Inequality constraints on marginalised DAGs.

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

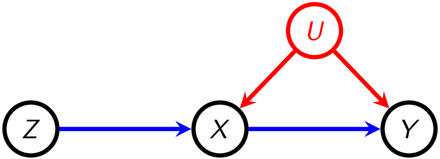
Statistical Laboratory, University of Cambridge

UK CIM, Manchester
15th May 2013

Outline

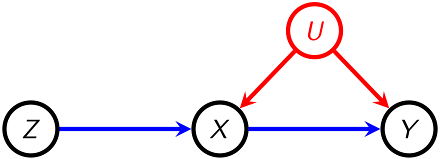
DAG Models

We like to propose models based on **directed acyclic graphs** (DAGs).



DAG Models

We like to propose models based on **directed acyclic graphs** (DAGs).

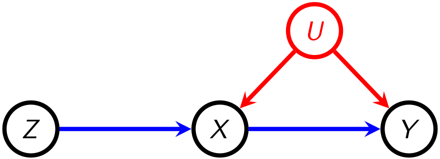


The DAG encodes a probability factorisation:

$$p(u, x, y, z) = p(u)\, p(z)\, p(x \mid z, u)\, p(y \mid u, x).$$

DAG Models

We like to propose models based on **directed acyclic graphs** (DAGs).



The DAG encodes a probability factorisation:

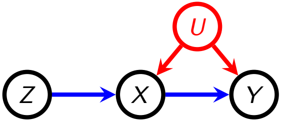
$$p(u, x, y, z) = p(u)\, p(z)\, p(x \mid z, u)\, p(y \mid u, x).$$

But sometimes we can only observe some of the random variables:

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

Call set of such distributions the **marginalised DAG model**.
We would like to test this model.

The Model

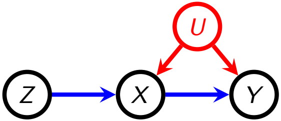


Marginalised DAG model

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

Assume all observed variables are discrete; no assumption made about latent variables.

The Model



Marginalised DAG model

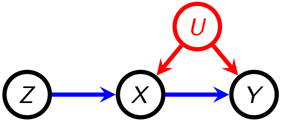
$$p(z, x, y) = \int p(u)\, p(z)\, p(x \mid z, u)\, p(y \mid x, u)\, du$$

Assume all observed variables are discrete; no assumption made about latent variables.

Can use Pearl's d-separation criterion to look for conditional independences:

$$U \perp\!\!\!\perp Z \qquad\qquad Y \perp\!\!\!\perp Z \mid X, U.$$

The Model



Marginalised DAG model

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

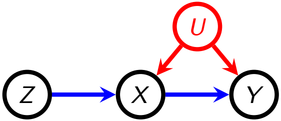
Assume all observed variables are discrete; no assumption made about latent variables.

Can use Pearl's d-separation criterion to look for conditional independences:

$$U \perp\!\!\!\perp Z \qquad\qquad Y \perp\!\!\!\perp Z \mid X, U.$$

These don't appear in the Z, X, Y marginal.

The Model



Marginalised DAG model

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

Assume all observed variables are discrete; no assumption made about latent variables.

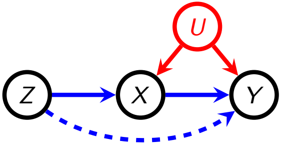
Can use Pearl's d-separation criterion to look for conditional independences:

$$U \perp\!\!\!\perp Z \qquad\qquad Y \perp\!\!\!\perp Z \mid X, U.$$

These don't appear in the Z, X, Y marginal.

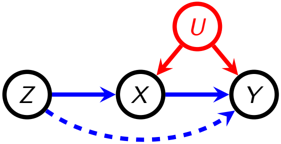
In fact, there are no equality constraints.

Instrumental Inequalities



The assumption $Z \not\to Y$ is important.
Can we check it?

Instrumental Inequalities



The assumption $Z \not\to Y$ is important.
Can we check it?

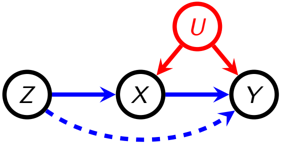
Pearl (1995) showed that if the observed variables are discrete,

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

This is the **instrumental inequality**, and can be empirically tested.

Instrumental Inequalities

The assumption $Z \not\to Y$ is important.
Can we check it?



Pearl (1995) showed that if the observed variables are discrete,

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

This is the **instrumental inequality**, and can be empirically tested.

If Z, X, Y are binary, then $(*)$ defines the marginalised DAG model (Bonet, 2001). e.g.

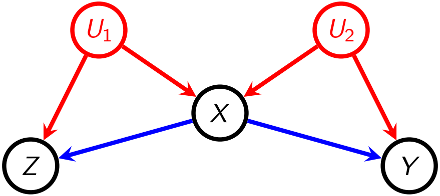
$$P(X = x, Y = 0 \mid Z = 0) + P(X = x, Y = 1 \mid Z = 1) \leq 1$$

We consider discrete models from here on.

The Problem

Question

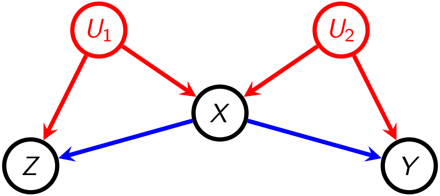
How can we determine if a general marginalised DAG model induces inequality constraints?



The Problem

Question

How can we determine if a general marginalised DAG model induces inequality constraints?

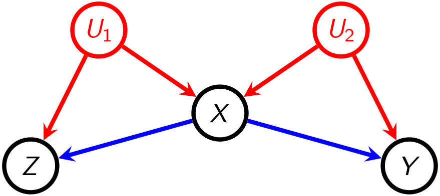


Pearl's proof of the instrumental inequality does not obviously generalise.

The Problem

Question

How can we determine if a general marginalised DAG model induces inequality constraints?



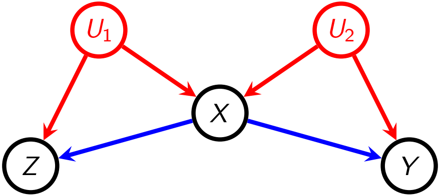
Pearl's proof of the instrumental inequality does not obviously generalise.

Computational linear algebra only works without adjacent latent variables. Also very computationally intensive.

The Problem

Question

How can we determine if a general marginalised DAG model induces inequality constraints?



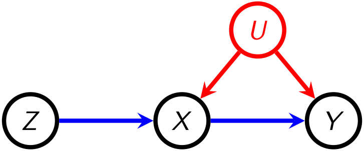
Pearl's proof of the instrumental inequality does not obviously generalise.

Computational linear algebra only works without adjacent latent variables. Also very computationally intensive.

Finding complete bounds in general is probably intractably hard.

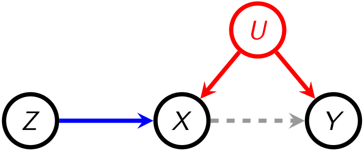
Outline

A New Proof



Have: $$p(x, y \mid z) = \int p(u)\, p(x \mid z, u)\, p(y \mid x, u)\, du.$$

A New Proof



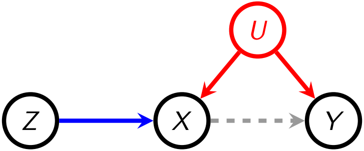
Have: $$p(x, y \mid z) = \int p(u)\, p(x \mid z, u)\, p(y \mid x, u)\, du.$$

Construct a **fictitious distribution** p^*:

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

Now Y behaves as though $X = 0$ regardless of X's actual value.

A New Proof



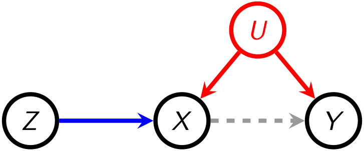
Have: $$p(x, y \mid z) = \int p(u)\, p(x \mid z, u)\, p(y \mid x, u)\, du.$$

Construct a **fictitious distribution** p^*:

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

Now Y behaves as though $X = 0$ regardless of X's actual value.
Causally, we can think of this as an **intervention** severing $X \to Y$.

A New Proof



Have: $$p(x, y \mid z) = \int p(u)\, p(x \mid z, u)\, p(y \mid x, u)\, du.$$

Construct a **fictitious distribution** p^*:

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

Now Y behaves as though $X = 0$ regardless of X's actual value.
Causally, we can think of this as an **intervention** severing $X \to Y$.

Can't observe p^* but:

- **Compatibility:** $p(0, y \mid z) = p^*(0, y \mid z)$ for each z, y; and
- **Independence:** $Y \perp\!\!\!\perp Z$ under p^*.

A New Proof

$$p(0, y \mid z) = p^*(0, y \mid z) \qquad\qquad Y \perp\!\!\!\perp Z \, [p^*]$$

Does such a distribution exist?

A New Proof

$$p(0, y \mid z) = p^*(0, y \mid z) \qquad\qquad Y \perp\!\!\!\perp Z\,[p^*]$$

Does such a distribution exist? Fix x.

$$p^*(y \mid z) = p^*(y)$$

A New Proof

$$p(0, y \mid z) = p^*(0, y \mid z) \qquad\qquad Y \perp\!\!\!\perp Z\,[p^*]$$

Does such a distribution exist? Fix x.

$$p^*(x = 0, y \mid z) \leq p^*(y \mid z) = p^*(y)$$

A New Proof

$$p(0, y \mid z) = p^*(0, y \mid z) \qquad\qquad Y \perp\!\!\!\perp Z\,[p^*]$$

Does such a distribution exist? Fix x.

$$\begin{aligned} p^*(x = 0, y \mid z) &\leq p^*(y \mid z) = p^*(y) \\ p(x = 0, y \mid z) &= \end{aligned}$$

A New Proof

$$p(0, y \mid z) = p^*(0, y \mid z) \qquad\qquad Y \perp\!\!\!\perp Z \, [p^*]$$

Does such a distribution exist? Fix x.

$$\begin{aligned} p^*(x = 0, y \mid z) &\leq p^*(y \mid z) = p^*(y) \\ p(x = 0, y \mid z) &= \end{aligned}$$

So clearly

$$\max_z p(x = 0, y \mid z) \leq p^*(y)$$

A New Proof

$$p(0, y \mid z) = p^*(0, y \mid z) \qquad\qquad Y \perp\!\!\!\perp Z \, [p^*]$$

Does such a distribution exist? Fix x.

$$\begin{aligned} p^*(x=0, y \mid z) &\leq p^*(y \mid z) = p^*(y) \\ p(x=0, y \mid z) &= \end{aligned}$$

So clearly

$$\max_z p(x=0, y \mid z) \leq p^*(y)$$

$$\sum_y \max_z p(x=0, y \mid z) \leq 1.$$

By varying x, the instrumental inequality follows.
(This condition is necessary and sufficient for existence of such a p^*)

A New Proof

$$p(0, y \mid z) = p^*(0, y \mid z) \qquad\qquad Y \perp\!\!\!\perp Z\,[p^*]$$

Does such a distribution exist? Fix x.

$$\begin{aligned} p^*(x = 0, y \mid z) &\leq p^*(y \mid z) = p^*(y) \\ p(x = 0, y \mid z) &= \end{aligned}$$

So clearly

$$\max_z p(x = 0, y \mid z) \leq p^*(y)$$

$$\sum_y \max_z p(x = 0, y \mid z) \leq 1.$$

By varying x, the instrumental inequality follows.
(This condition is necessary and sufficient for existence of such a p^*)

We say that the probabilities $p(x, y \mid z)$ are **compatible** with $Y \perp\!\!\!\perp Z$.

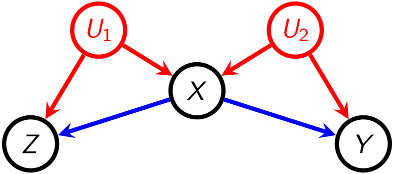
Generalising

How does this help us with other graphs?

Generalising

How does this help us with other graphs?

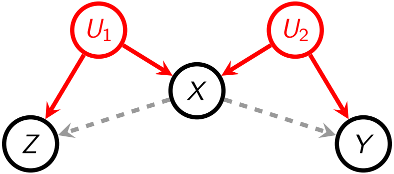
The argument works precisely because cutting edges led to a separation:



Generalising

How does this help us with other graphs?

The argument works precisely because cutting edges led to a separation:

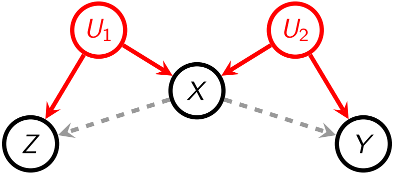


Z *is* d-separated from Y in the graph after cutting edges emanating from X.

Generalising

How does this help us with other graphs?

The argument works precisely because cutting edges led to a separation:



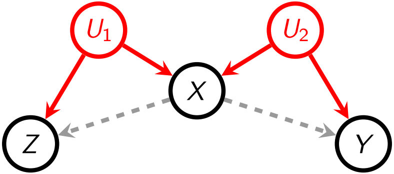
Z *is* d-separated from Y in the graph after cutting edges emanating from X.

So by the same argument, for fixed x, $p(x, y, z)$ must be compatible with a (fictitious) distribution p^* in which $Y \perp\!\!\!\perp Z$.

Generalising

How does this help us with other graphs?

The argument works precisely because cutting edges led to a separation:



Z *is* d-separated from Y in the graph after cutting edges emanating from X.

So by the same argument, for fixed x, $p(x, y, z)$ must be compatible with a (fictitious) distribution p^* in which $Y \perp\!\!\!\perp Z$.

[Note for the IV model, the conditional distribution $p(x, y \mid z)$ had to be compatible.]

Compatibility

Probabilities may not be compatible with independences.

Compatibility

Probabilities may not be compatible with independences.

Consider a partial probability table $p(x = 0, y, z)$:

	$Z = 0$	$Z = 1$
$Y = 0$	$1/3$	0
$Y = 1$	0	$1/3$

Compatibility

Probabilities may not be compatible with independences.

Consider a partial probability table $p(x = 0, y, z)$:

	$Z = 0$	$Z = 1$
$Y = 0$	$1/3$	0
$Y = 1$	0	$1/3$

There is no way to construct a joint distribution over X, Y, Z with these probabilities such that Y and Z are independent.

Compatibility

Probabilities may not be compatible with independences.

Consider a partial probability table $p(x = 0, y, z)$:

	$Z = 0$	$Z = 1$
$Y = 0$	$1/3$	0
$Y = 1$	0	$1/3$

There is no way to construct a joint distribution over X, Y, Z with these probabilities such that Y and Z are independent.

Most likely to happen if $p(x)$ is large for some value of x.

A Generalisation

For a DAG $\mathcal{G}$ and set of variables $\mathbf{W}$, let $\mathcal{G}[\mathbf{W}]$ be the graph after removing edges pointing away from $\mathbf{W}$.

A Generalisation

For a DAG $\mathcal{G}$ and set of variables $\mathbf{W}$, let $\mathcal{G}[\mathbf{W}]$ be the graph after removing edges pointing away from $\mathbf{W}$.

Theorem 1 (Evans, 2012)

Let p be a discrete distribution in marginalised DAG model for $\mathcal{G}$.
Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ be sets of variables in $\mathcal{G}$.

A Generalisation

For a DAG $\mathcal{G}$ and set of variables $\mathbf{W}$, let $\mathcal{G}[\mathbf{W}]$ be the graph after removing edges pointing away from $\mathbf{W}$.

Theorem 1 (Evans, 2012)

Let p be a discrete distribution in marginalised DAG model for $\mathcal{G}$. Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ be sets of variables in $\mathcal{G}$.

If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in $\mathcal{G}[\mathbf{W}]$, then for each fixed $\{\mathbf{W} = \mathbf{w}\}$ the probabilities

$$p(\mathbf{x}, \mathbf{y}, \mathbf{w} \mid \mathbf{z}), \qquad \mathbf{x}, \mathbf{y}, \mathbf{z}.$$

are **compatible with a distribution** p^*, in which $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \, [p^*]$.

A Generalisation

For a DAG $\mathcal{G}$ and set of variables $\mathbf{W}$, let $\mathcal{G}[\mathbf{W}]$ be the graph after removing edges pointing away from $\mathbf{W}$.

Theorem 1 (Evans, 2012)

Let p be a discrete distribution in marginalised DAG model for $\mathcal{G}$.
Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ be sets of variables in $\mathcal{G}$.

If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in $\mathcal{G}[\mathbf{W}]$, then for each fixed $\{\mathbf{W} = \mathbf{w}\}$ the probabilities

$$p(\mathbf{x}, \mathbf{y}, \mathbf{w} \,|\, \mathbf{z}), \qquad \mathbf{x}, \mathbf{y}, \mathbf{z}.$$

are **compatible with a distribution** p^*, in which $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \, [p^*]$.

If, in addition, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ and $\mathbf{X}_2, \mathbf{Y}_2$ are not descendants of $\mathbf{W}$, then

$$p(\mathbf{x}_1, \mathbf{y}_1, \mathbf{w} \,|\, \mathbf{x}_2, \mathbf{y}_2, \mathbf{z}) \qquad \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2, \mathbf{z}.$$

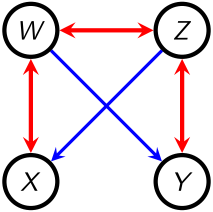
are **compatible with a distribution** p^*, in which $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \, [p^*]$.

Missing Edges Give Constraints

Corollary

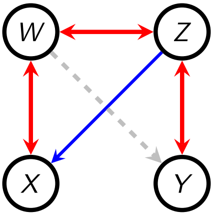
If X and Y are not joined by an edge, nor share a hidden common cause, then a constraint is always induced on the joint distribution.

Example 1



X and Y cannot be d-separated in this graph $\implies$ no independences.

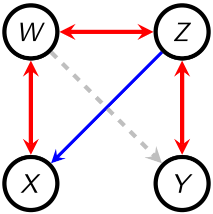
Example 1



X and Y cannot be d-separated in this graph $\implies$ no independences.

Remove edges emanating from W, see that now $X \perp\!\!\!\perp Y \mid Z$.
So $p(x, y, w \mid z)$ compatible with $X \perp\!\!\!\perp Y \mid Z$ for each w.

Example 1

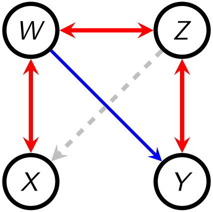


X and Y cannot be d-separated in this graph $\implies$ no independences.

Remove edges emanating from W, see that now $X \perp\!\!\!\perp Y \mid Z$.
So $p(x, y, w \mid z)$ compatible with $X \perp\!\!\!\perp Y \mid Z$ for each w.

In fact, Y not a descendant of Z, so $p(x, w \mid z, y)$ compatible.

Example 1



X and Y cannot be d-separated in this graph $\implies$ no independences.

Remove edges emanating from W, see that now $X \perp\!\!\!\perp Y \mid Z$.
So $p(x, y, w \mid z)$ compatible with $X \perp\!\!\!\perp Y \mid Z$ for each w.

In fact, Y not a descendant of Z, so $p(x, w \mid z, y)$ compatible.

By symmetry: $p(y, z \mid w, x)$ compatible with $X \perp\!\!\!\perp Y \mid W$ for each z.

Compatibility

Can we easily determine whether distributions are ‘compatible’ with independences?

Compatibility

Can we easily determine whether distributions are 'compatible' with independences?

Suppose we need $p(x, y, w \mid z)$ to be compatible with $X \perp\!\!\!\perp Y \mid Z \, [p^*]$.

Compatibility

Can we easily determine whether distributions are 'compatible' with independences?

Suppose we need $p(x, y, w \mid z)$ to be compatible with $X \perp\!\!\!\perp Y \mid Z \, [p^*]$.

In other words, for each z, w need a rank 1 matrix $B = (b_{xy})$ such that

$$b_{xy} \geq p(x, y, w \mid z) \qquad \text{and} \qquad \sum_{xy} b_{xy} \leq 1.$$

Compatibility

Can we easily determine whether distributions are 'compatible' with independences?

Suppose we need $p(x, y, w \mid z)$ to be compatible with $X \perp\!\!\!\perp Y \mid Z \, [p^*]$.

In other words, for each z, w need a rank 1 matrix $B = (b_{xy})$ such that

$$b_{xy} \geq p(x, y, w \mid z) \qquad \text{and} \qquad \sum_{xy} b_{xy} \leq 1.$$

Proposition 2

The existence of such a matrix is a **convex optimisation problem**.

Compatibility

Can we easily determine whether distributions are 'compatible' with independences?

Suppose we need $p(x, y, w \mid z)$ to be compatible with $X \perp\!\!\!\perp Y \mid Z \, [p^*]$.

In other words, for each z, w need a rank 1 matrix $B = (b_{xy})$ such that

$$b_{xy} \geq p(x, y, w \mid z) \qquad \text{and} \qquad \sum_{xy} b_{xy} \leq 1.$$

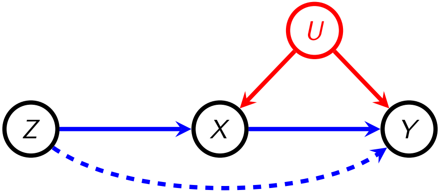
Proposition 2

The existence of such a matrix is a **convex optimisation problem**.

In general, Theorem 1 gives necessary but not sufficient conditions for p to be in the marginalised DAG model. Can be thought of as a convex relaxation of this intractable membership problem.

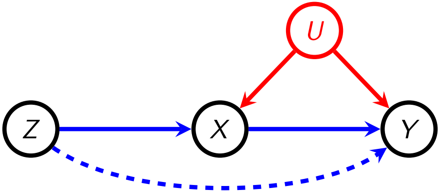
Outline

Causal Effects



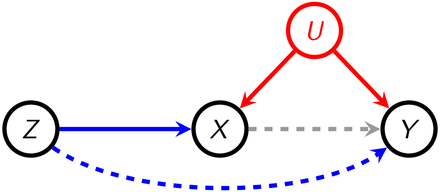
So far we've given inequalities which 'prove existence' for edges.

Causal Effects



So far we've given inequalities which 'prove existence' for edges.
Now we'd like to determine the strength of its causal effect.

Causal Effects

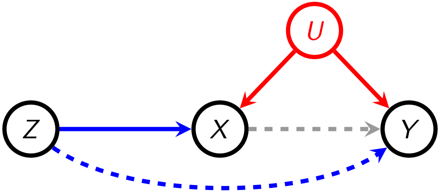


So far we've given inequalities which 'prove existence' for edges.
Now we'd like to determine the strength of its causal effect.

Construct p^* as before. Then

$$p(y \mid \mathrm{do}(x, z)) = p^*(y \mid z)$$

Causal Effects



So far we've given inequalities which 'prove existence' for edges.
Now we'd like to determine the strength of its causal effect.

Construct p^* as before. Then

$$\begin{aligned} p(y \mid \mathrm{do}(x, z)) &= p^*(y \mid z) \\ &= p(x, y \mid z) + \sum_{x' \neq x} p^*(x', y \mid z). \end{aligned}$$

Causal Bounds

This approach gives bounds on the interventional distributions (Evans, 2012) and, for example, the **average controlled direct effect**

$$\mathrm{ACDE}_{Z\to Y}(x) \equiv p(y=1 \mid \mathrm{do}(x, z=1)) - p(y=1 \mid \mathrm{do}(x, z=0)).$$

Causal Bounds

This approach gives bounds on the interventional distributions (Evans, 2012) and, for example, the **average controlled direct effect**

$$\mathrm{ACDE}_{Z \to Y}(x) \equiv p(y = 1 \mid \mathrm{do}(x, z = 1)) - p(y = 1 \mid \mathrm{do}(x, z = 0)).$$

Theorem 3(a)

Let $X \to Y$, but otherwise d-separated in the graph $\mathcal{G}[\mathbf{W}]$. Then an upper-bound on $\mathrm{ACDE}_{X \to Y}(w)$ is given by maximising

$$\frac{p(y = 1, x = 1, w) + \beta}{p(x = 1, w) + \beta} - \frac{p(y = 1, x = 0, w)}{p(x = 0, w) + 1 - p(w) - \beta}$$

over $0 \leq \beta \leq 1 - p(w)$.

Causal Bounds

This approach gives bounds on the interventional distributions (Evans, 2012) and, for example, the **average controlled direct effect**

$$\mathrm{ACDE}_{Z \to Y}(x) \equiv p(y = 1 \mid \mathrm{do}(x, z = 1)) - p(y = 1 \mid \mathrm{do}(x, z = 0)).$$

Theorem 3(a)

Let $X \to Y$, but otherwise d-separated in the graph $\mathcal{G}[\mathbf{W}]$. Then an upper-bound on $\mathrm{ACDE}_{X \to Y}(w)$ is given by maximising

$$\frac{p(y = 1, x = 1, w) + \beta}{p(x = 1, w) + \beta} - \frac{p(y = 1, x = 0, w)}{p(x = 0, w) + 1 - p(w) - \beta}$$

over $0 \leq \beta \leq 1 - p(w)$.

This is just a quadratic equation. There is an analogous lower-bound.

Bounds: Special Case

Theorem 3(b)

Let $X \to Y$, but otherwise d-separated in the graph $\mathcal{G}[\mathbf{W}]$, and that X is not a descendant of any variable in $\mathbf{W}$. Then

$$\begin{aligned} p(y=0,\mathbf{w} \mid x=0) + p(y=1,\mathbf{w} \mid x=1) - 1 \\ \leq \mathrm{ACDE}(\mathbf{w}) \leq \\ 1 - p(y=0,\mathbf{w} \mid x=1) - p(y=1,\mathbf{w} \mid x=0). \end{aligned}$$

For the IV model, this is the tight bound given by Cai et al (2008).

Bounds: Special Case

Theorem 3(b)

Let $X \to Y$, but otherwise d-separated in the graph $\mathcal{G}[\mathbf{W}]$, and that X is not a descendant of any variable in $\mathbf{W}$. Then

$$
\begin{aligned}
p(y=0,\mathbf{w}\,|\,x=0) + p(y=1,\mathbf{w}\,|\,x=1) - 1 \\
\leq \mathrm{ACDE}(\mathbf{w}) \leq \\
1 - p(y=0,\mathbf{w}\,|\,x=1) - p(y=1,\mathbf{w}\,|\,x=0).
\end{aligned}
$$

For the IV model, this is the tight bound given by Cai et al (2008).

If bounds exclude zero then models violate Theorem 1 compatibility.

Outline

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;
- key is 'compatibility' of probabilities with conditional independences;

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;
- key is 'compatibility' of probabilities with conditional independences;
- can be seen as a convex relaxation of full constraints;

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;
- key is 'compatibility' of probabilities with conditional independences;
- can be seen as a convex relaxation of full constraints;
- inversion of problem leads to causal bounds.

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;
- key is 'compatibility' of probabilities with conditional independences;
- can be seen as a convex relaxation of full constraints;
- inversion of problem leads to causal bounds.

Some current limitations:

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;
- key is 'compatibility' of probabilities with conditional independences;
- can be seen as a convex relaxation of full constraints;
- inversion of problem leads to causal bounds.

Some current limitations:

- performing inference for finite samples is non-trivial;

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;
- key is 'compatibility' of probabilities with conditional independences;
- can be seen as a convex relaxation of full constraints;
- inversion of problem leads to causal bounds.

Some current limitations:

- performing inference for finite samples is non-trivial;
- could we get MLEs with these inequalities? (Ramsahai and Lauritzen, 2011)

Conclusions

- Pearl's instrumental inequality generalises to many marginalised DAG models;
- key is 'compatibility' of probabilities with conditional independences;
- can be seen as a convex relaxation of full constraints;
- inversion of problem leads to causal bounds.

Some current limitations:

- performing inference for finite samples is non-trivial;
- could we get MLEs with these inequalities? (Ramsahai and Lauritzen, 2011)
- asymptotic distributions of LR statistics are complex.

References

Bonet – Instrumentality tests revisited, *UAI*, 2001.

Cai, Kuroki, Pearl and Tian – Bounds on direct effects in the presence of confounded intermediate variables, *Biometrics*, 64(3):695–701, 2008.

Evans – Graphical methods for inequality constraints in marginalized DAGs, *MLSP*, 2012.

Kang and Tian – Inequality constraints in causal models with hidden variables, *UAI*, 2006.

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

Pitowsky – Correlation polytopes: their geometry and complexity, *Mathematical Programming*, 50:395–414, 1991.

Ramsahai and Lauritzen – Likelihood analysis of the binary instrumental variable model*Biometrika*, 98(4):987-994, 2011.

Shpitser, Richardson, Robins and Evans – Parameter and structure learning in nested Markov models, *UAI*, 2012.

Ver Steeg and Galstyan – A sequence of relaxations constraining hidden variable models, *UAI*, 2011.

d-Separation

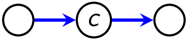
A **path** is a sequence of edges in the graph; vertices may not be repeated.

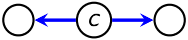
d-Separation

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:



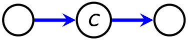


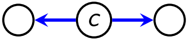
d-Separation

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

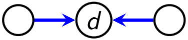
A path from v to w is **blocked** by $C \subseteq V \setminus \{v, w\}$ if either

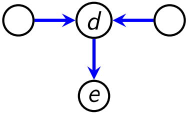
(i) any non-collider is in C:





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



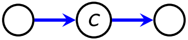


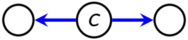
d-Separation

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

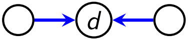
A path from v to w is **blocked** by $C \subseteq V \setminus \{v, w\}$ if either

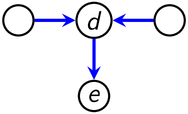
(i) any non-collider is in C:





(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.