Hidden Variables

Congyuan Duan

February 21, 2022

Contents

Contents

Causally sufficient

- A set of variables $\mathbf{X}$ is usually said to be **causally sufficient** if there is no hidden common cause $C \notin \mathbf{X}$ that is causing more than one variable in $\mathbf{X}$.
- A variable C is a common cause of X and Y if there is a directed path from C to X and Y that does not include Y and X, respectively.
- Common causes are also called confounders.

Interventional sufficiency

Definition 9.1 (Interventional sufficiency) *We call a set* $\mathbf{X}$ *of variables interventionally sufficient if there exists an SCM over* $\mathbf{X}$ *that cannot be falsified as an interventional model; that is, it induces observational and intervention distributions that coincide with what we observe in practice.*

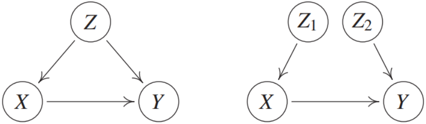
- Simpson's paradox shows that two variables are not causally sufficient if there exists a latent common cause, and in general these two variables are not interventionally sufficient either.
- There are also examples that the set of variables is interventionally sufficient but causally insufficient.

Example

Consider the following SCM

$$
\begin{aligned}
Z &:= N_Z \\
X &:= \mathbf{1}_{Z \geq 2} + N_X \\
Y &:= Z \bmod 2 + X + N_Y
\end{aligned}
$$

with $N_Z \sim U(\{0, 1, 2, 3\})$ and $N_X, N_Y \overset{iid}{\sim} N(0, 1)$.



- While variables X and Y are clearly causally insufficient, they interventionally sufficient.
- There cannot be a solely graphical criterion for determining whether a subset of the variables are interventionally sufficient.

Interventional sufficiency and causal sufficiency

In general, we have the following relationship between causal and interventional sufficiency.

Proposition 9.3 (Interventional sufficiency and causal sufficiency) *Let $\mathfrak{C}$ be an SCM for the variables $\mathbf{X}$ that cannot be falsified as an interventional model.*

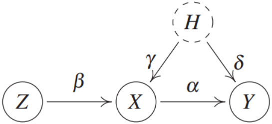
(i) If a subset $\mathbf{O} \subseteq \mathbf{X}$ is causally sufficient, then it is interventionally sufficient.

(ii) In general, the converse is false; that is, there are examples of interventionally sufficient sets $\mathbf{O} \subseteq \mathbf{X}$ that are not causally sufficient.

Contents

Instrumental Variables

Consider a linear Gaussian SCM with graph

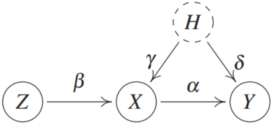


Here, the coefficient α in the structural assignment

$$Y := \alpha X + \delta H + N_Y$$

is the average causal effect.
However, simply regressing Y on X results in a biased estimator.

Instrumental Variables



Since (H, N_X) is independent of Z, we can have

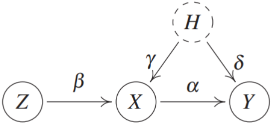
$$X := \beta Z + \gamma H + N_X$$

Because of

$$Y := \alpha X + \delta H + N_Y = \alpha(\beta Z) + (\alpha\gamma + \delta)H + N_Y$$

We can then consistently estimate α by regressing Y on βZ.

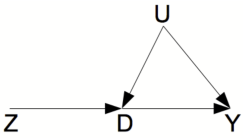
Instrumental Variables



We call a variable Z in an SCM an instrumental variable for (X, Y) if

- Z is independent of H.
- Z is not independent of X.
- Z effects Y only through X.

Instrumental Variables Under Potential Outcomes



Z:encouragement(randomly assigned) D:treatment Y:result

- As originally formulated, the potential outcomes $D_i(Z)$ and $Y_i(D)$ are fixed but unknown values partially observed through the assignment of treatments to units.
- We restricted both Z and D to have only two levels.

Instrumental Variables Under Potential Outcomes

Assumptions:

- Random Assignment: $Z_i \perp\!\!\!\perp \{D_i(1), D_i(0), Y_i(0), Y_i(1)\}$
- Exclusion Restriction: $Y(Z, D) = Y(Z^{'}, D)$
- Nonzero Average Causal Effect: $E[D_i(1) - D_i(0)]$ is not equalto zero
- Monotonicity: $D_i(1) \geq D_i(0)$ for all $i = 1, \cdots, N$

A variable Z is an instrumental variable for the causal effect of D on Y if the above assumptions hold.

Instrumental Variables Under Potential Outcomes

At the unit level, we have

$$
\begin{aligned}
& Y_i(1, D_i(1)) - Y_i(0, D_i(0)) \\
& \quad = Y_i(D_i(1)) - Y_i(D_i(0)) \\
& \quad = [Y_i(1) \cdot D_i(1) + Y_i(0) \cdot (1 - D_i(1))] \\
& \qquad - [Y_i(1) \cdot D_i(0) + Y_i(0) \cdot (1 - D_i(0))] \\
& \quad = (Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0)).
\end{aligned}
$$

Thus the causal effect of Z on Y for person i is the product of (i)the causal effect of D on Y and (ii)the causal effect of Z on D.

Instrumental Variables Under Potential Outcomes

We can therefore write the average causal effect of Z on Y as:

$$
\begin{aligned}
&E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))] \\
&\quad = E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \\
&\quad = E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = 1] \\
&\qquad\qquad\qquad \cdot P[D_i(1) - D_i(0) = 1] \\
&\qquad - E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = -1] \\
&\qquad\qquad\qquad \cdot P[D_i(1) - D_i(0) = -1]. \qquad (10)
\end{aligned}
$$

Use the monotonicity assumption,

$$
\begin{aligned}
&E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)] \\
&\quad = E[(Y_i(1) - Y_i(0)) | D_i(1) - D_i(0) = 1] \\
&\qquad\qquad\qquad \cdot P[D_i(1) - D_i(0) = 1].
\end{aligned}
$$

Instrumental Variables Under Potential Outcomes

The causal interpretation of the IV estimator is

$$\begin{aligned}
\hat{\beta}^{IV} &= E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \\
&= \frac{E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)]}{E[D_i(1) - D_i(0)]} \\
&= \frac{ACE(Z \rightarrow Y)}{ACE(Z \rightarrow D)}
\end{aligned}$$

We call this the Local Average Treatment Effect or Complier Average Causal Effect.

Contents

Difficulty

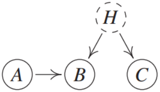
We want to obtain identifiability results for an SCM $\mathfrak{C}$ over variables $\mathbf{X} = (\mathbf{O}, \mathbf{H})$ that includes observed variables $\mathbf{O}$ and hidden variables $\mathbf{H}$.
This comes with additional difficulties when searching over the space of DAGs with latent variables.

- We do not know the size of $\mathbf{H}$, so there is an infinite number of graphical candidates that we have to search over.
- The set of distributions that are Markovian and faithful with respect to a DAG forms a curved exponential family, which justifies the use of the BIC. The set of distributions that are Markovian and faithful with respect to a DAG with latent variables, however, does not.

Difficulty

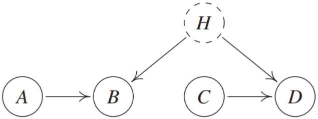
Can we assume that the entailed distribution P_O is Markovian and faithful with respect to a DAG without hidden variables instead?

- Representing the set of conditional independences with a DAG over the observed variables can lead to causal misinterpretations.
- The set of distributions whose pattern of independences correspond to the d-separation statements in a DAG is not closed under marginalization.



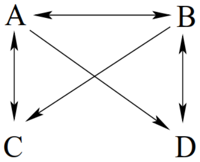
true DAG

DAG
(PC output)



Maximal Ancestral Graph

- An inducing path relative to L is a path on which every vertex not in L (except for the endpoints) is a collider on the path and every collider is an ancestor of an endpoint of the path.
- We refer to inducing paths relative to the empty set simply as inducing paths.
- A directed cycle occurs in $\mathcal{G}$ when $Y \rightarrow X$ is in $\mathcal{G}$ and $X \in An(Y)$.
- An almost directed cycle occurs when $Y \leftrightarrow X$ is in $\mathcal{G}$ and $X \in An(Y)$.

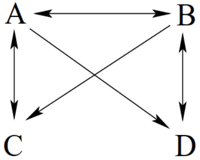


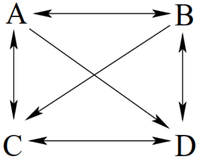
Maximal Ancestral Graph

Definition 1 (MAG) *A mixed graph is called a maximal ancestral graph (MAG) if*

i. the graph does not contain any directed or almost directed cycles (ancestral); and

ii. there is no inducing path between any two non-adjacent vertices (maximal).





m-separation

It is straightforward to extend d-separation to mixed graphs.

Definition 2 (m-separation) *In a mixed graph, a path p between vertices X and Y is active (or m-connecting) relative to a (possibly empty) set of vertices $\mathbf{Z}$ ($X, Y \notin \mathbf{Z}$) if*

i. every non-collider on p is not a member of $\mathbf{Z}$;

ii. every collider on p is an ancestor of some member of $\mathbf{Z}$.

X and Y are said to be m-separated by $\mathbf{Z}$ if there is no active path between X and Y relative to $\mathbf{Z}$.

Two disjoint sets of variables $\mathbf{X}$ and $\mathbf{Y}$ are m-separated by $\mathbf{Z}$ if every variable in $\mathbf{X}$ is m-separated from every variable in $\mathbf{Y}$ by $\mathbf{Z}$.

This captures exactly the conditional independence relations entailed by a MAG according to the Markov condition.

Maximal Ancestral Graph

MAGs represent the marginal independence models of DAGs: given any DAG $\mathcal{G}$ over $V = O \cup H$, there is a MAG over O alone such that for any disjoint $X, Y, Z \subseteq O$, X and Y are d-separated by Z in $\mathcal{G}$ if and only if they are m-separated by Z in the MAG.
The following construction gives us such a MAG $M_{\mathcal{G}}$:

- For each pair of variables $A, B \in O$, A and B are adjacent in $M_{\mathcal{G}}$ if and only if there is an inducing path between them relative to H in $\mathcal{G}$.
- For each pair of adjacent variables $A, B \in M_{\mathcal{G}}$, orient the edge as $A \rightarrow B$ if A is an ancestor of B in $\mathcal{G}$; orient it as $A \leftarrow B$ if B is an ancestor of A; orient it as $A \leftrightarrow B$ otherwise.

It can be shown that $M_{\mathcal{G}}$ is a MAG and represents the marginal independence model over O.

Maximal Ancestral Graph

- Different causal DAGs may correspond to the same causal MAG.
- The MAG retains the causal relationships among the observed variables.
- Alternative definition: an ancestral graph is said to be maximal if, for every pair of non-adjacent nodes X, Y there exists a set Z such that X and Y are m-separated conditional on Z.

Just as DAGs, different MAGs can entail the exact same constraints by the m-separation criterion. This motivates the following representation of equivalence classes of MAGs.

Partial Ancestral Graph

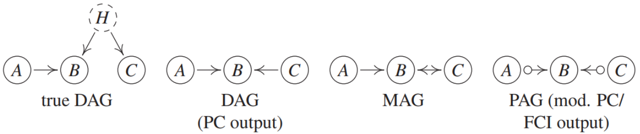
Definition 3 (PAG) *Let $[\mathcal{M}]$ be the Markov equivalence class of an arbitrary MAG $\mathcal{M}$. The partial ancestral graph (PAG) for $[\mathcal{M}]$, $\mathcal{P}_{[\mathcal{M}]}$, is a partial mixed graph such that*

i. *$\mathcal{P}_{[\mathcal{M}]}$ has the same adjacencies as $\mathcal{M}$ (and any member of $[\mathcal{M}]$) does;*

ii. *A mark of arrowhead is in $\mathcal{P}_{[\mathcal{M}]}$ if and only if it is shared by all MAGs in $[\mathcal{M}]$; and*

iii. *A mark of tail is in $\mathcal{P}_{[\mathcal{M}]}$ if and only if it is shared by all MAGs in $[\mathcal{M}]$.*[8]

In PAGs, edges can end with a circle, which represents both possibilities of an arrow's head and tail.



Summary

Graphical object	DAG (without hiddens)	MAG	IPG	ADMG (with nested Markov)
Type of edges directed / undir. / bidir. / combination	✓/ - / - / -	✓/ ✓/ ✓/ -	✓/ - / ✓/ -	✓/ - / ✓/ ✓
Correct causal interpretation	✗	✓	✓	✓
Graphical separation for global Markov	d-seperation	m-seperation	m-seperation	m-seperation
Criterion for valid adjustment sets	✓	✓	?	✓
Algorithm for identification of intervention distribution	✓	?	?	✓
Representation of equivalence class	CPDAG (Markov)	PAG (Markov)	POIPG (Markov)	? (nested Markov)
Independence-based method for learning	PC, IC, SGS	FCI	FCI	-
Score-based method for learning	GDS, GES	for linear/binary/ discrete SCMs	?	for binary/ discrete SCMs
Can encode all equality constraints	✗	✗	✗	✓ (if obs. var. are discrete)
Can encode all constraints	✗	✗	✗	✗

Table 9.1: Consider an SCM over (observed) variables $\mathbf{O}$ and (hidden) variables $\mathbf{H}$ that induces a distribution $P_{\mathbf{O},\mathbf{V}}$. How do we model the observed distribution $P_{\mathbf{O}}$? We would like to use an SCM with (arbitrarily many) latent variables. This model class, however, has bad properties for causal learning. This table summarizes some alternative model classes (current research focuses especially on MAGs and ADMGs).

Reference

Angrist J D, Imbens G W, Rubin D B. Identification of causal effects using instrumental variables[J]. Journal of the American statistical Association, 1996, 91(434): 444-455.
Zhang J. Causal reasoning with ancestral graphs[J]. Journal of Machine Learning Research, 2008, 9: 1437-1474.