Towards Robust Causal Effect Identification Beyond Markov Equivalence

Kai Teh, Kayvan Sadeghi, Terry Soo
Department of Statistical Science, University College London

UCL

Keywords Causal Effect Identification, Observational Data, Graphical Models, Robustness

1. Motivation

Causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ depends on causal DAG $\mathcal{G}$!

$$\underbrace{p^{\mathrm{do}_{\mathcal{G}}(\boldsymbol{x})}(\boldsymbol{y})}_{\text{causal effect}} = \underbrace{\int \prod_{V_i \in \boldsymbol{V} \setminus \boldsymbol{X}} p_{V_i}(v_i \mid \mathrm{pa}_{\mathcal{G}}(v_i))\, d\overline{\boldsymbol{v}}}_{\text{truncated factorisation formula}}$$

(marginalised over $\overline{\mathbf{V}} = \mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})$)

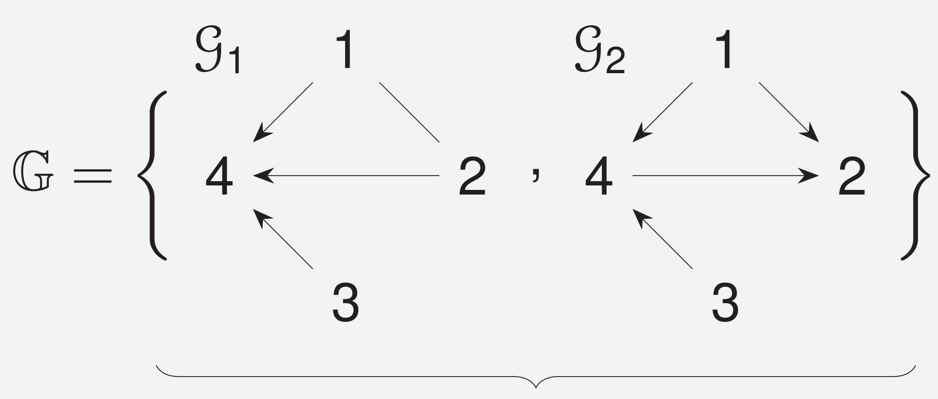
WHAT IF?? $\mathcal{G}$ is not specified up to Markov equivalence class? Can we still identify causal effects not knowing causal DAG $\mathcal{G}$??

Example †: Given distribution P with the conditional independencies:

$$\left\{ 1 \perp\!\!\!\perp 2, \quad 1 \perp\!\!\!\perp 2 \mid \{3,4\}, \quad 3 \perp\!\!\!\perp 4 \mid \{1,2\} \right\}$$

Running the Sparsest Permutation causal discovery algorithm (Raskutti and Uhler, 2018) with input P, does **not** return a unique output.

The set $\mathbb{G}$ of possible outputs is:



Non-Markov equivalent $\mathcal{G}_1$ and $\mathcal{G}_2$

and causal DAG $\mathcal{G}$ can be represented by either CPDAG $\mathcal{G}_1$ or $\mathcal{G}_2$ in $\mathbb{G}$.

Can happen in general when the faithfulness assumption is violated! (Teh, Sadeghi and Soo, 2024)

2. Problem Statement

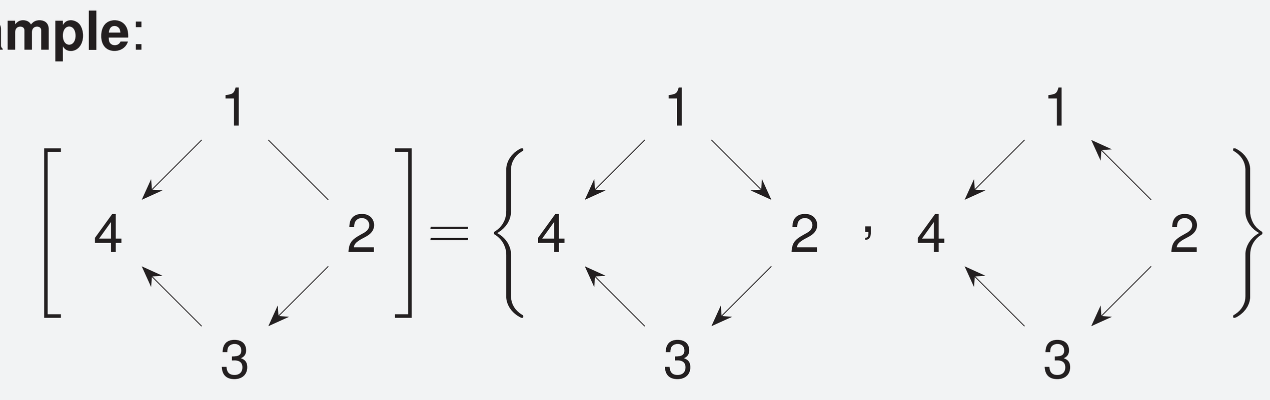
Given:
$\mathbb{G}$: a set of MPDAGs over nodes $\boldsymbol{V}$, and $\boldsymbol{X}, \boldsymbol{Y} \subseteq \boldsymbol{V}$.

- What are ***conditions*** on $\mathbb{G}$ such that

$$p^{\mathrm{do}_{\mathcal{G}_1}(\boldsymbol{x})}(\boldsymbol{y}) = p^{\mathrm{do}_{\mathcal{G}_2}(\boldsymbol{x})}(\boldsymbol{y})$$

  for every observational density p Markovian to all MPDAGs in $\mathbb{G}$, and every DAG $\mathcal{G}_1, \mathcal{G}_2 \in \bigcup_{\mathcal{G} \in \mathbb{G}} [\mathcal{G}]$?

  $[\mathcal{G}]$ is the set of DAGs represented by MPDAG $\mathcal{G}$.
  Example:



  MPDAG — Represented DAGs (obtained by directing undirected edges in the MPDAG.)

  i.e. for every possible observational density with $\mathbb{G}$, casual effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ does not depend on the choice of DAG chosen from ***all*** DAGs represented by ***all*** MPDAGs in $\mathbb{G}$; the causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is ***simultaneously identifiable***.

- If simultaneously identifiable, ***formula*** for the causal effect?

3. Results

The **conditions** on $\mathbb{G}$:
For all $\mathcal{G} \in \mathbb{G}$,

1. there are no **semi-directed paths** from $\boldsymbol{X}$ to $\boldsymbol{Y}$ with only the first node in $\boldsymbol{X}$, that starts with $-$, **and**

$$\underbrace{X_1 - \cdot \rightarrow \cdot - \cdots \rightarrow Y_i}_{\text{semi-directed path (no } \leftarrow \text{ edges)}}$$

Note: same condition as Perkovic 2020

For all $\mathcal{G}_1, \mathcal{G}_2 \in \mathbb{G}$,

2a. for all $\boldsymbol{X}_j$ (maximal subsets (that are connected via $-$ in $\mathcal{G}_1$) of $\boldsymbol{X}$ that are also ancestors of $\boldsymbol{Y}$ in $\mathcal{G}_1$),

$$\text{parents of } X_j \text{ in } \mathcal{G}_1 \quad = \quad \text{parents of } X_j \text{ in } \mathcal{G}_2$$

and likewise with the roles of $\mathcal{G}_1$ and $\mathcal{G}_2$ switched.

Intuition: reweighting used to obtain the truncated factorisation formula is the same for both $\mathcal{G}_1$ and $\mathcal{G}_2$

or

2b. $RM(\mathcal{G}_1; \boldsymbol{X}, \boldsymbol{Y})$ and $RM(\mathcal{G}_2; \boldsymbol{X}, \boldsymbol{Y})$ are Markov equivalent

Intuition: both $\mathcal{G}_1$ and $\mathcal{G}_2$ implies the same truncated factorisation formula over $\boldsymbol{Y}$.

$RM(\mathcal{G}; \boldsymbol{X}, \boldsymbol{Y})$: ancestral margin of $\boldsymbol{Y}$ in graph $\mathcal{G}$ after intervening on $\boldsymbol{X}$.

Theorem

Simultaneous identifiability if 1 + 2a **or** 1 + 2b holds.

The **formula** is the same as Perkovic 2020, since the **conditions** ensure equal evaluation of the causal effect across $\mathbb{G}$.

TL; DR

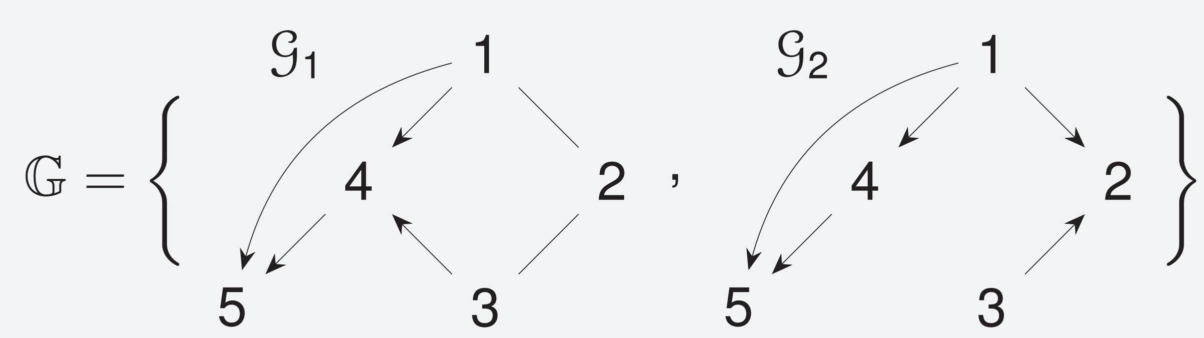
We propose sufficient conditions for **identifying causal effects** when the **Markov equivalence class of the causal graph cannot be uniquely identified**.

This can happen during causal discovery when faithfulness is violated or when edge detection lacks power.

Faithfulness is violated

Consider the set $\mathbb{G}$ in Example † from Motivation. The causal effect of 3 on 2 is simultaneously identifiable since 1 + 2a holds.

Uncertain edge 3 → 4



$RM(\mathcal{G}_1; \{4\}, \{5\}) = RM(\mathcal{G}_2; \{4\}, \{5\}) = 5 \leftarrow 4 \quad 1$

Condition 1 + 2b holds, the causal effect of 4 on 5 is simultaneously identifiable.

References
1. Perkovic, E. Identifying causal effects in maximally oriented partially directed acyclic graphs, 2020.
2. Raskutti, G. and Uhler, C. Learning directed acyclic graph models based on sparsest permutations, 2018.
3. Teh, K. Z., Sadeghi, K., and Soo, T. A general framework on conditions for constraint-based causal learning, 2024.

ICML
International Conference On Machine Learning