Constraints on marginalised DAGs.

Robin J. Evans
www.statslab.cam.ac.uk/∼rje42

Statistical Laboratory, University of Cambridge

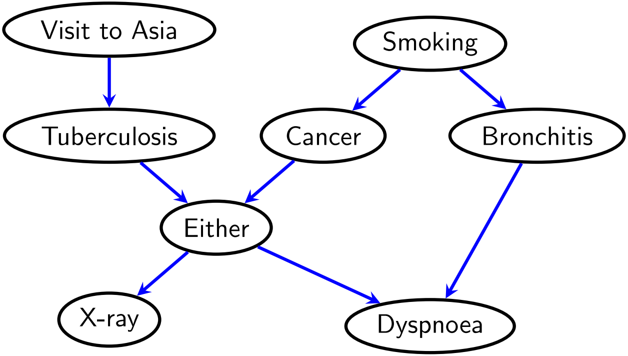
Fields Institute, Toronto
17th April 2012

Outline

Outline

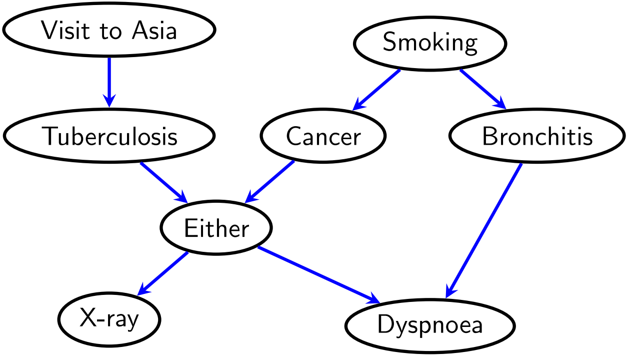
Implications of Models

It is common (especially in causal settings) to hypothesise models based on DAGs:



Implications of Models

It is common (especially in causal settings) to hypothesise models based on DAGs:



This encodes the assumption that the joint distribution factorises as:

$$p(A)\,p(T \mid A)\,p(S)\,p(C \mid S)\,p(B \mid S)\,p(E \mid T, C)\,p(X \mid E)\,p(D \mid E, B).$$

d-Separation

The factorisation criterion

$$p(x_V) = \prod_{v \in V} p(x_v \mid x_{\mathrm{pa}_{\mathcal{G}}(v)})$$

is equivalent to the global Markov property:

$$A \text{ d-separated from } B \text{ by } C \implies X_A \perp\!\!\!\perp X_B \mid X_C \, [P].$$

d-Separation

The factorisation criterion

$$p(x_V) = \prod_{v \in V} p(x_v \mid x_{\mathrm{pa}_{\mathcal{G}}(v)})$$

is equivalent to the global Markov property:

$$A \text{ d-separated from } B \text{ by } C \implies X_A \perp\!\!\!\perp X_B \mid X_C \, [P].$$

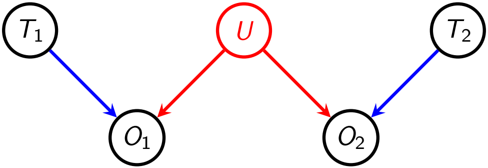
In particular, all constraints on DAGs are conditional independences.

Marginalisation

We often cannot observe all variables.

Marginalisation

We often cannot observe all variables. Consider a medical trial on pairs of siblings:



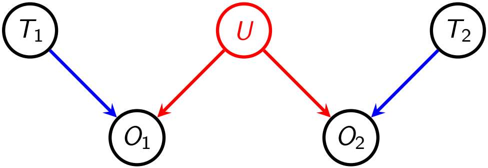
T_i (randomised) treatment for sibling $i = 1, 2$

O_i recorded outcome for sibling i

U **unmeasured** confounding (genetics etc.)

Marginalisation

We often cannot observe all variables. Consider a medical trial on pairs of siblings:



T_i (randomised) treatment for sibling $i = 1, 2$

O_i recorded outcome for sibling i

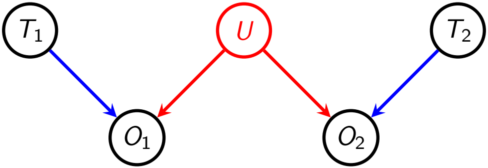
U **unmeasured** confounding (genetics etc.)

Can only observe independences which don't involve U:

$$T_1 \perp\!\!\!\perp O_2, T_2 \qquad\qquad T_2 \perp\!\!\!\perp O_1, T_1.$$

Marginalisation

We often cannot observe all variables. Consider a medical trial on pairs of siblings:



T_i (randomised) treatment for sibling $i = 1, 2$

O_i recorded outcome for sibling i

U **unmeasured** confounding (genetics etc.)

Can only observe independences which don't involve U:

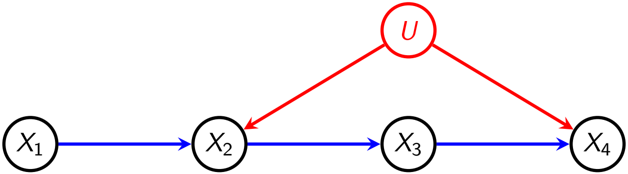
$$T_1 \perp\!\!\!\perp O_2, T_2 \qquad\qquad T_2 \perp\!\!\!\perp O_1, T_1.$$

Is this all?

Outline

Example 1: Verma Graph

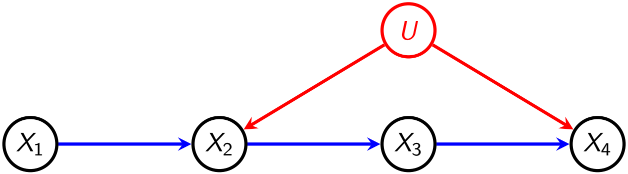
Consider the following DAG on 5 variables (Verma and Pearl, 1990).



$$X_1 \perp\!\!\!\perp U, \qquad X_3 \perp\!\!\!\perp X_1, U \mid X_2, \qquad X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3, U.$$

Example 1: Verma Graph

Consider the following DAG on 5 variables (Verma and Pearl, 1990).

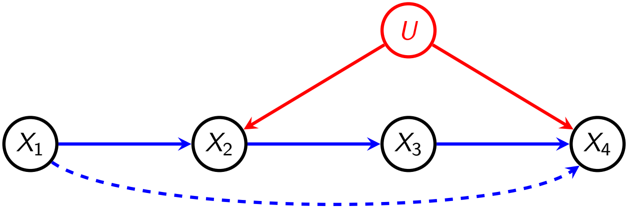


$$X_1 \perp\!\!\!\perp U, \qquad X_3 \perp\!\!\!\perp X_1, U \mid X_2, \qquad X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3, U.$$

If U is latent, we can only observe $X_3 \perp\!\!\!\perp X_1 \mid X_2$.

Example 1: Verma Graph

Consider the following DAG on 5 variables (Verma and Pearl, 1990).

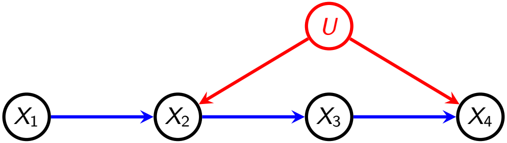


$$X_1 \perp\!\!\!\perp U, \qquad X_3 \perp\!\!\!\perp X_1, U \mid X_2, \qquad X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3, U.$$

If U is latent, we can only observe $X_3 \perp\!\!\!\perp X_1 \mid X_2$.

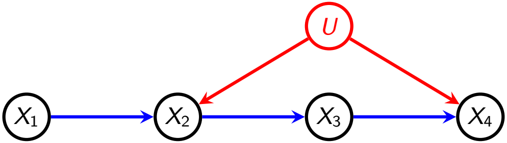
But if we add an arrow $X_1 \to X_4$, we still have $X_3 \perp\!\!\!\perp X_1 \mid X_2$.
So can we detect that $X_1 \not\to X_4$?

The Verma Constraint



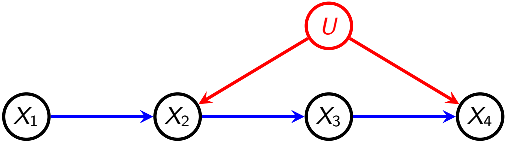
$$f(x_1, x_2, x_3, x_4) = \int f(u)\, f(x_1)\, f(x_2 \mid x_1, u)\, f(x_3 \mid x_2)\, f(x_4 \mid x_3, u)\, du$$

The Verma Constraint



$$
\begin{aligned}
f(x_1, x_2, x_3, x_4) &= \int f(u)\, f(x_1)\, f(x_2 \mid x_1, u)\, f(x_3 \mid x_2)\, f(x_4 \mid x_3, u)\, du \\
&= f(x_1)\, f(x_3 \mid x_2) \int f(u)\, f(x_2 \mid x_1, u)\, f(x_4 \mid x_3, u)\, du \\
&= f(x_1)\, f(x_3 \mid x_2)\, f^*(x_2, x_4 \mid x_1, x_3).
\end{aligned}
$$

The Verma Constraint



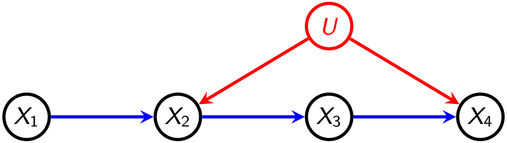
$$f(x_1, x_2, x_3, x_4) = \int f(u)\, f(x_1)\, f(x_2 \mid x_1, u)\, f(x_3 \mid x_2)\, f(x_4 \mid x_3, u)\, du$$
$$= f(x_1)\, f(x_3 \mid x_2) \int f(u)\, f(x_2 \mid x_1, u)\, f(x_4 \mid x_3, u)\, du$$
$$= f(x_1)\, f(x_3 \mid x_2)\, f^*(x_2, x_4 \mid x_1, x_3).$$

Note that

$$\int f^*(x_2, x_4 \mid x_1, x_3)\, dx_2 = f(x_4 \mid x_3)$$

is independent of x_1, precisely because $X_1 \not\to X_4$.

The Verma Constraint



$$
\begin{aligned}
f(x_1, x_2, x_3, x_4) &= \int f(u)\, f(x_1)\, f(x_2 \mid x_1, u)\, f(x_3 \mid x_2)\, f(x_4 \mid x_3, u)\, du \\
&= f(x_1)\, f(x_3 \mid x_2) \int f(u)\, f(x_2 \mid x_1, u)\, f(x_4 \mid x_3, u)\, du \\
&= f(x_1)\, f(x_3 \mid x_2)\, f^*(x_2, x_4 \mid x_1, x_3).
\end{aligned}
$$

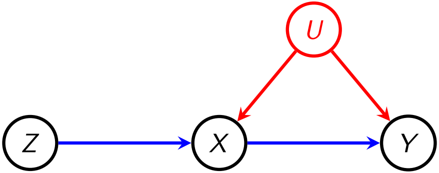
Note that

$$\int f^*(x_2, x_4 \mid x_1, x_3)\, dx_2 = f(x_4 \mid x_3)$$

is independent of x_1, precisely because $X_1 \not\to X_4$.

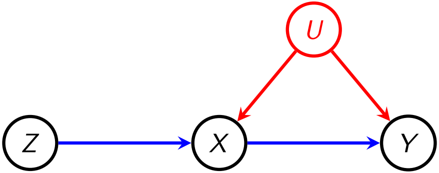
This is the **Verma constraint**, and provides a non-parametric test for the presence of $X_1 \to X_4$.

Example 2: Instrumental Variables



Conditional independences all involve U, and so can't be observed.

Example 2: Instrumental Variables



Conditional independences all involve U, and so can't be observed.

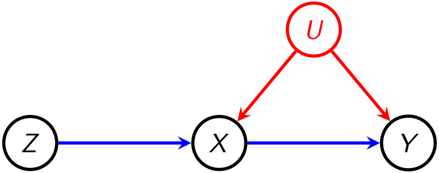
Can we detect that $Z \not\to Y$? Pearl (1995) showed that for discrete Z, X and Y,

$$\max_x \sum_y \max_z P(X = x,\ Y = y, \mid Z = z) \leq 1.$$

So, for example

$$P(X = 0,\ Y = 0 \mid Z = 0) + P(X = 0,\ Y = 1 \mid Z = 1) \leq 1.$$

Example 2: Instrumental Variables



Conditional independences all involve U, and so can't be observed.

Can we detect that $Z \not\to Y$? Pearl (1995) showed that for discrete Z, X and Y,

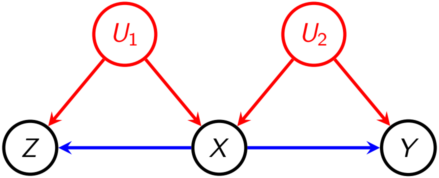
$$\max_x \sum_y \max_z P(X = x,\ Y = y, \,|\, Z = z) \leq 1.$$

So, for example

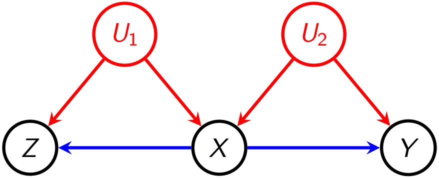
$$P(X = 0,\ Y = 0 \,|\, Z = 0) + P(X = 0,\ Y = 1 \,|\, Z = 1) \leq 1.$$

Inequalities for the discrete IV model can be derived using linear programs (`Porta`, `cdd`).

Example 3: Unrelated Confounding

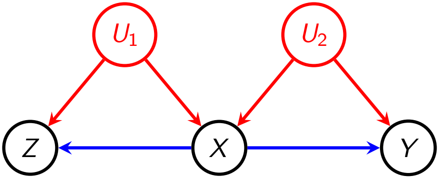


Example 3: Unrelated Confounding



Again there are no independences (in fact this contains the IV model).

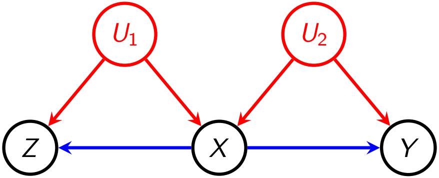
Example 3: Unrelated Confounding



Again there are no independences (in fact this contains the IV model).

Computational algebra doesn't seem to be useful here.

Example 3: Unrelated Confounding



Again there are no independences (in fact this contains the IV model).

Computational algebra doesn't seem to be useful here.

However, there are inequalities (as we will see).

So...

This raises the following question:

- what constraints does a DAG with latent variables imply in general?

So...

This raises the following question:

- what constraints does a DAG with latent variables imply in general?

We know this is a very hard problem (even for IV model).

So...

This raises the following question:

- what constraints does a DAG with latent variables imply in general?

We know this is a very hard problem (even for IV model).

Perhaps easier:

- can we find an equivalence class of these models?
- what graphs do we need to represent these models?

Prior Work

Conditional independences from marginalised DAGs can be captured by larger classes of graphs (ADMGs, summary graphs, MC graphs, LMGs, ...).

Prior Work

Conditional independences from marginalised DAGs can be captured by larger classes of graphs (ADMGs, summary graphs, MC graphs, LMGs, ...).

Richardson et al. (2012) deal with the same problem but also encodes Verma constraints (nested Markov property) with ADMGs.

Prior Work

Conditional independences from marginalised DAGs can be captured by larger classes of graphs (ADMGs, summary graphs, MC graphs, LMGs, ...).

Richardson et al. (2012) deal with the same problem but also encodes Verma constraints (nested Markov property) with ADMGs.

Pearl (1995) first gave inequality constraints for IV model. Bonet (2001) used linear programming to derive tight bounds.

Outline

Simplifications

We will consider marginalised DAGs where no assumption is made about the hidden variables.

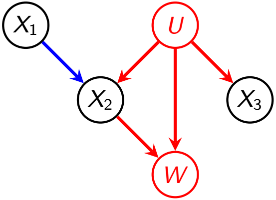
It turns out that we need not consider arbitrarily large numbers of latents.

Simplifications

We will consider marginalised DAGs where no assumption is made about the hidden variables.

It turns out that we need not consider arbitrarily large numbers of latents.

Simplification 1. Latents with no children can be ignored.

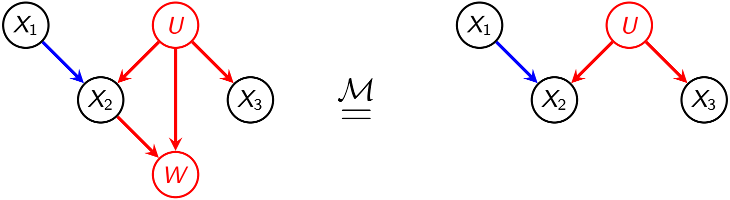


Simplifications

We will consider marginalised DAGs where no assumption is made about the hidden variables.

It turns out that we need not consider arbitrarily large numbers of latents.

Simplification 1. Latents with no children can be ignored.

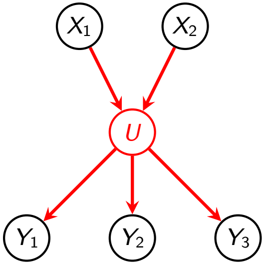


Simplifications

Simplification 2. Latents with parents can be transformed.

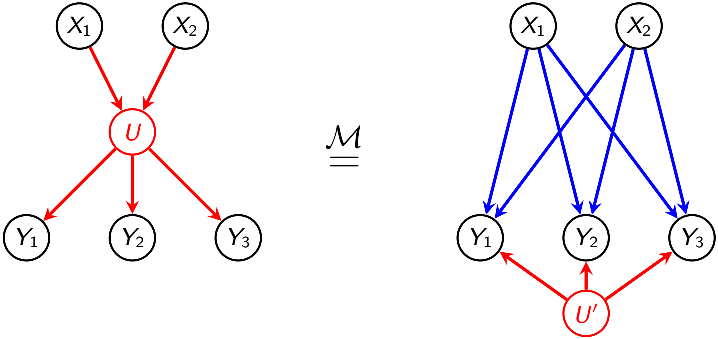
Simplifications

Simplification 2. Latents with parents can be transformed.



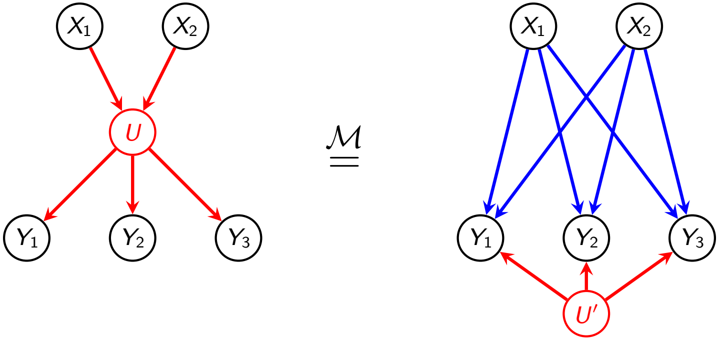
Simplifications

Simplification 2. Latents with parents can be transformed.



Simplifications

Simplification 2. Latents with parents can be transformed.

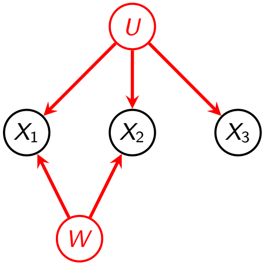


Hence we only need to consider latents with no parents.

Of course this is not true if we assume, e.g. latents are binary!

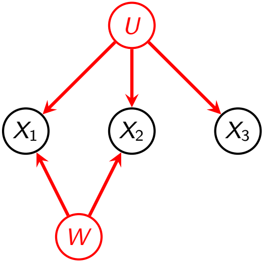
Simplifications

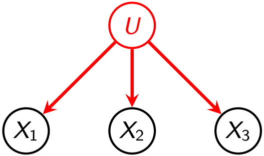
Simplification 3. If U, W are latent with $\text{ch}_{\mathcal{G}}(W) \subseteq \text{ch}_{\mathcal{G}}(U)$, then we don't need W.



Simplifications

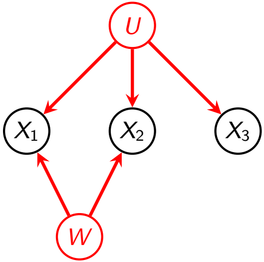
Simplification 3. If U, W are latent with $\mathrm{ch}_{\mathcal{G}}(W) \subseteq \mathrm{ch}_{\mathcal{G}}(U)$, then we don't need W.



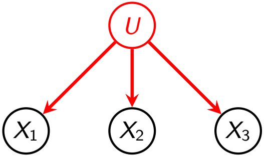


Simplifications

Simplification 3. If U, W are latent with $\text{ch}_{\mathcal{G}}(W) \subseteq \text{ch}_{\mathcal{G}}(U)$, then we don't need W.

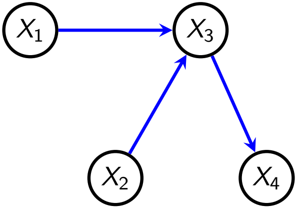


$$\overset{\mathcal{M}}{=}$$



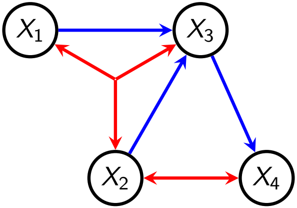
mDAGs

An **mDAG** is a DAG over vertices V, together with a collection B of inclusion maximal subsets of V.



mDAGs

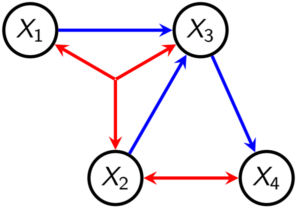
An **mDAG** is a DAG over vertices V, together with a collection B of inclusion maximal subsets of V.



In this case $B = \{\{1, 2, 3\}, \{2, 4\}\}$.

mDAGs

An **mDAG** is a DAG over vertices V, together with a collection B of inclusion maximal subsets of V.

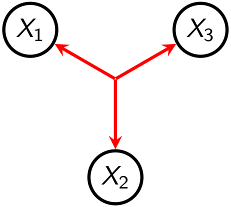


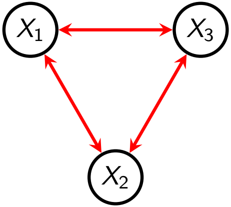
In this case $B = \{\{1,2,3\}, \{2,4\}\}$.

ADMGs are special case where B only contains subsets of size 2.

ADMGs are not sufficient

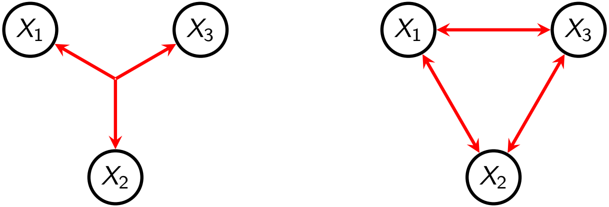
In general we need to distinguish between $\{1, 2, 3\}$ and $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.





ADMGs are not sufficient

In general we need to distinguish between $\{1, 2, 3\}$ and $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.



The model on the right is not saturated. Still true if we dichotomise.

Outline

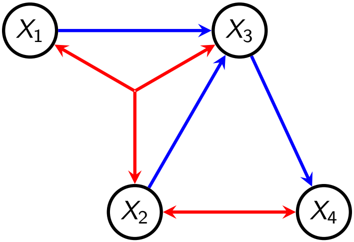
Complete 'Markov' Property

Given an mDAG $\mathcal{G}$ and statespace $\mathfrak{X}_V$, want to find $\mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$, the collection of distributions which could be generated from the mDAG.

Complete 'Markov' Property

Given an mDAG $\mathcal{G}$ and statespace $\mathfrak{X}_V$, want to find $\mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$, the collection of distributions which could be generated from the mDAG.

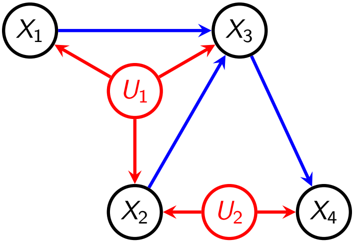
More precisely: with every mDAG $\mathcal{G}$ we can associate a DAG $\bar{\mathcal{G}}$ which includes those hidden variables U.



Complete 'Markov' Property

Given an mDAG $\mathcal{G}$ and statespace $\mathfrak{X}_V$, want to find $\mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$, the collection of distributions which could be generated from the mDAG.

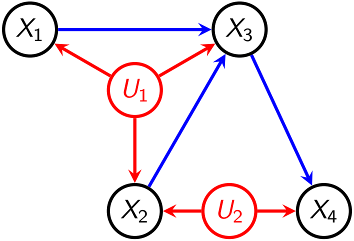
More precisely: with every mDAG $\mathcal{G}$ we can associate a DAG $\bar{\mathcal{G}}$ which includes those hidden variables U.



Complete 'Markov' Property

Given an mDAG $\mathcal{G}$ and statespace $\mathfrak{X}_V$, want to find $\mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$, the collection of distributions which could be generated from the mDAG.

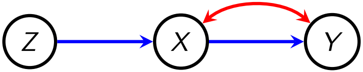
More precisely: with every mDAG $\mathcal{G}$ we can associate a DAG $\bar{\mathcal{G}}$ which includes those hidden variables U.



Then $P \in \mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$ if there exists some product space $\mathfrak{X}_U$ (and σ-algebra) and some distribution $\bar{P} \in \mathcal{M}(\bar{\mathcal{G}}, \mathfrak{X}_V \times \mathfrak{X}_U)$ such that P is the V-margin of $\bar{P}$.

Example

Instrumental Variables mDAG $\mathcal{G}$:

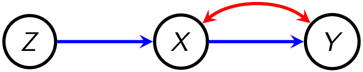


Let $\mathfrak{X}_V = \{0,1\}^3$; then Pearl (1995) shows

$$\mathcal{M}(\mathcal{G}, \mathfrak{X}_V) \subseteq \left\{ P \,\middle|\, \max_x \sum_y \max_z P(x, y \mid z) \leq 1 \right\}.$$

Example

Instrumental Variables mDAG $\mathcal{G}$:



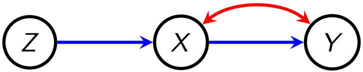
Let $\mathfrak{X}_V = \{0, 1\}^3$; then Pearl (1995) shows

$$\mathcal{M}(\mathcal{G}, \mathfrak{X}_V) \subseteq \left\{ P \,\middle|\, \max_x \sum_y \max_z P(x, y \mid z) \leq 1 \right\}.$$

With linear programming one can show equality holds (Bonet, 2001).

In the general discrete case (especially for increasing statespace of Z) these inequalities are not sufficient.

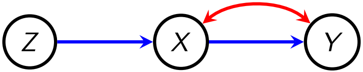
Instrumental Inequality: Alternative Interpretation



Let X be discrete.

$$f(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x, u)\, du.$$

Instrumental Inequality: Alternative Interpretation

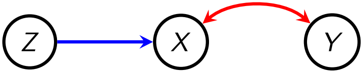


Let X be discrete.

$$f(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x, u)\, du.$$

Can think of X as a selector for how Y responds to U.

Instrumental Inequality: Alternative Interpretation



Let X be discrete.

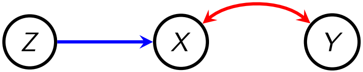
$$f(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x, u)\, du.$$

Can think of X as a selector for how Y responds to U. Suppose we turn off this selection:

$$f^*(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x = 0, u)\, du,$$

so Y behaves *as if* $X = 0$.

Instrumental Inequality: Alternative Interpretation



Let X be discrete.

$$f(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x, u)\, du.$$

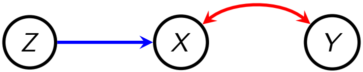
Can think of X as a selector for how Y responds to U. Suppose we turn off this selection:

$$f^*(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x = 0, u)\, du,$$

so Y behaves *as if* $X = 0$.

Clearly $Y \perp\!\!\!\perp Z\, [f^*]$, and $f^*(0, y \mid z) = f(0, y \mid z)$.

Instrumental Inequality: Alternative Interpretation



Let X be discrete.

$$f(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x, u)\, du.$$

Can think of X as a selector for how Y responds to U. Suppose we turn off this selection:

$$f^*(x, y \mid z) = \int f(u)\, f(x \mid z, u)\, f(y \mid x = 0, u)\, du,$$

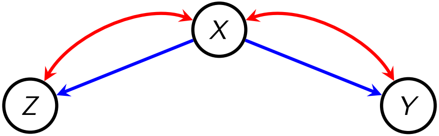
so Y behaves *as if* $X = 0$.

Clearly $Y \perp\!\!\!\perp Z\,[f^*]$, and $f^*(0, y \mid z) = f(0, y \mid z)$.

So $f(0, y \mid z)$ for y, z must be **compatible** with a distribution under which $Y \perp\!\!\!\perp Z$. This gives Pearl's instrumental inequality.

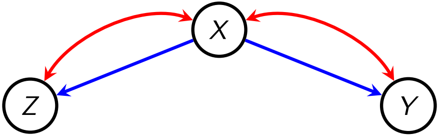
Applying to Other Graphs

Consider the 'unrelated confounding' model.



Applying to Other Graphs

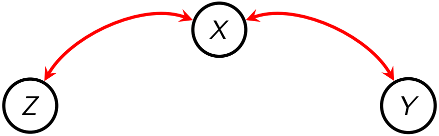
Consider the 'unrelated confounding' model.



Again no independences; in fact strictly contains IV models on (Z, X, Y) and (Y, X, Z).

Applying to Other Graphs

Consider the 'unrelated confounding' model.

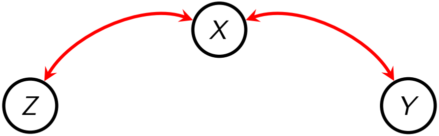


Again no independences; in fact strictly contains IV models on (Z, X, Y) and (Y, X, Z).

But by same argument, probabilities $f(0, y, z)$, must be compatible with $Y \perp\!\!\!\perp Z$.

Applying to Other Graphs

Consider the 'unrelated confounding' model.



Again no independences; in fact strictly contains IV models on (Z, X, Y) and (Y, X, Z).

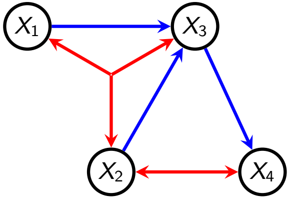
But by same argument, probabilities $f(0, y, z)$, must be compatible with $Y \perp\!\!\!\perp Z$.

So e.g.

$$(1 - f(0, y, z))^2 + (1 - f(0, 1 - y, 1 - z))^2 \geq 1.$$

Skeleton

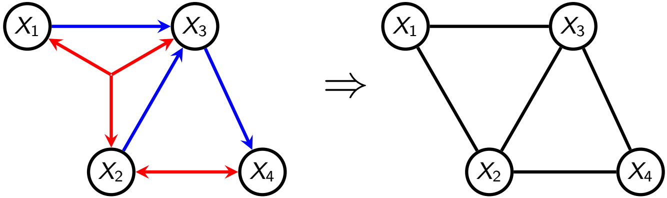
Call the **skeleton** of an mDAG the undirected graph given by joining any vertices in the same edge.



$\Rightarrow$

Skeleton

Call the **skeleton** of an mDAG the undirected graph given by joining any vertices in the same edge.



Distinction by Skeleton

The instrumental inequality can be generalised for discrete graphs.

Theorem

Let $\mathcal{G}$, $\mathcal{G}'$ be mDAGs and $\mathfrak{X}_V$ a discrete statespace. If

- $\mathcal{G}' \subseteq \mathcal{G}$; and
- $\mathcal{G}'$ and $\mathcal{G}$ have different skeletons,

then $\mathcal{M}(\mathcal{G}', \mathfrak{X}_V) \subsetneq \mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$. In other words, a constraint is always induced.

Distinction by Skeleton

The instrumental inequality can be generalised for discrete graphs.

Theorem

Let $\mathcal{G}$, $\mathcal{G}'$ be mDAGs and $\mathfrak{X}_V$ a discrete statespace. If

- $\mathcal{G}' \subseteq \mathcal{G}$; and
- $\mathcal{G}'$ and $\mathcal{G}$ have different skeletons,

then $\mathcal{M}(\mathcal{G}', \mathfrak{X}_V) \subsetneq \mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$. In other words, a constraint is always induced.

The additional constraints could be independences, Verma constraints, or inequalities (or something else!).

Distinction by Skeleton

The instrumental inequality can be generalised for discrete graphs.

Theorem

Let $\mathcal{G}$, $\mathcal{G}'$ be mDAGs and $\mathfrak{X}_V$ a discrete statespace. If

- $\mathcal{G}' \subseteq \mathcal{G}$; and
- $\mathcal{G}'$ and $\mathcal{G}$ have different skeletons,

then $\mathcal{M}(\mathcal{G}', \mathfrak{X}_V) \subsetneq \mathcal{M}(\mathcal{G}, \mathfrak{X}_V)$. In other words, a constraint is always induced.

The additional constraints could be independences, Verma constraints, or inequalities (or something else!).

The proof of this result is constructive (in that it produces inequalities).

Causal Effects

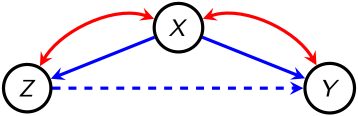
Corollary

We can derive non-trivial inequalities for an Average Controlled Direct Effect (ACDE) between any $Z \rightarrow Y$ as long as Z and Y are not directly confounded.

Causal Effects

Corollary

We can derive non-trivial inequalities for an Average Controlled Direct Effect (ACDE) between any $Z \to Y$ as long as Z and Y are not directly confounded.



Unrelated confounding model, e.g.

$$\text{ACDE}(x) \leq \frac{1 + P(Y = 1, x, z) - P(x)}{P(x, z)} - \frac{P(Y = 1, x, 1 - z)}{1 - P(x, z)}$$

for each x, z.

Model Equality

Lemma

Let $\mathcal{G}'$ and $\mathcal{G}$ be ADMGs, with $\mathcal{G}'$ a subgraph of $\mathcal{G}$. Then $\mathcal{M}(\mathcal{G}') \subseteq \mathcal{M}(\mathcal{G})$

Model Equality

Lemma

Let $\mathcal{G}'$ and $\mathcal{G}$ be ADMGs, with $\mathcal{G}'$ a subgraph of $\mathcal{G}$. Then $\mathcal{M}(\mathcal{G}') \subseteq \mathcal{M}(\mathcal{G})$

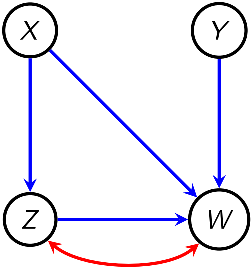
Theorem

Let $\mathcal{G}$ be an ADMG with an edge $a \leftrightarrow b$ such that

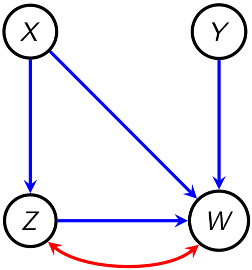
- $\text{pa}_{\mathcal{G}}(a) \subseteq \text{pa}_{\mathcal{G}}(b)$;
- $\text{sp}_{\mathcal{G}}(a) = \{b\}$.

Then if $\mathcal{G}'$ is equal to $\mathcal{G}$ except that $a \to b$ and $a \not\leftrightarrow b$, we have $\mathcal{M}(\mathcal{G}') = \mathcal{M}(\mathcal{G})$.

Examples (1)

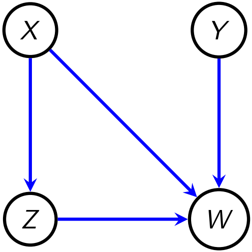


Examples (1)



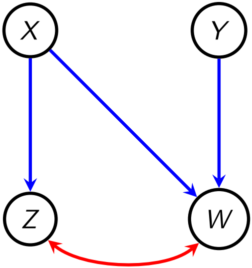
m-separation implies $X, Z \perp\!\!\!\perp Y$...

Examples (1)

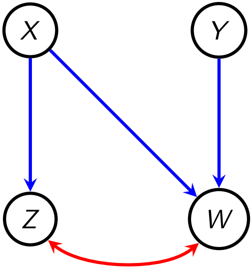


m-separation implies $X, Z \perp\!\!\!\perp Y$... but we get precisely this without the bidirected edge anyway.

Examples (2)

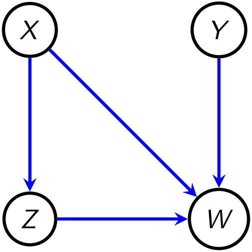


Examples (2)



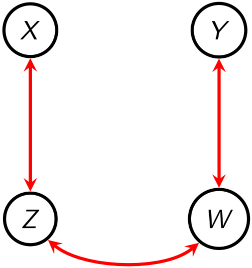
The edge $Z \leftrightarrow W$ fulfils the conditions of the theorem.

Examples (2)

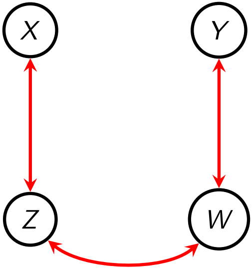


The edge $Z \leftrightarrow W$ fulfils the conditions of the theorem.

Examples (3)

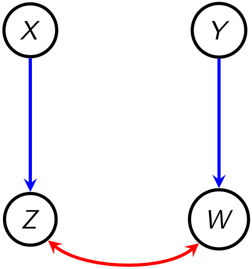


Examples (3)



Other graphs can merely be simplified.

Examples (3)



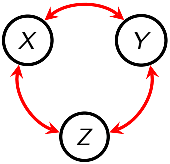
Other graphs can merely be simplified.

Here we have $X \perp\!\!\!\perp W, Y$ and $Y \perp\!\!\!\perp X, Z$ as well as Bell's inequalities.

Models on Three Observed Variables

40 unlabelled ADMGs on 3 variables (48 mDAGs).

$\perp\!\!\!\perp \{X, Y, Z\}$	complete independence	1
$X \perp\!\!\!\perp Y, Z$	joint independence	3
$X \perp\!\!\!\perp Y \mid Z$	conditional independence	5
$X \perp\!\!\!\perp Y$	marginal independence	6
$\mathrm{IV}(X, Y, Z)$	instrumental variable	3
$\mathrm{UC}(X, Y, Z)$	unrelated confounding	1
	unrestricted	20
	3-cycle	1



Larger Models

There are 1567 ADMGs over 4 variables. At least 509 are equivalent to a DAG.

After applying Theorem on equivalence, at most 671 distinct models not equivalent to DAGs.

Can reduce to 543 by splitting into districts.

Outline

Summary

We have seen that:

- mDAGs provide the most general necessary framework for representing DAGs under marginalisation (ADMGs are not sufficient);

Summary

We have seen that:

- mDAGs provide the most general necessary framework for representing DAGs under marginalisation (ADMGs are not sufficient);
- Pearl's IV bounds have a nice interpretation in terms of marginal independence;

Summary

We have seen that:

- mDAGs provide the most general necessary framework for representing DAGs under marginalisation (ADMGs are not sufficient);
- Pearl's IV bounds have a nice interpretation in terms of marginal independence;
- this interpretation leads to constructive bounds for other models;

Summary

We have seen that:

- mDAGs provide the most general necessary framework for representing DAGs under marginalisation (ADMGs are not sufficient);
- Pearl's IV bounds have a nice interpretation in terms of marginal independence;
- this interpretation leads to constructive bounds for other models;
- the absence of an edge in *any* mDAG can (in principle) be refuted;

Summary

We have seen that:

- mDAGs provide the most general necessary framework for representing DAGs under marginalisation (ADMGs are not sufficient);
- Pearl's IV bounds have a nice interpretation in terms of marginal independence;
- this interpretation leads to constructive bounds for other models;
- the absence of an edge in *any* mDAG can (in principle) be refuted;
- consequently causal bounds can be constructed for any unconfounded variables.

Questions

Some outstanding questions:

- Do equalities other than CI and Verma constraints exist?

Questions

Some outstanding questions:

- Do equalities other than CI and Verma constraints exist?
- How tight can we get the inequalities to be?

Questions

Some outstanding questions:

- Do equalities other than CI and Verma constraints exist?
- How tight can we get the inequalities to be?
- How powerful are the inequalities?

Questions

Some outstanding questions:

- Do equalities other than CI and Verma constraints exist?
- How tight can we get the inequalities to be?
- How powerful are the inequalities?
- What is the complete equivalence class of models?

Questions

Some outstanding questions:

- Do equalities other than CI and Verma constraints exist?
- How tight can we get the inequalities to be?
- How powerful are the inequalities?
- What is the complete equivalence class of models?
- Are the models smooth?

Questions

Some outstanding questions:

- Do equalities other than CI and Verma constraints exist?
- How tight can we get the inequalities to be?
- How powerful are the inequalities?
- What is the complete equivalence class of models?
- Are the models smooth?
- What about conditioning?

Thank you!

References

Bonet, B. – Instrumentality test revisited, *UAI-01*, 2001.

Pearl, J. – On the testability of causal models with latent and instrumental variables, *UAI-95*, 1995.

Richardson, T. S. – Markov properties for acyclic directed mixed graphs, *Scan. J. Statist.*, **30**, 145–157, 2003.

Richardson, T. S. – A factorization criterion for acyclic directed mixed graphs, *UAI-09*, 2009.

Verma, T. and Pearl, J. – Equivalence and synthesis of causal models, *UAI-90*, 1990.

d-Separation

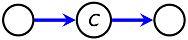
A **path** is a sequence of edges in the graph; vertices may not be repeated.

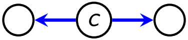
d-Separation

A **path** is a sequence of edges in the graph; vertices may not be repeated.

A path from v to w is **blocked** by $C \subseteq V \setminus \{v, w\}$ if either

(i) any non-collider is in C:





d-Separation

A **path** is a sequence of edges in the graph; vertices may not be repeated.

A path from v to w is **blocked** by $C \subseteq V \setminus \{v, w\}$ if either

(i) any non-collider is in C:

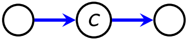
(ii) or any collider is not in C, nor has descendants in C:

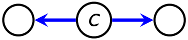
d-Separation

A **path** is a sequence of edges in the graph; vertices may not be repeated.

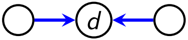
A path from v to w is **blocked** by $C \subseteq V \setminus \{v, w\}$ if either

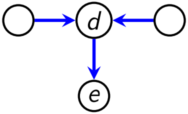
(i) any non-collider is in C:





(ii) or any collider is not in C, nor has descendants in C:





Two vertices v and w are **d-separated** given $C \subseteq V \setminus \{v, w\}$ if **all** paths are blocked.

Bonet's Inequalities

Suppose Z ternary and X, Y binary for IV model on Z, Y, X. Then

$$p(x_0, y_1 \mid z_1) + p(x_0, y_0 \mid z_2) + p(x_0, y_1 \mid z_0) + p(x_1, y_1 \mid z_1) + p(x_1, y_0 \mid z_0) \leq 2$$

ADMGs are not sufficient

Lemma

Let $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$ be mutually independent σ-algebrae (so that $\mathcal{F} \perp\!\!\!\perp \mathcal{G} \vee \mathcal{H}$ and so on), and let X, Y and Z be random variables such that

(i) X is $\mathcal{F} \vee \mathcal{G}$-measureable;

(ii) Y is $\mathcal{G} \vee \mathcal{H}$-measureable;

(iii) Z is $\mathcal{F} \vee \mathcal{H}$-measureable.

Then $P(X = Y = Z) > 1 - \epsilon$ implies

$$\operatorname{Var} X < 3\epsilon.$$