Graphs for margins of Bayesian networks.

Robin J. Evans
www.stats.ox.ac.uk/~evans

Department of Statistics, University of Oxford

ERCIM, Pisa
December 2014
Outline

1 Introduction

2 DAGs

3 Margins of DAG Models

4 mDAGs

5 Markov Equivalence

6 Summary
Correlation does not imply causation

Dr Matthew Hobbs, head of research for Diabetes UK, said there was no proof that napping actually caused diabetes.
How a short nap can raise the risk of diabetes: Study finds people who have a siesta are more likely to have high blood pressure and high cholesterol

- Napping for more than 30 minutes at a time can raise the risk of diabetes, according to a new study.
- It can also increase likelihood of high blood pressure and high cholesterol.

By PAT HAGAN

PUBLISHED: 01:04, 21 September 2013 | UPDATED: 10:34, 21 September 2013

They were much favoured by Margaret Thatcher, Albert Einstein and Winston Churchill.

But while afternoon naps may revitalise tired brains, they can also increase the risk of diabetes, according to new research.
Correlation does not imply causation

How a short nap can raise diabetes: Study finds siesta are more likely to elevate blood pressure and high cholesterol

- Napping for more than 30 minutes at a time is associated with a higher risk for impaired fasting plasma glucose and diabetes mellitus in older adults: results from the Dongfeng–Tongji cohort of retired workers

By PAT HAGAN

Published: 01:04, 21 September 2013 | Updated: 10:34, 21 September

598 shares

They were much favoured by Margaret Thatcher, Albert Einstein and Martin Luther King Jr.

But while afternoon naps may revitalise tired brains, they can also increase the likelihood of high blood pressure and diabetes, a new study revealed.

A team of researchers from Fudan University and Huazhong University of Science and Technology in China studied the influence of afternoon napping on blood pressure and glucose levels in 9,834 retired workers.

They found that adults who napped for at least 30 minutes in the afternoon were at greater risk of diabetes and higher blood pressure than those who did not nap.

Longer habitual afternoon napping is associated with a higher risk for impaired fasting plasma glucose and diabetes mellitus in older adults: results from the Dongfeng–Tongji cohort of retired workers

Weimin Fang, Zhongliang Li, Li Wu, Zhongqiang Cao, Yuan Liang, Handong Yang, Youjie Wang, Tangchun Wu

- Ministry of Education Key Laboratory of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, China
- Department of Maternal and Child Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, China
- Department of Social Medicine, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, China
- Dongfeng General Hospital, Dongfeng Motor Corporation and Hubei University of Medicine, China

Abstract

Objectives

Afternoon napping is a common habit in China. We used data obtained from the Dongfeng–Tongji cohort to examine if duration of habitual afternoon napping was associated with risks for impaired fasting plasma glucose (IFG) and diabetes mellitus (DM) in a Chinese elderly population.

Methods
Correlation does not imply causation

How a short nap can raise diabetes: Study finds people who siesta are more likely to have high blood pressure and high cholesterol.

- Napping for more than 30 minutes at a time, according to a new study.
- It can also increase the likelihood of high blood pressure and diabetes.

“Dr Matthew Hobbs, head of research for Diabetes UK, said there was no proof that napping actually caused diabetes.”

Original Article

Longer habitual afternoon napping is associated with a higher risk for impaired fasting plasma glucose and diabetes mellitus in older adults: results from the Dongfeng-Tongji cohort of retired workers

Weimin Fang, Zhongliang Li, Li Wu, Zhongqiang Cao, Yuan Liang, Handong Yang, Youjie Wang, Tangchun Wu

Department of Maternal and Child Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, China

Abstract

Objectives

Afternoon napping is a common habit in China. We used data obtained from the Dongfeng-Tongji cohort to examine if duration of habitual afternoon napping was associated with risks for impaired fasting plasma glucose (IFG) and diabetes mellitus (DM) in a Chinese elderly population.

Methods
Distinguishing Between Causal Models

In order to compare the models, we need to understand in what ways causal models will differ, both observationally; under interventions.
Distinguishing Between Causal Models

In order to compare the models, we need to understand in what ways causal models will differ, both observationally; under interventions.

Diagram:
- **Gene** → **Diabetes** → **Nap**
- **Lifestyle** influences **Diabetes** and **Nap**

4 / 36
Distinguishing Between Causal Models

In order to compare the models, we need to understand in what ways causal models will differ, both:

- observationally;
- under interventions.
Outline

1. Introduction
2. DAGs
3. Margins of DAG Models
4. mDAGs
5. Markov Equivalence
6. Summary
Outline

1 Introduction

2 DAGs

3 Margins of DAG Models

4 mDAGs

5 Markov Equivalence

6 Summary
Directed Acyclic Graphs

vertices

edges

If \( w \rightarrow v \) then \( w \) is a parent of \( v \):
\[ \text{pa}_G(4) = \{1, 2\} \]

If \( w \rightarrow \cdots \rightarrow v \) then \( w \) is an ancestor of \( v \).
Directed Acyclic Graphs

vertices

edges

no directed cycles

If $w \rightarrow v$ then $w$ is a parent of $v$: $\text{pa}_G(4) = \{1, 2\}$.

If $w \rightarrow \cdots \rightarrow v$ then $w$ is an ancestor of $v$. 
Directed Acyclic Graphs

vertices

edges

no directed cycles

directed acyclic graph (DAG), $G$

If $w \rightarrow v$ then $w$ is a parent of $v$: $\text{pa}_G(4) = \{1, 2\}$.

If $w \rightarrow \cdots \rightarrow v$ then $w$ is an ancestor of $v$. 

$\text{pa}_G(4) = \{1, 2\}$.
Directed Acyclic Graphs

vertices

edges

no directed cycles

If \( w \rightarrow v \) then \( w \) is a parent of \( v \): \( \text{pa}_G(4) = \{1, 2\} \).

If \( w \rightarrow \cdots \rightarrow v \) then \( w \) is a ancestor of \( v \)
DAG Models

vertex $a$ $\iff$ random variable $X_a$
DAG Models

vertex

random variable

\[ a \xlongleftarrow{} X_a \]

graph \( G \)

model \( \mathcal{M}(G) \)

\[ P : X_i \perp \!\!\!\!\perp X_{\text{pre}(i)} \mid X_{\text{pa}(i)}[P] \]

ordered local Markov property
DAG Models

vertex \[ a \] \iff \text{random variable} \quad X_a

graph \( G \)

So in example above:

\[
X_2 \perp \perp X_1 \\
X_4 \perp \perp X_3 \mid X_1, X_2
\]

model \( M(G) \)

\[
P : X_i \perp \perp X_{\text{pre}(i)} \mid X_{\text{pa}(i)}[P]
\]

ordered local Markov property

\[
X_3 \perp \perp X_1 \mid X_2 \\
X_5 \perp \perp X_1, X_2 \mid X_3, X_4
\]
Global Markov Property

$P$ satisfies the **global Markov property** if for all sets $A, B, C,$

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

$P$ satisfies the **global Markov property** if for all sets $A, B, C$,

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

**Theorem (Lauritzen et al, 1990)**

$P$ satisfies the global Markov property if and only if it satisfies the ordered local Markov property.
Global Markov Property

$P$ satisfies the **global Markov property** if for all sets $A, B, C$,

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

**Theorem (Lauritzen et al, 1990)**

$P$ satisfies the global Markov property if and only if it satisfies the ordered local Markov property.

**Point:** the model can be defined in terms of ‘paths of information’.
Causal Interventions

If we interpret the DAG as representing structural assumptions, then if we intervene on a node, the graph of the resulting model is just locally altered:
Causal Interventions

If we interpret the DAG as representing structural assumptions, then if we intervene on a node, the graph of the resulting model is just locally altered:

so if we force people to stop napping...
Outline

1 Introduction

2 DAGs

3 Margins of DAG Models

4 mDAGs

5 Markov Equivalence

6 Summary
Latent Projection

Can preserve conditional independences and causal coherence even with latents by considering paths. DAG $\mathcal{G}$ on vertices $V = O \cup U$, define latent projection $\mathcal{G}(O)$ as follows: (Verma and Pearl, 1992)
Latent Projection

Can preserve conditional independences and causal coherence even with latents by considering paths. DAG $\mathcal{G}$ on vertices $V = O \cup U$, define latent projection $\mathcal{G}(O)$ as follows: (Verma and Pearl, 1992)

Whenever there is a path of the form

```
        x
         ↗
        u_1
         ↗
        ...  
         ↗
        u_k
         ↗
y
```

add

```
        x
         ↗
        y
```

Whenever there is a path of the form
**Latent Projection**

Can preserve conditional independences and causal coherence even with latents by considering paths. DAG $\mathcal{G}$ on vertices $V = O \cup U$, define **latent projection** $\mathcal{G}(O)$ as follows: (Verma and Pearl, 1992)

Whenever there is a path of the form

![Diagram](https://via.placeholder.com/150)

add

![Diagram](https://via.placeholder.com/150)

Whenever there is a path of the form

![Diagram](https://via.placeholder.com/150)

add

![Diagram](https://via.placeholder.com/150)
Latent Projection

Can preserve conditional independences and causal coherence even with latents by considering paths. DAG $G$ on vertices $V = O \cup U$, define **latent projection** $G(O)$ as follows: (Verma and Pearl, 1992)

Whenever there is a path of the form

$$x \rightarrow u_1 \rightarrow \cdots \rightarrow u_k \rightarrow y$$

add

$$x \leftrightarrow y$$

Whenever there is a path of the form

$$x \leftarrow u_1 \leftarrow \cdots \leftarrow u_k \rightarrow y$$

add

$$x \leftrightarrow y$$

Then remove the latent variables $U$ from the graph.
Latent projection leads to an acyclic directed mixed graph (ADMG) (equivalent to summary graph without undirected edges). Can read off independences with d/m-separation. Like an ancestral graph, these are precisely observable independences from the original DAG. See Richardson (2003) for more details. In addition, can see that projection preserves the causal structure; Verma and Pearl (1992).
Latent projection leads to an **acyclic directed mixed graph** (ADMG) (equivalent to summary graph without undirected edges).
Latent projection leads to an **acyclic directed mixed graph** (ADMG) (equivalent to summary graph without undirected edges).

Can read off independences with d/m-separation. Like an ancestral graph, these are precisely observable independences from the original DAG. See Richardson (2003) for more details.

In addition, can see that projection preserves the causal structure; Verma and Pearl (1992).
Causal Coherence

If we intervene on some observed variables, this ‘breaks’ their dependence upon their parents.
Causal Coherence

If we intervene on some observed variables, this ‘breaks’ their dependence upon their parents.

![Diagram showing causal coherence]

1. If we intervene on some observed variables, this ‘breaks’ their dependence upon their parents.

![Diagram showing causal coherence]
Causal Coherence

If we intervene on some observed variables, this ‘breaks’ their dependence upon their parents.
Causal Coherence

If we intervene on some observed variables, this ‘breaks’ their dependence upon their parents.

\[ u \rightarrow 3 \rightarrow w \rightarrow 4 \]

↓ project

\[ \rightarrow \text{intervene} \]

\[ u \rightarrow 3 \rightarrow w \rightarrow 4 \]

↓ project

\[ \rightarrow \text{intervene} \]
Causal Coherence

If we intervene on some observed variables, this ‘breaks’ their dependence upon their parents.
ADMGs are not sufficient

These have the same latent projection: $x, y, z$

But the model over $(x, y, z)$ in $G_2$ is not saturated. Still true if we dichotomize.
ADMGs are not sufficient

$G_1$

$G_2$

These have the same latent projection:
ADMGs are not sufficient

\[ G_1 \]

\[ G_2 \]

These have the same latent projection:

But the model over \((x, y, z)\) in \(G_2\) is not saturated. Still true if we dichotomize.
The Problem

- Verma and Pearl’s latent projection only uses paths, which are inherently ‘pairwise’;
- the paths are objects suited to conditional independence, but not all constraints on margins are conditional independences;
- ADMGs are not a sufficiently rich class of graphs to capture the different models one can obtain.
Outline

1. Introduction
2. DAGs
3. Margins of DAG Models
4. mDAGs
5. Markov Equivalence
6. Summary
Structural Equation Model View

There is another way to think about DAG models (e.g. Lauritzen et al, 1990).

\( P \in \mathcal{M}(G) \) iff there exist functions \( f_i \) and independent variables \( E_i \) such that recursively setting

\[
X_i = f_i(X_{pa(i)}, E_i)
\]

gives \( X_V \) the distribution \( P \).
Structural Equation Model View

There is another way to think about DAG models (e.g. Lauritzen et al, 1990).

\[ P \in \mathcal{M}(G) \text{ iff there exist functions } f_i \text{ and independent variables } E_i \text{ such that recursively setting} \]

\[ X_i = f_i(X_{pa(i)}, E_i) \]

gives \( X_V \) the distribution \( P \).

\[
\begin{align*}
X_1 &= f_1(E_1) \\
X_2 &= f_2(E_2) \\
X_3 &= f_3(X_2, E_3) \\
X_4 &= f_4(X_1, X_2, E_4) \\
X_5 &= f_5(X_3, X_4, E_5).
\end{align*}
\]
Simplifications

Simplification 1. WLOG latent vertices have no parents.
Simplifications

**Simplification 1.** WLOG latent vertices have no parents.
Simplifications

Simplification 1. WLOG latent vertices have no parents.

(Of course, this is not true if we assume a specific state-space: e.g. phylogenetic model)
Simplification 2. If $u, w$ are latent with $\text{ch}_G(w) \subseteq \text{ch}_G(u)$, then we don’t need $w$.
Simplifications

**Simplification 2.** If $u, w$ are latent with $\text{ch}_G(w) \subseteq \text{ch}_G(u)$, then we don’t need $w$. 

\[ x_1 \xrightarrow{u} x_2 \xrightarrow{w} x_3 \]

\[ x_1 \xrightarrow{u} x_2 \xrightarrow{x_3} \]

$\mathcal{M} =$
mDAGs

So we only need to consider models like this:
mDAGs

So we only need to consider models like this:

...which we represent with a hyper-graph called an mDAG.
mDAGs

So we only need to consider models like this:

...which we represent with a hyper-graph called an mDAG.

Formally, an mDAG on $V$ is a DAG (in blue), together with some inclusion maximal collection subsets of size at least 2 (red).
mDAGs

So we only need to consider models like this:

...which we represent with a hyper-graph called an mDAG.

Formally, an mDAG on $V$ is a DAG (in blue), together with some inclusion maximal collection subsets of size at least 2 (red).

Going backwards and replacing bidirected edges with latents gives us the canonical DAG $\tilde{G}$. 
Given an mDAG $\mathcal{G}$ and distribution $P$, say $P$ is in the complete model for $\mathcal{G}$, or $P \in M(\mathcal{G})$ if it is the margin of some distribution in the model for the canonical DAG $M(\overline{\mathcal{G}})$. 
Given an mDAG $\mathcal{G}$ and distribution $P$, say $P$ is in the complete model for $\mathcal{G}$, or $P \in \mathcal{M}(\mathcal{G})$ if it is the margin of some distribution in the model for the canonical DAG $\mathcal{M}(\bar{\mathcal{G}})$.

There are other (weaker) properties Shpitser et al. (2014).
Latent Projection

For mDAG $\mathcal{G}$ and subset of vertices $O$, form latent projection $p(\mathcal{G}, O)$ by:
Whenever there is a path of the form

![Diagram](x \rightarrow u_1 \rightarrow \cdots \rightarrow u_k \rightarrow y)

add

![Diagram](x \rightarrow y)
Latent Projection

For mDAG $\mathcal{G}$ and subset of vertices $O$, form latent projection $p(\mathcal{G}, O)$ by:
Whenever there is a path of the form

```
\begin{tikzpicture}
  \path[->, red, thick] (0,0) node[draw, circle] (x) {x} -- (1,0) node[draw, circle] (u1) {u_1} -- (2,0) node[draw, circle] (uk) {u_k} -- (3,0) node[draw, circle] (y) {y};
\end{tikzpicture}
```

add

```
\begin{tikzpicture}
  \path[->, blue, thick] (0,0) node[draw, circle] (x) {x} -- (1,0) node[draw, circle] (y) {y};
\end{tikzpicture}
```

Whenever there is a maximal set $B = \{x_1, x_2, \ldots, x_k\}$ such that these variables share a hidden common cause, add hyper-edge $B$.
Then remove the latent variables $U$ from the graph.
Latent Projection

For mDAG $\mathcal{G}$ and subset of vertices $O$, form latent projection $p(\mathcal{G}, O)$ by:
Whenever there is a path of the form $x \xrightarrow{} u_1 \rightarrow \cdots \rightarrow u_k \rightarrow y$
add $x \rightarrow y$
Whenever there is a maximal set $B = \{x_1, x_2, \ldots, x_k\}$ such that these variables share a hidden common cause, add hyper-edge $B$.
Then remove the latent variables $U$ from the graph.
Outline

1 Introduction

2 DAGs

3 Margins of DAG Models

4 mDAGs

5 Markov Equivalence

6 Summary
Results

The mDAG latent projection preserves the distinction between models.

Theorem (Evans, 2014)
If \( p(G, O) = p(G', O) \) then the models induced by \( M(G) \) and \( M(G') \) on the margin \( O \) are the same.

So the problem which arises with ADMGs never occurs for mDAGs.
Results

The mDAG latent projection preserves the distinction between models.

Theorem (Evans, 2014)

If \( p(G, O) = p(G', O) \) then the models induced by \( M(G) \) and \( M(G') \) on the margin \( O \) are the same.

So the problem which arises with ADMGs never occurs for mDAGs.

Theorem (Evans, 2014)

If \( C \subseteq O \) then \( p(G_{\overline{C}}, O) = p(G, O)_{\overline{C}} \); i.e. the projection respects causal interventions.
Instrumental Variables

The Instrumental Variables model assumes causally exogenous variable $z$ affects the treatment $x$. 

\[ z \rightarrow x \rightarrow y \]
Instrumental Variables

The Instrumental Variables model assumes causally exogenous variable $z$ affects the treatment $x$.

But it’s well known that this is observationally indistinguishable from a hidden common cause for $x$ and $z$ (e.g. Didelez and Sheehan, 2007).
Instrumental Variables

To see this, imagine $z$ is an exact copy of $z'$. 

$$z \quad \quad z' \quad \quad x \quad \quad y$$
Instrumental Variables

To see this, imagine $z$ is an exact copy of $z'$. Doesn’t really matter whether $x$ gets information from $z$ or $z'$. 
To see this, imagine \( z \) is an exact copy of \( z' \).

Doesn’t really matter whether \( x \) gets information from \( z \) or \( z' \).

Very hard to see this equivalence with conditional independence.
Instrument Generalisation

Let $G$ have bidirected edge $B = C \cup D$ with:

(i) every $c \in C$ contained in no other bidirected edge;

(ii) $\text{pa}_G(d) \supseteq \text{pa}_G(C)$ for each $d \in D$. 

Instrument Generalisation

Let $\mathcal{G}$ have bidirected edge $B = C \cup D$ with:

(i) every $c \in C$ contained in no other bidirected edge;

(ii) $\text{pa}_G(d) \supseteq \text{pa}_G(C)$ for each $d \in D$.

Can ‘split’ $B$ into $C$ and $D$ and add edges $c \rightarrow d$ where necessary.
Instrument Generalisation

Let $\mathcal{G}$ have bidirected edge $B = C \cup D$ with:

(i) every $c \in C$ contained in no other bidirected edge;
(ii) $\text{pa}_G(d) \supseteq \text{pa}_G(C)$ for each $d \in D$.

Can ‘split’ $B$ into $C$ and $D$ and add edges $c \rightarrow d$ where necessary.
Skeletons

Define the **skeleton** of two mDAGs as the undirected graph with \( v \rightarrow w \) whenever \( v \) and \( w \) are contained in some edge together.

Theorem mDAGs with different skeletons induce different models in general. (Consequence of Theorem 4.2 of Evans, 2012)
Skeletons

Define the **skeleton** of two mDAGs as the undirected graph with $v \sim w$ whenever $v$ and $w$ are contained in some edge together.

![Diagram showing two mDAGs with different skeletons.](image)

Theorem mDAGs with different skeletons induce different models in general. (Consequence of Theorem 4.2 of Evans, 2012)
Skeletons

Define the **skeleton** of two mDAGs as the undirected graph with $v \rightarrow w$ whenever $v$ and $w$ are contained in some edge together.

![Diagram showing two mDAGs with different skeletons](image)

**Theorem**

mDAGs with different skeletons induce different models in general.

(Consequence of Theorem 4.2 of Evans, 2012)
Combining the previous results, there are 8 Markov equivalence classes on three variables.
But Not on Four!

On four variables, it’s still not clear whether or not the following models are saturated: (they are of full dimension in the discrete case)
Outline

1. Introduction
2. DAGs
3. Margins of DAG Models
4. mDAGs
5. Markov Equivalence
6. Summary
We have seen that:

- graphs with ‘ordinary’ edges can give a causally coherent representation of marginal models;
- **but**: ordinary mixed graphs are not rich enough to represent all models;
- mDAGs provide the most general necessary framework for representing causal DAGs under marginalization;
- general Markov equivalence in this class is hard, but we’re getting there!
Thank you!
Main References


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 $a$ to $b$ is blocked by $C \subseteq V \setminus \{a, b\}$ if either

(i) any non-collider is in $C$:

\[
\begin{array}{c}
\circ \rightarrow C \rightarrow \circ \\
\end{array}
\quad\quad
\begin{array}{c}
\circ \leftarrow C \rightarrow \circ \\
\end{array}
\]
**d-Separation**

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

A path from \( a \) to \( b \) is **blocked** by \( C \subseteq V \setminus \{a, b\} \) if either

(i) any non-collider is in \( C \):

\[
\begin{align*}
\text{\includegraphics[width=0.5\textwidth]{path1.png}}
\end{align*}
\]

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

\[
\begin{align*}
\text{\includegraphics[width=0.5\textwidth]{path2.png}}
\end{align*}
\]

Two vertices \( a \) and \( b \) are **d-separated** given \( C \subseteq V \setminus \{a, b\} \) if all paths are blocked.
**d-Separation**

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

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

(i) any non-collider is in $C$:

```
    O---c---O                   O---c---O
```

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

```
    O---d---O                   O---d---O
    ^                        ^
     ^                        ^
     v                        v
     e
```

Two vertices $a$ and $b$ are **d-separated** given $C \subseteq V \setminus \{a, b\}$ if **all** paths are blocked.
Inequality Results

Let \( p^\ast(x, y | z) \equiv \int p(u) p(x | z, u) \cdot p(y | x, u) \, du \)

Can't observe \( p^\ast \) but:

- **Compatibility:** \( p(0, y | z) = p^\ast(0, y | z) \) for each \( z, y \);

- **Independence:** \( Y \perp \perp Z \) under \( p^\ast \).

This 'compatibility' requirement turns out to place an inequality restriction on \( p^\ast \):
\[
\max_x \sum_y \max_z p(x, y | z) \leq 1.
\]
Inequality Results

\[ p(x, y \mid z) = \int p(u) p(x \mid z, u) \cdot p(y \mid x, u) \, du \]
Inequality Results

\[
p(x, y \mid z) = \int p(u) p(x \mid z, u) \cdot p(y \mid x, u) \, du
\]

Let \( p^*(x, y \mid z) \equiv \int p(u) p(x \mid z, u) \cdot p(y \mid x = 0, u) \, du \)
Inequality Results

\[ p(x, y \mid z) = \int p(u) p(x \mid z, u) \cdot p(y \mid x, u) \, du \]

Let \( p^*(x, y \mid z) \equiv \int p(u) p(x \mid z, u) \cdot p(y \mid x = 0, u) \, du \)

Can’t observe \( p^* \) but:

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

\[ p(x, y \mid z) = \int p(u) p(x \mid z, u) \cdot p(y \mid x, u) \, du \]

Let \( p^*(x, y \mid z) \equiv \int p(u) p(x \mid z, u) \cdot p(y \mid x = 0, u) \, du \)

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 \( p^* \).

This 'compatibility' requirement turns out to place an inequality restriction on \( p \):

\[ \max_x \sum_y \max_z p(x, y \mid z) \leq 1. \]
Inequality Results

Generalizing this argument, we find a rich theory of results on inequalities (Evans, 2012).
Inequality Results

Generalizing this argument, we find a rich theory of results on inequalities (Evans, 2012).

However these results are **not exhaustive**! Finding **all** inequality constraints in marginal models is probably an NP hard problem.
Generalizing this argument, we find a rich theory of results on inequalities (Evans, 2012).

However these results are not exhaustive! Finding all inequality constraints in marginal models is probably an NP hard problem.

Additionally:

- fitting models with inequality constraints is not trivial;
- the usual asymptotic results do not necessarily apply.
Inequality Results

Generalizing this argument, we find a rich theory of results on inequalities (Evans, 2012).

However these results are not exhaustive! Finding all inequality constraints in marginal models is probably an NP hard problem.

Additionally:

- fitting models with inequality constraints is not trivial;
- the usual asymptotic results do not necessarily apply.

Maybe the nested model is a good compromise!
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 dichotomize.
ADMGs are not sufficient

Lemma

Let $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$ be mutually independent $\sigma$-algebrae (so that $\mathcal{F} \perp \mathcal{G} \vee \mathcal{H}$ 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

$$\text{Var } X < 3\epsilon.$$
Causal Equivalence

The two mDAGs below are Markov equivalent, and lead to the same graph under any ordinary causal intervention.