A General Framework for Constraint-based Causal Learning

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

Abstract

By representing any constraint-based causal learning algorithm via a placeholder property, we decompose the consistency condition into a part relating the distribution and true causal graph and a part that depends solely on the distribution. This provides a general framework to obtain consistency conditions for causal learning, from which we:

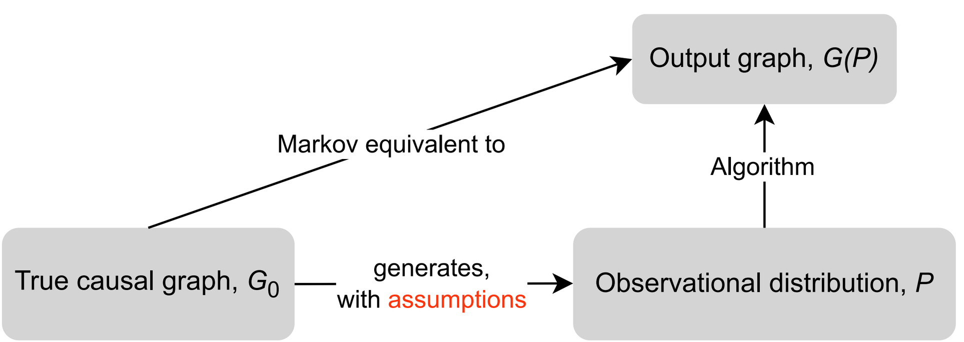
1. Show that the Sparsest Markov Representation (SMR) condition is the weakest amongst existing notions of minimality.
2. Provide exact consistency conditions for the PC algorithm, which are then related to some existing conditions.
3. Derive an algorithm that works beyond faithfulness under different conditions from the Sparsest Permutation algorithm.

Keywords Causal Discovery, Graphical Models, Faithfulness Assumption

Problem Setting and Objective

The setting and assumptions are as follows:

1. Purely *observational* setting, not accounting for interventional data.
2. Constraint-based, assume access to a *conditional independence oracle*.
3. Graph-based, assume P is generated by a true causal graph G_0.



- The algorithm is *consistent* if the output $G(P)$ is the 'same' as the true causal graph.
- The most common generative assumption is the *faithfulness* assumption.

Faithfulness is a strong assumption, and there have been works in causal learning *beyond* faithfulness, such as:

1. Sparsest Permutation (SP) algorithm.
2. Natural structure learning algorithm.

Objective Provide a general framework to study consistency conditions in causal learning.

Framework

$$\mathcal{A}(P, G_0) = \top \text{ and } U_{\mathcal{A}}(P) = \top \quad \Rightarrow \quad \text{algorithm s.t. } \mathcal{A}(P, G(P)) = \top \text{ is consistent.}$$

- $\mathcal{A}(P, G) = \top$ - P satisfies property $\mathcal{A}$ w.r.t. graph G.
  For example, if $\mathcal{A}$ is the faithfulness assumption, $\mathcal{A}(P, G) = \top$ can be read as P is faithful to G.
- $U_{\mathcal{A}}(P) = \top$ - P is $\mathcal{A}$-unique:
  All graphs G s.t. $\mathcal{A}(P, G) = \top$ are Markov equivalent.

! Reverse implication holds if property $\mathcal{A}$ is a *class property*:
For G_1 Markov equivalent to G_2, $\quad \mathcal{A}(P, G_1) = \top \iff \mathcal{A}(P, G_2) = \top$

Interpretation

Property $\mathcal{A}$ - Generative assumptions
$\mathcal{A}$ and $U_{\mathcal{A}}$ - Consistency conditions

Examples

$\mathcal{A}(P, G) = \top$	$\mathcal{A}(P, G) = \top$ and $U_{\mathcal{A}}(P) = \top$
P is faithful to G	P is faithful to G
G is the sparsest Markov graph to P	P satisfies the SMR assumption w.r.t. G
P is Markovian to G	G is complete and P is not Markovian to any subgraph

In general: $\mathcal{A}(P, G) = \top \Rightarrow \mathcal{B}(P, G) = \top$, does **not** imply:

$$\mathcal{A}(P, G) = \top \text{ and } U_{\mathcal{A}}(P) = \top \quad \Rightarrow \text{ or } \Leftarrow \quad \mathcal{B}(P, G) = \top \text{ and } U_{\mathcal{B}}(P) = \top$$

SMR is 'best' amongst Minimality

For P and G: denote

- $\mathcal{M}_1(P, G) = \top$ if P is minimally Markovian w.r.t. G.
- $\mathcal{M}_2(P, G) = \top$ if G is the sparsest Markov graph of P.
- $\mathcal{M}_3(P, G) = \top$ if G is P-minimal w.r.t. P.
- $\mathcal{M}_4(P, G) = \top$ if P is causally minimal w.r.t. G.

Then we have:

$\mathcal{M}_1 \Downarrow \mathcal{M}_2 \Rightarrow \mathcal{M}_3 \Rightarrow \mathcal{M}_4$ (Results from literature) $\longrightarrow$ $\mathcal{M}_1$ and $U_{\mathcal{M}_1}$ $\Downarrow$ $\mathcal{M}_2$ and $U_{\mathcal{M}_2}$ (SMR) $\Leftarrow$ $\mathcal{M}_3$ and $U_{\mathcal{M}_3}$ $\Leftarrow$ $\mathcal{M}_4$ and $U_{\mathcal{M}_4}$ (Obtained from framework)

Exact Consistency Conditions for PC

Depending on the computational implementation, PC may use different orientation rules:

$$i - k - j \quad \xrightarrow{\forall/\exists \;\; C,\, i \perp\!\!\!\perp j \mid C,\;\; k \in C} \text{non-collider} \qquad \xrightarrow{\forall/\exists \;\; C,\, i \perp\!\!\!\perp j \mid C,\;\; k \notin C} \text{collider}$$

with different combinations of the quantifiers for the orientation rule. Different orientation rule have a different corresponding property $\mathcal{V}$:

Orientation rule		$\mathcal{V}(P, G(P)) = \top$	
Collider	Non-collider	Collider	Non-collider
$\exists$	$\forall$	$\forall$	$\exists$
$\forall$	$\exists$	$\exists$	$\forall$
$\exists$	$\exists$	$\forall$	$\forall$

From the framework, we obtain:

$$\mathcal{V}(P, G_0) = \top \iff \text{PC using the corresponding orientation rule is consistent}$$

Property $\mathcal{A}$ being V-OUS and collider-stable

$\mathcal{A}(P, G) = \top$ if:

1. P is adjacency faithful to G.
2. P is V-OUS and collider-stable w.r.t. G:
   - **V-OUS**

     $$\text{non-collider } i - k - j \quad \Rightarrow \quad \forall C, \quad i \perp\!\!\!\perp j \mid C \Rightarrow i \perp\!\!\!\perp j \mid C \cup k$$

   - **Collider-stable**

     $$\text{collider } i - k - j \quad \Rightarrow \quad \exists C, k \notin C, \quad i \perp\!\!\!\perp j \mid C$$

We can see that $\mathcal{A}$ and $U_{\mathcal{A}}$:

1. neither implies or is implied by the SMR assumption
2. strictly weaker than *restricted faithfulness*.

Me-LoNS Algorithm

Modifi**e**d V-stable **Lo**calised **N**atural **S**tructure Learning algorithm:

1. Construct the skeleton

$$i - j \quad \xrightarrow{\exists C,\, i \perp\!\!\!\perp j \mid C} \quad i \cdots j$$

2. Orient the v-configurations

$$i - k - j \quad \xrightarrow{\exists C, \quad i \perp\!\!\!\perp j \mid C,\, i \not\perp\!\!\!\perp j \mid C \cup k} \text{collider}$$

$$\xrightarrow{\forall C,\, i \perp\!\!\!\perp j \mid C, \quad k \in C} \text{non-collider}$$

$$\xrightarrow{\text{otherwise}} \text{unassigned}$$

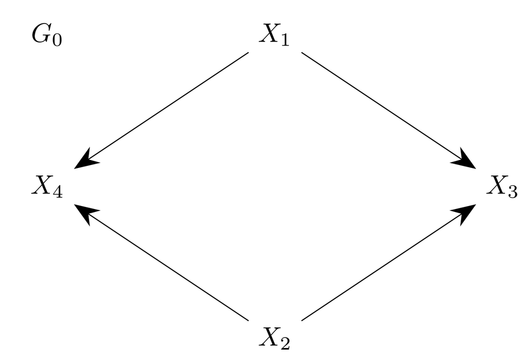
3. Solve for DAG

Simulation Comparisons

The setting for our simulations are as follows:

- We obtain 10,0000 samples from each SCM 100 times.
- We then implement Me-LoNS using Python package `causal-learn`.

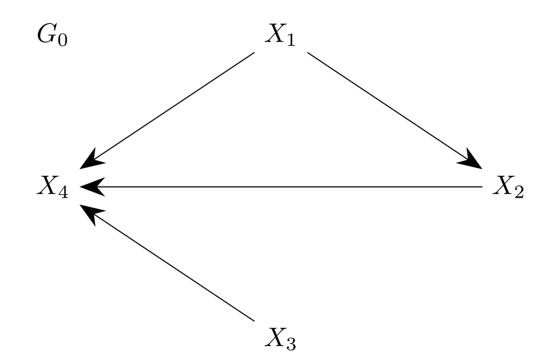
Comparison with PC



$$\epsilon_i \overset{\text{i.i.d.}}{\sim} N(0,1), i = 1, 2, 3, 4$$
$$X_1 = \epsilon_1$$
$$X_2 = \epsilon_2$$
$$X_3 = -6X_1 + 2X_2 + \epsilon_3$$
$$X_4 = 3X_1 + 4X_2 + \epsilon_4$$

PC	Me-LoNS
8%	90%

Comparison with SP



$$\epsilon_i, \phi_j \overset{\text{i.i.d.}}{\sim} Bern(\tfrac{1}{2}), i = 1, .., 4, j = 1, ..., 5$$
$$X_1 = (\phi_1, \phi_2, \epsilon_1)$$
$$X_2 = (X_1^1, \phi_3, \epsilon_2)$$
$$X_3 = (\phi_4, \phi_5, \epsilon_3)$$
$$X_4 = (X_1^1 + X_3^1, X_2^1 + X_3^2, X_2^2, \epsilon_4)$$

GRaSP	Me-LoNS
56%	94%

Conclusion

Generative assumptions

$\mathcal{A}(P, G_0) = \top$ and $U_{\mathcal{A}}(P) = \top$ $\Rightarrow$ Algorithm s.t. $\mathcal{A}(P, G(P)) = \top$ is consistent

Consistency conditions for algorithm — Output conditions for algorithm

By considering $\mathcal{A}$ and $U_{\mathcal{A}}$:

- SMR is the 'best' amongst minimality
- Exact conditions for PC

Sub $\mathcal{A}$ as:

1. Adjacency faithfulness
2. V-OUS + collider-stable

Me-LoNS algorithm (weak consistency conditions)