1}^2 \rightarrow X_t^3$ e $X_{t-1}^i \rightarrow X_t^i$ para $i = 1, 2, 3$; teremos então um modelo no formato

$$\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = \mathbf{A}(1) \begin{bmatrix} X_{t-1}^1 \\ X_{t-1}^2 \\ X_{t-1}^3 \end{bmatrix} + \mathbf{N}_t,$$

onde $\mathbf{A}(1)$ é do tipo:

$$\mathbf{A}(1) = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

A multiplicação nos dariam equações lineares conforme a modelagem proposta. O coeficiente a_{ij} representa o efeito causal da série j na série i no tempo t considerado.

A existência de uma ordem causal garante então que se a ordem das entradas de \mathbf{X}_t for aleatória, existe uma matriz de permutação \mathbf{P} tal que $\mathbf{P}\mathbf{A}(\tau)\mathbf{P}^T$ seja triangular inferior. Sendo assim é possível recuperar a ordem causal via permutação das matrizes de coeficientes, se todas as hipóteses do modelo são satisfeitas. O autor de [4] disponibiliza seus códigos em R online ² e utiliza um algoritmo de força bruta neste caso, para obter uma permutação ótima e retornar a ordem causal.

Vamos agora apresentar o método de estimação das matrizes de coeficientes:

Passo 1: Estimar um modelo VAR (sem efeitos instantâneos) como abaixo utilizando implementações padrões de otimização quadrática:

$$\mathbf{X}_t = \sum_{\tau=1}^p \mathbf{M}(\tau) \mathbf{X}_{t-\tau} + \mathbf{n}_t \quad (3)$$

Denotemos por $\hat{\mathbf{M}}(\tau)$ as estimativas de matrizes obtidas.

²<https://sites.google.com/site/dorisentner/publications/VARLiNGAM>

Passo 2: Estime os resíduos:

$$\hat{\mathbf{n}}_t = \mathbf{X}_t - \sum_{\tau=1}^p \hat{\mathbf{M}}(\tau) \mathbf{X}_{t-\tau}$$

Passo 3: Estime $\mathbf{A}(0)$ com a solução do modelo de causalidade instantânea usando modelo LiNGAM:

$$\hat{\mathbf{n}}_t = \mathbf{A}(0)\hat{\mathbf{n}}_t + \hat{\mathbf{N}}_t$$

Passo 4: Estimacão de $\mathbf{A}(\tau), \tau > 0$. De 2, isolando \mathbf{X}_t no primeiro membro, temos:

$$(\mathbf{I} - \mathbf{A}(0))\mathbf{X}_t = \sum_{\tau=1}^p \mathbf{A}(\tau)\mathbf{X}_{t-\tau} + \mathbf{N}_t,$$

com isso, multiplicando por $(\mathbf{I} - \mathbf{A}(0))^{-1}$ temos:

$$\mathbf{X}_t = \sum_{\tau=1}^p (\mathbf{I} - \mathbf{A}(0))^{-1} \mathbf{A}(\tau) \mathbf{X}_{t-\tau} + (\mathbf{I} - \mathbf{A}(0))^{-1} \mathbf{N}_t$$

O passo 3 se justifica comparando $\hat{\mathbf{n}}_t$ de (3) com o termo de ruído $(\mathbf{I} - \mathbf{A}(0))^{-1} \mathbf{N}_t$

O passo 4 consiste em estimar $\hat{\mathbf{A}}(\tau), \tau > 0$ com a fórmula abaixo, que se justifica comparando (3) com os termos do somatório acima.

$$\hat{\mathbf{A}}(\tau) = (\mathbf{I} - \hat{\mathbf{A}}(0))\hat{\mathbf{M}}(\tau)$$

Todo este processo está codificado nos scripts disponibilizados por [4].

A interpretação de causalidade no caso deste modelo estende a noção de Granger, e inclui agora os efeitos instantâneos em $\mathbf{A}(0)$. Diremos que X^i causa X^j se pelo menos um dos coeficientes de $\mathbf{A}(\tau)$ na posição (j, i) é significantemente diferente de zero para algum $\tau \geq 0$.

4 Causalidade TiMINo

Definição 3. [5] (*TiMINo - Time Series model with Independent Noise*) Sejam as séries temporais $\mathbf{X}_t = (X_t^i)_{i \in V}$, dizemos que elas satisfazem um Modelo Temporal com Ruído Independente (TiMINo) se existe $p > 0$ e $\forall i \in V$ existem conjuntos $\mathbf{PA}_0^i \subseteq X^{V \setminus \{i\}}, \mathbf{PA}_k^i \subseteq X^V$ tais que, $\forall t$

$$X_t^i = f_i((\mathbf{PA}_p^i)_{t-p}, \dots, (\mathbf{PA}_1^i)_{t-1}, (\mathbf{PA}_0^i)_t, N_t^i), \quad (4)$$

com N_t^i conjuntamente independentes sobre i e t para todo i , e identicamente distribuídos em t .

A notação \mathbf{PA}^i vem do livro [3] e quer dizer *parents*, ou seja, nós que causam i diretamente na estrutura do grafo. No caso \mathbf{PA}_0^i é um conjunto de séries a serem tomadas de $X^{V \setminus \{i\}}$ como candidatas a efeitos causais. Dizer que $X_t^i = f_i((\mathbf{PA}_0^i)_t)$ é dizer que no tempo t as séries de \mathbf{PA}_0^i são causa instantânea de X^i . Os conjuntos \mathbf{PA}_k^i são tomados sobre todo o espaço de séries como candidatos a efeitos com *lags*.

Em resumo o que a expressão (4) diz é que em cada t existem registros passados e possivelmente efeitos instantâneos que causam X^i como função dessas causas mais um ruído independente.

O **grafo temporal completo** consiste em setas partindo das estradas de f_i e chegando em X_t^i . O Teorema 1 de [5] discute sob que condições o grafo temporal reduzido é acíclico. Além disso o artigo propõe um algoritmo que recebe amostras de séries temporais e devolve o DAG estimado correspondente ao grafo temporal reduzido. A ideia é eleger uma classe de modelos f_i e tentar ajustar um Modelo Temporal com Ruído Independente (TiMINo) nos dados; caso tal independência de ruídos não seja alcançável, o algoritmo não retorna nada, evitando conclusões incorretas. Segue abaixo o pseudocódigo:

Entradas:

- $(\mathbf{X}_1, \dots, \mathbf{X}_T)$: Amostras de uma série temporal v -dimensional de tamanho T cada;
- p : ordem máxima dos *lags*;

Defina $S := (1, \dots, v)$

Enquanto $\#S \neq 1$ repita:

—— **Para** cada k em S faça:

———— Estime um modelo TiMINo para X_t^k usando $X_{t-p}^k, \dots, X_{t-1}^k$ e $X_{t-p}^i, \dots, X_{t-1}^i$ para $i \in S \setminus \{k\}$

———— Teste se os resíduos são independentes de $X^i, i \in S$

—— Escolha k^* como o k com a menor medida de dependência. Caso não haja k com independência, pare o algoritmo.

—— Redefina $S := S \setminus \{k^*\}$ e faça $pa(k^*) := S$

Para cada k remova todos os nós pais desnecessários para obter resíduos independentes.

Saída: $(pa(1), \dots, pa(v))$

Para o ajuste do TiMINo devemos prover ao algoritmo um modelo de séries temporais, neste texto vamos nos restringir ao VAR. Para testar a independência entre os resíduos e as séries os autores propõem o HSIC (*Hilbert-Schmidt independence criterion*) e um método de correlação cruzada, que não detalharemos aqui. Na prática também passaremos como entrada uma significância α para esses testes.

A codificação do algoritmo em R está disponível no site de um dos autores³.

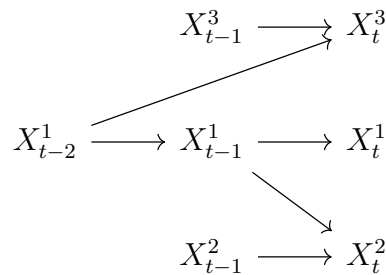
³<http://web.math.ku.dk/peters/code.html>

5 Experimento: Influência de Confundidores

Para testar os algoritmos apresentados, geramos o seguinte esquema temporal:

$$\begin{aligned}X_t^1 &= 0.3X_{t-1}^1 + N_t^1 \\X_t^2 &= 0.8X_{t-1}^2 + 0.4X_{t-1}^1 + N_t^2 \\X_t^3 &= -0.6X_{t-1}^3 + 0.7X_{t-2}^1 + N_t^3,\end{aligned}$$

onde $N^i \sim \mathcal{N}_T(0, \mathbf{I})$, $i = 1, 2, 3$, ou seja o vetor de ruídos segue uma normal padrão multivariada independente. Graficamente podemos representar o grafo temporal completo (cortado em t) da seguinte forma:



o grafo temporal reduzido seria simplesmente $X^2 \leftarrow X^1 \rightarrow X^3$

Abaixo segue o código em R de geração dos dados:

```

1 set.seed(1)
2 T <- 1000
3 X1 <- rep(0, T)
4 X2 <- rep(0, T)
5 X3 <- rep(0, T)
6
7 N1 <- rnorm(T)
8 N2 <- rnorm(T)
9 N3 <- rnorm(T)
10
11 for(t in 3:n)
12 {
13   X1[t] <- 0.3*X1[t-1]+N1[t]
14   X2[t] <- 0.8*X2[t-1]+0.4*X1[t-1]+N2[t]
15   X3[t] <- -0.6*X3[t-1]+0.7*x1[t-2]+N3[t]
16 }
```

A matriz de adjacência do grafo temporal reduzido segue abaixo, se $(i, j) = 1$, X^i causa X^j no sentido apresentado na introdução.

$$\text{DAG Real} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (5)$$

Aplicamos então os algoritmos de G-causalidade, VAR-LiNGAM e TiMINo na amostra temporal $(\mathbf{X}_t)_{t=1,\dots,T}$, com $\mathbf{X}_t = (X_t^1, X_t^2, X_t^3)$, na tentativa de recuperar a matriz de adjacência acima. O modelo LiNGAM não retorna diretamente o DAG, mas sim as matrizes de coeficientes autorregressivos $\mathbf{A}(\tau)$ e a ordem causal obtida a partir da permutação ótima. Note que ordens causais válidas neste caso seriam X^1, X^2, X^3 ou ainda X^1, X^3, X^2 .

Como não temos efeitos instantâneos, a matriz $\mathbf{A}(0)$ é identicamente nula, e no caso de $\mathbf{A}(1)$ e $\mathbf{A}(2)$ teremos:

$$\mathbf{A}(1) = \begin{bmatrix} 0.3 & 0 & 0 \\ 0.4 & 0.8 & 0 \\ 0 & 0 & -0.6 \end{bmatrix}, \mathbf{A}(2) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.7 & 0 & 0 \end{bmatrix}$$

Abaixo segue a aplicação dos algoritmos utilizando as as funções disponibilizadas pelos autores.

```

1 granger_dag <- granger_dag_partial(cbind(X1,X2,X3), alpha = 0.05, max_lag = 2,
  output = FALSE)
2
3 X_can <- tsdata2canonicalform(cbind(X1,X2,X3), 2)
4 lingam_model <- VARLiNGAM(X_can, "ols", pruning=FALSE, ntests=FALSE)
5 timino_dag <- timino_dag(cbind(x1,x2,x3), alpha = 0.05, max_lag = 2, model =
  traints_linear, indtest = indtestts_hsic, output = TRUE)
6
7 A0 <- lingam_model$Bhat[[1]]
8 A1 <- lingam_model$Bhat[[2]]
9 A2 <- lingam_model$Bhat[[3]]
10 causal_order <- lingam_model$var_order

```

Os objetos *granger_dag* e *timino_dag* são diretamente as matrizes de adjacência. Em ambos os casos a matriz correta (5) foi retornada. No caso do Modelo LiNGAM, a saída *lingam_model* contém vários objetos como as matrizes de coeficientes e ordem causal. As matrizes estimadas foram próximas das reais, apesar de os ruídos serem gaussianos quebrando uma das hipóteses deste modelo, o algoritmo exibe um aviso em relação à isso alertando que

as matrizes não são aproximadamente triangulares⁴:

$$\hat{A}(0) \approx \begin{bmatrix} 0 & 0 & 0 \\ 0.009 & 0 & 0.034 \\ 0.044 & 0 & 0 \end{bmatrix}, \hat{A}(1) \approx \begin{bmatrix} 0.258 & -0.011 & 0.012 \\ 0.404 & 0.801 & 0.008 \\ -0.039 & -0.023 & -0.537 \end{bmatrix},$$

$$\hat{A}(2) \approx \begin{bmatrix} -0.028 & -0.004 & 0.005 \\ 0.039 & 0.027 & -0.049 \\ 0.667 & -0.013 & 0.049 \end{bmatrix}$$

Veja que $\hat{A}(0)$ é aproximadamente nula, e $\hat{A}(1), \hat{A}(2)$ são boas aproximações das matrizes reais dos coeficientes.

o objeto `causal_order` retornou um vetor 1, 2, 3 que é consistente com a ordem causal válida X^1, X^2, X^3 .

Como observamos anteriormente, X^1 é uma variável confundidora. Supondo que esta série não fosse observada, repetidos os mesmos experimentos, omitindo-a do conjunto de dados. O algoritmo de Granger infere de forma errônea que $X^2 \rightarrow X^3$ e $X^3 \rightarrow X^2$, isso pois incluir uma das séries ajuda a prever a outra por conta da variável latente X^1 . O algoritmo VAR-LiNGAM não consegue recuperar os coeficientes e infere também erroneamente a ordem causal X^3, X^2 . Já o algoritmo TiMINo, infere corretamente a matriz de adjacência nula $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Abaixo segue o output gerado durante o processo de estimação. Ao dizer "fit 1 with the help of 2" por exemplo, 1 está representando a primeira série, que no caso é X^2 e 2 representa a segunda série que é X^3 . Ao tentar ajustar uma série usando outra, a hipótese de resíduos independentes foi rejeitada, mas ao ajustar as séries com seus próprios lags (with the help of NOTHING) a hipótese de ruídos independentes falha em ser rejeitada, e com isso os nós não são criados.

```

1 > timino_dag2 <- timino_dag(cbind(X2,X3), alpha = 0.05, max_lag = 2, model =
   traints_linear, indtest = indtestts_hsic, output = TRUE)
2 [1] "fit 1 with the help of 2 ..."
3 [1] "Independence rejected: p-value = 1.96550755421354e-05"
4 [1] "fit 2 with the help of 1 ..."
5 [1] "Modelling the first time series without any other leads to independent
   residuals."
6 [1] "Independence not rejected: p-value = 0.0500001"
7 [1] "Possible sink node found: 2"
8 [1] "causal order (beginning at sink): 2"
9 [1] "causal order (beginning at sink): 2 1"
10 [1] "removing unnecessary edges..."
11 [1] 1
12 [1] "...and performing final independence test using HSIC..."
13 [1] "fitting 2 with the help of NOTHING and testing independence against
   1"
```

⁴Ao trocar a distribuição dos ruídos para $\text{Uniforme}(-\frac{1}{2}, \frac{1}{2})$ o aviso deixa de aparecer

```

14 [1] "Test statistic: 0.549 and critical value: 0.742 and p-value 3.63e-01"
15 [1] "all correct..."
16 [1] "final summary time graph:"
17   [,1] [,2]
18 [1,]    0    0
19 [2,]    0    0
20 >

```

6 Conclusão

Este estudo teve como objetivo final mostrar a importância da análise causal em problemas estatísticos. No contexto de séries temporais, se ignorássemos a estrutura causal das séries e prosseguíssemos com a implementação de algoritmos padrões, poderíamos chegar em conclusões equivocadas caso a estrutura causal real do processo gerador de dados for diferente daquela que é assumida pelos modelos. A presença de variáveis confundidoras por exemplo pode afetar profundamente os processos de estimação, caso não sejam observadas.

Apresentamos assim então três métodos de recuperação de DAG's a partir de amostras temporais, cada um deles com suas hipóteses e particularidades. Vimos que o TiMINo performou bem na presença de confundidores, evitando falsas conclusões. Existem adaptações não lineares desses métodos que não exploramos aqui, mas que dariam uma flexibilidade de ajuste maior à tais métodos. O intuito foi simplesmente estudar G-causalidade, modelo LiNGAM e o TiMINo, e por isso não nos preocupamos com essas generalizações.

Como trabalho futuro seria interessante estudar as extensões não lineares desses algoritmos, e aplicá-los em dados reais com diferentes estruturas de causalidade e ruídos.

A Códigos

Algoritmo que testa G-causalidade linear de duas séries x, y com respeito ao conjunto universo (x, y, z) .

```
1 granger_partial<-function(x,y,z,alpha,max_lag,output = FALSE)
2 {
3   #tests, whether y is causing x given the information from z.
4   #mod_2 = mod_full
5   #mod_1 = mod_restr
6   #choose the order of mod_full to be the order of mod_restr
7
8   mod2_fit<-ar(cbind(x,y,z),aic=TRUE,order.max=max_lag)
9   order2<-max(mod2_fit$order,1)
10  mod_fit<-ar(cbind(x,z),aic=FALSE,order.max=order2)
11  order1<-max(mod_fit$order,1)
12
13  r1<-mod_fit$resid[-c(1:order1),1]
14  r2<-mod2_fit$resid[-c(1:order2),1]
15  RSS1<-t(r1)%*%r1
16  RSS2<-t(r2)%*%r2
17
18  dim_cond_set<-dim(z)[2]
19  if(length(z)==length(x)){dim_cond_set=1}
20
21  T_num <- (RSS1-RSS2)/order2
22  T_den <- RSS2/(length(x)-((dim_cond_set+1)*order2+1))
23  T <- T_num/T_den
24
25  quan <- qf(1-alpha,order2,length(x)-((dim_cond_set+1)*order2+1))
26  Tquan <- c(T,quan)
27  if(output)
28  {
29    show("test statistic and critical value are:")
30    show(Tquan)
31    if(Tquan[1]<Tquan[2])
32    {
33      show("Thus, H0 (no reduction) is not rejected: 2nd var DOES NOT CAUSE 1
34        st var.")
35    }
36    else
37    {
38      show("Thus, H0 (no reduction) is rejected: 2nd var CAUSES 1st var.")
39    }
40  }
41  Tquan <- Tquan
42 }
```

Algoritmo que reconstrói DAG pelo método de G-causalidade linear

```
1 granger_dag_partial<-function(M,alpha,max_lag,output = TRUE)
2 {
3   #M (n x p) should consists of p cols, each of which is the realization of one
4   #variable, that has n data points.
5   #(i,j)==1 in the answer matrix means causal link i->j
6   #(i,j)==0 in the answer matrix means no causal link i->j
7   np<-dim(M)
8   CC<-matrix(0,np[2],np[2])
9   for(i in 1:(np[2]-1))
10  {
11    for(j in (i+1):np[2])
12    {
13      Fc<-granger_partial(M[,j],M[,i],M[, -c(i,j)],alpha,max_lag,output)
14      ifelse(Fc[1]<Fc[2],
15        CC[i,j]<-0,
16        CC[i,j]<-1)
17      Fc<-granger_partial(M[,i],M[,j],M[, -c(i,j)],alpha,max_lag,output)
18      ifelse(Fc[1]<Fc[2],
19        CC[j,i]<-0,
20        CC[j,i]<-1)
21    }
22  }
23  return(CC)
24 }
```

Referência dos códigos de G-causalidade e TiMINo:

<http://web.math.ku.dk/~peters/code.html>

Referência dos códigos do modelo VAR-LiNGAM:

<https://sites.google.com/site/dorisentner/publications/VARLiNGAM>

Referências

- [1] Shohei Shimizu et al. “A Linear Non-Gaussian Acyclic Model for Causal Discovery”. Em: *Journal of Machine Learning Research* 7.72 (2006), pp. 2003–2030. URL: <http://jmlr.org/papers/v7/shimizu06a.html>.
- [2] Aapo Hyvärinen, Shohei Shimizu e Patrik Hoyer. “Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-Gaussianity”. Em: jan. de 2008, pp. 424–431. DOI: [10.1145/1390156.1390210](https://doi.org/10.1145/1390156.1390210).
- [3] J. Pearl. *Causality*. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2009. ISBN: 9780521895606.
- [4] Aapo Hyvärinen et al. “Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity”. Em: *Journal of Machine Learning Research* 11.56 (2010), pp. 1709–1731. URL: <http://jmlr.org/papers/v11/hyvarinen10a.html>.
- [5] J. Peters. “Restricted Structural Equation Models for Causal Inference”. <http://dx.doi.org/10.3929/ethz-a-007597940>. Tese de dout. ETH Zurich e MPI for Intelligent Systems, 2012.
- [6] S. Palachy. *Infering causality in time series data*. 2019. URL: <https://towardsdatascience.com/infering-causality-in-time-series-data-b8b75fe52c46>.