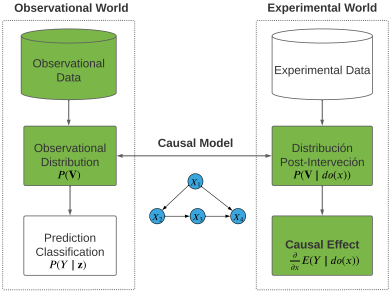
Estimating Bounds on Causal Effects Considering Unmeasured Common Causes

Sebastián Bejos (PhD Student),
Dr. Enrique Sucar & Dr. Eduardo Morales

National Institute of Astrophysics, Optics and Electronics

Introduction



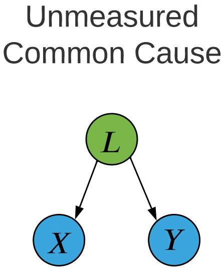
- We address the problem of **estimating bound of causal effect** between pairs of variables (X, Y) in a system, i.e., **measuring how much** one variable Y changes after another variable X is intervened, **relaxing the causal sufficiency assumption** and using only **observational data**.

Motivation

- Most works consider the **causal sufficiency assumption**, i.e., that there are no latent variables in the system that are **common direct causes** of measured variables. This is often **unrealistic in applied contexts**.
- On **Maximal Ancestral Graphs (MAGs)**, used to model insufficient systems, it is not always possible to estimate the causal effect directly between some pairs of variables (X, Y).

Main Goal

Develop and validate a way to **approximate the causal effect** between a pair of variables when it **cannot be estimated directly** in a causal system with **unmeasured common causes**.

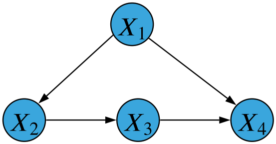


Causal Graphical Model

i $(\mathcal{G}, \Phi)$, $\mathcal{G}$ causal structure and Φ model parameters.

ii Φ Structural Equations (SE), $X_i = f_i(\boldsymbol{pa}(X_i), U_i)$.

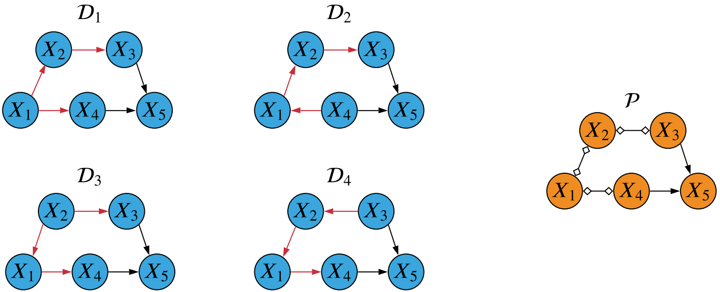
DAG Causal



SE Lineales Gaussianas

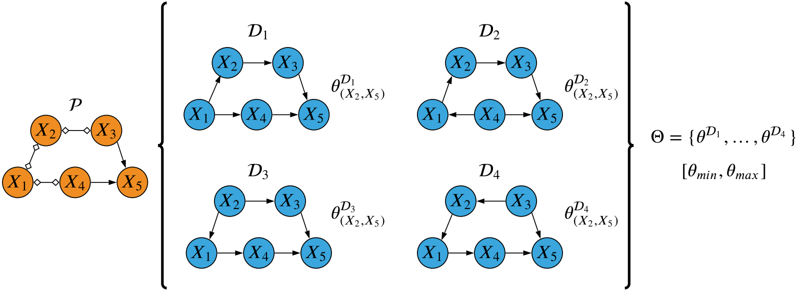
$$
\begin{aligned}
X_1 &= \epsilon_1 \\
X_2 &= 2X_1 + \epsilon_2 \\
X_3 &= 3X_2 + \epsilon_3 \\
X_4 &= X_1 + 2X_3 + \epsilon_4 \\
\epsilon_1, \ldots, \epsilon_4 &\sim \mathcal{N}(\mu, \sigma)
\end{aligned}
$$

Markov Equivalence Class (MEC)



- Partial structure $\mathcal{P}$ representing a MEC.
- Causal discovery algorithms find these partial structures $\mathcal{P}$.

IDA Algorithm
[Maathuis et al., 2009]

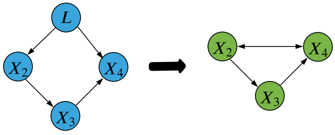


- They assume the system $\boldsymbol{V} = \{X_1, \ldots, X_p\}$ is causal sufficiency and jointly Gaussian.

Causal Insufficiency and MAGs

i **Causal Sufficiency**: There are no latent variables in $\boldsymbol{V}$ that are common direct causes of measured variables.

ii Maximal Ancestral Graphs (**MAGs**), *m*-separation.

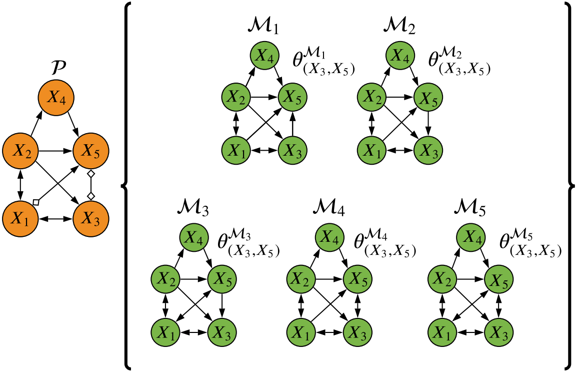


$\mathcal{D}$ $\quad\quad\quad\quad$ $\mathcal{M}$

iii A MAG represents a **DAG with latent variables preserving all the conditional independence relationships** among the **measured variables**.

Algoritmo LV-IDA

[Malinsky and Spirtes, 2017]



$$\Theta = \{\theta^{\mathcal{M}_1}, \ldots, \theta^{\mathcal{M}_5}\}$$

$$[\theta_{min}, \theta_{max}]$$

Post-Intervention Distribution

Covariate Adjustment

i Given a **causal structure** $\mathcal{G}$.

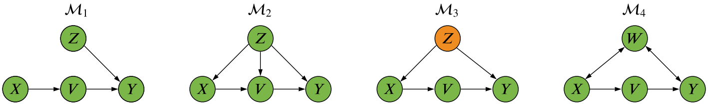
ii $\mathcal{G}$ is used to find a subset of variables $\boldsymbol{Z}$, called the **adjustment set**.

iii Using this set $\boldsymbol{Z}$, the post-intervention distribution is computed as:

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \begin{cases} P(\boldsymbol{y} \mid \boldsymbol{x}), & \text{if } \boldsymbol{Z} = \emptyset \\ \int_{\boldsymbol{Z}} P(\boldsymbol{y} \mid \boldsymbol{xz}) P(\boldsymbol{z}) d\boldsymbol{z}, & \text{otherwise.} \end{cases}$$

LV-IDA Limitations

- There is not always an **adjustment set** in a MAG for some pairs of variables.



- **NAs Simples**

$$\Theta = \{NA, NA, \theta^{\mathcal{M}_3}, \theta^{\mathcal{M}_4}\}. \quad ¿[\theta_{min}, \theta_{max}]?$$

- **NAs Extremos**

$$\Theta = \{NA, NA, NA, NA\}. \quad ¿[\theta_{min}, \theta_{max}]?$$

Problem Definition

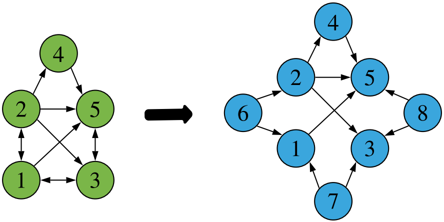
Is there any **plausible approximation** for the value of the causal effect between a pair of variables (X, Y) when it is not possible to find an adjustment set in an **insufficient causal system** $\boldsymbol{V} = \{X_1, \dots X_p\}$ modeled with a MAG?

$$\Theta = \{NA, NA, \theta^{\mathcal{M}_3}, \theta^{\mathcal{M}_4}\}. \quad ¿[\theta_{min}, \theta_{max}]?$$

$$\Theta = \{NA, NA, NA, NA\}. \quad ¿[\theta_{min}, \theta_{max}]?$$

Hypothesis

Estimating the causal effect for a pair of variables (X, Y) by **covariate adjustment in the canonical DAG** associated to a MAG, obtaining its **full parameterization by means of Expectation Maximization (EM)** techniques, can be a good approximation for cases in which the causal effect on the pair (X, Y) is not identifiable in the MAG.

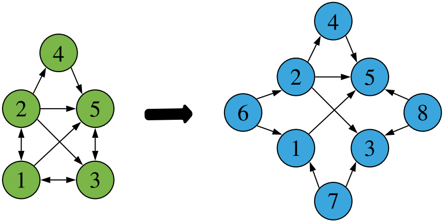


DAGs canónicos

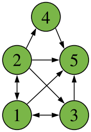
i Sea $\boldsymbol{L}_{\mathcal{D}} = \{\lambda_{XY} \mid X \leftrightarrow Y \text{ in } \mathcal{M}\}$.

ii El **DAG canónico** $\mathcal{D}$ asociado a un MAG $\mathcal{M}$ tiene conjunto de vértices $\boldsymbol{V} \cup \boldsymbol{L}_{\mathcal{D}(\mathcal{M})}$ y conjunto de aristas dado por:

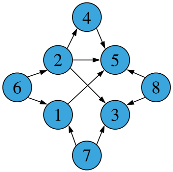
$$\text{Si } \left\{ \begin{array}{l} X \to Y \\ X \leftrightarrow Y \end{array} \right\} \text{en } \mathcal{M}, \text{ entonces } \left\{ \begin{array}{l} X \to Y \\ X \leftarrow \lambda_{XY} \to Y \end{array} \right\} \text{en } \mathcal{D}.$$



Causal Effects Matrices

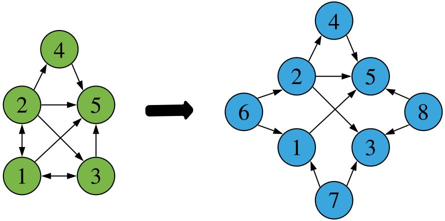


$$CE_{\mathcal{M}_w} = \begin{bmatrix} 1.00 & NA & NA & 0.01 & NA \\ NA & 1.00 & NA & 0.55 & NA \\ NA & NA & 1.00 & 0.00 & NA \\ 0.05 & NA & 0.01 & 1.00 & NA \\ NA & NA & NA & NA & 1.00 \end{bmatrix}$$



$$CE_{\mathcal{D}_w} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.01 & 0.67 & 0.03 & 0.15 & 0.00 \\ 0.00 & 1.00 & 0.80 & 0.55 & 1.69 & 0.06 & 0.00 & 0.00 \\ 0.00 & 0.48 & 1.00 & 0.00 & 0.00 & 0.00 & 0.03 & 0.00 \\ 0.05 & 0.42 & 0.01 & 1.00 & 0.61 & 0.00 & 0.00 & 0.00 \\ 0.14 & 0.35 & 0.00 & 0.32 & 1.00 & 0.00 & 0.00 & 0.01 \\ 0.04 & 0.06 & 0.05 & 0.03 & 0.14 & 1.00 & 0.00 & 0.00 \\ 0.15 & 0.00 & 0.05 & 0.00 & 0.01 & 0.00 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.05 & 0.00 & 0.00 & 1.00 \end{bmatrix}$$

Causal Effects Matrices $CE^{+}_{\mathcal{M}_w}$



$$CE^{+}_{\mathcal{M}_w} = CE_{\mathcal{M}_w} \oplus CE_{\mathcal{D}_w} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.01 & 0.67 \\ 0.00 & 1.00 & 0.80 & 0.55 & 1.69 \\ 0.00 & 0.48 & 1.00 & 0.00 & 0.00 \\ 0.05 & 0.42 & 0.01 & 1.00 & 0.61 \\ 0.14 & 0.35 & 0.00 & 0.32 & 1.00 \end{bmatrix}$$

LV-IDA+ algorithm

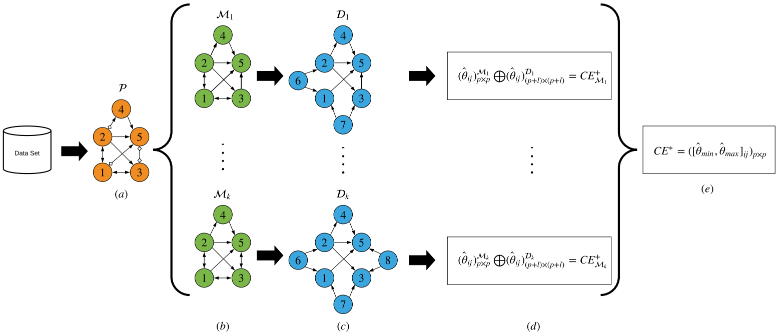


Figure 1:
(d) $CE_{\mathcal{M}_w} = (\hat{\theta}_{ij})^{\mathcal{M}_w}_{p\times p} \oplus CE_{\mathcal{D}_w} = (\hat{\theta}_{ij})^{\mathcal{D}_w}_{(p+l_w)\times(p+l_w)} = CE^{+}_{\mathcal{M}_w} = (\hat{\theta}_{ij})_{p\times p}$.
(e) $CE^{*} = ([\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij})_{p\times p}$

Matriz de Intervalos de Efectos Causales CE^* en LV-IDA+

$$\boxed{CE^+_{\mathcal{M}_1}}$$

$$\vdots$$

$$\boxed{CE^+_{\mathcal{M}_k}}$$

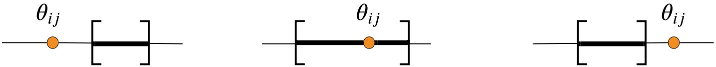
$$CE^* = \begin{bmatrix} [1.00, 1.00] & [0.00, 0.00] & [0.00, 0.16] & [0.01, 0.01] & [0.00, 0.67] \\ [0.00, 0.00] & [1.00, 1.00] & [0.68, 0.80] & [0.55, 0.55] & [1.69, 2.18] \\ [0.00, 0.00] & [0.43, 0.48] & [1.00, 1.00] & [-0.04, 0.00] & [0.00, 1.02] \\ [0.05, 0.05] & [0.42, 0.42] & [0.02, 0.15] & [1.00, 1.00] & [0.61, 0.64] \\ [0.00, 0.14] & [0.34, 0.36] & [0.00, 0.32] & [0.29, 0.33] & [1.00, 1.00] \end{bmatrix}$$

Experimental Evaluation

- The **LV-IDA+ algorithm was implemented in the R programming language** and the LV-IDA implementation given in [Malinsky and Spirtes, 2017] was used.
- 80 synthetic causal models were generated, varying the number of variables $p \in \{5, 8, 11, 14\}$, 20 different models for each p with samples of size 1000.
- For a total of **7,360 causal effect estimations**, with each of the two algorithms: LV-IDA+ and LV-IDA.

Evaluation Metric I

$$\text{IntMSE} = \begin{cases} 0 & \text{if } \theta_{ij} \in [\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij} \\ \min\{|\theta_{ij} - (\hat{\theta}_{min})_{ij}|, |\theta_{ij} - (\hat{\theta}_{max})_{ij}|\} & \text{otherwise,} \end{cases}$$



$$CEReal = (\theta_{ij}) = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.67 \\ 0.00 & 1.00 & 0.79 & 0.58 & 1.74 \\ 0.00 & 0.30 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.53 & 0.00 & 1.00 & 0.62 \\ 0.40 & 0.31 & 0.00 & 0.25 & 1.00 \end{bmatrix}$$

Evaluation Metric II

- Let $IEM = (\epsilon_{ij})$ be the Interval Error Matrix where the ϵ_{ij} are defined as the IntMSE between the true total effect θ_{ij} in *CEReal*, and the estimate interval of causal effect $[\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij}$ in CE^*.
- Let Σ be the sum of all the elements ϵ_{ij} of the Interval Error Matrix $IEM = (\epsilon_{ij})$. The Average Interval Mean Square Error (AIntMSE), is defined as:

$$\text{AIntMSE} = \frac{\Sigma}{p^2 - (p + \text{eNA})},$$

where p is the number of variables and eNA is the number of extreme NAs.

Results

	AIntMSE	**SimpleNAs**	**ExtremeNAs**	**Time (sec)**
$\boldsymbol{p = 5}$				
LV-IDA+	**0.0223** ± **0.0182**	**0**	**0**	6.6 ± 2.4
LV-IDA	0.0357 ± 0.0413	28.88 ± 16.04	3.50 ± 2.27	**1.4** ± **0.7**
$\boldsymbol{p = 8}$				
LV-IDA+	**0.0441** ± **0.0333**	**0**	**0**	14.4 ± 7.5
LV-IDA	0.0710 ± 0.0611	107.8 ± 28.32	20.3 ± 4.44	**4.6** ± **1.6**
$\boldsymbol{p = 11}$				
LV-IDA+	**0.1040** ± **0.1130**	**0**	**0**	17.0 ± 11.1
LV-IDA	0.1797 ± 0.2133	177.3 ± 78.2	49.4 ± 26.6	**16.4** ± **7.1**
$\boldsymbol{p = 14}$				
LV-IDA+	**0.2353** ± **0.2028**	**0**	**0**	18.4 ± 18.2
LV-IDA	0.4101 ± 0.3605	340.3 ± 98.69	74.1 ± 8.37	**12.2** ± **21.6**

Conclusions

1. Employing **EM techniques** can effectively solve the problem of obtaining a **full parameterization** when a **DAG with latent variables** is selected to **approximate causal effects** that cannot be calculated directly on the MAG.
2. The **canonical DAG** associated with a MAG is a way of representing the set of DAGs included in a MAG, that can be **efficiently constructed** and **adds few latent variables** with respect to the number of variables in the MAG.

Conclusions

3. The experiments show **better accuracy** in the **intervals estimated by LV-IDA+** than those estimated by **LV-IDA** in the case where the data were **generated by a canonical DAG**.
4. This is because the **non-estimated causal effects** caused by the missing values, i.e., **simple NAs**, make the LV-IDA estimates **misleading**.

Future Work

1. **Extend the spectrum of the type of synthetic models** in experimentation to **explore the generality of our algorithm**, e.g., perform tests with simulated data from **anti-canonical DAGs**.
2. **Extend the experimentation** for LV-IDA+ on **PAGs** generated by the **FCI**([Zhang, 2008b]) and **GFCI** ([Ogarrio et al., 2016]) algorithms, with synthetic data and **real Gaussian Bayesian networks**.
3. Extend the work on [Montero-Hernandez et al., 2018] to insufficient system. In which **intervals of causal effects** are used to point to a single model of the Markov equivalence class.

Thank you.
Any questions?

Referencias I

Glymour, C., Zhang, K., and Spirtes, P. (2019).
Review of causal discovery methods based on graphical models.
Frontiers in Genetics, 10:524.

Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., and Bühlmann, P. (2012).
Causal inference using graphical models with the R package pcalg.
Journal of Statistical Software, 47(11).

Koller, D. and Friedman, N. (2009).
Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning.
The MIT Press.

Maathuis, M. H. and Colombo, D. (2015).
A generalized back-door criterion.
Ann. Statist., 43(3):1060–1088.

Referencias II

Maathuis, M. H., Kalisch, M., and Bühlmann, P. (2009).
Estimating high-dimensional intervention effects from observational data.
Ann. Statist., 37(6A):3133–3164.

Malinsky, D. and Spirtes, P. (2017).
Estimating bounds on causal effects in high-dimensional and possibly confounded systems.
International Journal of Approximate Reasoning, 88:371–384.

Montero-Hernandez, S., Orihuela-Espina, F., and Sucar, L. E. (2018).
Intervals of causal effects for learning causal graphical models.
In *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, volume 72 of *Proceedings of Machine Learning Research*, pages 296–307, Prague, Czech Republic. PMLR.

Ogarrio, J. M., Spirtes, P., and Ramsey, J. (2016).
A Hybrid Causal Search Algorithm for Latent Variable Models.
JMLR workshop and conference proceedings, 52:368–379.

Referencias III

Pearl, J. (2009).
Causality: Models, Reasoning and Inference.
Cambridge University Press, USA, 2nd edition.

Perković, E., Textor, J., Kalisch, M., and Maathuis, M. H. (2018).
Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs.
Journal of Machine Learning Research, 18:1–62.

Richardson, T. and Spirtes, P. (2002).
Ancestral graph markov models.
Ann. Statist., 30(4):962–1030.

Sucar, L. E. (2015).
Probabilistic Graphical Models: Principles and Applications.
Springer Publishing Company, Incorporated.

Referencias IV

van der Zander, B., Liśkiewicz, M., and Textor, J. (2019).
Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework.
Artificial Intelligence, 270:1–40.

Zhang, J. (2008a).
Causal reasoning with ancestral graphs.
J. Mach. Learn. Res., 9:1437–1474.

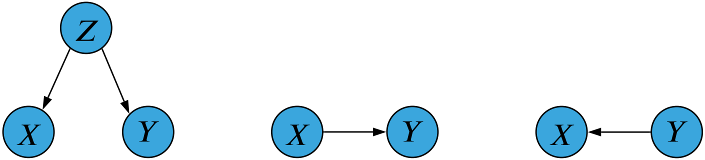
Zhang, J. (2008b).
On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias.
Artificial Intelligence, 172(16-17):1873–1896.

Referencias V

Zhang, J. and Spirtes, P. (2005).
A transformational characterization of markov equivalence for directed acyclic graphs with latent variables.
In *Proceedings of the Twenty-First Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pages 667–674, Arlington, Virginia. AUAI Press.

Principio de Causa Común

El **principio de causa común** dice que toda **correlación** entre un par de variables es debida a una una relación de **causa directa** entre las variables, o debida a un tercer factor llamado **causa común**.

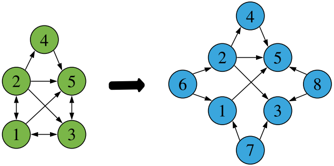


No hay correlación sin causalidad.

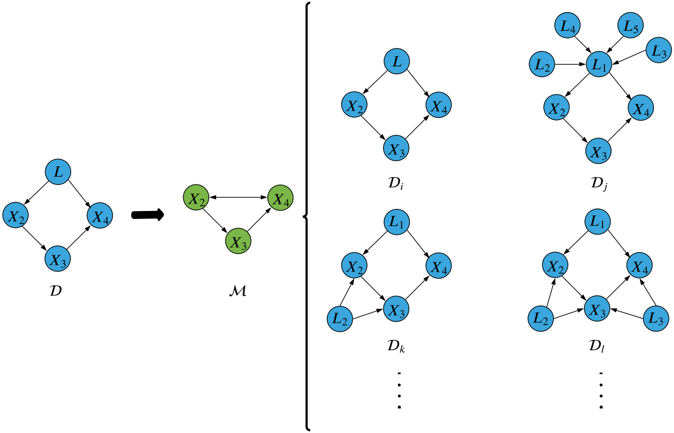
DAGs canónicos

- i Sea $\boldsymbol{L}_{\mathcal{D}} = \{\lambda_{XY} \mid X \leftrightarrow Y \text{ in } \mathcal{M}\}$.
- ii El **DAG canónico** $\mathcal{D}$ asociado a un MAG $\mathcal{M}$ tiene conjunto de vértices $\boldsymbol{V} \cup \boldsymbol{L}_{\mathcal{D}(\mathcal{M})}$ y conjunto de aristas dado por:

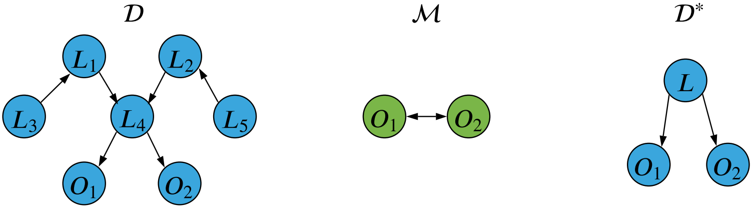
$$\text{Si } \left\{ \begin{array}{l} X \to Y \\ X \leftrightarrow Y \end{array} \right\} \text{en } \mathcal{M}, \text{ entonces } \left\{ \begin{array}{l} X \to Y \\ X \leftarrow \lambda_{XY} \to Y \end{array} \right\} \text{en } \mathcal{D}.$$



Fundamento Lógico

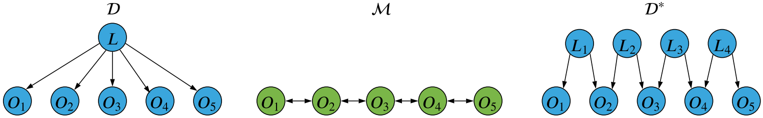


Sub-estructuras Complejas de Variables Latentes



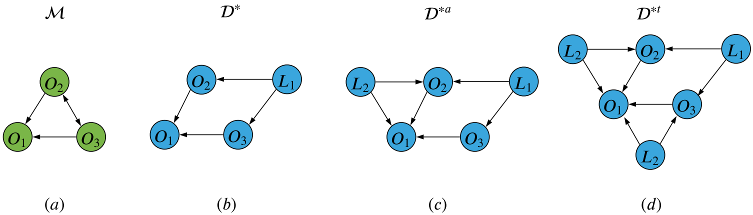
(a) (b) (c)

Sub-estructuras con Causas Comunes Compartidas



(a) (b) (c)

Limitantes LV-IDA+ / Patrones Anti-canónicos



Objetivos I

1. Implementar una forma de calcular un DAG canónico a partir de un MAG y obtener una completa parametrización del mismo utilizando técnicas de maximización de expectativas para poder estimar los efectos causales entre pares de variables en este DAG canónico.
2. Extender el algoritmo LV-IDA ([Malinsky and Spirtes, 2017]) para estimar los límites de los efectos causales dados un PAG calculando los efectos causales sobre los DAG canónicos asociados con los MAG en el PAG, como aproximaciones para los casos en los que no se pueden calcular directamente sobre los MAG.

Objetivos II

1. Presentar un proceso de generación de datos en el que primero se construye un PAG aleatorio, luego se selecciona aleatoriamente un MAG dentro de la MEC representado por este PAG, y finalmente se computa el DAG canónico asociado a este MAG para generar los datos de acuerdo con este DAG.
2. Evaluar nuestra propuesta de aproximar los efectos causales para los casos en los que no se pueden calcular directamente en los MAG, comparando la calidad de las cotas estimadas de los efectos causales con nuestra extensión al algoritmo LV-IDA que implementa nuestra hipótesis, con el algoritmo original LV-IDA.

Aprendizaje de Parámetros con Variables Ocultas

i. Obtener los CPD para todas las variables no ocultas en base a un estimador de ML.
ii. Inicializar los parámetros desconocidos con valores aleatorios.
iii. Considerando los parámetros reales, estime los valores de los nodos ocultos basados en las variables conocidas mediante inferencia probabilística.
iv. Utilice los valores estimados de los nodos ocultos para completar / actualizar el conjunto de datos.
v. Vuelva a estimar los parámetros para los nodos ocultos con los datos actualizados.

Aprendizaje de Parámetros con Variables Ocultas

vi Repita 3–5 hasta que converjan, es decir, no se observen cambios significativos en los parámetros.

vii El algoritmo EM optimiza los parámetros desconocidos y da un máximo local, las estimaciones finales dependen de la inicialización.

Intervenciones sobre CBNs

Fórmula Truncada de Factorización:

$$f(\boldsymbol{v} \mid do(\boldsymbol{x})) = \begin{cases} \prod_{\{i \mid X_i \in \boldsymbol{V} \setminus \boldsymbol{X}\}} f(x_i \mid \boldsymbol{pa}(x_i)) & \text{si } \boldsymbol{v} \text{ es consistente con } \boldsymbol{x}, \\ 0 & \text{en cualquier otro caso,} \end{cases} \tag{1}$$

donde $\boldsymbol{v}$ consistente con $\boldsymbol{x}$ significa que $\boldsymbol{v}$ y $\boldsymbol{x}$ asignan los mismos valores a las variables en $\boldsymbol{V} \cap \boldsymbol{X}$.

Distribución Post-Intervención

Ajuste por Covariantes

i Dada una **estructura causal** $\mathcal{G}$.

ii $\mathcal{G}$ es usada para encontrar un subconjunto de variables, llamado **conjunto de ajuste**.

iii Utilizando este conjunto se calcula la distribución post-intervención.

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \begin{cases} P(\boldsymbol{y} \mid \boldsymbol{x}), & \text{if } \boldsymbol{Z} = \emptyset \\ \int_{\boldsymbol{Z}} P(\boldsymbol{y} \mid \boldsymbol{xz}) P(\boldsymbol{z}) d\boldsymbol{z}, & \text{otherwise.} \end{cases}$$

Efecto Causal Total

El **efecto causal total** se define como $\frac{\partial}{\partial x} E(Y \mid do(x))$

- Utilizando la definición de conjunto de ajuste suponiendo una distribución conjunta Gaussiana, se tiene que:

$$E(Y \mid do(x)) = \alpha + \beta x + \gamma^T E(\boldsymbol{Z}).$$

- Bajo este esquema el efecto causal de X sobre Y está dado por β, i.e., el coeficiente de X en la regresión anterior.

Multivariate Gaussian & Linear Regression

- For **multivariate Gaussian densities**, we can use the fact that this kind of densities are fully defined by **expectations**.
- For conditional independencies $P(Y \mid x, \boldsymbol{z}) = P(Y \mid \boldsymbol{z})$ we can use **conditional expectations** $E(Y \mid x, \boldsymbol{z}) = E(Y \mid \boldsymbol{z})$.
- Since conditional expectations are linear in a multivariate Gaussian distribution, allows to use **regression** for estimate $E(Y \mid x, \boldsymbol{z}) = \alpha + \beta x + \gamma^T \boldsymbol{z}$, for some $\alpha, \beta \in \mathbb{R}$ and $\gamma \in \mathbb{R}^{|\boldsymbol{z}|}$.

Total Causal Effect

Definition

The **total causal effect** of X on Y for continuous random variables setting is defined as $\frac{\partial}{\partial x} E(Y \mid do(x))$ (see [Maathuis et al., 2009]).

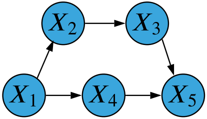
- Since by the adjustment set definition, we have that

$$E(Y \mid do(x)) = \int_{\boldsymbol{z}} E(Y \mid x, \boldsymbol{z}) f(\boldsymbol{z}) d\boldsymbol{z} = \alpha + \beta x + \gamma^T E(\boldsymbol{Z}).$$

- The **total causal effect** of X on Y in this setting is β, that is, the coefficient of X in the regression of Y on X and the adjustment set $\boldsymbol{Z}$

Redes Bayesianas

i. $(\mathcal{G}, \phi)$, **Mapeo de Independencias Condicionales** a partir de datos observacionales (*d*-separación).

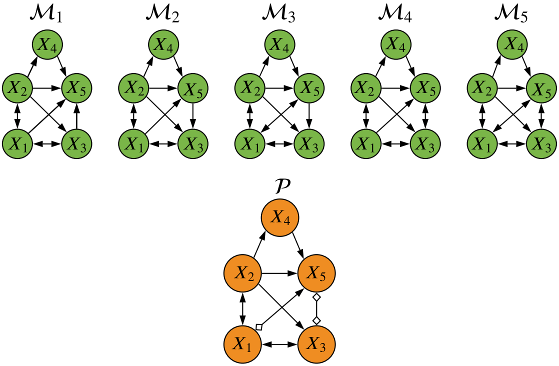


ii. Distribuciones de Probabilidad Conjunta **(CPDs)** $\phi = \{P(x_i \mid \boldsymbol{pa}(x_i))\}$.

iii.
$$P(X_1, \ldots, X_p) = \prod_{i=1}^{p} P(x_i \mid \boldsymbol{pa}(x_i)).$$

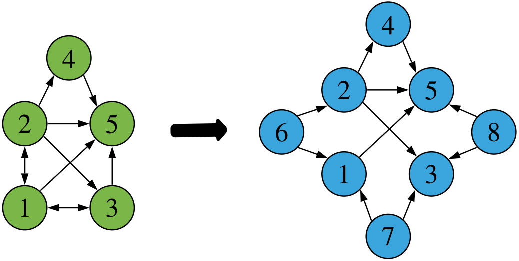
Algoritmos FCI y ZML

([Zhang, 2008b, Malinsky and Spirtes, 2017])



- Partial Ancestral Graphs (**PAGs**).

Parameterization of the canonical DAG



$$P(X_4 \mid X_2) = \mathcal{N}(\beta_4 + \beta_{24} X_2; \sigma_4^2)$$

$$P(X_3 \mid X_2, X_7) = \mathcal{N}(\beta_3 + \beta_{23} X_2 + \beta_{73} X_7 + \beta_{83} X_8; \sigma_3^2)$$
$$P(X_7) = \mathcal{N}(\mu_7; \sigma_7^2)$$

Data generation process

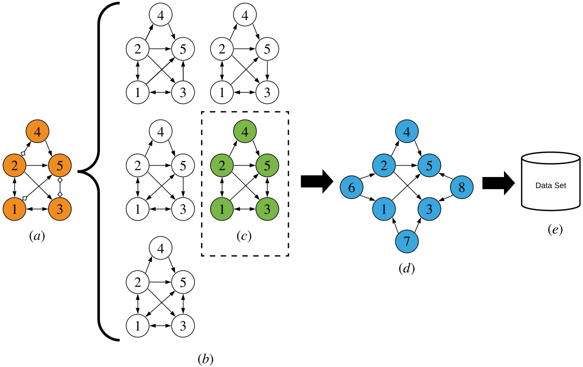
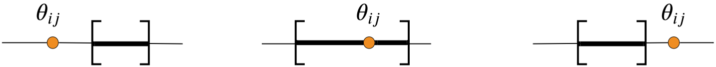


Figure 2: Linear Gaussian structural equations, coefficients distributed as $\pm\mathcal{U}([0.5, 1.5])$ and random perturbations distributed as $\mathcal{N}(0, \mathcal{U}([1, 3]))$.

Evaluation Metric I

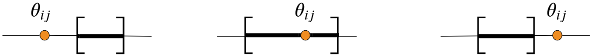
$$\text{IntMSE} = \begin{cases} 0 & \text{if } \theta_{ij} \in [\hat{\theta}_{min}, \hat{\theta}_{max}]_{ij} \\ \min\{|\theta_{ij} - (\hat{\theta}_{min})_{ij}|, |\theta_{ij} - (\hat{\theta}_{max})_{ij}|\} & \text{otherwise,} \end{cases}$$



$$CEReal = (\theta_{ij}) = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.67 \\ 0.00 & 1.00 & 0.79 & 0.58 & 1.74 \\ 0.00 & 0.30 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.53 & 0.00 & 1.00 & 0.62 \\ 0.40 & 0.31 & 0.00 & 0.25 & 1.00 \end{bmatrix}$$

Evaluation Metrics II

$$\begin{array}{r} \text{LV-IDA+} \\ CE^* = \end{array} \begin{bmatrix} [1.00, 1.00] & [0.00, 0.00] & [0.00, 0.16] & [0.01, 0.01] & [0.00, 0.67] \\ [0.00, 0.00] & [1.00, 1.00] & [0.68, 0.80] & [0.55, 0.55] & [1.69, 2.18] \\ [0.00, 0.00] & [0.43, 0.48] & [1.00, 1.00] & [-0.04, 0.00] & [0.00, 1.02] \\ [0.05, 0.05] & [0.42, 0.42] & [0.02, 0.15] & [1.00, 1.00] & [0.61, 0.64] \\ [0.00, 0.14] & [0.34, 0.36] & [0.00, 0.32] & [0.29, 0.33] & [1.00, 1.00] \end{bmatrix}$$



$$\begin{array}{r} \text{LV-IDA} \\ CE^* = \end{array} \begin{bmatrix} [1.00, 1.00] & [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.00] \\ [0.00, 0.00] & [1.00, 1.00] & \text{NA} & [0.55, 0.55] & \text{NA} \\ [0.00, 0.00] & [0.00, 0.00] & [1.00, 1.00] & [0.00, 0.00] & [0.00, 0.00] \\ [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.00] & [1.00, 1.00] & \text{NA} \\ [0.00, 0.00] & [0.00, 0.00] & [0.00, 0.27] & [0.00, 0.00] & [1.00, 1.00] \end{bmatrix}$$

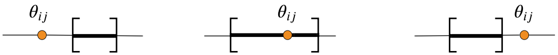
Métrica de Evaluación AIntMSE

LV-IDA+

$$IEM = (\epsilon_{ij}) = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.03 & 0.00 \\ 0.00 & 0.13 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.11 & 0.02 & 0.00 & 0.00 \\ 0.26 & 0.03 & 0.00 & 0.04 & 0.00 \end{bmatrix}$$

$$\text{AIntMSE} = \frac{\Sigma}{p^2-(p+\text{eNA})}$$

$$= \frac{0.68}{25-(5+0)} = \frac{.68}{20} = 0.034$$



LV-IDA

$$IEM = (\epsilon_{ij}) = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.67 \\ 0.00 & 0.00 & \text{NA} & 0.03 & \text{NA} \\ 0.00 & 0.30 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.53 & 0.00 & 0.00 & \text{NA} \\ 0.40 & 0.31 & 0 & 0.25 & 0.00 \end{bmatrix}$$

$$\text{AIntMSE} = \frac{\Sigma}{p^2-(p+\text{eNA})}$$

$$= \frac{2.49}{25-(5+3)} = \frac{2.49}{17} = 0.146$$

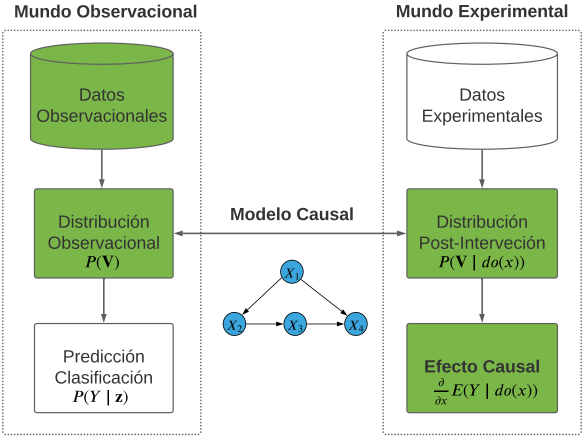
Contribution

1. La contribución principal es un **método para aproximar el efecto causal** en los casos en que otros trabajos se limitan a responder que **no es posible encontrar el efecto causal**, entre algunos pares de variables, en un sistema que **no asume suficiencia causal**.

Future Work

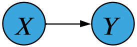
3. Explorar alternativas a **Maximización de Esperanza (EM)** en el proceso de aprendizaje de los parámetros con **variables ocultas** de los DAGs canonicos, ej. Descenso de Gradiente.
4. Extender el trabajo en [Montero-Hernandez et al., 2018] a sistema insuficientes. En el cual se utilizan **intervalos de efectos causales** para señalar un modelo único de la clase de equivalencia de Markov.

Context: Causal Inference

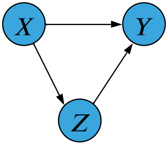


Causal Relations

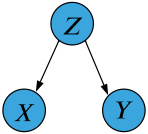
Direct Cause



Mediator



Common
Cause



Latent Common
Cause