Estimating causal effects with ancestral graph Markov models

Daniel Malinsky and Peter Spirtes

Carnegie Mellon University
malinsky@cmu.edu

September 6th, 2016

Causal structures and causal effects

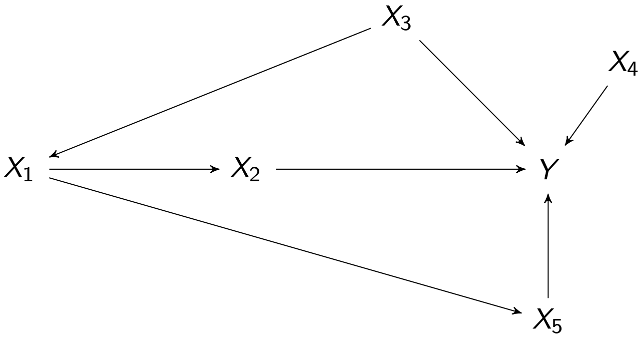

Figure: A Directed Acyclic Graph (DAG)

Want to estimate the causal effect (a.k.a. intervention effect) of X_i on Y from passively observed (non-experimental) data.

The total causal effect of $\text{do}(X_i = x_i')$ on Y is $\frac{\partial}{\partial x}\mathbf{E}(Y|\text{do}(X_i = x))|_{x=x'}$.

What if the causal structure is underdetermined by the data?

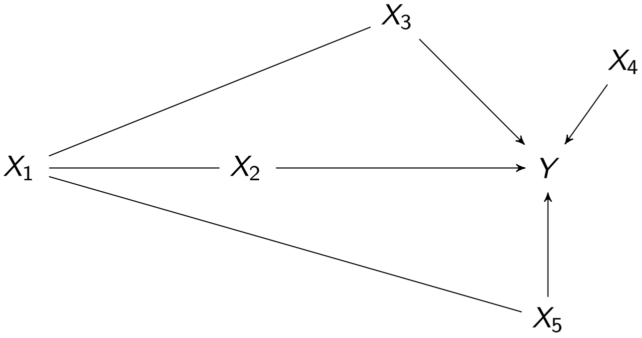

Figure: A Pattern (or CPDAG) representing a Markov equivalence class of DAGs

The IDA Method

Maathuis, Kalisch, and Bühlmann (2009)

Goal: To estimate the total causal effect of each covariate X_i $\{i = 1, ..., p\}$ on some target variable Y.

Motivation: We'd like to pick out candidates for intervention by (for example) looking at the covariates with the largest (absolute value) casual effects. MKB were doing this for genetic data, where the number of variables is very large.

Assumptions

- They assume the data is generated from an unknown DAG over the measured variables
  - no directed cycles (no "feedback")
  - no unmeasured common causes (a.k.a. causal sufficiency)*
- Causal Markov Condition and Faithfulness.
- They assume variables are jointly Gaussian** (and thus linear***).
  - $Y = \alpha X_1 + \beta X_2 + \gamma X_3 + ... + \varepsilon$

*This assumption will be relaxed by our method

Assumptions

- They assume the data is generated from an unknown DAG over the measured variables
  - no directed cycles (no "feedback")
  - no unmeasured common causes (a.k.a. causal sufficiency)*
- Causal Markov Condition and Faithfulness.
- They assume variables are jointly Gaussian** (and thus linear***).
  - $Y = \alpha X_1 + \beta X_2 + \gamma X_3 + ... + \varepsilon$

*This assumption will be relaxed by our method

**This assumption can be relaxed relatively easily, with conditional independence tests for non-Gaussian data

Assumptions

- They assume the data is generated from an unknown DAG over the measured variables
  - no directed cycles (no "feedback")
  - no unmeasured common causes (a.k.a. causal sufficiency)*
- Causal Markov Condition and Faithfulness.
- They assume variables are jointly Gaussian** (and thus linear***).
  - $Y = \alpha X_1 + \beta X_2 + \gamma X_3 + ... + \varepsilon$

*This assumption will be relaxed by our method
**This assumption can be relaxed relatively easily, with conditional independence tests for non-Gaussian data
***This assumption can be relaxed with some more work, can use non-parameteric regression techniques

General picture of their method

Input: An equivalence class of DAGs, covariance matrix for $X_1, ..., X_p, Y$
Output: The causal effect of each X_i on Y, for each DAG in the equivalence class

- Their (global) algorithm begins by running PC to search for a Pattern (or CPDAG), which is an equivalence class of DAGs.
- Then enumerate all the DAGs represented by that CPDAG.
- For each DAG, perform the appropriate regression to calculate the causal effect.
- They also have a **local** method which does not require enumerating all the DAGs, which is much faster.

The backdoor criterion

The reason we estimate causal effects using graphical models and not just simple regression is because we do not know in advance **which covariates to adjust for**, i.e., which variables to condition on in our estimate.

The backdoor criterion

The reason we estimate causal effects using graphical models and not just simple regression is because we do not know in advance **which covariates to adjust for**, i.e., which variables to condition on in our estimate.

For a given causal graph, we can estimate the causal effect of X on Y by regressing Y on X along with its **backdoor set**. This is due to Pearl (2009).

The backdoor criterion

The reason we estimate causal effects using graphical models and not just simple regression is because we do not know in advance **which covariates to adjust for**, i.e., which variables to condition on in our estimate.

For a given causal graph, we can estimate the causal effect of X on Y by regressing Y on X along with its **backdoor set**. This is due to Pearl (2009).

In a DAG, we can estimate the causal effect of X on Y by regressing Y on X and the parents of X. The parents of X are the backdoor set.

The backdoor criterion

The reason we estimate causal effects using graphical models and not just simple regression is because we do not know in advance **which covariates to adjust for**, i.e., which variables to condition on in our estimate.

For a given causal graph, we can estimate the causal effect of X on Y by regressing Y on X along with its **backdoor set**. This is due to Pearl (2009).

In a DAG, we can estimate the causal effect of X on Y by regressing Y on X and the parents of X. The parents of X are the backdoor set.

The local IDA algorithm exploits the fact that we do not need to know the entire graphical structure to calculate a causal effect, we only need to know which set satisfies the backdoor criterion in each graph in the equivalence class.

Relaxing Causal Sufficiency

Causal Sufficiency = the assumption that there are no unmeasured common causes

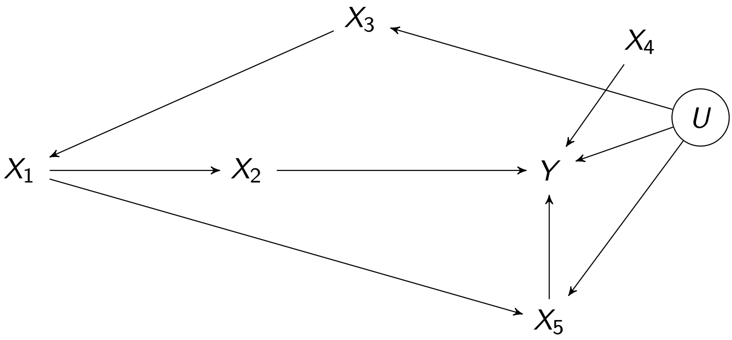

Figure: A DAG which includes the latent (unmeasured) variable U

Relaxing Causal Sufficiency Makes a Difference

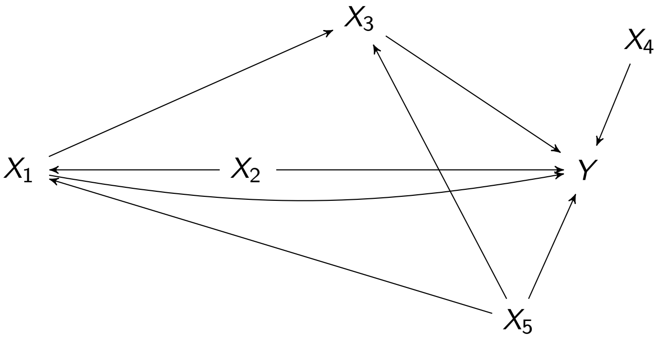

Figure: What you get running PC

Latent Variable IDA: using ancestral graphical models

The FCI algorithm[1] does not assume causal sufficiency. The output of FCI is an equivalence class of MAGs (Maximal Ancestral Graphs).

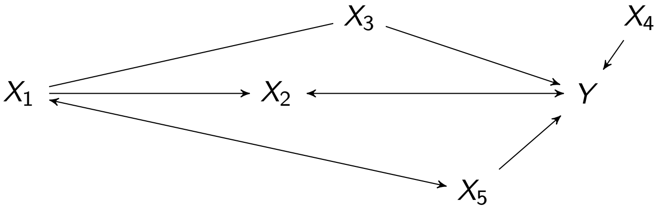

Figure: A MAG with with directed, undirected, and double-headed edges

[1]See also RFCI and GFCI. The latter algorithm will be discussed later in this session.

Reasoning with PAGs

A PAG (Partial Ancestral Graph) is represents a Markov equivalence class of MAGs.

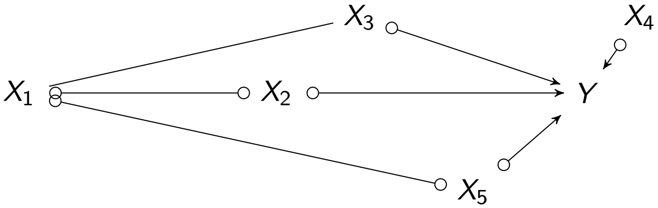

Figure: A PAG with with directed, undirected, partially directed, and double-headed edges

Backdoor sets for MAGs and PAGs

Finding the backdoor sets in a MAG (or a PAG) is trickier. Colombo and Maathuis (2015) provide general conditions for backdoor sets which are sufficient for adjustment.

Backdoor sets for MAGs and PAGs

Finding the backdoor sets in a MAG (or a PAG) is trickier. Colombo and Maathuis (2015) provide general conditions for backdoor sets which are sufficient for adjustment.

Though we can sometimes identify or rule out latent variables on the basis of observed independence facts, sometimes we cannot. Because there may be latent variables lurking, not all the causal effects will be identifiable.

Backdoor sets for MAGs and PAGs

Finding the backdoor sets in a MAG (or a PAG) is trickier. Colombo and Maathuis (2015) provide general conditions for backdoor sets which are sufficient for adjustment.

Though we can sometimes identify or rule out latent variables on the basis of observed independence facts, sometimes we cannot. Because there may be latent variables lurking, not all the causal effects will be identifiable.

LV-IDA carries out the same logic as IDA: it begins with an equivalence class of (ancestral) graphs, and determines which covariates are in the backdoor set for each graph in the equivalence class to estimate a causal effect.

LV-IDA

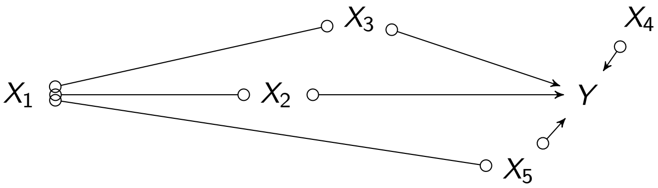

Figure: A PAG, which is an equivalence class of MAGs

Enumerate the MAGs and calculate causal effects

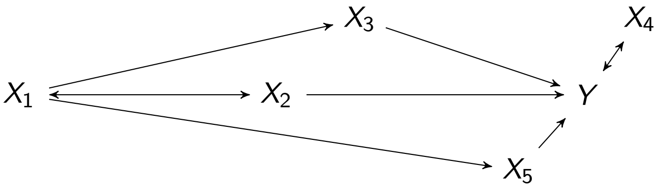

Figure: A MAG in the equivalence class

Local version

There is a Local version of the algorithm which takes advantage of the same principle as before: we only really need to know the possible backdoor sets to calculate causal effects.

Local version

There is a Local version of the algorithm which takes advantage of the same principle as before: we only really need to know the possible backdoor sets to calculate causal effects.

To do that, we only need to enumerate all possible MAGs over a **subset** of the full variable set (those variables in the original PAG which may constitute a possible backdoor set).

Local version

There is a Local version of the algorithm which takes advantage of the same principle as before: we only really need to know the possible backdoor sets to calculate causal effects.

To do that, we only need to enumerate all possible MAGs over a **subset** of the full variable set (those variables in the original PAG which may constitute a possible backdoor set).

This is an improvement in terms of speed, but still could take a long time if the graph is very big and very highly connected. In very high-dimensional settings, the algorithm could be applied to the Markov blanket of a variable of interest.

Simulations: infinite sample limit

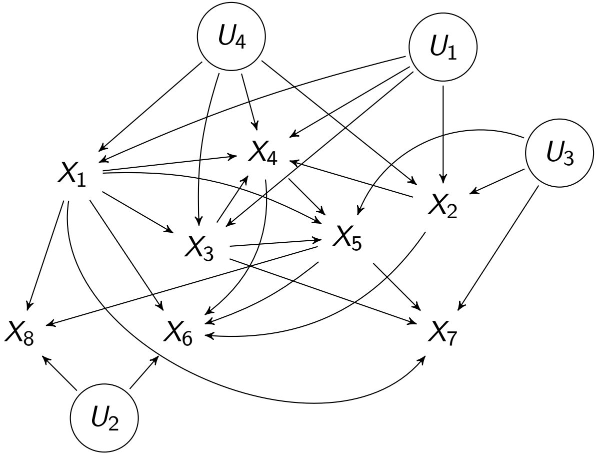

Figure: The true causal effects of X_5 on X_6 and X_5 on X_7 are 0.894 and 1.143, respectively. LV-IDA produces the estimates $\{NA, 0.894, 1.345, 1.707\}$ and $\{NA, 0, 1.143, 1.662\}$, respectively. IDA produces the estimates $\{1.345, 1.481\}$ and $\{1.603, 1.662\}$, respectively.

Simulations: finite samples

We generated 100 random sparse DAGs with 15 variables, 4 or 5 of which are latent. Linear Gaussian parameterization. $n = 1000$. Used GFCI for structure search.

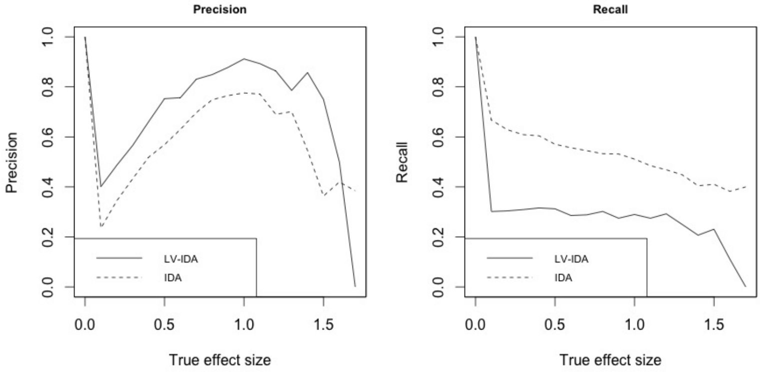

Figure: Precision and recall for finding large causal effects using LV-IDA and IDA.

Conclusion

- We combine structure learning for ancestral graph Markov models (using FCI, RFCI, or GFCI) with a procedure which enumerates graphs in an equivalence class, and determines appropriate adjustment sets for estimating causal effects.
- In the end, we estimate a sets of causal effects which can be used as bounds on the true effect.
- The set may include an indication ("NA") that the effect is not identifiable in some/all graphs in the equivalence class.
- The procedure can be used to find candidates for follow-up experiments by (for example) focusing on variables with large minimum effect values.

Thank you!

R code for LV-IDA is available online: `https://github.com/dmalinsk/lv-ida`

References:

M. H. Maathuis and D. Colombo. A generalized back-door criterion. The Annals of Statistics, 43 (3):10601088, 2015.

M. H. Maathuis, M. Kalisch, and P. Bühlmann. Estimating high-dimensional intervention effects from observational data. The Annals of Statistics, 37(6A):31333164, 2009.

J. M. Ogarrio, P. Spirtes, and J. D. Ramsey. A hybrid causal search algorithm for latent variable models. JMLR: W&CP (PGM 16), 52:368379, 2016.

J. Pearl. *Causality: Models, reasoning, and inference* (Second ed.). Cambridge: Cambridge University Press, 2009.

T. Richardson and P. Spirtes. Ancestral graph Markov models. The Annals of Statistics, 30(4): 9621030, 2002.

P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search.* MIT Press, 2000.

Appendix

Unidentifiable causal effects

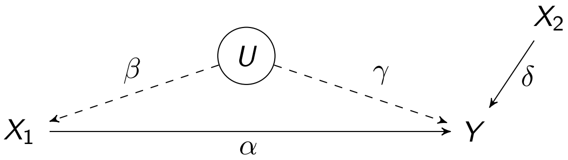

Figure: The $X_1 \to Y$ edge is **invisible** and the model is compatable with an unmeasured confounder U. LV-IDA would not identify the intervention effect α; it would return "NA".

Markov equivalence

The following three DAGs represent the same conditional independence facts: $X_1 \rightarrow Y \rightarrow X_2$, $X_1 \leftarrow Y \leftarrow X_2$, and $X_1 \leftarrow Y \rightarrow X_2$. They are represented by the Pattern $X_1 - Y - X_2$.

$X_1 \rightarrow Y \leftarrow X_2$ is not Markov equivalent, because this graph implies that X_1 is dependent on X_2 conditional on Y.

Definition

(*Visible and invisible edges*) *All directed edges in DAGs and CPDAGs are said to be visible. Given a MAG* $\mathcal{M}$ / *PAG* $\mathcal{P}$*, a directed edge* $A \to B$ *in* $\mathcal{M}$ / $\mathcal{P}$ *is visible if there is a vertex* C *not adjacent to* B*, such that there is an edge between* C *and* A *that is into* A*, or there is a collider path between* C *and* A *that is into* A *and every non-endpoint vertex on the path is a parent of* B*. Otherwise* $A \to B$ *is said to be invisible.*

Definition

($D\text{-}SEP(X, Y, \mathcal{G})$) *Let* X *and* Y *be two distinct vertices in mixed graph* $\mathcal{G}$*. We say that* $V \in D\text{-}SEP(X, Y, \mathcal{G})$ *if* $V \neq X$ *and there is a collider path between* X *and* V *in* $\mathcal{G}$*, such that every vertex on this path is an ancestor of* X *or* Y *in* $\mathcal{G}$*.*

Definition

($\mathcal{R}$ and $\mathcal{R}_{\underline{X}}$) *Let X be a vertex in $\mathcal{G}$, where $\mathcal{G}$ represents a causal DAG, CPDAG, MAG, or PAG. Let $\mathcal{R}$ be a DAG or MAG represented by $\mathcal{G}$, in the following sense. If $\mathcal{G}$ is a DAG or MAG, we simply let $\mathcal{R} = \mathcal{G}$. If $\mathcal{G}$ is a CPDAG/PAG, we let $\mathcal{R}$ be a DAG/MAG in the Markov equivalence class described by $\mathcal{G}$ with the same number of edges into X as $\mathcal{G}$. Let $\mathcal{R}_{\underline{X}}$ be the graph obtained from $\mathcal{R}$ by removing all directed edges out of X that are visible in $\mathcal{P}$.*

Theorem

(*Back-door Set*) *Let X and Y be two distinct vertices in a causal DAG, CPDAG, MAG, or PAG $\mathcal{G}$. Let $\mathcal{R}$ and $\mathcal{R}_{\underline{X}}$ be defined as above. If $Y \in adj(X, \mathcal{R}_{\underline{X}})$ or $D\text{-}SEP(X, Y, \mathcal{R}_{\underline{X}}) \cap possibleDe(X, \mathcal{G}) \neq \emptyset$, then $f(y|do(x))$ is not identifiable via the generalized back-door criterion. Otherwise $D\text{-}SEP(X, Y, \mathcal{R}_{\underline{X}})$ satisfies the generalized back-door criterion relative to (X, Y) and $\mathcal{G}$.*

IDA

Algorithm 0.1: IDA(“*global*”)

Input: CPDAG $\mathcal{G}$, conditional dependencies of $X_1, ..., X_p, Y$
Output: Matrix Θ of possible causal effects
1. List the DAGs $\mathcal{G}_1, ..., \mathcal{G}_m$ in the equivalence class of $\mathcal{G}$.
2. **for** $j = 1$ **to** m
3.   **for** $i = 1$ **to** p
4. $\theta_{ij} = \beta_{i|pa(X_i,\mathcal{G}_j)}$
5.   **end**
6. **end**

IDA (local)

Algorithm 0.2: IDA(“*local*”)

Input: CPDAG $\mathcal{G}$, conditional dependencies of $X_1, ..., X_p, Y$
Ouput: Multisets Θ_i^L, $i = 1, ..., p$
1. **for** $i = 1$ **to** p
2. $\Theta_i^L = \emptyset$
3. **for each** subset S of $sib(X_i, \mathcal{G})$
4. **if** $\mathcal{G}_{S \to i}$ is locally valid (i.e., has no new v-structure with collider X_i
5. **then** add $\beta_{i|pa(X_i,\mathcal{G}) \cup S}$ to Θ_i^L
6. **end**
7. **end**

A straightforward analog to ("global") IDA

Algorithm 0.3: LV-IDA("*global*")

Input: PAG $\mathcal{P}$, conditional dependencies of $X_1, ..., X_p, Y$

Output: Matrix Θ of possible causal effects

1. List the MAGs $\mathcal{M}_1, ..., \mathcal{M}_n$ in the equivalence class of $\mathcal{P}$.

2. **for** $j = 1$ **to** n

3.   **for** $i = 1$ **to** p

4.   **if** $Y \notin De(X_i, \mathcal{M}_j)$ **then** $\theta_{ij} = 0$

5.   **if** $Y \in adj(X_i, \mathcal{M}_{j,\underline{X_i}})$ or $\text{D-SEP}(X_i, Y, \mathcal{M}_{j,\underline{X_i}}) \cap De(X_i, \mathcal{M}_j) \neq \emptyset$

6.   **then** $\theta_{ij} =$ "NA"

7.   **else** $\begin{cases} S = \text{D-SEP}(X_i, Y, \mathcal{M}_{j,\underline{X_i}}) \\ \theta_{ij} = \beta_{i|S} \end{cases}$

8.   **end**

9. **end**

Algorithm 0.4: LV-IDA(*"local"*)

Input: PAG $\mathcal{P}$, conditional dependencies of $X_1, ..., X_p, Y$
Output: Multisets $\Theta_i^L, i = 1, ..., p$
1. **for** $i = 1$ **to** p
2. Form the set $\mathbf{Z_i} = possibleDe(X_i, \mathcal{P}) \cup pds(X_i, Y, \mathcal{P})$.
3. Form $\mathcal{P}^*$, the subgraph of $\mathcal{P}$ over vertices $\mathbf{Z_i}$.
4. List the MAGs $\mathcal{M}_1, ..., \mathcal{M}_m$ represented by $\mathcal{P}^*$.
5. **for** $k = 1$ **to** m
6. **if** $Y \notin De(X_i, \mathcal{M}_k)$ **then** add $\theta_{ik} = 0$ to Θ_i^L
7. **if** $Y \in adj(X_i, \mathcal{M}_{k,\underline{X_i}})$
or D-SEP$(X_i, Y, \mathcal{M}_{k,\underline{X_i}}) \cap De(X_i, \mathcal{M}_k) \neq \emptyset$
8. **then** add $\theta_{ik} =$ "NA" to Θ_i^L
9. **else** $\begin{cases} S = \text{D-SEP}(X_i, Y, \mathcal{M}_{k,\underline{X_i}}) \\ \text{add } \theta_{ik} = \beta_{i|S} \text{ to } \Theta_i^L \end{cases}$
10. **end**
11. **end**