Non-parametric causal models

Robin J. Evans    Thomas S. Richardson

Oxford and Univ. of Washington

UAI Tutorial
12th July 2015
Structure

- Part One: Causal DAGs with latent variables
- Part Two: Statistical Models arising from DAGs with latents
Outline for Part One

- Intervention distributions
- The general identification problem
- Tian’s ID Algorithm
- Fixing: generalizing marginalizing and conditioning
- Non-parametric constraints aka Verma constraints
Intervention distributions (I)

Given a causal DAG $\mathcal{G}$ with distribution:

$$p(V) = \prod_{v \in V} p(v \mid \text{pa}(v))$$

we wish to compute an intervention distribution via truncated factorization:

$$p(V \setminus X \mid \text{do}(X = x)) = \prod_{v \in V \setminus X} p(v \mid \text{pa}(v)).$$
Example

\[ p(X, L, M, Y) = p(L) p(X \mid L) p(M \mid X)p(Y \mid L, M) \]
Example

\[ p(X, L, M, Y) = p(L) \cdot p(X | L) \cdot p(M | X) \cdot p(Y | L, M) \]

\[ p(L, M, Y \mid \text{do}(X = \tilde{x})) = p(L) \times p(M | \tilde{x}) \cdot p(Y | L, M) \]
Intervention distributions (II)

Given a causal DAG $G$ with distribution:

$$p(V) = \prod_{v \in V} p(v \mid \text{pa}(v))$$

we wish to compute an intervention distribution via truncated factorization:

$$p(V \setminus X \mid \text{do}(X = x)) = \prod_{v \in V \setminus X} p(v \mid \text{pa}(v)).$$

Hence if we are interested in $Y \subset V \setminus X$ then we simply marginalize:

$$p(Y \mid \text{do}(X = x)) = \sum_{w \in V \setminus (X \cup Y)} \prod_{v \in V \setminus X} p(v \mid \text{pa}(v)).$$

This is the ‘g-computation’ formula of Robins (1986).
Intervention distributions (II)

Given a causal DAG $G$ with distribution:

$$p(V) = \prod_{v \in V} p(v | \text{pa}(v))$$

we wish to compute an intervention distribution via truncated factorization:

$$p(V \setminus X | \text{do}(X = x)) = \prod_{v \in V \setminus X} p(v | \text{pa}(v)).$$

Hence if we are interested in $Y \subset V \setminus X$ then we simply marginalize:

$$p(Y | \text{do}(X = x)) = \sum_{w \in V \setminus (X \cup Y)} \prod_{v \in V \setminus X} p(v | \text{pa}(v)).$$

This is the ‘g-computation’ formula of Robins (1986).

Note: $p(Y | \text{do}(X = x))$ is a sum over a product of terms $p(v | \text{pa}(v))$. 
Example

\[ p(X, L, M, Y) = p(L)p(X \mid L)p(M \mid X)p(Y \mid L, M) \]

\[ p(L, M, Y \mid \text{do}(X = \tilde{x})) = p(L)p(M \mid \tilde{x})p(Y \mid L, M) \]

\[ p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l, m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l, M = m) \]
Example

\[
p(X, L, M, Y) = p(L)p(X \mid L)p(M \mid X)p(Y \mid L, M)
\]

\[
p(L, M, Y \mid \text{do}(X = \tilde{x})) = p(L)p(M \mid \tilde{x})p(Y \mid L, M)
\]

\[
p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l,m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l, M = m)
\]

Note that \( p(Y \mid \text{do}(X = \tilde{x})) \neq p(Y \mid X = \tilde{x}) \).
Example: no effect of $M$ on $Y$

\[ p(X, L, M, Y) = p(L)p(X | L)p(M | X)p(Y | L, M) \]
Example: no effect of $M$ on $Y$

\[
p(X, L, M, Y) = p(L)p(X \mid L)p(M \mid X)p(Y \mid L)
\]
Example: no effect of $M$ on $Y$

$$p(X, L, M, Y) = p(L)p(X \mid L)p(M \mid X)p(Y \mid L)$$

$$p(L, M, Y \mid \text{do}(X = \tilde{x})) = p(L)p(M \mid \tilde{x})p(Y \mid L)$$
Example: no effect of $M$ on $Y$

\[ p(X, L, M, Y) = p(L)p(X \mid L)p(M \mid X)p(Y \mid L) \]

\[ p(L, M, Y \mid do(X = \tilde{x})) = p(L)p(M \mid \tilde{x})p(Y \mid L) \]

\[ p(Y \mid do(X = \tilde{x})) = \sum_{l,m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l) \]
Example: no effect of $M$ on $Y$

$$p(X, L, M, Y) = p(L)p(X \mid L)p(M \mid X)p(Y \mid L)$$

$$p(L, M, Y \mid \text{do}(X = \tilde{x})) = p(L)p(M \mid \tilde{x})p(Y \mid L)$$

$$p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l,m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l)$$

$$= \sum_{l} p(L = l)p(Y \mid L = l)$$

since $X \not\perp \! \! \! \perp Y$. 'Correlation is not Causation'.
Example: no effect of $M$ on $Y$

$$p(X, L, M, Y) = p(L)p(X \mid L)p(M \mid X)p(Y \mid L)$$

$$p(L, M, Y \mid \text{do}(X = \tilde{x})) = p(L)p(M \mid \tilde{x})p(Y \mid L)$$

$$p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l, m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l)$$

$$= \sum_{l} p(L = l)p(Y \mid L = l)$$

$$= p(Y) \neq P(Y \mid \tilde{x})$$

since $X \not\perp \!\!\!\!\perp Y$. ‘Correlation is not Causation’.
Example with $M$ unobserved

$$p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l,m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l, M = m)$$
Example with $M$ unobserved

\[ p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l,m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l, M = m) \]

\[ = \sum_{l,m} p(L = l)p(M = m \mid \tilde{x}, L = l)p(Y \mid L = l, M = m, X = \tilde{x}) \]

Here we have used that $M \perp \perp L \mid X$ and $Y \perp \perp X \mid L, M$. 
Example with $M$ unobserved

\[
p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l,m} p(L = l)p(M = m \mid \tilde{x})p(Y \mid L = l, M = m)
\]

\[
= \sum_{l,m} p(L = l)p(M = m \mid \tilde{x}, L = l)p(Y \mid L = l, M = m, X = \tilde{x})
\]

\[
= \sum_{l,m} p(L = l)p(Y, M = m \mid L = l, X = \tilde{x})
\]
Example with $M$ unobserved

\[ p(Y \mid \text{do}(X = \tilde{x})) = \sum_{l,m} p(L = l) p(M = m \mid \tilde{x}) p(Y \mid L = l, M = m) \]

\[ = \sum_{l,m} p(L = l) p(M = m \mid \tilde{x}, L = l) p(Y \mid L = l, M = m, X = \tilde{x}) \]

\[ = \sum_{l,m} p(L = l) p(Y, M = m \mid L = l, X = \tilde{x}) \]

\[ = \sum_{l} p(L = l) p(Y \mid L = l, X = \tilde{x}). \]

⇒ can find $p(Y \mid \text{do}(X = \tilde{x}))$ even if $M$ not observed.

This is an example of the ‘back door formula’.
Example with $L$ unobserved

$$p(Y \mid \text{do}(X = \tilde{x}))$$
Example with $L$ unobserved

\[
p(Y \mid \text{do}(X = \tilde{x})) = \sum_{m} p(M = m \mid \text{do}(X = \tilde{x})) p(Y \mid \text{do}(M = m))
\]
Example with $L$ unobserved

$$p(Y \mid \text{do}(X = \tilde{x}))$$

$$= \sum_m p(M = m \mid \text{do}(X = \tilde{x}))p(Y \mid \text{do}(M = m))$$

$$= \sum_m p(M = m \mid X = \tilde{x})p(Y \mid \text{do}(M = m))$$
Example with \( L \) unobserved

\[
\begin{align*}
p(Y \mid \text{do}(X = \tilde{x})) &= \sum_m p(M = m \mid \text{do}(X = \tilde{x}))p(Y \mid \text{do}(M = m)) \\
&= \sum_m p(M = m \mid X = \tilde{x})p(Y \mid \text{do}(M = m)) \\
&= \sum_m p(M = m \mid X = \tilde{x}) \left( \sum_{x^*} p(X = x^*)p(Y \mid M = m, X = x^*) \right)
\end{align*}
\]
Example with $L$ unobserved

\[
p(Y \mid \text{do}(X = \tilde{x})) = \sum_m p(M = m \mid \text{do}(X = \tilde{x}))p(Y \mid \text{do}(M = m))
\]

\[
= \sum_m p(M = m \mid X = \tilde{x})p(Y \mid \text{do}(M = m))
\]

\[
= \sum_m p(M = m \mid X = \tilde{x}) \left( \sum_{x^*} p(X = x^*)p(Y \mid M = m, X = x^*) \right)
\]

⇒ can find $p(Y \mid \text{do}(X = \tilde{x}))$ even if $L$ not observed.

This is an example of the ‘front door formula’.
But with both $L$ and $M$ unobserved....

...we are out of luck!
But with both $L$ and $M$ unobserved....

...we are out of luck!

Given $P(X, Y)$, absent further assumptions we cannot distinguish:
Given: a latent DAG $G(O \cup H)$, where $O$ are observed, $H$ are hidden, and disjoint subsets $X, Y \subseteq O$.

Q: Is $p(Y | \text{do}(X))$ identified given $p(O)$?
General Identification Question

Given: a latent DAG $\mathcal{G}(O \cup H)$, where $O$ are observed, $H$ are hidden, and disjoint subsets $X, Y \subseteq O$.

Q: Is $p(Y \mid \text{do}(X))$ identified given $p(O)$?

A: Provide either an identifying formula that is a function of $p(O)$

or report that $p(Y \mid \text{do}(X))$ is not identified.
Latent Projection

Can preserve conditional independences and causal coherence with latents using paths. DAG $G$ on vertices $V = O \cup H$, define latent projection as follows: (Verma and Pearl, 1992)
Latent Projection

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

Whenever there is a path of the form

![Diagram](image)

add

![Diagram](image)
Latent Projection

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

Whenever there is a path of the form

$$x \rightarrow h_1 \rightarrow \cdots \rightarrow h_k \rightarrow y$$

add

$$x \rightarrow y$$

Whenever there is a path of the form

$$x \leftarrow h_1 \leftarrow \cdots \leftarrow h_k \rightarrow y$$

add

$$x \leftrightarrow y$$
Latent Projection

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

Whenever there is a path of the form

\[ x \rightarrow h_1 \rightarrow \cdots \rightarrow h_k \rightarrow y \]

add

\[ x \rightarrow y \]

Whenever there is a path of the form

\[ x \leftarrow h_1 \leftarrow \cdots \leftarrow h_k \rightarrow y \]

add

\[ x \leftrightarrow y \]

Then remove all latent variables $H$ from the graph.
Latent projection leads to an acyclic directed mixed graph (ADMG). Can read off independences with d/m-separation. The projection preserves the causal structure; Verma and Pearl (1992).
Latent projection leads to an **acyclic directed mixed graph** (ADMG)
Latent projection leads to an **acyclic directed mixed graph** (ADMG)

Can read off independences with d/m-separation.

The projection preserves the causal structure; Verma and Pearl (1992).
‘Conditional’ Acyclic Directed Mixed Graphs

An ‘conditional’ acyclic directed mixed graph (CADMG) is a bi-partite graph $G(V, W)$, used to represent structure of a distribution over $V$, indexed by $W$, for example $P(V \mid \text{do}(W))$.

We require:

(i) The induced subgraph of $G$ on $V$ is an ADMG;
(ii) The induced subgraph of $G$ on $W$ contains no edges;
(iii) Edges between vertices in $W$ and $V$ take the form $w \rightarrow v$.

We represent $V$ with circles, $W$ with squares:

Here $V = \{L_1, Y\}$ and $W = \{A_0, A_1\}$. 
Ancestors and Descendants

In a CADMG $G(V, W)$ for $v \in V$, let the set of ancestors, descendants of $v$ be:

$$\text{an}_G(v) = \{ a | a \rightarrow \cdots \rightarrow v \text{ or } a = v \text{ in } G, a \in V \cup W \},$$

$$\text{de}_G(v) = \{ d | d \leftarrow \cdots \leftarrow v \text{ or } d = v \text{ in } G, d \in V \cup W \},$$
Ancestors and Descendants

In a CADMG $\mathcal{G}(V, W)$ for $v \in V$, let the set of ancestors, descendants of $v$ be:

$$\text{an}_G(v) = \{ a \mid a \rightarrow \cdots \rightarrow v \text{ or } a = v \text{ in } G, a \in V \cup W \},$$

$$\text{de}_G(v) = \{ d \mid d \leftarrow \cdots \leftarrow v \text{ or } d = v \text{ in } G, d \in V \cup W \},$$

In the example above:

$$\text{an}(y) = \{ a_0, l_1, a_1, y \}.$$
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

![Diagram of connected nodes](image)

Districts are called 'c-components' by Tian.

\[
\sum_{u, v} p(x_1 | u) p(x_2 | u) = \sum_{u, v} p(x_1 | u) p(x_2 | u) \cdot \sum_{u, v} p(x_3 | x_1, v) p(x_4 | x_2, v) p(x_5 | x_3) = \prod_i q_{D_i}(x_{D_i} | x_{pa(D_i)} - D_i)
\]
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

![Diagram](image-url)
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

\[
\sum_{u,v} p(u) p(x_1 \mid u) p(x_2 \mid u) \ p(v) p(x_3 \mid x_1, v) p(x_4 \mid x_2, v) \ p(x_5 \mid x_3)
\]

Districts are called ‘c-components’ by Tian.
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

\[
\sum_{u,v} p(u) p(x_1 | u) p(x_2 | u) p(v) p(x_3 | x_1, v) p(x_4 | x_2, v) p(x_5 | x_3)
\]
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

\[
\sum_{u,v} p(u) p(x_1 | u) p(x_2 | u) \cdot p(v) p(x_3 | x_1, v) p(x_4 | x_2, v) \cdot p(x_5 | x_3)
\]

\[
= \sum_{u} p(u) p(x_1 | u) p(x_2 | u) \sum_{v} p(v) p(x_3 | x_1, v) p(x_4 | x_2, v) \cdot p(x_5 | x_3)
\]

Districts are called 'c-components' by Tian.
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

\[
\sum_{u,v} p(u) p(x_1 \mid u) p(x_2 \mid u) \cdot p(v) p(x_3 \mid x_1, v) p(x_4 \mid x_2, v) \cdot p(x_5 \mid x_3) = q(x_1, x_2) \cdot q(x_3, x_4 \mid x_1, x_2) \cdot q(x_5 \mid x_3)
\]
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

\[
\sum_{u,v} p(u) p(x_1 | u) p(x_2 | u) \cdot p(v) p(x_3 | x_1, v) p(x_4 | x_2, v) \cdot p(x_5 | x_3)
\]

\[
= \sum_u p(u) p(x_1 | u) p(x_2 | u) \sum_v p(v) p(x_3 | x_1, v) p(x_4 | x_2, v) \cdot p(x_5 | x_3)
\]

\[
= q(x_1, x_2) \cdot q(x_3, x_4 | x_1, x_2) \cdot q(x_5 | x_3).
\]

\[
= \prod_i q_{D_i}(x_{D_i} | x_{pa(D_i) \setminus D_i})
\]
Districts

Define a **district** in a C/ADMG to be maximal sets connected by bi-directed edges:

\[
\sum_{u,v} p(u) \, p(x_1 | u) \, p(x_2 | u) \, p(v) \, p(x_3 | x_1, v) \, p(x_4 | x_2, v) \, p(x_5 | x_3) \\
= \sum_u p(u) \, p(x_1 | u) \, p(x_2 | u) \sum_v p(v) \, p(x_3 | x_1, v) \, p(x_4 | x_2, v) \, p(x_5 | x_3) \\
= q(x_1, x_2) \cdot q(x_3, x_4 | x_1, x_2) \cdot q(x_5 | x_3) \\
= \prod_i q_{D_i}(x_{D_i} | x_{\text{pa}(D_i) \setminus D_i})
\]

Districts are called ‘c-components’ by Tian.
Edges between districts

There is no ordering on vertices such that parents of a district precede every vertex in the district.

(Cannot form a ‘chain graph’ ordering.)
Notation for Districts

In a CADMG $\mathcal{G}(V, W)$ for $v \in V$, the district of $v$ is:

$$\text{dis}_{\mathcal{G}}(v) = \{ d \mid d \leftrightarrow \cdots \leftrightarrow v \text{ or } d = v \text{ in } \mathcal{G}, d \in V \}.$$ 

Only variables in $V$ are in districts.
In a CADMG $\mathcal{G}(V, W)$ for $v \in V$, the district of $v$ is:

$$\text{dis}_\mathcal{G}(v) = \{ d | d \leftrightarrow \cdots \leftrightarrow v \text{ or } d = v \text{ in } \mathcal{G}, d \in V \}.$$ 

Only variables in $V$ are in districts.

In example above:

$$\text{dis}(y) = \{ l_0, l_1, y \}, \quad \text{dis}(a_1) = \{ a_1 \}.$$
Notation for Districts

In a CADMG $\mathcal{G}(V, W)$ for $v \in V$, the district of $v$ is:

$$\text{dis}_\mathcal{G}(v) = \{ d \mid d \leftrightarrow \cdots \leftrightarrow v \text{ or } d = v \text{ in } \mathcal{G}, \ d \in V \}.$$ 

Only variables in $V$ are in districts.

In example above:

$$\text{dis}(y) = \{ l_0, l_1, y \}, \quad \text{dis}(a_1) = \{ a_1 \}.$$ 

We use $\mathcal{D}(\mathcal{G})$ to denote the set of districts in $\mathcal{G}$.

In example $\mathcal{D}(\mathcal{G}) = \{ \{ l_0, l_1, y \}, \{ a_1 \} \}$.
(A) Re-express the query as a sum over a product of intervention distributions on districts:

\[ p(Y \mid \text{do}(X)) = \sum \prod_{i} p(D_i \mid \text{do}(\text{pa}(D_i) \setminus D_i)). \]
Tian’s ID algorithm for identifying $P(Y \mid \text{do}(X))$

(A) Re-express the query as a sum over a product of intervention distributions on districts:

$$p(Y \mid \text{do}(X)) = \sum \prod_i p(D_i \mid \text{do}(\text{pa}(D_i) \setminus D_i)).$$

(B) Check whether each term: $p(D_i \mid \text{do}(\text{pa}(D_i) \setminus D_i))$ is identified.
Tian’s ID algorithm for identifying $P(Y \mid \text{do}(X))$

(A) Re-express the query as a sum over a product of intervention distributions on districts:

$$p(Y \mid \text{do}(X)) = \sum \prod_{i} p(D_i \mid \text{do(pa}(D_i) \setminus D_i)).$$

(B) Check whether each term: $p(D_i \mid \text{do(pa}(D_i) \setminus D_i))$ is identified.

This is clearly sufficient for identifiability.

Necessity follows from results of Shpitser (2006).
(A) Decomposing the query

4. Remove edges into $X$:
   Let $G[V \setminus X]$ denote the graph formed by removing edges with an arrowhead into $X$. 
(A) Decomposing the query

1. **Remove edges into $X$:**
   Let $G[V \setminus X]$ denote the graph formed by removing edges with an arrowhead into $X$.

2. **Restrict to variables that are (still) ancestors of $Y$:**
   Let $T = \text{an}_{G[V \setminus X]}(Y)$ be vertices that lie on directed paths between $X$ and $Y$ (after intervening on $X$).
(A) Decomposing the query

1. Remove edges into $X$:
   Let $G[V \setminus X]$ denote the graph formed by removing edges with an arrowhead into $X$.

2. Restrict to variables that are (still) ancestors of $Y$:
   Let $T = \text{an}_{G[V \setminus X]}(Y)$ be vertices that lie on directed paths between $X$ and $Y$ (after intervening on $X$).
   Let $G^*$ be formed from $G[V \setminus X]$ by removing vertices not in $T$. 

(A) Decomposing the query

1. Remove edges into $X$:
   Let $G[V \setminus X]$ denote the graph formed by removing edges with an arrowhead into $X$.

2. Restrict to variables that are (still) ancestors of $Y$:
   Let $T = \text{an}_{G[V \setminus X]}(Y)$ be vertices that lie on directed paths between $X$ and $Y$ (after intervening on $X$).
   Let $G^*$ be formed from $G[V \setminus X]$ by removing vertices not in $T$.

3. Find the districts:
   Let $D_1, \ldots, D_s$ be the districts in $G^*$.
(A) Decomposing the query

1. **Remove edges into X:**
   Let $\mathcal{G}[V \setminus X]$ denote the graph formed by removing edges with an arrowhead into $X$.

2. **Restrict to variables that are (still) ancestors of Y:**
   Let $T = \text{an}_{\mathcal{G}[V \setminus X]}(Y)$ be vertices that lie on directed paths between $X$ and $Y$ (after intervening on $X$).
   Let $\mathcal{G}^*$ be formed from $\mathcal{G}[V \setminus X]$ by removing vertices not in $T$.

3. **Find the districts:**
   Let $D_1, \ldots, D_s$ be the districts in $\mathcal{G}^*$.

Then:

$$P(Y \mid \text{do}(X)) = \sum_{T \setminus (X \cup Y)} \prod_{D_i} p(D_i \mid \text{do}(\text{pa}(D_i) \setminus D_i)).$$
Example: front door graph

\[ G \]

\[ p(Y \mid \text{do}(X)) \]
Example: front door graph

\[ G \]

\[ G_{[V \setminus \{X\}]} = G^* \]

\[ \begin{align*}
X & \quad M \quad Y \\
p(Y \mid \text{do}(X)) &
\end{align*} \]

\[ \begin{align*}
X & \quad M \quad Y \\
T = \{X, M, Y\} &
\end{align*} \]
Example: front door graph

\[ p(Y \mid \text{do}(X)) = \sum_{M} p(M \mid \text{do}(X))p(Y \mid \text{do}(M)) \]

Districts in \( T \setminus \{A_0, A_1\} \) are \( D_1 = \{M\}, D_2 = \{Y\} \).

\[ \mathcal{G}_{[\mathcal{V}\setminus\{X\}]} = \mathcal{G}^* \]
Example: The Verma Graph

\[ G = (V, E) \]

\[ p(Y \mid \text{do}(A_0, A_1)) \]
Example: The Verma Graph

\[ G \]

\[ A_0 \xrightarrow{} L_1 \xrightarrow{} A_1 \xrightarrow{} Y \]

\[ p(Y \mid \text{do}(A_0, A_1)) \]

\[ G_{[V \setminus \{A_0, A_1\}] } \]

\[ A_0 \xrightarrow{} L_1 \xrightarrow{} A_1 \xrightarrow{} Y \]

\[ T = \{ A_0, A_1, Y \} \]
Example: The Verma Graph

\[ G \]

\[ \begin{align*}
A_0 & \xrightarrow{} L_1 & \xrightarrow{} A_1 & \xrightarrow{} Y \\
\end{align*} \]

\[ p(Y \mid do(A_0, A_1)) \]

\[ G[V \setminus \{A_0, A_1\}] \]

\[ \begin{align*}
A_0 & \xrightarrow{} L_1 & A_1 & \xrightarrow{} Y \\
\end{align*} \]

\[ T = \{A_0, A_1, Y\} \]

\[ G^* \]

\[ \begin{align*}
A_0 & \xrightarrow{} A_1 & \xrightarrow{} Y \\
\end{align*} \]

\[ D_1 = \{Y\} \]
Example: The Verma Graph

\[ G \]

\[ A_0 \rightarrow L_1 \rightarrow A_1 \rightarrow Y \]

\[ p(Y \mid do(A_0, A_1)) \]

\[ G_{[V \backslash \{A_0, A_1\}]} \]

\[ A_0 \rightarrow L_1 \rightarrow A_1 \rightarrow Y \]

\[ T = \{A_0, A_1, Y\} \]

\[ G^* \]

\[ A_0 \rightarrow A_1 \rightarrow Y \]

\[ D_1 = \{Y\} \]

(Here the decomposition is trivial since there is only one district and no summation.)
Finding if $P(D \mid \text{do}(\text{pa}(D) \setminus D))$ is identified

Idea: Find an ordering $r_1, \ldots, r_p$ of $O \setminus D$ such that:

If $P(O \setminus \{r_1, \ldots, r_{t-1}\} \mid \text{do}(r_1, \ldots, r_{t-1}))$ is identified

Then $P(O \setminus \{r_1, \ldots, r_t\} \mid \text{do}(r_1, \ldots, r_t))$ is also identified.
Finding if $P(D \mid \text{do}(\text{pa}(D) \setminus D))$ is identified

Idea: Find an ordering $r_1, \ldots, r_p$ of $O \setminus D$ such that:

If $P(O \setminus \{r_1, \ldots, r_{t-1}\} \mid \text{do}(r_1, \ldots, r_{t-1}))$ is identified

Then $P(O \setminus \{r_1, \ldots, r_t\} \mid \text{do}(r_1, \ldots, r_t))$ is also identified.

Sufficient for identifiability of $P(D \mid \text{do}(\text{pa}(D) \setminus D))$, since:

- $P(O)$ is identified
- $D = O \setminus \{r_1, \ldots, r_p\}$, so
- $P(O \setminus \{r_1, \ldots, r_p\} \mid \text{do}(r_1, \ldots, r_p)) = P(D \mid \text{do}(\text{pa}(D) \setminus D))$. 
(B) Finding if $P(D \mid \text{do}(\text{pa}(D) \setminus D))$ is identified

Idea: Find an ordering $r_1, \ldots, r_p$ of $O \setminus D$ such that:

If $P(O \setminus \{r_1, \ldots, r_{t-1}\} \mid \text{do}(r_1, \ldots, r_{t-1}))$ is identified

Then $P(O \setminus \{r_1, \ldots, r_t\} \mid \text{do}(r_1, \ldots, r_t))$ is also identified.

Sufficient for identifiability of $P(D \mid \text{do}(\text{pa}(D) \setminus D))$, since:

$P(O)$ is identified

$D = O \setminus \{r_1, \ldots, r_p\}$, so

$P(O \setminus \{r_1, \ldots, r_p\} \mid \text{do}(r_1, \ldots, r_p)) = P(D \mid \text{do}(\text{pa}(D) \setminus D))$.

Such a vertex $r_t$ will said to be ‘fixable’, given that we have already ‘fixed’ $r_1, \ldots, r_{t-1}$:

‘fixing’ differs from ‘do’/intervening since the latter does not preserve identifiability.
(B) Finding if \( P(D \mid \text{do(pa}(D) \setminus D)) \) is identified

Idea: Find an ordering \( r_1, \ldots, r_p \) of \( O \setminus D \) such that:

If \( P(O \setminus \{r_1, \ldots, r_{t-1}\} \mid \text{do}(r_1, \ldots, r_{t-1})) \) is identified

Then \( P(O \setminus \{r_1, \ldots, r_t\} \mid \text{do}(r_1, \ldots, r_t)) \) is also identified.

Sufficient for identifiability of \( P(D \mid \text{do(pa}(D) \setminus D)) \), since:

\( P(O) \) is identified

\( D = O \setminus \{r_1, \ldots, r_p\} \), so

\( P(O \setminus \{r_1, \ldots, r_p\} \mid \text{do}(r_1, \ldots, r_p)) = P(D \mid \text{do(pa}(D) \setminus D)). \)

Such a vertex \( r_t \) will said to be ‘fixable’, given that we have already ‘fixed’ \( r_1, \ldots, r_{t-1} \):

‘fixing’ differs from ‘do’/intervening since the latter does not preserve identifiability.

To do:

- Give a graphical characterization of ‘fixability’;
- Construct the identifying formula.
The set of fixable vertices

Given a CADMG $G(V, W)$ we define the set of fixable vertices,

$$F(G) \equiv \{v \mid v \in V, \text{dis}_G(v) \cap \text{de}_G(v) = \{v\}\}.$$ 

In words, a vertex $v \in V$ is fixable in $G$ if there is no (proper) descendant of $v$ that is in the same district as $v$ in $G$. 

Note that the set of fixable vertices is a subset of $V$, and contains at least one vertex from each district in $G$. 

The set of fixable vertices

Given a CADMG $G(V, W)$ we define the set of fixable vertices,

$$F(G) \equiv \{v \mid v \in V, \text{dis}_G(v) \cap \text{de}_G(v) = \{v\}\}.$$ 

In words, a vertex $v \in V$ is fixable in $G$ if there is no (proper) descendant of $v$ that is in the same district as $v$ in $G$.

Thus $v$ is fixable if there is no vertex $y \neq v$ such that

$$v \leftrightarrow \cdots \leftrightarrow y \quad \text{and} \quad v \rightarrow \cdots \rightarrow y \quad \text{in} \ G.$$

Note that the set of fixable vertices is a subset of $V$, and contains at least one vertex from each district in $G$. 
Example: front door graph

\[ F(G) = \{ M, Y \} \]

\( X \) is not fixable since \( Y \) is a descendant of \( X \) and

\( Y \) is in the same district as \( X \)
Example: The Verma Graph

Here $F(\mathcal{G}) = \{A_0, A_1, Y\}$.

$L_1$ is not fixable since $Y$ is a descendant of $L_1$ and $Y$ is in the same district as $L_1$. 
The \textit{graphical} operation of fixing vertices

Given a CADMG $G(V, W, E)$, for every $r \in F(G)$ we associate a transformation $\phi_r$ on the pair $(G, P(X_V \mid X_W))$: 

$$\phi_r(G) \equiv G^{\dagger}(V \setminus \{r\}, W \cup \{r\}),$$

where in $G^{\dagger}$ we remove from $G$ any edge that has an arrowhead at $r$. 

The *graphical* operation of fixing vertices

Given a CADMG $\mathcal{G}(V, W, E)$, for every $r \in F(\mathcal{G})$ we associate a transformation $\phi_r$ on the pair $(\mathcal{G}, P(X_V \mid X_W))$:

$$\phi_r(\mathcal{G}) \equiv \mathcal{G}^\dagger(V \setminus \{r\}, W \cup \{r\}),$$

where in $\mathcal{G}^\dagger$ we remove from $\mathcal{G}$ any edge that has an arrowhead at $r$.

The operation of ‘fixing $r$’ simply transfers $r$ from ‘$V$’ to ‘$W$’, and removes edges $r \leftrightarrow$ or $r \leftarrow$. 
Example: front door graph

\[ F(\mathcal{G}) = \{M, Y\} \]

\[ F(\phi_M(\mathcal{G})) = \{X, Y\} \]

Note that \(X\) was not fixable in \(\mathcal{G}\),
but it is fixable in \(\phi_M(\mathcal{G})\) after fixing \(M\).
Example: The Verma Graph

Here $F(G) = \{A_0, A_1, Y\}$.

Notice $F(\phi_{A_1}(G)) = \{A_0, L_1, Y\}$.

Thus $L_1$ was not fixable prior to fixing $A_1$, but $L_1$ is fixable in $\phi_{A_1}(G)$ after fixing $A_1$. 
The probabilistic operation of fixing vertices

Given a distribution $P(V \mid W)$ we associate a transformation:

$$\phi_r(P(V \mid W); G) \equiv P(V \mid W)/P(r \mid \text{mb}_G(r)).$$

Here

$$\text{mb}_G(r) = \{ y \neq r \mid (r \leftarrow y) \text{ or } (r \leftrightarrow o \cdots o \leftrightarrow y) \text{ or } (r \leftrightarrow o \cdots o \leftrightarrow o \leftarrow y) \}.$$  

In words: we divide by the conditional distribution of $r$ given the other vertices in the district containing $r$, and the parents of the vertices in that district.
The *probabilistic* operation of fixing vertices

Given a distribution $P(V \mid W)$ we associate a transformation:

$$
\phi_r(P(V \mid W); G) \equiv P(V \mid W)/P(r \mid mb_G(r)).
$$

Here

$$
mb_G(r) = \{y \neq r \mid (r \leftarrow y) \text{ or } (r \leftrightarrow o \cdots o \leftrightarrow y) \text{ or } (r \leftrightarrow o \cdots o \leftrightarrow o \leftarrow y)\}.
$$

In words: *we divide by the conditional distribution of $r$ given the other vertices in the district containing $r$, and the parents of the vertices in that district.*

It can be shown that if $r$ is fixable in $G$ then:

$$
\phi_r(P(V \mid \text{do}(W)); G) = P(V \setminus \{r\} \mid \text{do}(W \cup \{r\})).
$$

as required.

Note: If $r$ is fixable in $G$ then $mb_G(r)$ is the ‘Markov blanket’ of $r$ in $\text{an}_G(\text{dis}_G(r))$. 
Unifying Marginalizing and Conditioning

Some special cases:

- If $\text{mb}_G(r) = (V \cup W) \setminus \{r\}$ then fixing corresponds to marginalizing:

  $$\phi_r(P(V \mid W); G) = \frac{P(V \mid W)}{P(r \mid (V \cup W) \setminus \{r\})} = P(V \setminus \{r\} \mid W)$$

- If $\text{mb}_G(r) = W$ then fixing corresponds to ordinary conditioning:

  $$\phi_r(P(V \mid W); G) = \frac{P(V \mid W)}{P(r \mid W)} = P(V \setminus \{r\} \mid W \cup \{r\})$$

- In the general case fixing corresponds to re-weighting, so

  $$\phi_r(P(V \mid W); G) = P^*(V \setminus \{r\} \mid W \cup \{r\}) \neq P(V \setminus \{r\} \mid W \cup \{r\})$$
Composition of fixing operations

We use $\circ$ to indicate composition of operations in the natural way, so that:

$$\phi_r \circ \phi_s(G) \equiv \phi_r(\phi_s(G))$$
$$\phi_r \circ \phi_s(P(V \mid W); G) \equiv \phi_r(\phi_s(P(V \mid W); G); \phi_s(G))$$
Example: front door graph ($D_1$)

\[ F(\mathcal{G}) = \{M, Y\} \]

\[ F(\phi_Y(\mathcal{G})) = \{X, M\} \]

\[ \phi_X \circ \phi_Y(\mathcal{G}) \]

This proves that $p(M \mid \text{do}(X))$ is identified.
Example: front door graph \((D_2)\)

\[
\begin{align*}
F(G) &= \{M, Y\} \\
F(\phi_M(G)) &= \{X, Y\}
\end{align*}
\]

This proves that \(p(Y \mid \text{do}(M))\) is identified.
Example: The Verma Graph

This establishes that $P(Y \mid \text{do}(A_0, A_1))$ is identified.
(A) Re-express the query as a sum over a product of intervention distributions on districts:

\[ p(Y \mid \text{do}(X)) = \sum \prod_i p(D_i \mid \text{do(\text{pa}(D_i) \setminus D_i)}). \]

- Cut edges into \( X \);
- Restrict to vertices that are (still) ancestors of \( Y \);
- Find the set of districts \( D_1, \ldots, D_p \).
Review: Tian’s ID algorithm via fixing

(A) Re-express the query as a sum over a product of intervention distributions on districts:

\[ p(Y \mid \text{do}(X)) = \sum \prod_i p(D_i \mid \text{do}(\text{pa}(D_i) \setminus D_i)). \]

- Cut edges into \( X \);
- Restrict to vertices that are (still) ancestors of \( Y \);
- Find the set of districts \( D_1, \ldots, D_p \).

(B) Check whether each term: \( p(D_i \mid \text{do}(\text{pa}(D_i) \setminus D_i)) \) is identified.

- Iteratively find a vertex that \( r_t \) that is fixable in \( \phi_{r_{t-1}} \circ \cdots \circ \phi_{r_1}(G) \), with \( r_t \notin D_i \);
- If no such vertex exists then \( P(D_i \mid \text{do}(\text{pa}(D_i) \setminus D_i)) \) is not identified.
Not identified example

\[ F(G) = \{ Y \} \]

We see that \( p(Y \mid \text{do}(M)) \) is not identified since the only fixable vertex is \( Y \).
Reachable subgraphs of an ADMG

A CADMG $G(V, W)$ is reachable from ADMG $G^*(V \cup W)$ if there is an ordering of the vertices in $W = \langle w_1, \ldots, w_k \rangle$, such that for $j = 1, \ldots, k$,

$$w_1 \in F(G^*) \text{ and for } j = 2, \ldots, k,$$

$$w_j \in F(\phi_{w_{j-1}} \circ \cdots \circ \phi_{w_1}(G^*)).$$

Thus a subgraph is reachable if, under some ordering, each of the vertices in $W$ may be fixed, first in $G^*$, and then in $\phi_{w_1}(G^*)$, then in $\phi_{w_2}(\phi_{w_1}(G^*))$, and so on.
Intrinsic sets

A set $D$ is said to be *intrinsic* if it forms a *district* in a *reachable* subgraph.

If $D$ is intrinsic in $G$ then $p(D \mid \text{do}(\text{pa}(D) \setminus D))$ is identified.

The intervention distributions $p(D \mid \text{do}(\text{pa}(D) \setminus D))$ for intrinsic $D$ play the same role as $P(v \mid \text{do}(\text{pa}(v))) = p(v \mid \text{pa}(v))$ in the simple fully observed case.

Given an ADMG $G$ we let $\mathcal{I}(G)$ denote the intrinsic sets in $G$. 
Intrinsic sets and ‘hedges’

Shpitser (2006) provided a characterization in terms of graphical structures called ‘hedges’ of those interventional distributions that were not identified.

It may be shown that if a $\leftrightarrow$-connected set is not intrinsic then there exists a hedge, hence we have:

$\leftrightarrow$-connected set $S$ is intrinsic iff $p(S \mid \text{do}(\text{pa}(S) \setminus S))$ is identified.

It follows that intrinsic sets may thus also be defined in terms of the non-existence of a hedge.
Deriving constraints via fixing

Let $p(O)$ be the observed margin from a DAG with latents $G(O \cup H)$,

**Idea:** If $r \in O$ is fixable then $\phi_r(p(O); G)$ will obey the Markov property for the graph $\phi_r(G)$.

... and this can be iterated.

This gives non-parametric constraints that are not independences, that are implied by the latent DAG.
Example: The Verma Constraint

Here $F(G) = \{A_0, A_1, Y\}$. 
Example: The Verma Constraint

Here $F(\mathcal{G}) = \{A_0, A_1, Y\}$.

$\phi_{A_1}(\mathcal{G})$

$\phi_{A_1}(p(A_0, L_1, A_1, Y)) = \frac{p(A_0, L_1, A_1, Y)}{p(A_1 \mid A_0, L_1)}$

$A_0 \perp \perp Y \mid A_1 \quad [\phi_{A_1}(p(A_0, L_1, A_1, Y); \mathcal{G})]$
References