Causal Structure Learning from Multivariate Time Series in Settings with Unmeasured Confoundings

Modelos Gráficos Causales

Sebastián Bejos

8 de julio de 2019

Instituto Nacional de Astrofísica Óptica y Electrónica

Introducción

- Se modifica el **algoritmo FCI** para buscar una **clase de equivalencia de MAGs dinámicos**.
- Suponiendo que el **proceso de generación de datos es un SVAR** (posiblemente no lineal) con variables latentes, o
- Equivalentemente, un **DAG dinámico con componentes latentes**.

DAG causal

- Un DAG es un grafo que contiene solo aristas dirigidas ($\rightarrow$) y es acíclico. En un **DAG causal**, $X_i \rightarrow X_j$ si y solo si X_i es una causa directa de X_j relativo al conjunto de variables $\mathbf{V}$.
- Existe una correspondencia directa entre la causalidad directa en los DAG causales y los **modelos de ecuaciones estructurales (no paramétricos)**.
- Las **relaciones de independencia condicional** implicadas por un modelo de ecuaciones estructurales pueden obtenerse a partir del DAG causal correspondiente utilizando el conocido criterio de **d-separación**.

Grafos Ancestrales Maximales

En entornos donde puede haber co-factores no medidos (causas comunes latentes), uno puede representar las relaciones entre las variables medidas por un **Grafo Ancestral Máximal** (Maximal Ancestral Graphs) (MAG).

Definición

Un MAG es un grafo mixto que puede tener aristas dirigida ($\rightarrow$) *y aristas bi-dirigidas* ($\leftrightarrow$)

Grafos Ancestrales Maximales

- Un MAG representa un DAG (o un conjunto de DAG que comparten características comunes) **después de que todas las variables latentes se han marginado** y **conserva todas las relaciones de independencia condicional** entre las variables medidas implicadas en el DAG subyacente.
- Las cuales pueden ser enumeradas por medio del criterio gráfico de **m-separación**.

Grafos Ancestrales Maximales

- En un MAG $\mathcal{M}$, una **cola** sobre X_i (e.g., $X_i \rightarrow X_j$) significa que X_i es un ancestro de X_j en todas las DAGs representadas por $\mathcal{M}$.
- Una **punta de flecha** sobre X_i (e.g., $X_i \leftarrow X_j$ o $X_i \leftrightarrow X_j$) significa que X_i no es un ancestro de X_j en todas las DAGs representadas por $\mathcal{M}$.
- Una arista bi-dirigidas ($\leftrightarrow$) entre dos variables indica que ninguna de las variables es ancentro de la otra.

Grafos Ancestrales Maximales

- Una **Clase de Equivalencia de Markov** de MAGs es un conjunto de MAGs que implican los mismos hechos de *m* separación.
- Esta clase de equivalencia es representada por un **Grafo Ancestral Parcial (PAG)**, en donde son posibles aristas adicionales de **marca de círculo** $\circ$ (e.g., $X_i \circ\!\!\rightarrow X_j$).
- Estas marcas de circulo indican que en alguna MAG dentro de la clase de equivalencia existe una **punta de flecha** sobre sobre X_i (e.g., $X_i \leftarrow X_j$ o $X_i \leftrightarrow X_j$) y en algunas otras MAGs hay una **cola** sobre X_i (e.g., $X_i \rightarrow X_j$).
- Entonces las **PAG**s que se consideran pueden tener los siguientes tipos de aristas: $\rightarrow, \circ\!\!\rightarrow, \circ\!\!-$ y $\leftrightarrow$.

DAGs Dinámicos con Variables Latentes

- El esquema de correspondencia entre DAG y las ecuaciones estructurales se puede extender a **sistemas dinámicos**.
- En donde los vértices representan elementos de procesos estocásticos de tiempo discreto, en lugar de observaciones transversal y las ecuaciones son **Vectores de Auto-Regresión Estructurales (SVAR)**, (posiblemente no lineales).
- Este trabajo se enfoca en procesos estocásticos $\{\mathbf{X}_t\}_{t\in\mathbb{N}}$ que son generados a partir de SVARs que incluyen variables latentes.

Definición

Un procesos estocástico SVAR k-dimensional de orden p esta dado por:

$$X_{i,t} = f_i(\mathbf{X}_t^{-i}, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-p}, \epsilon_{i,t}), \forall i \in \{1, \ldots k\}, \forall t \in \mathbb{N},$$

donde $\mathbf{X}_t$ *es un vector k-dimensional de variables de series de tiempo* $(X_{1,t}, \ldots, X_{k,t})$, $\mathbf{X}_t^{-i} = X_t \setminus \{X_{i,t}\}$ *y los* $\epsilon_{i,t}$ *son mutuamente independientes e independientes en serie.*

DAGs Dinámicos con Variables Latentes

- Al igual que para los modelos de ecuaciones estructurales, las f_i pueden ser funciones medible arbitrarias.
- En el caso lineal es más común escribir un SVAR en notación matricial:

$$\mathbf{\Gamma}_0\mathbf{X}_t = \mathbf{\Gamma}_1\mathbf{X}_{t-1} + \ldots + \mathbf{\Gamma}_p\mathbf{X}_{t-p} + \boldsymbol{\epsilon}_t, \forall t \in \mathbb{N},$$

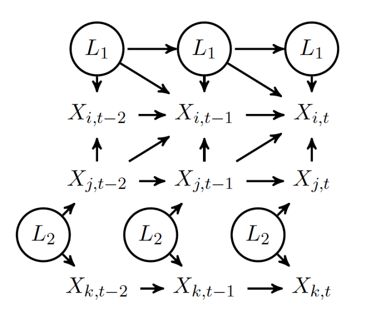
  donde las $\mathbf{\Gamma}_j$ son matrices de $k \times k$ de coeficientes constantes.

DAGs Dinámicos con Variables Latentes

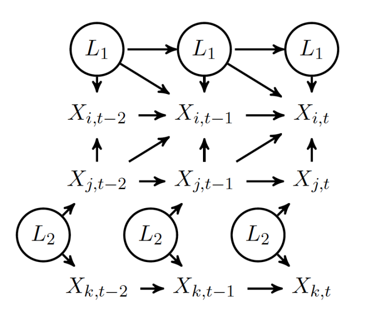
- Se permite que el proceso de generación de datos tenga cualquier número de **componentes latentes**.
- Estos a veces se representan explícitamente reemplazando X_t en con $\tilde{\boldsymbol{X}}_t = (\boldsymbol{L}_t', \boldsymbol{X}_t')'$, siendo $\boldsymbol{L}_t$ un vector de variables de serie de tiempo no medidas y $\boldsymbol{X}_t$ las variables observadas.
- Se asume que el proceso es estable y por lo tanto **estacionario**, i.e., todos los momentos del proceso estocástico son **invariantes con respeto al tiempo**.

DAGs Dinámicos con Variables Latentes

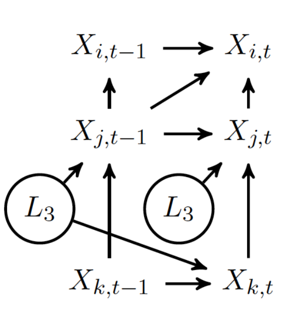
- Un proceso de generación de datos de este tipo corresponde a un **DAG Dinámico con Variables Latentes**, también denominado Red Bayesiana Dinámica (DBN).
- Si la función f_i no es una función trivial de $X_{j,s}, (s \leq t)$, entonces $X_{j,s} \rightarrow X_{i,t}$ en el grafo G.



- Tenga en cuenta que los grafos considerados se pueden llamar **repetidos** dado que $X_{i,t-h} \to X_{j,t}$ si y solo si $X_{i,t-h-m} \to X_{j,t-m}$ $\forall h \geq 0,\ m \in \mathbb{N}$.
- Además, estos grafos son **(semi-)infinitos**, i.e., $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{E})$, donde $|\boldsymbol{V}| = |\mathbb{N} \times \tilde{\boldsymbol{X}}_t|$.
- En la práctica solo se representan segmentos de estos grafos ya que se repiten.



- Un DAG dinámico con procesos latentes L_1 y L_2. L_1 se le llama **co-factor de retraso automático** y a L_2 se le llama **co-factor contemporaneo**.



- Un DAG dinámico con proceso latente L_3. L_3 se le llama **co-factor de retraso cruzado**.

Dados los valores iniciales $(\boldsymbol{X}_{-p+1}, \ldots \boldsymbol{X}_0)'$, la factorización de Markov para un proceso estocástico $\{\boldsymbol{X}_t\}_{t\in\mathbb{N}}$, k-dimensional de orden p, se puede escribir como sigue:

$$P(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_{t-p}) = \prod_{i\in[k], s\in\{t,\ldots t-p\}} P(X_{i,s} \mid pa(X_{i,s}, \boldsymbol{\mathcal{G}})), \tag{1}$$

$\forall t \in \mathbb{N}$, donde $\boldsymbol{\mathcal{G}}$ es la DAG dinámica infinita.

Grafos Ancestrales de Markov Dinámicos

- Para el segmento del DAG dinámico $\mathcal{G}$ correspondiente a todo el proceso, se puede derivar el segmento de un MAG dinámico $\mathcal{M}$ sobre las variables observadas, únicamente.

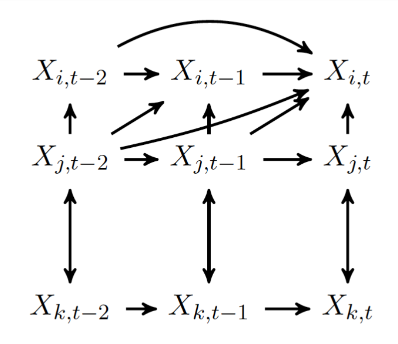


Figura 1: El MAG dinámico implicado del primer DAG dinámico.

Grafos Ancestrales de Markov Dinámicos

$$
\begin{tikzcd}
X_{i,t-1} \arrow[r] & X_{i,t} \\
X_{j,t-1} \arrow[u] \arrow[r] \arrow[ur] \arrow[dr, leftrightarrow] & X_{j,t} \arrow[u] \\
X_{k,t-1} \arrow[u] \arrow[r] & X_{k,t} \arrow[u]
\end{tikzcd}
$$

Figura 2: El MAG dinámico implicado del segundo DAG dinámico.

Los algoritmos introducidos en este articulo tiene como propósito aprender la clase de equivalencia de Markov de $\mathcal{M}$, i.e., un segmento de un PAG dinámico $\mathcal{P}$.

El procedimiento propuesto es llamado SVAR-FCI y realiza las siguientes modificaciones a FCI:

1. El SVAR-FCI respeta el orden temporal de las variables al restringir los posibles condicionamientos entre las variables en los segmentos de tiempo presente o pasado y prohíbe las orientaciones hacia atrás en el tiempo.
2. SVAR-FCI aplica la estructura de repetición del DAG dinámico subyacente para determinar las adyacencias y las orientaciones de la estructura del modelo.