An Introduction to Causal Graphical Models

Spencer Gordon

Simons Institute Causality Bootcamp

Handout available at
`https://tinyurl.com/causalitybootcamp`

Table of Contents

Table of Contents

Our viewpoint

- *"Smoking causes lung cancer."*

Our viewpoint

- *"Smoking causes lung cancer."* **Not always.**

Our viewpoint

- *"Smoking causes lung cancer."* **Not always.** We use probabilities to capture uncertainty/indeterminacy.

Our viewpoint

- *"Smoking causes lung cancer."* **Not always.** We use probabilities to capture uncertainty/indeterminacy.
- We will start with probabilistic causal models.

Our viewpoint

- *"Smoking causes lung cancer."* **Not always.** We use probabilities to capture uncertainty/indeterminacy.
- We will start with probabilistic causal models.
- We will (mostly) work with causal Bayesian networks.

Probabilistic Causal Models

A tuple $M = \langle U, V, F, P(U) \rangle$ where

1. U is a set of background random variables, which can't be observed or manipulated.

Probabilistic Causal Models

A tuple $M = \langle U, V, F, P(U) \rangle$ where

1. U is a set of background random variables, which can't be observed or manipulated.
2. $V = \{X_1, \ldots, X_n\}$ is set of observable variables. Each is a function of some subset of $U \cup V$.

Probabilistic Causal Models

A tuple $M = \langle U, V, F, P(U) \rangle$ where

1. U is a set of background random variables, which can't be observed or manipulated.
2. $V = \{X_1, \ldots, X_n\}$ is set of observable variables. Each is a function of some subset of $U \cup V$.
3. F is set of functions $\{f_i\}$ such that each f_i maps from a subset of $U \cup \{X_1, \ldots, X_{i-1}\}$ to X_i.

Probabilistic Causal Models

A tuple $M = \langle U, V, F, P(U) \rangle$ where

1. U is a set of background random variables, which can't be observed or manipulated.
2. $V = \{X_1, \ldots, X_n\}$ is set of observable variables. Each is a function of some subset of $U \cup V$.
3. F is set of functions $\{f_i\}$ such that each f_i maps from a subset of $U \cup \{X_1, \ldots, X_{i-1}\}$ to X_i.
4. $P(U)$ is a joint distribution over U.

Probabilistic Causal Models

A tuple $M = \langle U, V, F, P(U) \rangle$ where

1. U is a set of background random variables, which can't be observed or manipulated.
2. $V = \{X_1, \ldots, X_n\}$ is set of observable variables. Each is a function of some subset of $U \cup V$.
3. F is set of functions $\{f_i\}$ such that each f_i maps from a subset of $U \cup \{X_1, \ldots, X_{i-1}\}$ to X_i.
4. $P(U)$ is a joint distribution over U.

Together $P(U)$ and F induce a distribution on V, $P(V)$.

Probabilistic Causal Models

A tuple $M = \langle U, V, F, P(U) \rangle$ where

1. U is a set of background random variables, which can't be observed or manipulated.
2. $V = \{X_1, \ldots, X_n\}$ is set of observable variables. Each is a function of some subset of $U \cup V$.
3. F is set of functions $\{f_i\}$ such that each f_i maps from a subset of $U \cup \{X_1, \ldots, X_{i-1}\}$ to X_i.
4. $P(U)$ is a joint distribution over U.

Together $P(U)$ and F induce a distribution on V, $P(V)$.

$$P(v) = \sum_{u \in D_U} \prod_{i=1}^{n} P(x_i \mid \mathrm{parents}(x_i)) P(u)$$

An example

- $U = \{U_1, \ldots, U_5\}$.

An example

- $U = \{U_1, \ldots, U_5\}$.
- $V = \{X_1, \ldots, X_5\}$.

An example

- $U = \{U_1, \ldots, U_5\}$.
- $V = \{X_1, \ldots, X_5\}$.
- F is given below:

An example

- $U = \{U_1, \ldots, U_5\}$.
- $V = \{X_1, \ldots, X_5\}$.
- F is given below:
- U_i are independent.

An example

- $U = \{U_1, \ldots, U_5\}$.
- $V = \{X_1, \ldots, X_5\}$.
- F is given below:
- U_i are independent.
- $U_1 \sim \mathsf{Unif}(\{\mathrm{Wi}, \mathrm{Sp}, \mathrm{Su}, \mathrm{Fa}\})$

An example

- $U = \{U_1, \ldots, U_5\}$.
- $V = \{X_1, \ldots, X_5\}$.
- F is given below:
- U_i are independent.
- $U_1 \sim \mathsf{Unif}(\{\mathrm{Wi}, \mathrm{Sp}, \mathrm{Su}, \mathrm{Fa}\})$
- $\mathsf{supp}(U_i) = \{-1, 0, 1\}$ for $i = 2, 3, 4, 5$.

An example

- $U = \{U_1, \ldots, U_5\}$.
- $V = \{X_1, \ldots, X_5\}$.
- F is given below:
- U_i are independent.
- $U_1 \sim \mathsf{Unif}(\{\mathrm{Wi}, \mathrm{Sp}, \mathrm{Su}, \mathrm{Fa}\})$
- $\mathsf{supp}(U_i) = \{-1, 0, 1\}$ for $i = 2, 3, 4, 5$.

$$
\begin{aligned}
\text{SEASON}: &\quad X_1 \coloneqq U_1 \\
\text{RAIN}: &\quad X_2 \coloneqq (X_1 \in \{\mathrm{Wi}, \mathrm{Fa}\} \vee U_2 = 1) \wedge (U_2 > -1) \\
\text{SPRINKLER}: &\quad X_3 \coloneqq (X_3 \in \{\mathrm{Su}, \mathrm{Sp}\} \vee U_3 = 1) \wedge (U_2 > -1) \\
\text{WET}: &\quad X_4 \coloneqq (X_2 \vee X_3 \vee U_4 = 1) \wedge (U_4 > -1) \\
\text{SLIPPERY}: &\quad X_5 \coloneqq (X_4 \vee U_5 = 1) \wedge (U_5 > -1)
\end{aligned}
$$

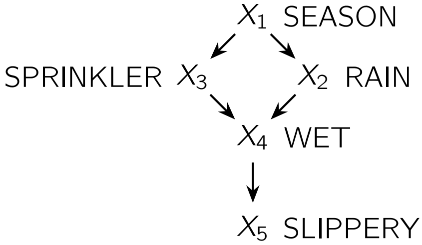
An example

- $U = \{U_1, \ldots, U_5\}$.
- $V = \{X_1, \ldots, X_5\}$.
- F is given below:
- U_i are independent.
- $U_1 \sim \mathsf{Unif}(\{\mathrm{Wi}, \mathrm{Sp}, \mathrm{Su}, \mathrm{Fa}\})$
- $\mathsf{supp}(U_i) = \{-1, 0, 1\}$ for $i = 2, 3, 4, 5$.

$$
\begin{aligned}
\text{SEASON}: &\quad X_1 \coloneqq U_1 \\
\text{RAIN}: &\quad X_2 \coloneqq (X_1 \in \{\mathrm{Wi}, \mathrm{Fa}\} \vee U_2 = 1) \wedge (U_2 > -1) \\
\text{SPRINKLER}: &\quad X_3 \coloneqq (X_3 \in \{\mathrm{Su}, \mathrm{Sp}\} \vee U_3 = 1) \wedge (U_2 > -1) \\
\text{WET}: &\quad X_4 \coloneqq (X_2 \vee X_3 \vee U_4 = 1) \wedge (U_4 > -1) \\
\text{SLIPPERY}: &\quad X_5 \coloneqq (X_4 \vee U_5 = 1) \wedge (U_5 > -1)
\end{aligned}
$$

Is there any better way to understand this?

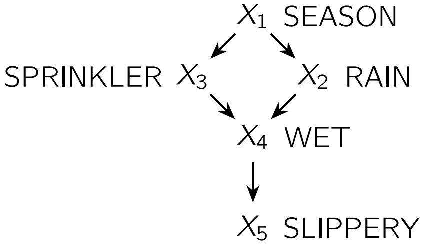
An example, continued



Each model induces a graph.

The graph has a vertex for each $X \in V$, an edge $X \to Y$ if f_Y depends on X.

An example, continued

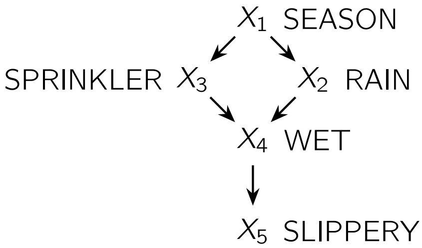


Each model induces a graph.

The graph has a vertex for each $X \in V$, an edge $X \to Y$ if f_Y depends on X.

- We will only be interested in models that induce *acyclic* graphs!

An example, continued

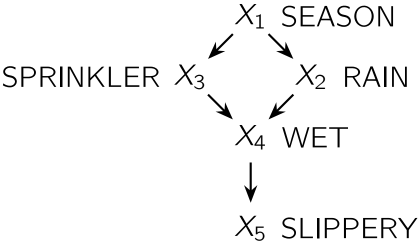


Each model induces a graph.

The graph has a vertex for each $X \in V$, an edge $X \to Y$ if f_Y depends on X.

- We will only be interested in models that induce *acyclic* graphs!
- What about confounders?

An example, continued



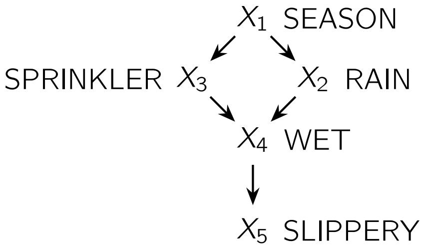
Each model induces a graph.

The graph has a vertex for each $X \in V$, an edge $X \to Y$ if f_Y depends on X.

- We will only be interested in models that induce *acyclic* graphs!
- What about confounders? If f_X, f_Y depend on a common U, we represent this with

$$X \leftarrow\dashrightarrow Y$$

Factorization



With no confounders the $P(V)$ induced by $P(U)$ factors according to G:

$$
\begin{aligned}
&P(X_1, X_2, X_3, X_4, X_5) \\
&\qquad = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)P(X_4 \mid X_2, X_3)P(X_5 \mid X_4)
\end{aligned}
$$

Interventions

Interventions correspond to changing the mechanism determining some X_i

$$
\begin{aligned}
\text{SEASON}: &\quad X_1 \coloneqq U_1 \\
\text{RAIN}: &\quad X_2 \coloneqq (X_1 \in \{\text{Wi}, \text{Fa}\} \vee U_2 = 1) \wedge (U_2 > -1) \\
\text{SPRINKLER}: &\quad X_3 \coloneqq (X_3 \in \{\text{Su}, \text{Sp}\} \vee U_3 = 1) \wedge (U_2 > -1) \\
\text{WET}: &\quad X_4 \coloneqq (X_2 \vee X_3 \vee U_4 = 1) \wedge (U_4 > -1) \\
\text{SLIPPERY}: &\quad X_5 \coloneqq (X_4 \vee U_5 = 1) \wedge (U_5 > -1)
\end{aligned}
$$

Interventions

Interventions correspond to changing the mechanism determining some X_i, e.g.,turning the sprinkler off.

$$\begin{aligned}
\text{SEASON}: &\quad X_1 \coloneqq U_1 \\
\text{RAIN}: &\quad X_2 \coloneqq (X_1 \in \{\text{Wi}, \text{Fa}\} \vee U_2 = 1) \wedge (U_2 > -1) \\
\text{SPRINKLER}: &\quad X_3 \coloneqq (X_3 \in \{\text{Su}, \text{Sp}\} \vee U_3 = 1) \wedge (U_2 > -1) \\
\text{WET}: &\quad X_4 \coloneqq (X_2 \vee X_3 \vee U_4 = 1) \wedge (U_4 > -1) \\
\text{SLIPPERY}: &\quad X_5 \coloneqq (X_4 \vee U_5 = 1) \wedge (U_5 > -1)
\end{aligned}$$

Interventions

Interventions correspond to changing the mechanism determining some X_i, e.g.,turning the sprinkler off.

$$
\begin{aligned}
\text{SEASON}: &\quad X_1 := U_1 \\
\text{RAIN}: &\quad X_2 := (X_1 \in \{\text{Wi}, \text{Fa}\} \vee U_2 = 1) \wedge (U_2 > -1) \\
\text{SPRINKLER}: &\quad X_3 \;{\color{red}:= \text{OFF}} \\
\text{WET}: &\quad X_4 := (X_2 \vee X_3 \vee U_4 = 1) \wedge (U_4 > -1) \\
\text{SLIPPERY}: &\quad X_5 := (X_4 \vee U_5 = 1) \wedge (U_5 > -1)
\end{aligned}
$$

Interventions

Interventions correspond to changing the mechanism determining some X_i, e.g.,turning the sprinkler off.

$$
\begin{aligned}
\text{SEASON}: \quad & X_1 := U_1 \\
\text{RAIN}: \quad & X_2 := (X_1 \in \{\text{Wi}, \text{Fa}\} \vee U_2 = 1) \wedge (U_2 > -1) \\
\text{SPRINKLER}: \quad & X_3 \ := \text{OFF} \\
\text{WET}: \quad & X_4 := (X_2 \vee X_3 \vee U_4 = 1) \wedge (U_4 > -1) \\
\text{SLIPPERY}: \quad & X_5 := (X_4 \vee U_5 = 1) \wedge (U_5 > -1)
\end{aligned}
$$

The induced graph and $P(V)$ change as well.

Interventions

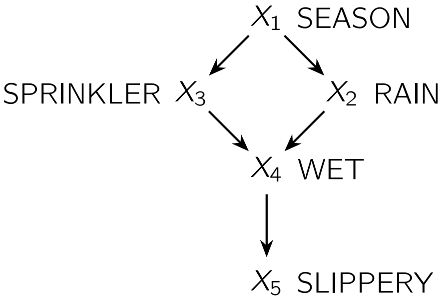
Interventions correspond to changing the mechanism determining some X_i, e.g.,turning the sprinkler off.

$$
\begin{aligned}
\text{SEASON}: \quad & X_1 := U_1 \\
\text{RAIN}: \quad & X_2 := (X_1 \in \{\text{Wi}, \text{Fa}\} \vee U_2 = 1) \wedge (U_2 > -1) \\
\text{SPRINKLER}: \quad & X_3 \ := \text{OFF} \\
\text{WET}: \quad & X_4 := (X_2 \vee X_3 \vee U_4 = 1) \wedge (U_4 > -1) \\
\text{SLIPPERY}: \quad & X_5 := (X_4 \vee U_5 = 1) \wedge (U_5 > -1)
\end{aligned}
$$

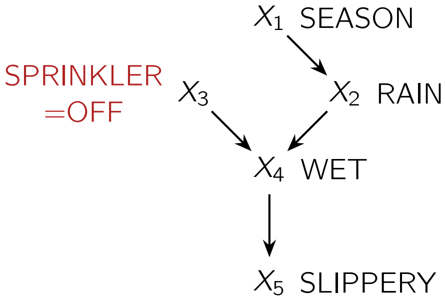
The induced graph and $P(V)$ change as well.

We write $P_x(V)$ for the distribution obtained by intervening to set $X := x$.

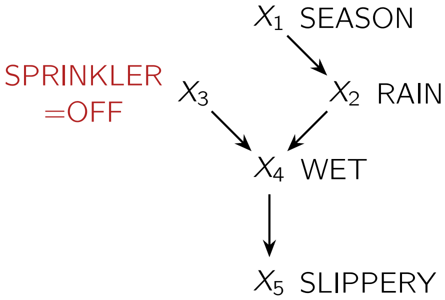
Interventions, continued



Interventions, continued



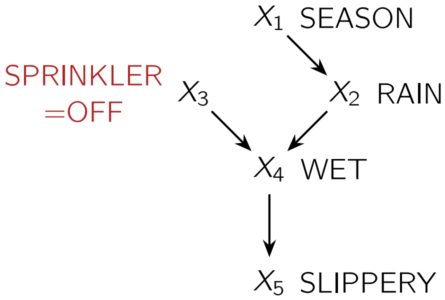
Interventions, continued



Let v be an assignment to V such that $X_3 = \text{OFF}$. Then

$$P_{X_3=\text{OFF}}(v)$$
$$= P(x_1)P(x_2 \mid x_1)P(x_4 \mid x_2, X_3 = \text{OFF})P(x_5 \mid x_4)$$

Interventions, continued



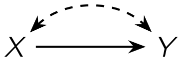
Let v be an assignment to V such that $X_3 = \text{OFF}$. Then

$$P_{X_3=\text{OFF}}(v)$$
$$= P(x_1)P(x_2 \mid x_1)P(x_4 \mid x_2, X_3 = \text{OFF})P(x_5 \mid x_4)$$

We can compute this from $P(V)$ alone. We don't need $P(U)$.

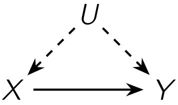
Interventions and confounders

Consider a model that induces this graph:



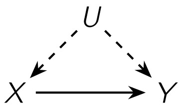
Interventions and confounders

Consider a model that induces this graph:



Interventions and confounders

Consider a model that induces this graph:

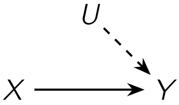


Then

$$P_x(Y) = \sum_u P(u) P(Y \mid X = x, u)$$

Interventions and confounders

Consider a model that induces this graph:



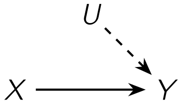
Then

$$P_x(Y) = \sum_u P(u)P(Y \mid X = x, u)$$

since the model after intervention induces the modified graph above.

Interventions and confounders

Consider a model that induces this graph:



Then

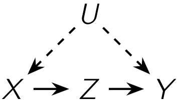
$$P_x(Y) = \sum_u P(u)P(Y \mid X = x, u)$$

since the model after intervention induces the modified graph above.

We can't compute $P_x(Y)$ with knowledge only of $P(V)$.

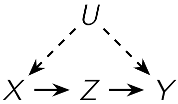
Causal inference with unobserved confounders

Consider a slightly different example:



Causal inference with unobserved confounders

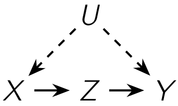
Consider a slightly different example:



Can we compute $P_x(y)$ without knowing $P(U)$?

Causal inference with unobserved confounders

Consider a slightly different example:

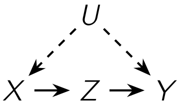


Can we compute $P_x(y)$ without knowing $P(U)$?

$$\begin{aligned} P_x(y) &= \sum_z P_x(z) P_x(y \mid z) \\ &= \sum_z P(z \mid x) P_x(y \mid z) \\ &= \sum_z P(z \mid x) \sum_{x'} P(y \mid z, x') P(x') \end{aligned}$$

Causal inference with unobserved confounders

Consider a slightly different example:



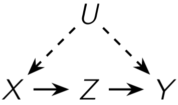
Can we compute $P_x(y)$ without knowing $P(U)$?

$$\begin{aligned} P_x(y) &= \sum_z P_x(z) P_x(y \mid z) \\ &= \sum_z P(z \mid x) P_x(y \mid z) \\ &= \sum_z P(z \mid x) \sum_{x'} P(y \mid z, x') P(x') \end{aligned}$$

Here $P(V)$ uniquely determines $P_x(y)$ in any causal model that induces G.

Causal inference with unobserved confounders

Consider a slightly different example:



Can we compute $P_x(y)$ without knowing $P(U)$?

$$
\begin{aligned}
P_x(y) &= \sum_z P_x(z) P_x(y \mid z) \\
&= \sum_z P(z \mid x) P_x(y \mid z) \\
&= \sum_z P(z \mid x) \sum_{x'} P(y \mid z, x') P(x')
\end{aligned}
$$

Here $P(V)$ uniquely determines $P_x(y)$ in any causal model that induces G. In this case we say that $P_x(y)$ is identifiable.

The big picture

The Shpitser-Pearl ID algorithm takes a graph G induced by a causal model, a distribution $P(V)$ for that model, and a target intervention $X, Y \subseteq V$, and returns

- a formula for $P_x(y)$ if it is identifiable from $P(V)$, or
- a proof that $P_x(y)$ is not identifable.

The big picture

The Shpitser-Pearl ID algorithm takes a graph G induced by a causal model, a distribution $P(V)$ for that model, and a target intervention $X, Y \subseteq V$, and returns

- a formula for $P_x(y)$ if it is identifiable from $P(V)$, or
- a proof that $P_x(y)$ is not identifable.

The agenda

- Understand the relationship between DAGs and distributions.
  - When do G_1 and G_2 correspond to the same set of possible distributions?
  - What conditional independencies are implied by a graph G?
- Understand the do-calculus, rules for manipulating interventional distributions.
- Understand the Shpitser-Pearl ID algorithm.

Table of Contents

Probability review

- X and Y are independent conditioned on Z if $\forall x \in D_X, y \in D_Y, z \in D_Z$,

  $$P(x \mid y, z) = P(x \mid z) \quad \text{if } P(y, z) > 0.$$

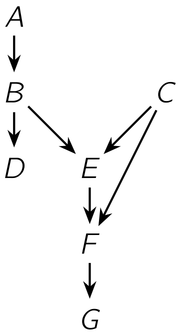
  Alternatively,

  $$P(x, y \mid z) = P(x \mid z) P(y \mid z).$$

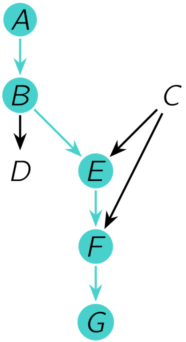
  We write:

  $$(X \perp\!\!\!\perp Y \mid Z)_P$$

Graph preliminaries



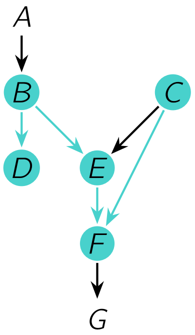
Graph preliminaries



- Directed paths

$$A \to B \to E \to F \to G \quad \text{(written } A \rightsquigarrow G\text{)}$$

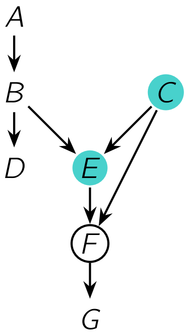
Graph preliminaries



- Directed paths
- Trails

$$D \leftarrow B \rightarrow E \rightarrow F \leftarrow C \quad (\text{written } D \;\circ\!\!\leadsto\!\!\circ\; C)$$

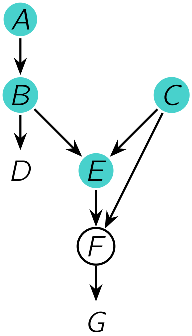
Graph preliminaries



- Directed paths
- Trails
- Parents, $\mathrm{Pa}(X)$.

$$\mathrm{Pa}(F) = \{C, E\}$$

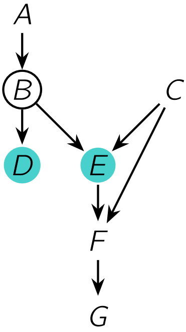
Graph preliminaries



- Directed paths
- Trails
- Parents, $\mathsf{Pa}(X)$.
- Ancestors, $\mathsf{An}(X)$.

$$\mathsf{An}(F) = \{A, B, C, E, F\}$$

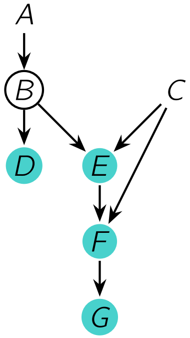
Graph preliminaries



- Directed paths
- Trails
- Parents, $\mathsf{Pa}(X)$.
- Ancestors, $\mathsf{An}(X)$.
- Children, $\mathsf{Ch}(X)$.

$$\mathsf{Ch}(B) = \{D, E\}$$

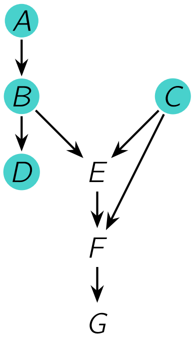
Graph preliminaries



- Directed paths
- Trails
- Parents, $\mathsf{Pa}(X)$.
- Ancestors, $\mathsf{An}(X)$.
- Children, $\mathsf{Ch}(X)$.
- Descendants, $\mathsf{De}(X)$.

$$\mathsf{De}(B) = \{B, D, E, F, G\}$$

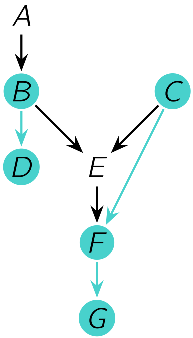
Graph preliminaries



- Directed paths
- Trails
- Parents, $\mathsf{Pa}(X)$.
- Ancestors, $\mathsf{An}(X)$.
- Children, $\mathsf{Ch}(X)$.
- Descendants, $\mathsf{De}(X)$.
- Upwards-closed set

$$\{A, B, C, D\}$$

Graph preliminaries



- Directed paths
- Trails
- Parents, $\mathsf{Pa}(X)$.
- Ancestors, $\mathsf{An}(X)$.
- Children, $\mathsf{Ch}(X)$.
- Descendants, $\mathsf{De}(X)$.
- Upwards-closed set
- Induced subgraph, $G[V']$

$$G[\{B, C, D, F, G\}]$$

Bayesian networks

A DAG $G = (V, E)$ along with a distribution $P(V)$ factoring as

$$P(V) = \prod_{X \in V} P(X \mid \mathsf{pa}(X)).$$

Bayesian networks

A DAG $G = (V, E)$ along with a distribution $P(V)$ factoring as

$$P(V) = \prod_{X \in V} P(X \mid \mathrm{pa}(X)).$$

We say that P is compatible with, or Markov relative to G.

Bayesian networks

A DAG $G = (V, E)$ along with a distribution $P(V)$ factoring as

$$P(V) = \prod_{X \in V} P(X \mid \mathsf{pa}(X)).$$

We say that P is compatible with, or Markov relative to G.

We write $\mathcal{P}(G)$ for all distributions compatible with G.

Bayesian networks

A DAG $G = (V, E)$ along with a distribution $P(V)$ factoring as

$$P(V) = \prod_{X \in V} P(X \mid \mathsf{pa}(X)).$$

We say that P is compatible with, or Markov relative to G.

We write $\mathcal{P}(G)$ for all distributions compatible with G.

Observation

If S is upwards-closed and P is compatible with G,

Bayesian networks

A DAG $G = (V, E)$ along with a distribution $P(V)$ factoring as

$$P(V) = \prod_{X \in V} P(X \mid \mathrm{pa}(X)).$$

We say that P is compatible with, or Markov relative to G.

We write $\mathcal{P}(G)$ for all distributions compatible with G.

Observation

If S is upwards-closed and P is compatible with G,

1. $P(S) = \sum_{v \setminus s} \prod_{X \in V} P(X \mid \mathrm{Pa}(X))$

Bayesian networks

A DAG $G = (V, E)$ along with a distribution $P(V)$ factoring as

$$P(V) = \prod_{X \in V} P(X \mid \mathsf{pa}(X)).$$

We say that P is compatible with, or Markov relative to G.

We write $\mathcal{P}(G)$ for all distributions compatible with G.

Observation

If S is upwards-closed and P is compatible with G,

1. $P(S) = \prod_{X \in S} P(X \mid \mathsf{Pa}(X))$ is compatible with $G[S]$.

Bayesian networks

A DAG $G = (V, E)$ along with a distribution $P(V)$ factoring as

$$P(V) = \prod_{X \in V} P(X \mid \mathrm{pa}(X)).$$

We say that P is compatible with, or Markov relative to G.

We write $\mathcal{P}(G)$ for all distributions compatible with G.

Observation

If S is upwards-closed and P is compatible with G,

1. $P(S) = \prod_{X \in S} P(X \mid \mathrm{Pa}(X))$ is compatible with $G[S]$.
2. $P(V \setminus S \mid S)$ is compatible with $G[V \setminus S]$.

Markov conditions

Ordered Markov Condition

P is compatible with $G \Leftrightarrow$ in any topological ordering $X_1, \ldots, X_n$, each X_i is independent of its predecessors given its parents.

Markov conditions

Ordered Markov Condition

P is compatible with $G \Leftrightarrow$ in any topological ordering $X_1, \ldots, X_n$, each X_i is independent of its predecessors given its parents.

Proof.

On board... □

Markov conditions

Ordered Markov Condition

P is compatible with $G \Leftrightarrow$ in any topological ordering $X_1, \ldots, X_n$, each X_i is independent of its predecessors given its parents.

Proof.

On board... □

Parental Markov Condition

P is compatible with $G \Leftrightarrow$ each X is independent of its nondescendants given its parents.

Markov conditions

Ordered Markov Condition

P is compatible with $G \Leftrightarrow$ in any topological ordering $X_1, \ldots, X_n$, each X_i is independent of its predecessors given its parents.

Proof.

On board. . . □

Parental Markov Condition

P is compatible with $G \Leftrightarrow$ each X is independent of its nondescendants given its parents.

Proof.

On board. . . □

Conditioning on common ancestors

Lemma

Fix any G and disjoint $X, Y, Z \subseteq V$. If $\mathsf{An}(X) \cap \mathsf{An}(Y) \subseteq Z$ and $\mathsf{An}(Z) \subseteq Z$, then

$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

in any distribution P compatible with G.

Conditioning on common ancestors

Lemma

Fix any G and disjoint $X, Y, Z \subseteq V$. If $\mathrm{An}(X) \cap \mathrm{An}(Y) \subseteq Z$ and $\mathrm{An}(Z) \subseteq Z$, then

$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$$

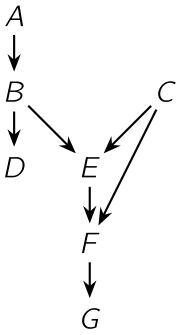
in any distribution P compatible with G.

Proof.

On board... □

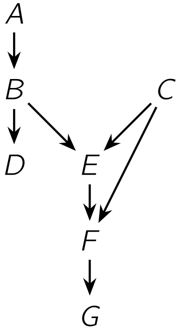
Conditional Independencies

What conditional independencies hold in any P compatible with G?



Conditional Independencies

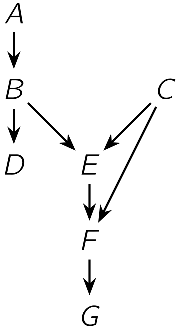
What conditional independencies hold in any P compatible with G?



- $(A \perp\!\!\!\perp F \mid C, E)_P$.

Conditional Independencies

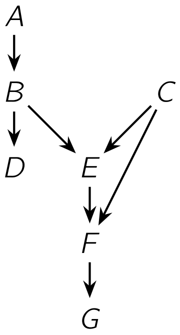
What conditional independencies hold in any P compatible with G?



- $(A \perp\!\!\!\perp F \mid C, E)_P$.
- $(B \perp\!\!\!\perp G \mid F)_P$.

Conditional Independencies

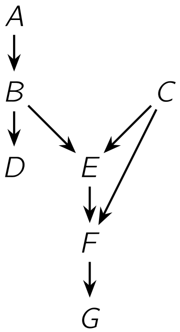
What conditional independencies hold in any P compatible with G?



- $(A \perp\!\!\!\perp F \mid C, E)_P$.
- $(B \perp\!\!\!\perp G \mid F)_P$.
- $(B \perp\!\!\!\perp F \mid E)_P$?

Conditional Independencies

What conditional independencies hold in any P compatible with G?

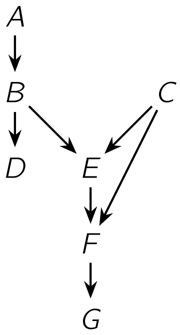


- $(A \perp\!\!\!\perp F \mid C, E)_P$.
- $(B \perp\!\!\!\perp G \mid F)_P$.
- $(B \perp\!\!\!\perp F \mid E)_P$?

Let $\mathcal{I}_{\text{prob}}(P) := \{(X, Y, Z) : (X \perp\!\!\!\perp Y \mid Z)_P\}$.

Blocked trails

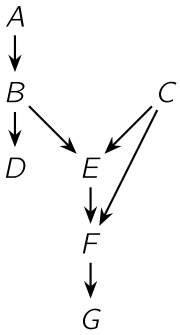
A trail in G is blocked by a set Z if it contains three consecutive vertices such that



Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

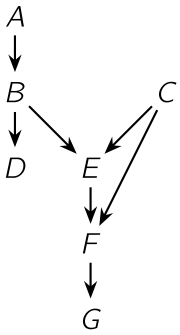
- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or



Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

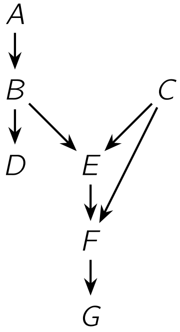


- Does $\{B\}$ block $D \leftarrow B \to E$?

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

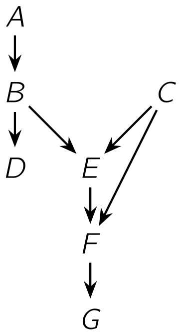


- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

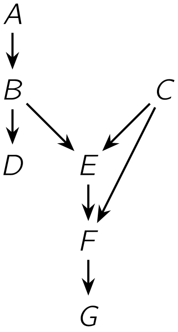


- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**
- Does $\{E\}$ block $B \to E \to F$?

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

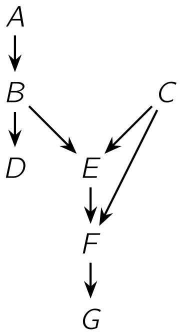


- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**
- Does $\{E\}$ block $B \to E \to F$? **Yes!**

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

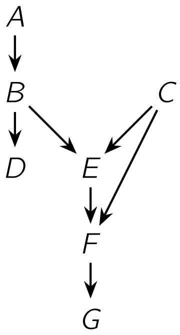


- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**
- Does $\{E\}$ block $B \to E \to F$? **Yes!**
- Does $\varnothing$ block $B \to E \leftarrow C$.

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

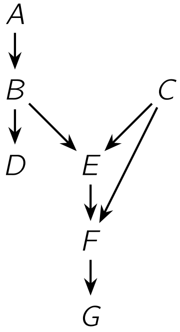


- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**
- Does $\{E\}$ block $B \to E \to F$? **Yes!**
- Does $\varnothing$ block $B \to E \leftarrow C$. **Yes!**

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \rightarrow B \rightarrow C$ is a chain or $A \leftarrow B \rightarrow C$ is a fork and $B \in Z$, or
- $A \rightarrow B \leftarrow C$ is a collider and no descendant of B is in Z.

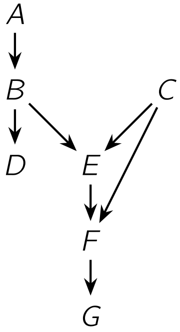


- Does $\{B\}$ block $D \leftarrow B \rightarrow E$? **Yes!**
- Does $\{E\}$ block $B \rightarrow E \rightarrow F$? **Yes!**
- Does $\varnothing$ block $B \rightarrow E \leftarrow C$. **Yes!**
- Does $\{E\}$ block $B \rightarrow E \rightarrow C \rightarrow F$?

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

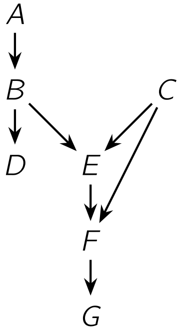


- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**
- Does $\{E\}$ block $B \to E \to F$? **Yes!**
- Does $\varnothing$ block $B \to E \leftarrow C$. **Yes!**
- Does $\{E\}$ block $B \to E \to C \to F$? **No!**

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.

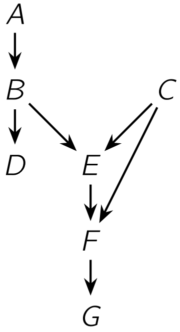


- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**
- Does $\{E\}$ block $B \to E \to F$? **Yes!**
- Does $\varnothing$ block $B \to E \leftarrow C$. **Yes!**
- Does $\{E\}$ block $B \to E \to C \to F$? **No!**
- Does $\{G\}$ block $B \to E \leftarrow C$?

Blocked trails

A trail in G is blocked by a set Z if it contains three consecutive vertices such that

- $A \to B \to C$ is a chain or $A \leftarrow B \to C$ is a fork and $B \in Z$, or
- $A \to B \leftarrow C$ is a collider and no descendant of B is in Z.



- Does $\{B\}$ block $D \leftarrow B \to E$? **Yes!**
- Does $\{E\}$ block $B \to E \to F$? **Yes!**
- Does $\varnothing$ block $B \to E \leftarrow C$. **Yes!**
- Does $\{E\}$ block $B \to E \to C \to F$? **No!**
- Does $\{G\}$ block $B \to E \leftarrow C$? **No!**

d-Separation

Let $X, Y, Z \subseteq V$ be disjoint. Then X is d-separated from Y by Z if every trail between any vertex in X and any vertex Y in G is blocked. We write

$$(X \perp\!\!\!\perp Y \mid Z)_G.$$

d-Separation

Let $X, Y, Z \subseteq V$ be disjoint. Then X is d-separated from Y by Z if every trail between any vertex in X and any vertex Y in G is blocked. We write

$$(X \perp\!\!\!\perp Y \mid Z)_G.$$

If there is a trail from a vertex in X to a vertex in Y that is not blocked, we say that X and Y are d-connected given Z.

d-Separation

Let $X, Y, Z \subseteq V$ be disjoint. Then X is d-separated from Y by Z if every trail between any vertex in X and any vertex Y in G is blocked. We write

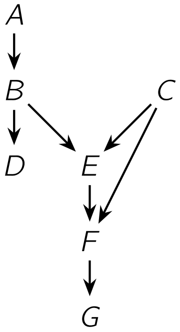
$$(X \perp\!\!\!\perp Y \mid Z)_G.$$

If there is a trail from a vertex in X to a vertex in Y that is not blocked, we say that X and Y are d-connected given Z.

Any trail that is not blocked is an active trail.

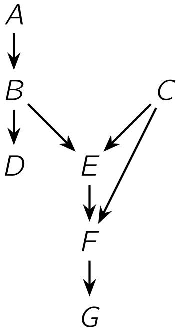
d-Separation examples

What d-separations hold in G?



d-Separation examples

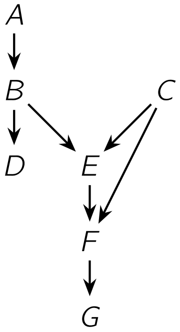
What d-separations hold in G?



- $(A \perp\!\!\!\perp F \mid C, E)_G$.

d-Separation examples

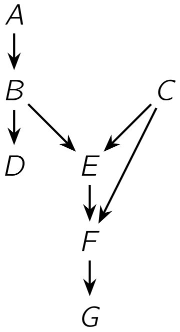
What d-separations hold in G?



- $(A \perp\!\!\!\perp F \mid C, E)_G$.
- $(B \perp\!\!\!\perp G \mid F)_G$.

d-Separation examples

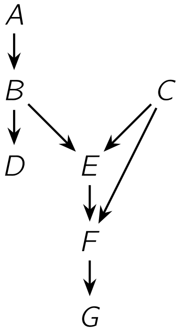
What d-separations hold in G?



- $(A \perp\!\!\!\perp F \mid C, E)_G$.
- $(B \perp\!\!\!\perp G \mid F)_G$.
- $(B \perp\!\!\!\perp F \mid E)_G$?

d-Separation examples

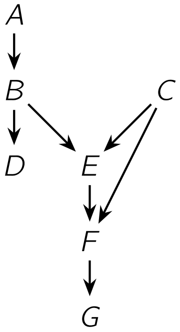
What d-separations hold in G?



- $(A \perp\!\!\!\perp F \mid C, E)_G$.
- $(B \perp\!\!\!\perp G \mid F)_G$.
- $(B \perp\!\!\!\perp F \mid E)_G$? **No!**

d-Separation examples

What d-separations hold in G?



- $(A \perp\!\!\!\perp F \mid C, E)_G$.
- $(B \perp\!\!\!\perp G \mid F)_G$.
- $(B \perp\!\!\!\perp F \mid E)_G$? **No!**

Let $\mathcal{I}_{\textsf{d-sep}}(G) \coloneqq \{(X, Y, Z) : (X \perp\!\!\!\perp Y \mid Z)_G\}$.

d-Separation and conditional independence

Theorem

$(X \perp\!\!\!\perp Y \mid Z)_G \implies (X \perp\!\!\!\perp Y \mid Z)_P$ in every distribution P compatible with G.

d-Separation and conditional independence

Theorem

$(X \perp\!\!\!\perp Y \mid Z)_G \implies (X \perp\!\!\!\perp Y \mid Z)_P$ *in every distribution P compatible with G.*

We'll prove this later.

d-Separation and conditional independence

Theorem

$(X \perp\!\!\!\perp Y \mid Z)_G \implies (X \perp\!\!\!\perp Y \mid Z)_P$ in every distribution P compatible with G.

We'll prove this later.

Theorem

If $\neg(X \perp\!\!\!\perp Y \mid Z)_G$, then there exists a distribution P compatible with G in which $\neg(X \perp\!\!\!\perp Y \mid Z)_P$.

d-Separation and conditional independence

Theorem

$(X \perp\!\!\!\perp Y \mid Z)_G \implies (X \perp\!\!\!\perp Y \mid Z)_P$ *in every distribution* P *compatible with* G*.*

We'll prove this later.

Theorem

If $\neg(X \perp\!\!\!\perp Y \mid Z)_G$*, then there exists a distribution* P *compatible with* G *in which* $\neg(X \perp\!\!\!\perp Y \mid Z)_P$*.*

We'll prove this later.

Table of Contents

Markov equivalence

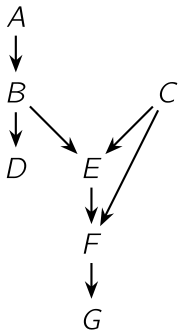
Which graphs have the same d-separations?

Markov equivalence

Which graphs have the same d-separations?

Theorem (Markov equivalence)

G_1 and G_2 have the same d-separations if and only if they have the same skeleton and the same immoralities.

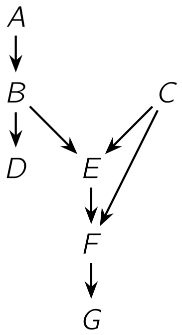


Markov equivalence

Which graphs have the same d-separations?

Theorem (Markov equivalence)

G_1 and G_2 have the same d-separations if and only if they have the same skeleton and the same immoralities.

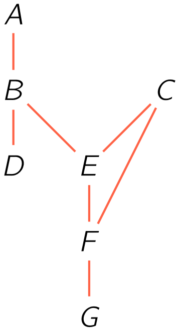


Markov equivalence

Which graphs have the same d-separations?

Theorem (Markov equivalence)

G_1 and G_2 have the same d-separations if and only if they have the same skeleton and the same immoralities.



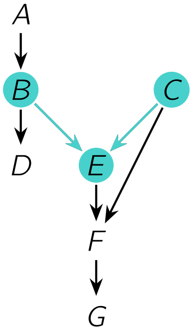
- The skeleton of a graph G is an undirected graph with the same adjacencies as in G.

Markov equivalence

Which graphs have the same d-separations?

Theorem (Markov equivalence)

G_1 and G_2 have the same d-separations if and only if they have the same skeleton and the same immoralities.



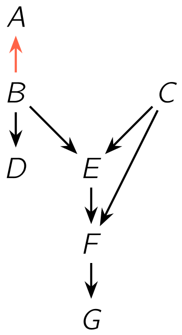
- The skeleton of a graph G is an undirected graph with the same adjacencies as in G.
- An immorality is a collider $X \to Y \leftarrow Z$ in which X and Z are not adjacent.

Markov equivalence

Which graphs have the same d-separations?

Theorem (Markov equivalence)

G_1 and G_2 have the same d-separations if and only if they have the same skeleton and the same immoralities.



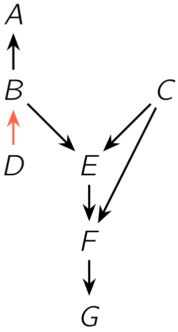
- The skeleton of a graph G is an undirected graph with the same adjacencies as in G.
- An immorality is a collider $X \to Y \leftarrow Z$ in which X and Z are not adjacent.
- Flipping these edges doesn't change d-separations.

Markov equivalence

Which graphs have the same d-separations?

Theorem (Markov equivalence)

G_1 and G_2 have the same d-separations if and only if they have the same skeleton and the same immoralities.



- The skeleton of a graph G is an undirected graph with the same adjacencies as in G.
- An immorality is a collider $X \to Y \leftarrow Z$ in which X and Z are not adjacent.
- Flipping these edges doesn't change d-separations.

Proving Markov equivalence

We need a preliminary lemma

Lemma

If X_i and X_j are not adjacent in G, then $(X_i \perp\!\!\!\perp X_j \mid \mathrm{Pa}_i, \mathrm{Pa}_j)_G$.

Proving Markov equivalence

We need a preliminary lemma

Lemma

If X_i and X_j are not adjacent in G, then $(X_i \perp\!\!\!\perp X_j \mid \mathrm{Pa}_i, \mathrm{Pa}_j)_G$.

Proof.

On board... □

Proving Markov equivalence, continued

Lemma

$\mathcal{I}_{\textsf{d-sep}}(G_1) = \mathcal{I}_{\textsf{d-sep}}(G_2) \implies G_1$ and G_2 have the same skeleton and immoralities.

Proving Markov equivalence, continued

Lemma

$\mathcal{I}_{\textsf{d-sep}}(G_1) = \mathcal{I}_{\textsf{d-sep}}(G_2) \implies G_1$ and G_2 have the same skeleton and immoralities.

Proof.

On board... □

Proving Markov equivalence, continued

Lemma

$\mathcal{I}_{\text{d-sep}}(G_1) = \mathcal{I}_{\text{d-sep}}(G_2) \implies G_1$ and G_2 have the same skeleton and immoralities.

Proof.

On board... □

Proving Markov equivalence, continued

Lemma

$\mathcal{I}_{\textsf{d-sep}}(G_1) = \mathcal{I}_{\textsf{d-sep}}(G_2) \implies G_1$ and G_2 have the same skeleton and immoralities.

Proof.

On board... □

Proving Markov equivalence, continued

Lemma

$\mathcal{I}_{\textsf{d-sep}}(G_1) = \mathcal{I}_{\textsf{d-sep}}(G_2) \implies G_1$ and G_2 have the same skeleton and immoralities.

Proof.

On board... □

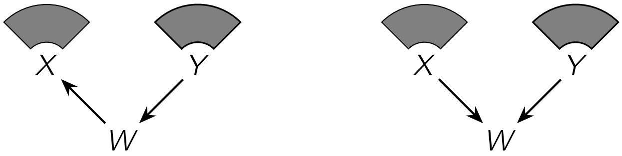
Proving Markov equivalence, continued

Lemma

$\mathcal{I}_{\text{d-sep}}(G_1) = \mathcal{I}_{\text{d-sep}}(G_2) \implies G_1$ and G_2 have the same skeleton and immoralities.

Proof.

On board... □



Tight active trails

An active trail is tight if...

Proposition

If X and Y are d-connected by Z, there is a tight active trail witnessing the connection.

Tight active trails, continued

Lemma

Let $T = (X = X_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ X_k = Y)$ be a tight active trail with observation set Z. Then for $i = 2, \ldots, k-1$, if X_{i-1} is adjacent to X_{i+1}, then $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$ and at least one of X_{i-1} or X_{i+1} is a collider in T.

Corollary

If X_i is a collider in T, then $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ is an immorality in G.

Proving Markov equivalence, continued

Lemma

If G_1 and G_2 with common vertex set V have the same skeleton and immoralities then $\mathcal{I}_{\text{d-sep}}(G_1) = \mathcal{I}_{\text{d-sep}}(G_2)$.

Proof.

On board... □

Table of Contents

d-Separation and conditional independence

Theorem

- *Completeness* *If* $\neg(X \perp\!\!\!\perp Y \mid Z)_G$ *then there exists a distribution* P *compatible with* G *such that* $\neg(X \perp\!\!\!\perp Y \mid Z)_P$.
- *Soundness* *If* $(X \perp\!\!\!\perp Y \mid Z)_G$ *then* $(X \perp\!\!\!\perp Y \mid Z)_P$ *in any distribution* P *compatible with* G.

Proof.

On board... □

Completeness of d-separation

Lemma

If $\neg(X \perp\!\!\!\perp Y \mid Z)_G$ then there exists a distribution P compatible with G such that $\neg(X \perp\!\!\!\perp Y \mid Z)_P$.

Proof.

Let $T = (X = V_1 \multimap\!\!\circ \cdots \multimap\!\!\circ V_k = Y)$ be an active path given Z.

Continued on board... □

Soundness of d-separation

Lemma

If $(X \perp\!\!\!\perp Y \mid Z)_G$ then $(X \perp\!\!\!\perp Y \mid Z)_P$ in any distribution P compatible with G.

Soundness of d-separation

Lemma

If $(X \perp\!\!\!\perp Y \mid Z)_G$ then $(X \perp\!\!\!\perp Y \mid Z)_P$ in any distribution P compatible with G.

Proof.

Let $(X \perp\!\!\!\perp Y \mid Z)_G$.

- Let $Z_1, \ldots, Z_k$ be a topological order of Z.
- Define $Z(j) \coloneqq \left\{Z_1, \ldots, Z_j\right\}$.

Continued... □

Soundness of d-separation

Lemma

If $(X \perp\!\!\!\perp Y \mid Z)_G$ then $(X \perp\!\!\!\perp Y \mid Z)_P$ in any distribution P compatible with G.

Proof.

Let $(X \perp\!\!\!\perp Y \mid Z)_G$.

- Let $Z_1, \ldots, Z_k$ be a topological order of Z.
- Define $Z(j) \coloneqq \{Z_1, \ldots, Z_j\}$.

Continued... □

We complete a DAG G by picking a topological order and adding all edges consistent with the order.

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 \coloneqq G$ with the subgraph $G[Z]$ completed.

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 \coloneqq G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \dots, G_k$.

- $G_0 := G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
  1. completing $G_{j-1}[A_j]$ where $A_j := \mathrm{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 := G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
    1. completing $G_{j-1}[A_j]$ where $A_j := \mathrm{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;
    2. reversing the edges in A_j;

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 := G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
  1. completing $G_{j-1}[A_j]$ where $A_j := \mathrm{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;
  2. reversing the edges in A_j; then
  3. completing $G_j[A_j \cup Z(j-1)]$.

The modification procedure

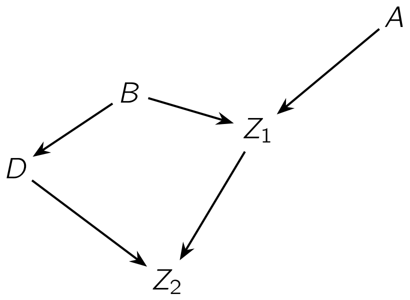
We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 \coloneqq G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
  1. completing $G_{j-1}[A_j]$ where $A_j \coloneqq \mathsf{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;
  2. reversing the edges in A_j; then
  3. completing $G_j[A_j \cup Z(j-1)]$.

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 := G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
  1. completing $G_{j-1}[A_j]$ where $A_j := \mathrm{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;
  2. reversing the edges in A_j; then
  3. completing $G_j[A_j \cup Z(j-1)]$.

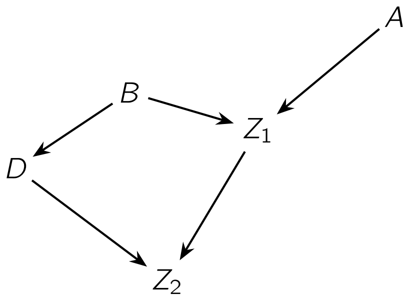


G

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \dots, G_k$.

- $G_0 \coloneqq G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
  1. completing $G_{j-1}[A_j]$ where $A_j \coloneqq \mathrm{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;
  2. reversing the edges in A_j; then
  3. completing $G_j[A_j \cup Z(j-1)]$.

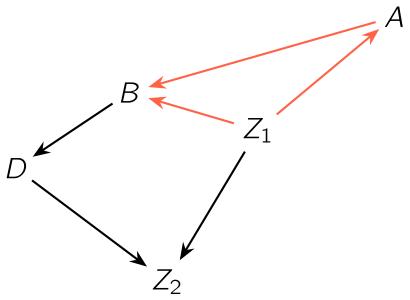


G_0

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 := G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
  1. completing $G_{j-1}[A_j]$ where $A_j := \mathrm{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;
  2. reversing the edges in A_j; then
  3. completing $G_j[A_j \cup Z(j-1)]$.

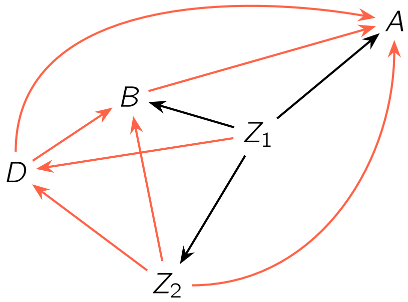


G_1

The modification procedure

We'll define a sequence of graphs: $G_0, G_1, \ldots, G_k$.

- $G_0 := G$ with the subgraph $G[Z]$ completed.
- G_j is obtained from G_{j-1} by
  1. completing $G_{j-1}[A_j]$ where $A_j := \mathrm{An}_{G_{j-1}}[Z_j] \setminus Z(j-1)$;
  2. reversing the edges in A_j; then
  3. completing $G_j[A_j \cup Z(j-1)]$.



G_2

Soundness of d-separation

Proposition

In G_j:

1. $Z(j)$ is upwards-closed.

Soundness of d-separation

Proposition

In G_j:

1. $Z(j)$ is upwards-closed.
2. $A_j \cup Z(j-1)$ is upwards-closed.

Soundness of d-separation

Proposition

In G_j:

1. $Z(j)$ is upwards-closed.
2. $A_j \cup Z(j-1)$ is upwards-closed.
3. G_j is acyclic.

Soundness of d-separation

Proposition

In G_j:

1. $Z(j)$ is upwards-closed.
2. $A_j \cup Z(j-1)$ is upwards-closed.
3. G_j is acyclic.
4. $G_j[A_j \cup Z(j-1)]$ is complete.

Soundness of d-separation

Proposition

In G_j:

1. $Z(j)$ is upwards-closed.
2. $A_j \cup Z(j-1)$ is upwards-closed.
3. G_j is acyclic.
4. $G_j[A_j \cup Z(j-1)]$ is complete.
5. $(X \perp\!\!\!\perp Y \mid Z)_{G_j} \iff (X \perp\!\!\!\perp Y \mid Z)_G$.

Soundness of d-separation

Proposition

In G_j:

1. $Z(j)$ is upwards-closed.
2. $A_j \cup Z(j-1)$ is upwards-closed.
3. G_j is acyclic.
4. $G_j[A_j \cup Z(j-1)]$ is complete.
5. $(X \perp\!\!\!\perp Y \mid Z)_{G_j} \iff (X \perp\!\!\!\perp Y \mid Z)_G$.
6. P is compatible with G_j.

Soundness of d-separation

Proposition

In G_j:

1. $Z(j)$ is upwards-closed.
2. $A_j \cup Z(j-1)$ is upwards-closed.
3. G_j is acyclic.
4. $G_j[A_j \cup Z(j-1)]$ is complete.
5. $(X \perp\!\!\!\perp Y \mid Z)_{G_j} \iff (X \perp\!\!\!\perp Y \mid Z)_G$.
6. P is compatible with G_j.

Now we can finish the proof!

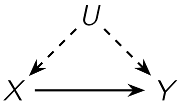
Table of Contents

Back to causal models

Recall: We model interventions in a causal model by swapping the mechanism used to set X with a constant function of our choice.

Back to causal models

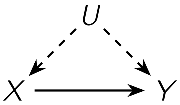
Recall: We model interventions in a causal model by swapping the mechanism used to set X with a constant function of our choice.



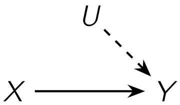
$$P(x, y) = \sum_u P(x \mid u) P(y \mid x, u) P(u)$$

Back to causal models

Recall: We model interventions in a causal model by swapping the mechanism used to set X with a constant function of our choice.



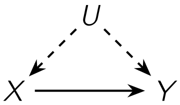
$$P(x, y) = \sum_u P(x \mid u) P(y \mid x, u) P(u)$$



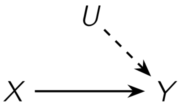
$$P_x(y) = \sum_u P(y \mid x, u) P(u)$$

Back to causal models

Recall: We model interventions in a causal model by swapping the mechanism used to set X with a constant function of our choice.



$$P(x, y) = \sum_u P(x \mid u)P(y \mid x, u)P(u)$$



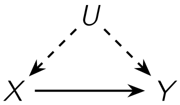
$$P_x(y) = \sum_u P(y \mid x, u)P(u)$$

We write $\mathsf{do}(x)$ for the intervention $X := x$ and define

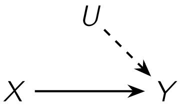
$$P(Y \mid \mathsf{do}(x)) := P_x(Y).$$

Back to causal models

Recall: We model interventions in a causal model by swapping the mechanism used to set X with a constant function of our choice.



$$P(x, y) = \sum_u P(x \mid u) P(y \mid x, u) P(u)$$



$$P_x(y) = \sum_u P(y \mid x, u) P(u)$$

We write $\mathsf{do}(x)$ for the intervention $X \coloneqq x$ and define

$$P(Y \mid \mathsf{do}(x)) \coloneqq P_x(Y).$$

The graph induced by $\mathsf{do}(x)$ is $G_{\overline{X}}$, obtained by removing all edges from $\mathsf{Pa}(X)$ to X.

The do-calculus

Rules for manipulating interventional distributions.

The do-calculus

Rules for manipulating interventional distributions.

P is compatible with $G \implies P_x$ is compatible with $G_{\bar{x}}$.

The do-calculus

Rules for manipulating interventional distributions.

P is compatible with $G \implies P_x$ is compatible with $G_{\bar{x}}$.

We can use d-separation to reason about interventional distributions!

Rule 1: Insertion/deletion of observations

Theorem (Insertion/deletion of observations)

$$P(y \mid \mathrm{do}(x), z, w) = P(y \mid \mathrm{do}(x), w)$$

if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$.

Rule 1: Insertion/deletion of observations

Theorem (Insertion/deletion of observations)

$$P(y \mid \mathrm{do}(x), z, w) = P(y \mid \mathrm{do}(x), w)$$

if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$.

Proof.

$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}} \implies (Y \perp\!\!\!\perp Z \mid X, W)_{P_x}$ since P_x is compatible with $G_{\overline{X}}$. □

Rule 2: Action/observation exchange

Theorem (Action/observation exchange)

Let $X, Y, Z, W \subseteq V$ be disjoint. Then

$$P(y \mid \mathrm{do}(x), \mathrm{do}(z), w) = P(y \mid \mathrm{do}(x), z, w)$$

if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$.

Lemma

Let $H = G_{\overline{X}\underline{Z}}$. Then

$$(Y \perp\!\!\!\perp Z \mid X, W)_H \iff (\hat{Z} \perp\!\!\!\perp Y \mid X, Z, W)_{\mathbf{Aug}(H,Z)}.$$

Rule 3: Insertion/deletion of actions

Theorem (Insertion/deletion of actions)

$$P(y \mid \mathrm{do}(x), \mathrm{do}(z), W) = P(y \mid \mathrm{do}(x), w)$$

if $(Y \perp\!\!\!\perp Z | X, W)_{G_{\overline{XZ(W)}}}$, *where* $Z(W) := Z \setminus \mathrm{An}_{G_{\overline{X}}}(W)$.

Lemma

Any trail in $\mathrm{Aug}(G_{\overline{X}}, Z)$ that is active given X, W and uses only edges present in $G_{\overline{XZ(W)}}$ is also active in $G_{\overline{XZ(W)}}$ given X, W, where $Z(W) = Z \setminus \mathrm{An}_{G_{\overline{X}}}(W)$.

Table of Contents

Identifiability

Which causal effects can be determined from the observed variables only?

Identifiability

Which causal effects can be determined from the observed variables only?

Definition (Identifiability)

The causal effect of an intervention $\mathsf{do}(x)$ on a set of variables $Y \subseteq V$ (for $Y \subseteq V \setminus X$) is *identifiable* from P in a DAG G if $P_x(y)$ is uniquely computable from $P(V)$ in any causal model that induces G.

The ID algorithm theorem

Theorem (Shpitser-Pearl)

The algorithm **ID** *will return an expression for* $P_x(Y)$ *whenever it is identifiable from a graph* G*, and will return a witness to nonidentifiability whenever* $P_x(Y)$ *is not identifiable.*

The ID algorithm theorem

Theorem (Shpitser-Pearl)

The algorithm **ID** *will return an expression for* $P_x(Y)$ *whenever it is identifiable from a graph* G*, and will return a witness to nonidentifiability whenever* $P_x(Y)$ *is not identifiable.*

Every line of the algorithm is an application of a rule of the do-calculus!

The ID algorithm

function $\mathbf{ID}(\mathbf{y}, \mathbf{x}, P, G)$

1: if $\mathbf{x} = \varnothing$, return $\sum_{\mathrm{v} \setminus \mathrm{y}} P(\mathbf{v})$.

2: if $\mathbf{V} \neq \mathrm{An}(\mathbf{Y})_G$,
return $\mathbf{ID}(\mathbf{y}, \mathbf{x} \cap \mathrm{An}(\mathbf{Y})_G, P(\mathrm{An}(\mathbf{Y})), \mathrm{An}(\mathbf{Y})_G)$.

3: let $\mathbf{W} = (\mathbf{V} \setminus \mathbf{X}) \setminus \mathrm{An}(\mathbf{Y})_{G_{\overline{\mathrm{X}}}}$.
if $\mathbf{W} \neq \varnothing$, return $\mathbf{ID}(\mathbf{y}, \mathbf{x} \cup \mathbf{w}, P, G)$.

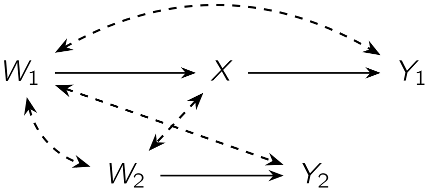
4: if $C(G \setminus \mathbf{X}) = \{S_1, \ldots, S_k\}$ (for $k \geq 2$),
return $\sum_{v \setminus (\mathrm{y} \cup \mathrm{x})} \prod_i \mathbf{ID}(s_i, \mathbf{v} \setminus s_i, P, G)$.
else if $C(G \setminus \mathbf{X}) = \{S\}$,

5: if $C(G) = \{G\}$, throw $\mathbf{FAIL}(G, S)$.

6: if $S \in C(G)$, return $\sum_{s \setminus \mathrm{y}} \prod_{V_i \in S} P(v_i \mid v_\pi^{(i-1)})$.

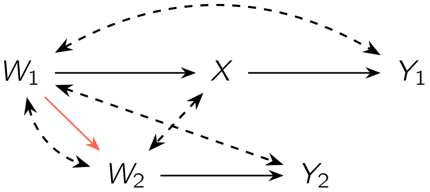
7: if $\exists S', S \subseteq S' \in C(G)$,
return
$\mathbf{ID}(\mathbf{y}, \mathbf{x} \cap S', \prod_{V_i \in S'} P(V_i \mid V_\pi^{(i-1)} \cap S', v_\pi^{(i-1)} \setminus S', S')$.

Two examples



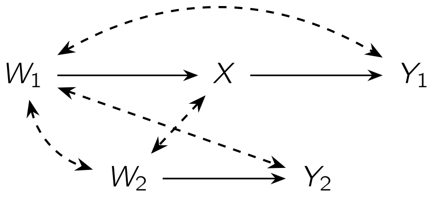
Is $P_x(y_1, y_2)$ identifiable?

Two examples



Is $P_x(y_1, y_2)$ identifiable? How about now?

A positive example



$$P_x(y_1, y_2) = \sum_{w_2} \left(\sum_{w_1} P(y_1|w_1, x)P(w_1) \right) P(y_2|w_2)P(w_2).$$

Hedges

Definition (C-component)

Let G be a semi-Markovian graph such that a subset of its bidirected edges form a spanning tree of V. Then G is a *C-component (confounded component)*.

Definition (Decomposition into C-components)

Any graph can be uniquely partitioned into a collection of subgraphs $C(G)$, each of which is a maximal C-component. (If G is itself a C-component, the partition is trivial.)

Definition (C-forest)

Let Y be the set of all sinks in a semi-Markovian graph G. Then G is a Y-rooted C-forest if G is a C-component and all vertices have at most one child.

Hedges and identifiability

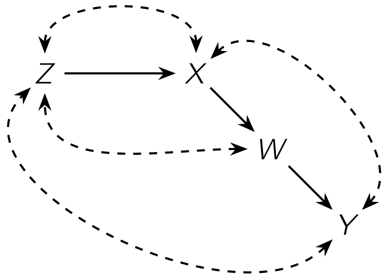
Definition (Hedge)

Let $X, Y \subseteq V$ in a graph G. Let F, F' be R-rooted C-forests such that $F \cap X \neq \varnothing$, $F' \cap X = \varnothing$, $F' \subseteq F$, and $R \subseteq \mathrm{An}(Y)_{G_{\overline{X}}}$. Then (F, F') form a *hedge* for $P_x(y)$ in G.

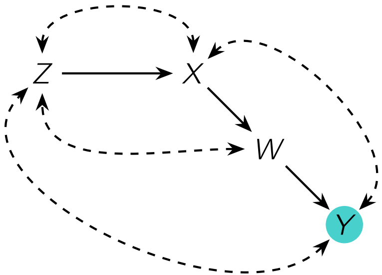
Theorem (Hedge Criterion for Identifiability)

$P_x(y)$ is identifiable if and only if there does not exists a hedge for $P_{x'}(y')$ in G for any $X' \subseteq X$, $Y' \subseteq Y$.

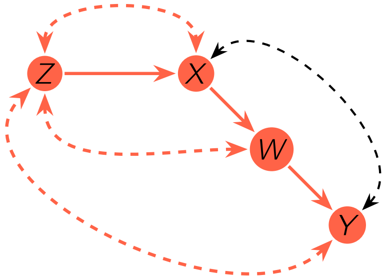
Hedges



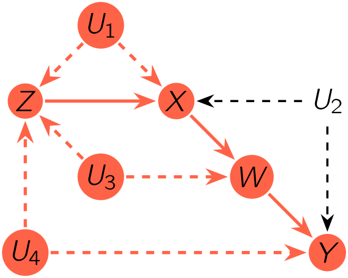
Hedges



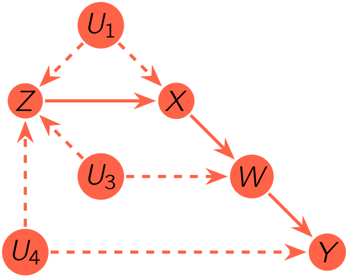
Hedges



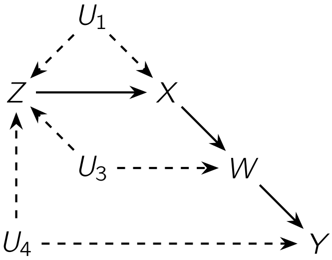
Hedges



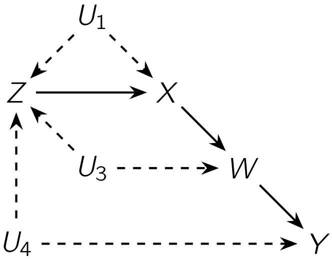
Hedges



Non-identifiability in hedges



Non-identifiability in hedges



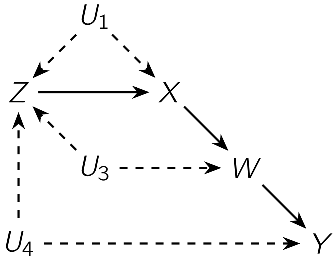
$\mathbf{M}^1$:		$\mathbf{M}^2$:	
U_i	$:= \text{Unif}(\{0,1\})$	U_i	$:= \text{Unif}(\{0,1\})$
Z	$:= U_1 \oplus U_3 \oplus U_4$	Z	$:= U_1 \oplus U_3 \oplus U_4$
X	$:= Z \oplus U_1$	X	$:= Z \oplus U_1$
W	$:= X \oplus U_3$	W	$:= X \oplus U_3$
Y	$:= W \oplus U_4$	Y	$:= 0$

Non-identifiability in hedges

In M_1 we also have $P^1(Y=0)=1$:

$$\begin{aligned} Y &= W \oplus U_4 \\ &= (X \oplus U_3) \oplus U_4 \\ &= (Z \oplus U_1) \oplus U_3 \oplus U_4 \\ &= (U_1 \oplus U_3 \oplus U_4) \oplus (U_1 \oplus U_3 \oplus U_4) \\ &= 0 \end{aligned}$$

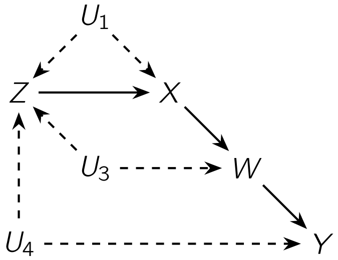
so $P^1(V) = P^2(V)$.



$\mathbf{M}^1$:		$\mathbf{M}^2$:	
U_i	$:= \text{Unif}(\{0,1\})$	U_i	$:= \text{Unif}(\{0,1\})$
Z	$:= U_1 \oplus U_3 \oplus U_4$	Z	$:= U_1 \oplus U_3 \oplus U_4$
X	$:= Z \oplus U_1$	X	$:= Z \oplus U_1$
W	$:= X \oplus U_3$	W	$:= X \oplus U_3$
Y	$:= W \oplus U_4$	Y	$:= 0$

Non-identifiability in hedges

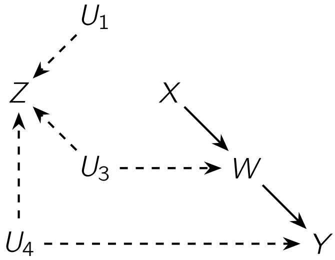
What happens when we intervene on X?



$$
\begin{array}{lll}
\mathbf{M}^1: & & \\
U_i & := \text{Unif}(\{0,1\}) \\
Z & := U_1 \oplus U_3 \oplus U_4 \\
X & := Z \oplus U_1 \\
W & := X \oplus U_3 \\
Y & := W \oplus U_4
\end{array}
\qquad
\begin{array}{lll}
\mathbf{M}^2: & & \\
U_i & := \text{Unif}(\{0,1\}) \\
Z & := U_1 \oplus U_3 \oplus U_4 \\
X & := Z \oplus U_1 \\
W & := X \oplus U_3 \\
Y & := 0
\end{array}
$$

Non-identifiability in hedges

What happens when we intervene on X?

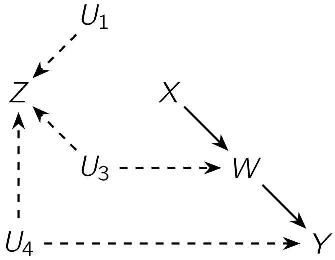


$\mathbf{M}^1$:		$\mathbf{M}^2$:	
U_i	$:= \text{Unif}(\{0,1\})$	U_i	$:= \text{Unif}(\{0,1\})$
Z	$:= U_1 \oplus U_3 \oplus U_4$	Z	$:= U_1 \oplus U_3 \oplus U_4$
X	$:= x$	X	$:= x$
W	$:= X \oplus U_3$	W	$:= X \oplus U_3$
Y	$:= W \oplus U_4$	Y	$:= 0$

Non-identifiability in hedges

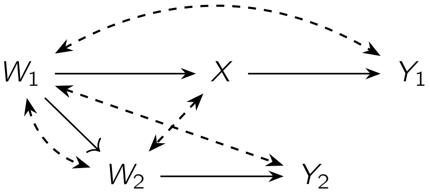
Then $Y = x \oplus U_3 \oplus U_4$. We have

$$P^1_x(Y) > 0, \quad P^2_x(Y = 1) = 0.$$

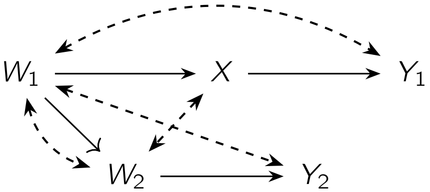


$\mathbf{M}^1$:		$\mathbf{M}^2$:	
U_i	$:= \text{Unif}(\{0, 1\})$	U_i	$:= \text{Unif}(\{0, 1\})$
Z	$:= U_1 \oplus U_3 \oplus U_4$	Z	$:= U_1 \oplus U_3 \oplus U_4$
X	$:= x$	X	$:= x$
W	$:= X \oplus U_3$	W	$:= X \oplus U_3$
Y	$:= W \oplus U_4$	Y	$:= 0$

Non-identifiability for the earlier example



Non-identifiability for the earlier example



In this example, $P_x(y_1, y_2)$ is unidentifiable because $\{W_1, W_2, Y_1, Y_2\}$ and $\{W_1, W_2, Y_1, Y_2, X\}$ form a hedge.