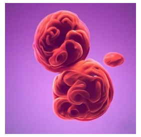
Bayesian Networks & Causal Graphs

Presenter: Hannah Chen

2023/06/28



University of Virginia

Overview

Bayesian Networks

Causal Axioms

Markov Equivalence

Statistical Modeling

$$P(x_1, x_2, ..., x_n) = P(x_1) \prod_{i=2}^{n} P(x_i \mid x_1, ..., x_{i-1})$$ (chain rule)

$$P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)P(x_4 \mid x_1, x_2, x_3)$$

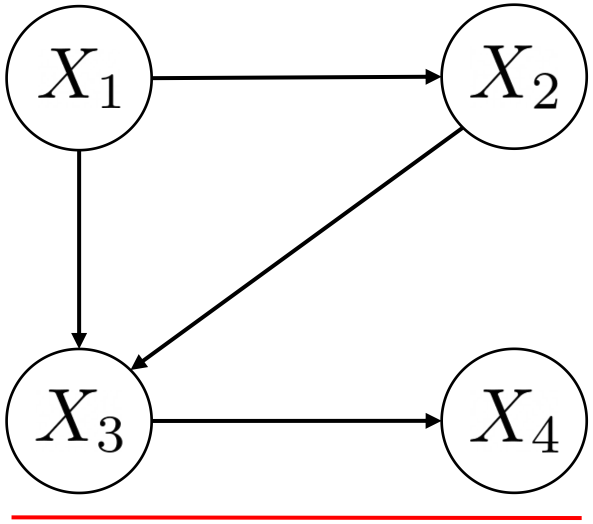
x_1	x_2	x_3	$P(x_4 \mid x_3, x_2, x_1)$
0	0	0	α_1
0	0	1	α_2
0	1	0	α_3
0	1	1	α_4
1	0	0	α_5
1	0	1	α_6
1	1	0	α_7
1	1	1	α_8

2^{n-1} parameters!

Bayesian Networks

Local Markov assumption: Given its parents in DAG, a node X is independent of all of its non-descendants.

$$P(x_1, x_2, ..., x_n) = \prod_i P(x_i \mid pa_i)$$



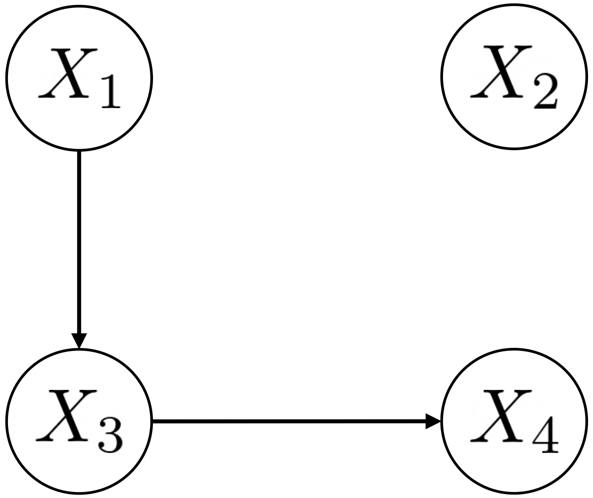
Bayesian $P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)P(\underline{x_4 \mid x_3})$

Statistical $P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)P(\underline{x_4 \mid x_1, x_2, x_3})$

Bayesian Networks

Local Markov assumption: Given its parents in DAG, a node X is independent of all of its non-descendants.

$$P(x_1, x_2, ..., x_n) = \prod_i P(x_i \mid pa_i)$$

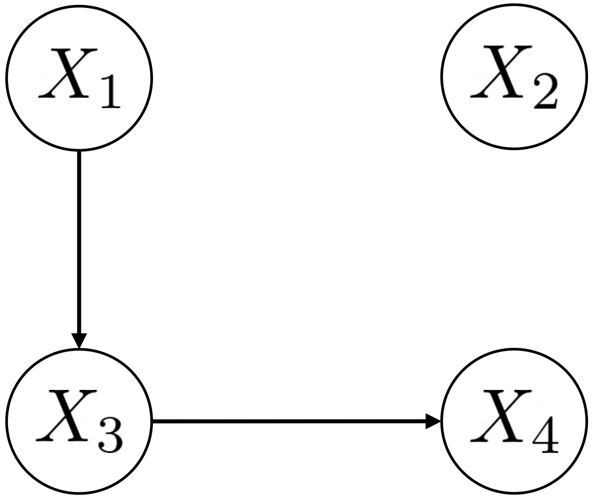


$$P(x_1, x_2, x_3, x_4) = ?$$

Bayesian Networks

Local Markov assumption: Given its parents in DAG, a node X is independent of all of its non-descendants.

$$P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i \mid pa_i)$$



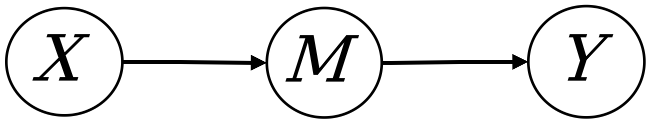
$$P(x_1, x_2, x_3, x_4) = P(x_1)P(x_2)P(x_3 \mid x_1)P(x_4 \mid x_3)$$

Global Markov Assumption

Given that P is Markov with respect to G (local Markov),

$$X \perp\!\!\!\perp_G Y \mid Z \Rightarrow X \perp\!\!\!\perp_P Y \mid Z$$

d-separation in graph G implies conditional independence in distribution P

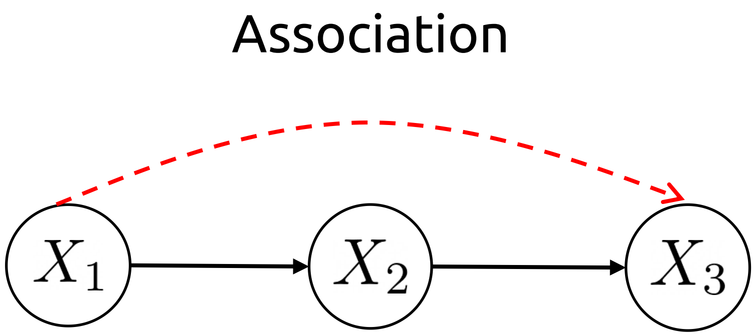


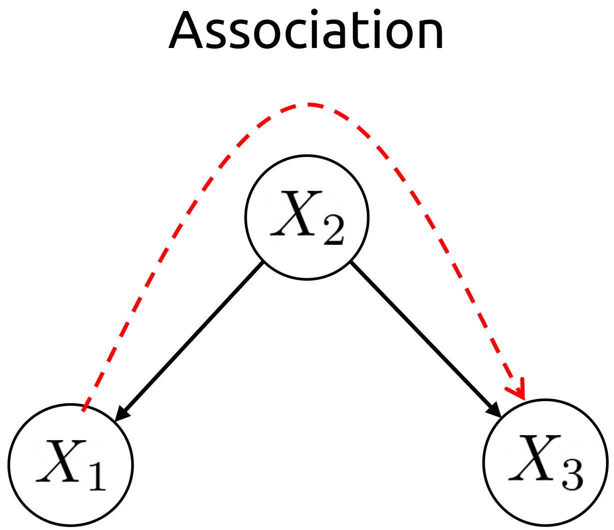
$$Z = \{M\}$$

Local Markov $\Longleftrightarrow$ Global Markov

Markov assumption

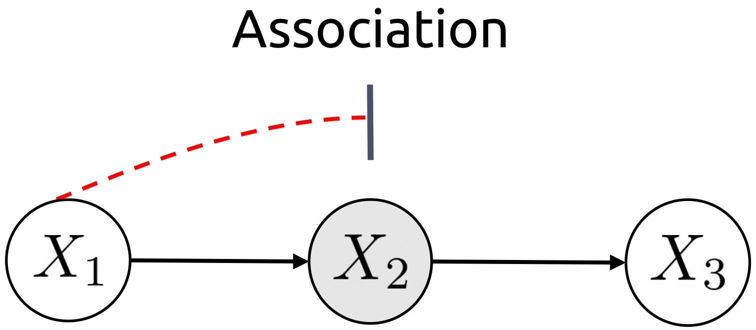
Flow of Association: Chains and Forks

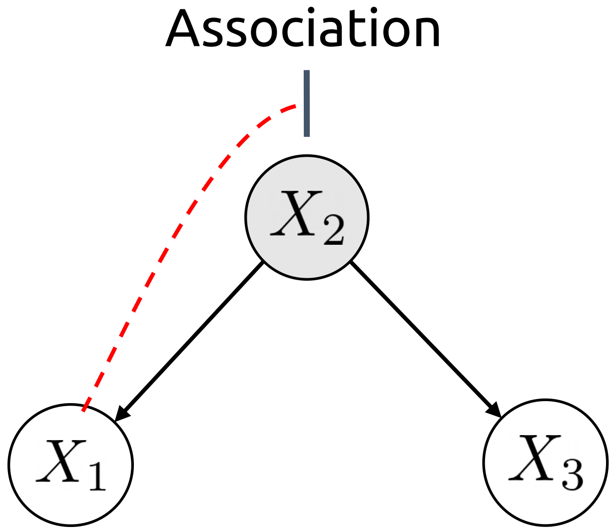




$$X_1 \not\perp\!\!\!\perp X_3$$

Flow of Association: Chains and Forks



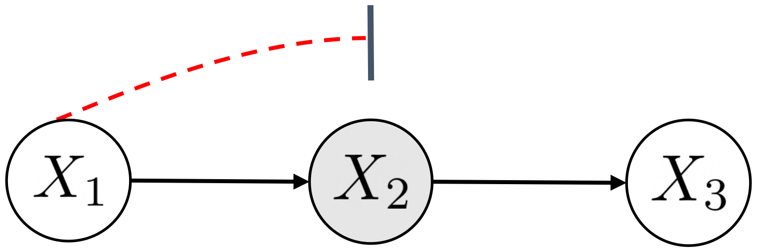


$$X_1 \perp\!\!\!\perp X_3 \mid X_2$$

Proof: Conditional Independence in Chains

Show: $P(x_1, x_3 \mid x_2) = \boxed{P(x_1 \mid x_2)} P(x_3 \mid x_2)$

$\equiv P(x_3 \mid x_1, x_2) = P(x_3 \mid x_2)$



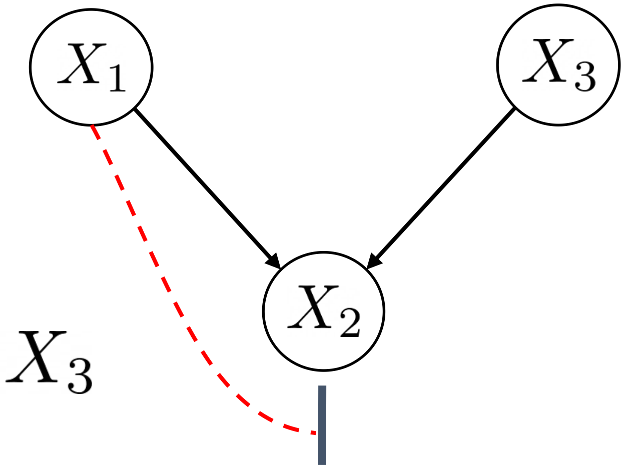
$$X_1 \perp\!\!\!\perp X_3 \mid X_2$$

1. Bayesian factorization $P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_2)$

2. Bayes' rule $P(x_1, x_3 \mid x_2) = \dfrac{P(x_1, x_2, x_3)}{P(x_2)} = \dfrac{\boxed{P(x_1)P(x_2 \mid x_1)}P(x_3 \mid x_2)}{\boxed{P(x_2)}}$

3. Bayes' rule again $P(x_1, x_3 \mid x_2) = \dfrac{P(x_1, x_2)}{P(x_2)} P(x_3 \mid x_2) = P(x_1 \mid x_2)P(x_3 \mid x_2)$

Flow of Association: Immoralities



$X_1 \perp\!\!\!\perp X_3$

Show: $P(x_1, x_3) = P(x_1)P(x_3)$

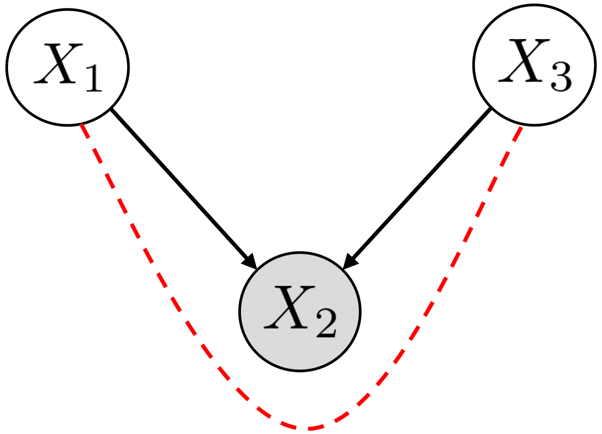
$$P(x_1, x_3) = \sum_{x_2} P(x_1, x_2, x_3)$$

$$= \sum_{x_2} P(x_1)P(x_3)P(x_2 \mid x_1, x_3) \qquad \text{(Bayesian factorization)}$$

$$= P(x_1)P(x_3) \overbrace{\sum_{x_2} P(x_2 \mid x_1, x_3)}^{1}$$

$$= P(x_1)P(x_3)$$

Flow of Association: Immoralities



Conditional dependence $X_1 \not\perp\!\!\!\perp X_3 \mid X_2$

$$P(x_1, x_3 \mid x_2) \propto P(x_2 \mid x_1, x_3) P(x_1) P(x_3)$$

$$\neq P(x_1 \mid x_2) P(x_3 \mid x_2)$$

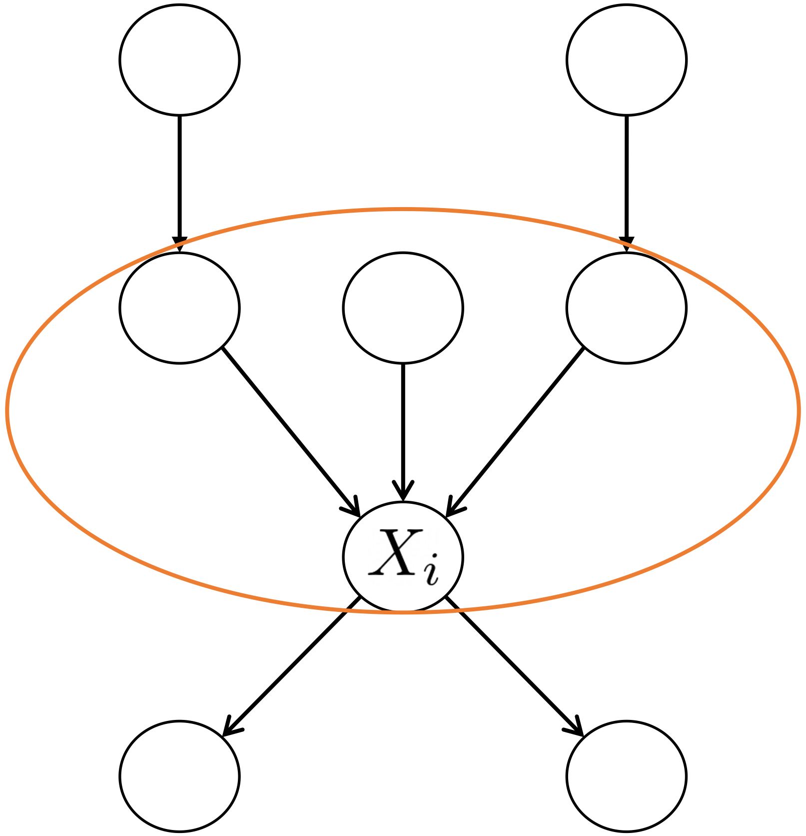
Intervention as Variables

Causal mechanism of X_i: $P(x_i \,|\, pa_i)$

Causal mechanism are modular

Interventions are local

$$do(X_i = x_i)$$



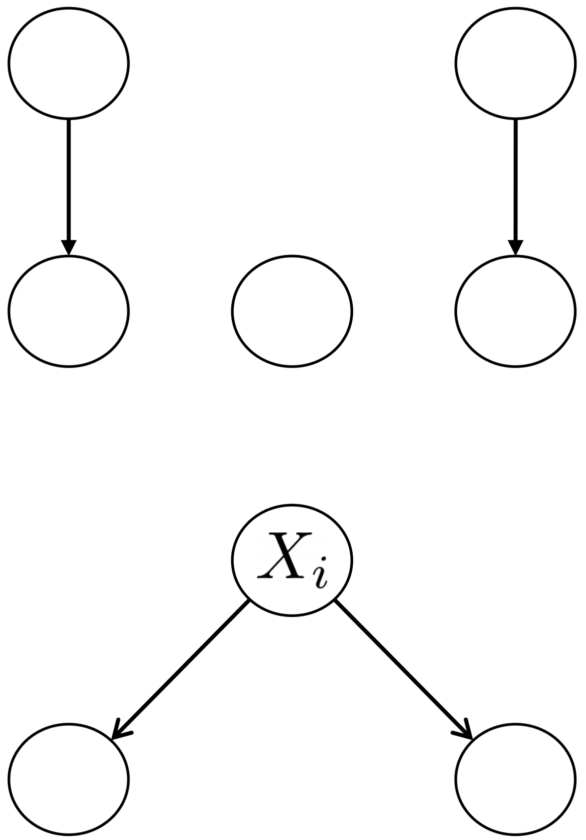
Intervention as Variables

Causal mechanism of X_i: $P(x_i \mid pa_i)$

Causal mechanism are modular

Interventions are local

$$do(X_i = x_i)$$



Modularity

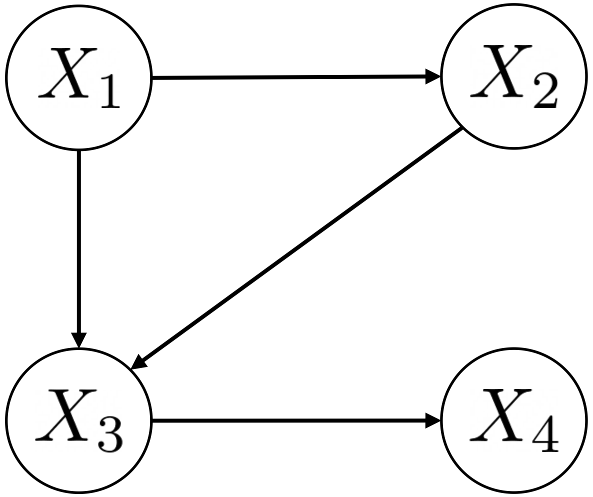
If we intervene on node X_i, then only $P(x_i|pa_i)$ changes. All other $P(x_j|pa_j)$ where $i \neq j$ remain unchanged.

More formally,

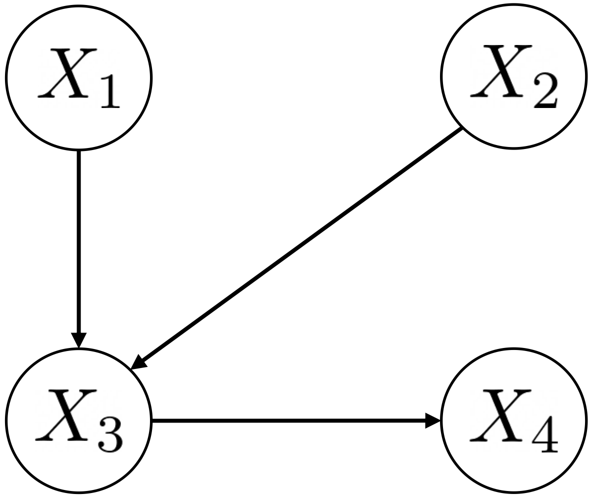
If we intervene on a set of nodes $S \subseteq [n]$, setting them to constants, then for all i, we have the following:

1. If $i \notin S$, then $P(x_i \mid \mathrm{pa}_i)$ remains unchanged.
2. If $i \in S$, then $P(x_i \mid \mathrm{pa}_i) = 1$ if x_i is the value that X_i was set to by the intervention; otherwise, $P(x_i \mid \mathrm{pa}_i) = 0$. consistent with the intervention

Modularity: Manipulated Graph



$do(X_2 = x_2)$



Observational distribution

$$P(x_1, x_2, x_3, x_4)$$
$$= P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)P(x_4 \mid x_3)$$

Interventional distribution

$$P(x_1, x_2, x_3, x_4 \mid do(X_2 = x_2))$$
$$= P(x_1) \cdot 1 \cdot P(x_3 \mid x_1, x_2)P(x_4 \mid x_3)$$

(Truncated factorization)

Interventions as Truncated Factorization

If x is consistent with the intervention (modularity):

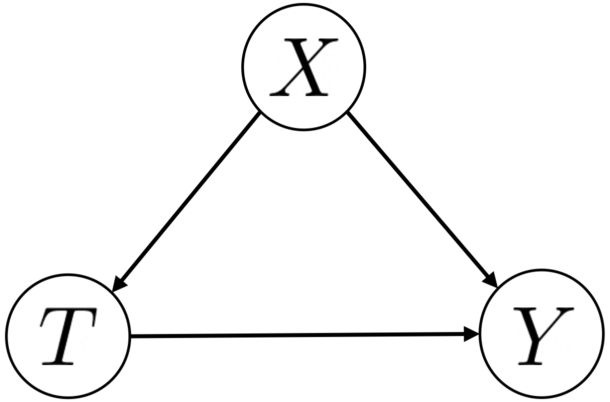
$$P(x_1, ..., x_n \mid do(S = s)) = \prod_{i:x_i \notin S} P(x_i \mid pa_i)$$

Otherwise,

$$P(x_1, ..., x_n \mid do(S = s)) = 0$$

Identification via Truncated Factorization

Identify $P(y \mid do(t))$



Bayesian factorization $P(y, t, x) = P(x)P(t \mid x)P(y \mid t, x)$

Truncated factorization $P(y, x \mid do(t)) = P(x)P(y \mid t, x)$

Marginalization $P(y \mid do(t)) = \sum_x P(y \mid t, x)P(x)$ → Backdoor adjustment!

Causal Axioms

Connecting probabilities with causal graphs:

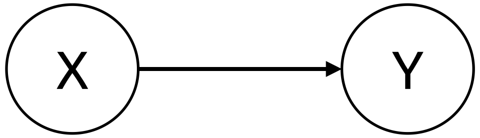
1. Markov Condition
2. Causal Minimality
3. Faithfulness Condition

Spirtes et al. (1999), Causation, Prediction, and Search", Chapter 3
Peters et al. (2017), Elements of Causal Inference: Foundations and Learning Algorithms, Section 6.5

MInimality Assumption

No subgraph of G also satisfies the Markov condition with respect to P

Markov assumption permits:



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

$$P(x, y) = P(x)P(y)$$

$\Longrightarrow$ Removing any edges from G, P would not be Markov with respect to G with the removed edges

Faithfulness Assumption

Markov Assumption: $X \perp\!\!\!\perp_G Y \mid Z \Rightarrow X \perp\!\!\!\perp_P Y \mid Z$

Causal graph ⟶ Data

Causal graph ⟵ Data

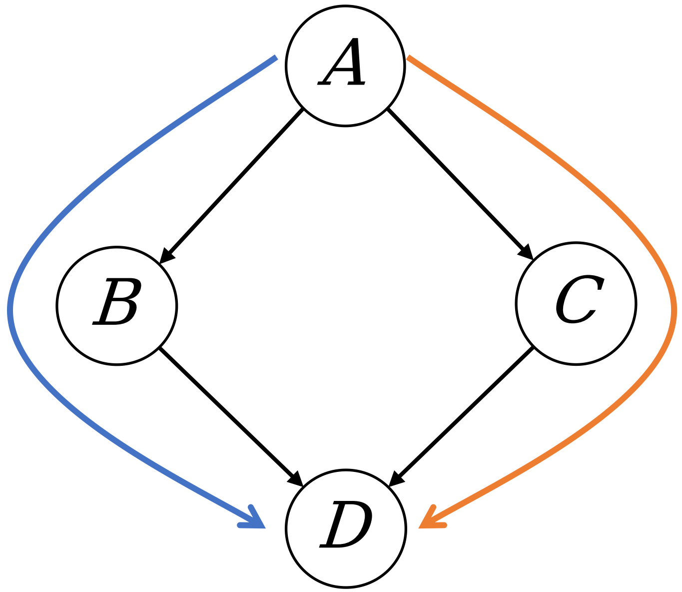
Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Leftarrow X \perp\!\!\!\perp_P Y \mid Z$

Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Leftarrow X \perp\!\!\!\perp_P Y \mid Z$

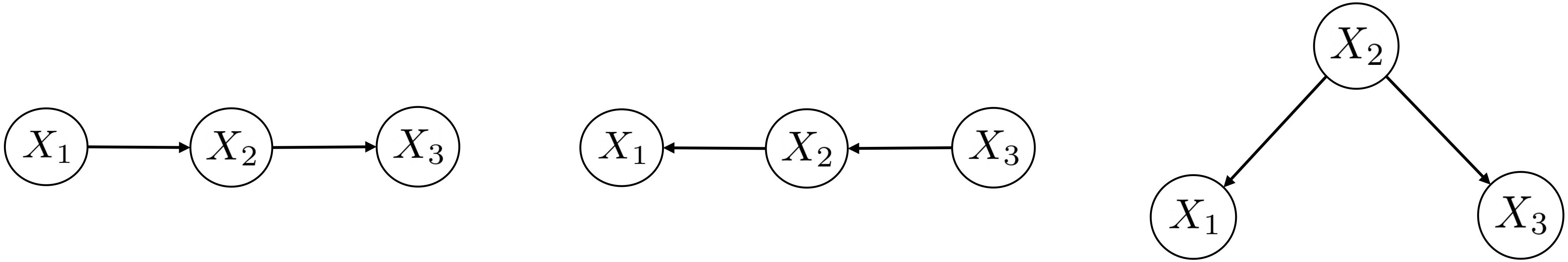
$$A \perp\!\!\!\perp D$$

but A and D are not d-separated



Two paths cancel each other

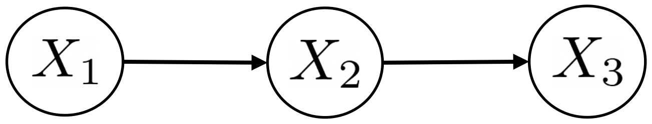
Markov Equivalence

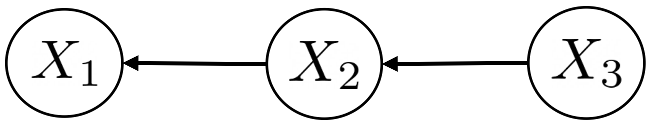


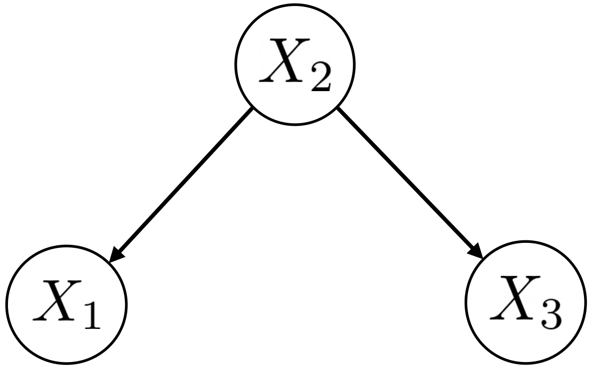
They're all $P(x_1, x_2, x_3)$

But different Markov factorization

Markov Equivalence







Markov: $X_1 \perp\!\!\!\perp X_3 \mid X_2$

Minimality: $X_1 \not\perp\!\!\!\perp X_2$ & $X_2 \not\perp\!\!\!\perp X_3$

Faithfulness: $X_1 \not\perp\!\!\!\perp X_3$

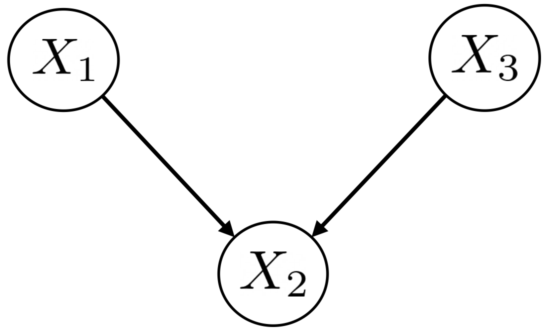
Markov equivalence class:

A set of DAGs that encode the same set of conditional independencies

Immoralities are Different

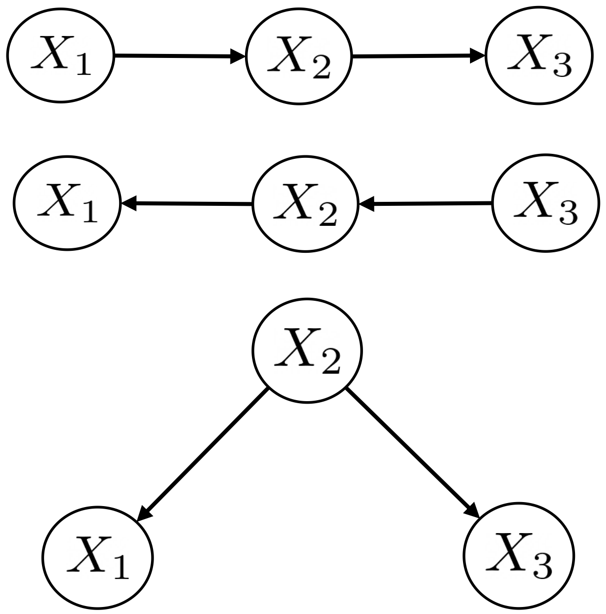
Markov equivalence class where

$X_1 \perp\!\!\!\perp X_3$ & $X_1 \not\perp\!\!\!\perp X_3 \mid X_2$



Markov equivalence class where

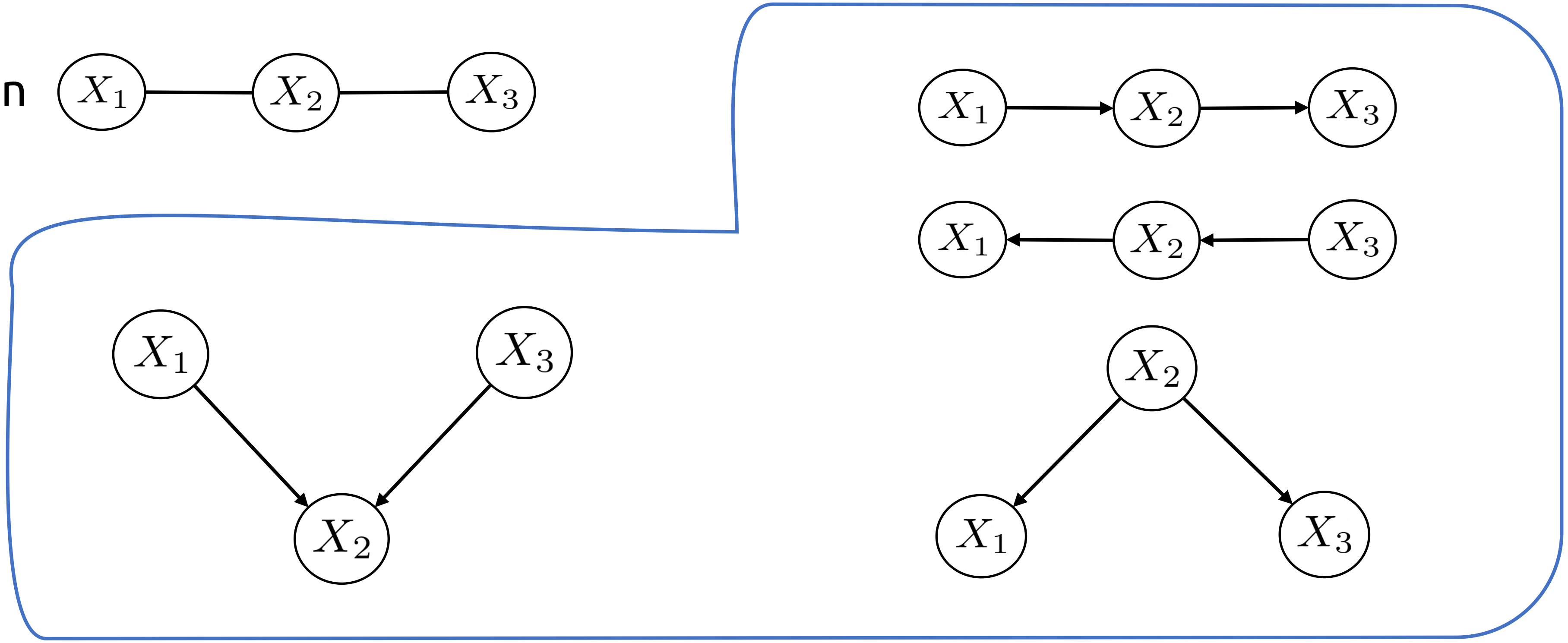
$X_1 \perp\!\!\!\perp X_3 \mid X_2$ & $X_1 \not\perp\!\!\!\perp X_3$



Markov Equivalence

[Verma & Pearl, 1991] Two DAGs are Markov equivalent if and only if they have the same **skeleton** and the same **immoralities** (v-structures).

Skeleton

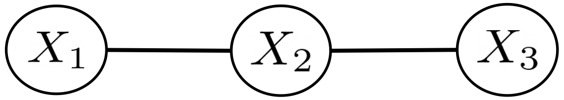


Markov Equivalence

Useful for causal discovery

[Verma & Pearl, 1991] Two DAGs are Markov equivalent if and only if they have the same **skeleton** and the same **immoralities** (v-structures).

Skeleton



Immorality