Causal Effects Identification with PAGs

Yimeng Chen

February 2020

Motivation

This work investigates the problem of known

- the observational distribution.
- a partial ancestral graph (PAG), describes a partial causal structure.

how to compute an experimental distribution (intervention distribution, $P_{\mathbf{x}}(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y} | do(\mathbf{X} = \mathbf{x}))$).

The reason to consider PAGs:

- a single, fully specified causal diagram is not always available in practical settings.
- PAGs can be consistently inferred from observational data.

Abstract

A brief introduction to

- A Graphical Criterion for Effect Identification in Equivalence Classes of Causal Diagrams, A. Jaber, JJ. Zhang, E. Bareinboim. IJCAI-18.
- Causal Identification under Markov Equivalence, A. Jaber, JJ. Zhang, E. Bareinboim. UAI-18. Best Student Paper Award (1 out of 337 papers).
- Causal Identification under Markov Equivalence: Completeness Results, A. Jaber, JJ. Zhang, E. Bareinboim. ICML-19.

based on

- On Causal Identification under Markov Equivalence, A. Jaber, JJ. Zhang, E. Bareinboim. IJCAI-19.

Main Contribution

- An atomic identification criterion.
- A novel graph-decomposition strategy.
- A complete algorithm to compute arbitrary *identifiable* $P_{\mathbf{x}}(\mathbf{y})$ from a PAG and observational data.

Contents

Structural Causal Models

Definition 1 (SCM)

An SCM M is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$.

1. $\mathbf{U}$ is a set of exogenous (latent) variables and $\mathbf{V}$ is a set of endogenous (measured) variables.
2. $\mathbf{F}$ represents a collection of functions $\{f_i\}$ s.t. for each variable $V_i \in \mathbf{V}$, $V_i = f_i(\mathbf{U_i} \cup \mathbf{Pa_i})$, $\mathbf{U_i} \subset \mathbf{U}, \mathbf{Pa_i} \subset \mathbf{V} \setminus \{V_i\}$.
3. $P(\mathbf{U})$ is a probability distribution over the latent variables.

- The marginal distribution induced over the measured variables $P(\mathbf{V})$ is called observational.
- The resulting distribution of $do(X = x)$ is denoted by P_x.

Causal diagram

Every SCM is associated with one acyclic causal diagram, where

- every variable $V_i \in \mathbf{V}$ is a node.
- there exists a directed edge from every node in $\mathbf{Pa_i}$ to V_i.
- for every pair $V_i, V_j \in \mathbf{V}$ such that $\mathbf{U_i} \cap \mathbf{U_j} \neq \varnothing$, there exists a bi-directed edge between V_i and V_j.

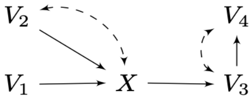

Figure 1: A causal diagram for example.

Ancestral Graphs

We now introduce graphical representations of equivalence classes of causal diagrams.
Starting from Mixed graphs, we introduce Maximal Ancestral Graph (MAG), and then the Markov equivalence class of MAGs called Partial Ancestral Graph (PAG).

- A MAG represents a set of causal diagrams with the same set of observed variables $\mathbf{V}$ that entail the same independence and ancestral relations among $\mathbf{V}$.
- A PAG represents a set of causal diagrams with the same observed variables $\mathbf{V}$ and independence model.

Mixed Graphs and Structures

A mixed graph can contain directed and bi-directed edges. The following are definitions of some structures in mixed graphs:

Notations: Structures in Mixed Graphs

- *spouse*: A is a spouse of B if $A \leftrightarrow B$ is present.
- *almost directed cycle*: happens when A is both a spouse and an ancestor of B.
- *inducing path*: a path on which every node (except for the endpoints) is a collider on the path and every collider is an ancestor of an endpoint of the path.

- *ancestral*: A mixed graph is *ancestral* if it does not contain a directed or almost directed cycle.
- *maximal*: if there is no inducing path between any two non-adjacent nodes.

MAG and PAG

Definition 2 (MAG)

A Maximal Ancestral Graph (MAG) is a mixed graph that is both ancestral and maximal.

The way to obtain a MAG from a causal DAG can be find in [1].

Definition 3 (PAG)

A partial ancestral graph (PAG) represents an equivalence class of MAGs $[\mathcal{M}]$, which shares the same adjacencies as every MAG in $[\mathcal{M}]$ and displays all and only the invariant edge marks.

A PAG is learnable from the conditional independence and dependence relations among the observed variables, and the FCI algorithm is a standard method to learn such an object.

Partial Ancestral Graph

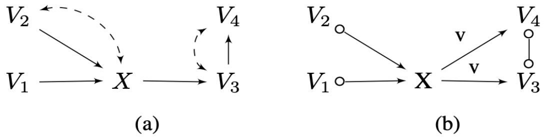

Figure 2: A causal diagram (left) and the inferred PAG (right).

- *circle*: indicates a edge mark that is not invariant.
- *visible*: A directed edge $X \to Y$ in a MAG or PAG is visible if every causal diagram in the corresponding equivalence class contains no inducing path between X and Y that is into X. *v* marks the visible edges.

Section conclusion

From this section, we know

- the relation between SCM and PAG:
  SCM $\Longrightarrow$ Causal Diagram $\Longrightarrow$ MAG $\Longrightarrow$ PAG
- The information declines along the arrows.
- PAG only keeps the observed variable set $\mathbf{V}$ and their independence relations.

In practice settings, we only know the observed variables $\mathbf{V}$, their distribution $P(\mathbf{V})$, and a PAG derived from the data.
Our target is to compute $P_{\mathbf{x}}(\mathbf{y})$ for abitrary $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$.

Why not to learn DAG

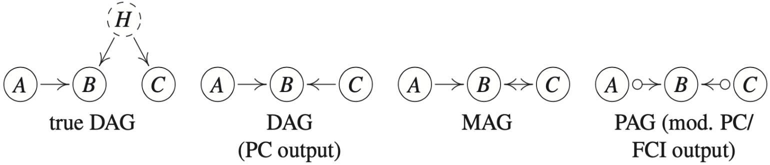

Figure 9.3: Starting with an SCM on the left-hand side, the three graphs on the right encode the set of conditional independences ($A \perp\!\!\!\perp C$). Due to an erroneous causal interpretation, the DAG is not desirable as an output of a causal learning method. In this example, the IPG and the latent projection (ADMG) are equal to the MAG.

Figure 3: From the book *Elements of Causal Inference*, Page 179. See page 180-184 for more information.

Section abstract

In this section, we introduce the notion of query distribution $Q[\mathbf{C}], \mathbf{C} \subset \mathbf{V}$, and a decomposition theorem for $Q[\mathbf{C}]$ based on the notion of *PC Components* and a property of them.

In fact, query distributions build our way to $P_{\mathbf{x}}(\mathbf{y})$:

Lemma 1

For arbitrary $\mathbf{X}, \mathbf{Y} \in \mathbf{V}$ and PAG $\mathcal{P}$,

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{d} \setminus \mathbf{y}} Q[\mathbf{D}]$$

where $\mathbf{D}$ is the set of possible ancestors of $\mathbf{Y}$ in $\mathcal{P}_{\mathbf{V} \setminus \mathbf{X}}$.

Subgraphs

Notations

- $\mathcal{D}_{\mathbf{C}}$: the (induced) subgraph of diagram $\mathcal{D}$ over $\mathbf{C}$.
- $\mathcal{P}_{\mathbf{C}}$: the induced subgraph of PAG $\mathcal{P}$ over $\mathbf{C}$.

Induced subgraphs of the original causal diagram play a critical role in identification. However, the subgraphs of a PAG is in general not a PAG.

In particular, if $\mathcal{D}$ is a diagram in the equivalence class represented by $\mathcal{P}$, $\mathcal{P}_{\mathbf{A}}$ is in general not the PAG that represents the equivalence class of $\mathcal{D}_{\mathbf{A}}$.

Read from Subgraph

In [2], they establish a few facts showing that some information about $\mathcal{D}_{\mathbf{A}}$ can still be read off from $\mathcal{P}_{\mathbf{A}}$.
The c-components of causal diagrams and their decomposition are the basis of Tian's identification algorithm.

Definition 4 (C Component)

In a causal diagram, two nodes are said to be in the same c-component if and only if they are connected by a bi-directed path, i.e., a path composed solely of bi-directed edges.

PC Component

As c-components in a causal diagram play a central role in identification, correspondingly, we consider pc-components in PAGs.

Definition 5 (PC Component)

In a MAG, a PAG, or any induced subgraph thereof, two nodes are in the same possible c-component (pc-component) if there is a path between them such that

- all non-endpoint nodes along the path are colliders,
- none of the edges is visible.

PC Component - Connection

The following proposition establishes a **graphical condition** in an induced sub-graph $\mathcal{P}_{\mathbf{A}}$ that is **sufficient** for two nodes **not** being in the same c-component in $\mathcal{D}_{\mathbf{A}}$ for any diagram $\mathcal{D}$ represented by $\mathcal{P}$.

Proposition 1

Let $\mathcal{P}$ be a PAG over $\mathbf{V}$, and $\mathcal{D}$ be any causal diagram in the equivalence class represented by $\mathcal{P}$. For any $X, Y \in \mathbf{A} \subset \mathbf{V}$, if X and Y are in the same c-component in $\mathcal{D}_{\mathbf{A}}$, then X and Y are in the same pc-component in $\mathcal{P}_{\mathbf{A}}$.

DC Component

As a special case of PC Component, we define the following notion, which will be used in the criterion Thm. 11.

Definition 6 (DC Component)

In a MAG, a PAG, or any induced subgraph thereof, two nodes are in the same definite c-component (dc-component) if they are connected with a bi-directed path, i.e. a path composed of bi-directed edges.

Query distribution

We introduce the notation of query distributions, which acts a key role in criterion 11 and the final algorithm.

Notation: Q[.]

$Q[\mathbf{C}]$: For any set $\mathbf{C} \subset \mathbf{V}$, the quantity $Q[\mathbf{C}]$ is defined to denote the post-intervention distribution of $\mathbf{C}$ under an intervention on $\mathbf{V} \setminus \mathbf{C}$, i.e. $P_{\mathbf{v} \setminus \mathbf{c}}(\mathbf{C})$.

Region

Next, we use the pc-component property to devise a de-composition for $Q[\mathbf{C}]$ in PAGs. We start by introducing the notion of a region.

Notation: bucket

If the edge marks on a path between X and Y are all circles, we call the path a *circle path*. We refer to the closure of nodes connected with circle paths as a *bucket*.

Definition 7 (Region $\mathcal{R}_{\mathbf{A}}^{\mathbf{C}}$)

Given a PAG or a MAG $\mathcal{G}$ over $\mathbf{V}$, and $\mathbf{A} \subset \mathbf{C} \subset \mathbf{V}$. The region of $\mathbf{A}$ with respect to $\mathbf{C}$, denoted $\mathcal{R}_{\mathbf{A}}^{\mathbf{C}}$, is defined as the union of the buckets that contain nodes in the pc-component of $\mathbf{A}$ in the induced subgraph $\mathcal{G}_{\mathbf{C}}$.

Decomposition

Using the construction of region, we derive the following decomposition theorem:

Theorem 8 (decomposition theorem)

Given a PAG $\mathcal{P}$ over $\mathbf{V}$ and set $\mathbf{C} \subset \mathbf{V}$, $Q[\mathbf{C}]$ decomposes as follows, where $\mathbf{A} \subset \mathbf{C}$ and $\mathcal{R}_{.} = \mathcal{R}^{\mathbf{C}}_{.}$.

$$Q[\mathbf{C}] = \frac{Q\left[\mathcal{R}_{\mathbf{A}}\right] \cdot Q\left[\mathcal{R}_{\mathbf{C} \setminus \mathcal{R}_{\mathbf{A}}}\right]}{Q\left[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C} \setminus \mathcal{R}_{\mathbf{A}}}\right]}$$

identifiable

First, we need to clarify which kind of $P_{\mathbf{x}}(\mathbf{y})$ is "identifiable" in our settings.

Definition 9 (identifiable in causal diagrams)

The causal effect of $\mathbf{X}$ on a disjoint set $\mathbf{Y}$ is said to be identifiable from a causal diagram $\mathcal{D}$ if $P_{\mathbf{x}}(\mathbf{y})$ can be computed uniquely from any positive probability of the observed variables $P(\mathbf{V})$.

Definition 10 (identifiable in PAGs)

Given a PAG $\mathcal{P}$ over $\mathbf{V}$ and a query $P_{\mathbf{x}}(\mathbf{y})$ where $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$, $P_{\mathbf{x}}(\mathbf{y})$ is identifiable given $\mathcal{P}$ iff $P_{\mathbf{x}}(\mathbf{y})$ is identifiable given every causal diagram $\mathcal{D}$ (represented by a MAG) in the Markov equivalence class represented by $\mathcal{P}$, and with the same expression.

An atomic identification criterion

Due to the absence of any causal information within a bucket in a PAG, the following criterion targets a bucket $\mathbf{X}$ rather than a single node.
A detailed discussion can be found in [1], [2].

An atomic identification criterion

Theorem 11 (criterion targets buckets)

Let $\mathcal{P}$ denote a PAG over $\mathbf{V}$, $\mathbf{T}$ be the union of a subset of the buckets in $\mathcal{P}$, and $\mathbf{X} \subset \mathbf{T}$ be a bucket. Given $\mathcal{P}$, and a partial topological order $\mathbf{B_1} < \cdots < \mathbf{B_m}$ with respect to $\mathcal{P}_{\mathbf{T}}$, $Q[\mathbf{T} \setminus \mathbf{X}]$ is identifiable iff, in $\mathcal{P}_{\mathbf{T}}$, $\nexists \mathbf{Z} \in \mathbf{X}$ s.t. $\mathbf{Z}$ has a possible child $C \notin \mathbf{X}$ that is in the pc-component of $\mathbf{Z}$. If identifiable, then

$$Q[\mathbf{T} \setminus \mathbf{X}] = \frac{P_{\mathbf{v} \setminus \mathbf{t}}}{\prod_{\{i | \mathbf{B_i} \subseteq S^{\mathbf{x}}\}} P_{\mathbf{v} \setminus \mathbf{t}} \left(\mathbf{B_i} | \mathbf{B^{(i-1)}} \right)} \sum_{\mathbf{x}} \prod_{\{i | \mathbf{B_i} \subseteq S^{\mathbf{X}}\}} P_{\mathbf{v} \setminus \mathbf{t}} \left(\mathbf{B_i} | \mathbf{B^{(i-1)}} \right)$$

where $S^{\mathrm{X}} = \bigcup_{Z \in \mathbf{X}} S^{Z}$, S^{Z} being the dc-component of Z in $\mathcal{P}_{\mathbf{T}}$, and $\mathbf{B}^{(i-1)}$ denoting the set of nodes preceding bucket $\mathbf{B}_i$ in the partial order.

Complete algorithm

Using the identification criterion in Thm. 11 and the decomposition in Thm. 8, a complete identification algorithm can be designed. Its completeness lies in

- For any identifiable $P_{\mathbf{x}}(\mathbf{y})$, the algorithm returns its expression.
- The algorithm returns `FAIL` iff $P_{\mathbf{x}}(\mathbf{y})$ is non-identifiable. In this case the effect is non-identifiable in some causal diagrams in the class.

See [3] for more information about the completeness and non-identifiability.

Algorithm Description

Key: recursive

1. After receiving the sets $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$, and a PAG $\mathcal{P}$, the algorithm computing $\mathbf{D}$, the set of possible ancestors of $\mathbf{Y}$ in $\mathcal{P}_{\mathbf{V} \setminus \mathbf{X}}$.
2. Calls the subroutine $\textsc{Identify}(\mathbf{D}, \mathbf{V}, P)$ over $\mathbf{D}$ to compute $Q[\mathbf{D}]$ from $P(\mathbf{V})$.
3. In subroutine $\textsc{Identify}(\mathbf{C}, \mathbf{T}, Q = Q[\mathbf{T}])$:
    1. Check the stop criterion: $\mathbf{C} = \emptyset$ or $\mathbf{T}$.
    2. For every bucket $\mathbf{B}$ in $\mathbf{T} \setminus \mathbf{C}$ Check the conditions of Theorem 11. If satisfied, compute $Q[\mathbf{T} \setminus \mathbf{B}]$, apply this subroutine on $\mathbf{T} \setminus \mathbf{B}$. (Shrink to a sub-graph)
    3. Check the condition of decomposition theorem 8. If satisfied, apply this subroutine to the compositions of $\mathbf{C}$. (Shrink to smaller query sets)

Notations in Algorithm

Notations

- *potentially directed path*: a path between X and Y is potentially directed (causal) from X to Y if there is no arrowhead on the path pointing towards X.
- $An(.)$: X is a possible ancestor of Y , i.e., $X \in An(Y)$, if there is a potentially directed path from X to Y.
- $Ch(.)$: Y is called a possible child of X, i.e. $Y \in Ch(X)$, if they are adjacent and the edge is not into X. For a set of nodes $\mathbf{X}$, we have $Ch(\mathbf{X}) = \bigcup_{X \in \mathbf{X}} Ch(X)$.

Algorithm

Algorithm 2 **IDP**$(\mathbf{x}, \mathbf{y})$ given PAG $\mathcal{P}$

Input: two disjoint sets $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$
Output: Expression for $P_{\mathbf{x}}(\mathbf{y})$ or `FAIL`

Let $\mathbf{D} = \texttt{An}(\mathbf{Y})_{\mathcal{P}_{\mathbf{V} \setminus \mathbf{x}}}$
$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{d} \setminus \mathbf{y}} \textsc{Identify}(\mathbf{D}, \mathbf{V}, P)$

function $\textsc{Identify}(\mathbf{C}, \mathbf{T}, Q = Q[\mathbf{T}])$
  if $\mathbf{C} = \emptyset$ **then return** 1
  if $\mathbf{C} = \mathbf{T}$ **then return** Q
    ▷ *In $\mathcal{P}_{\mathbf{T}}$, let $\mathbf{B}$ denote a bucket, and let $C^{\mathbf{B}}$ denote the pc-component of $\mathbf{B}$*
  if $\exists \mathbf{B} \subset \mathbf{T} \setminus \mathbf{C}$ such that $C^{\mathbf{B}} \cap \texttt{Ch}(\mathbf{B}) \subseteq \mathbf{B}$ **then**
    Compute $Q[\mathbf{T} \setminus \mathbf{B}]$ from Q (via Thm. 2)
    return $\textsc{Identify}(\mathbf{C}, \mathbf{T} \setminus \mathbf{B}, Q[\mathbf{T} \setminus \mathbf{B}])$
  else if $\exists \mathbf{B} \subset \mathbf{C}$ such that $\mathcal{R}_{\mathbf{B}} \neq \mathbf{C}$ **then**
    return $\frac{\textsc{Identify}(\mathcal{R}_{\mathbf{B}}, \mathbf{T}, Q) \cdot \textsc{Identify}(\mathcal{R}_{\mathbf{C} \setminus \mathcal{R}_{\mathbf{B}}}, \mathbf{T}, Q)}{\textsc{Identify}(\mathcal{R}_{\mathbf{B}} \cap \mathcal{R}_{\mathbf{C} \setminus \mathcal{R}_{\mathbf{B}}}, \mathbf{T}, Q)}$
  else
    throw `FAIL`

Figure 4: Identification procedure.

Reflections

- A potential problem in practice:
  Recall the definition of identifiable in PAGs: Is it a rare case?
- Whether a learned PAG ever contains non-trivial causal information?

For Further Reading

Identification of Conditional Causal Effects under Markov Equivalence, A. Jaber, J. Zhang, E. Bareinboim. NeurIPS-19.

Thanks for your attention!