# Multivariate Causal Models

Congyuan Duan

January 25, 2022

# Contents

Graph Terminology

Structural Causal Models

Interventions

Counterfactuals

Markov Property, Faithfulness, and Causal Minimality

# Contents

Graph Terminology

Structural Causal Models

Interventions

Counterfactuals

Markov Property, Faithfulness, and Causal Minimality

# Graph Terminology

- Consider finitely many random variables $X = (X_1, \cdots, X_d)$ with index set $V = \{1, \cdots, d\}$, joint distribution $P_X$, and density $p(x)$. The corresponding graph is $\mathcal{G} = (V, \mathcal{E})$.
- We say that there is an **undirected edge** between two adjacent nodes $i$ and $j$ if $(i, j) \in \mathcal{E}$ and $(j, i) \in \mathcal{E}$. An edge between two adjacent nodes is **directed** if it is not undirected. We call $\mathcal{G}$ **directed** if all its edges are directed.
- Three nodes are called an **immorality** or a **v-structure** if one node is a child of the two others that themselves are not adjacent.
- The **skeleton** of $\mathcal{G}$ does not take the directions of the edges into account.

# Graph Terminology

- A **path** in $\mathcal{G}$ is a sequence of distinct vertices $i_1, \cdots, i_m$, such that there is an edge between $i_k$ and $i_{k+1}$ for all $k = 1, \cdots, m-1$. If $i_{k-1} \rightarrow i_k$ and $i_{k+1} \rightarrow i_k$, $i_k$ is called a **collider relative to this path**.
- All ancestors of $i$ are denoted by $AN_i^{\mathcal{G}}$ and $i$ is not an ancestor of itself. We denote all descendants of $i$ by $DE_i^{\mathcal{G}}$ and all non-descendants of $i$, excluding $i$, by $ND_i^{\mathcal{G}}$. $ND_i^{\mathcal{G}}$ include the parents of i in graph $\mathcal{G}$.
- A graph $\mathcal{G}$ is called a **partially directed acyclic graph (PDAG)** if there is no directed cycle, that is, if there is no pair $(j, k)$ with directed paths from $j$ to $k$ and from $k$ to $j$. $\mathcal{G}$ is called a **directed acyclic graph (DAG)** if it is a PDAG and all edges are directed.

# Graph Terminology

**Definition 6.1 (Pearl's $d$-separation)** *In a DAG $\mathcal{G}$, a path between nodes $i_1$ and $i_m$ is **blocked by a set** $\mathbf{S}$ (with neither $i_1$ nor $i_m$ in $\mathbf{S}$) whenever there is a node $i_k$, such that one of the following two possibilities holds:*

*(i)* $i_k \in \mathbf{S}$ *and*

$$\begin{aligned} & i_{k-1} \to i_k \to i_{k+1} \\ \text{or } & i_{k-1} \leftarrow i_k \leftarrow i_{k+1} \\ \text{or } & i_{k-1} \leftarrow i_k \to i_{k+1} \end{aligned}$$

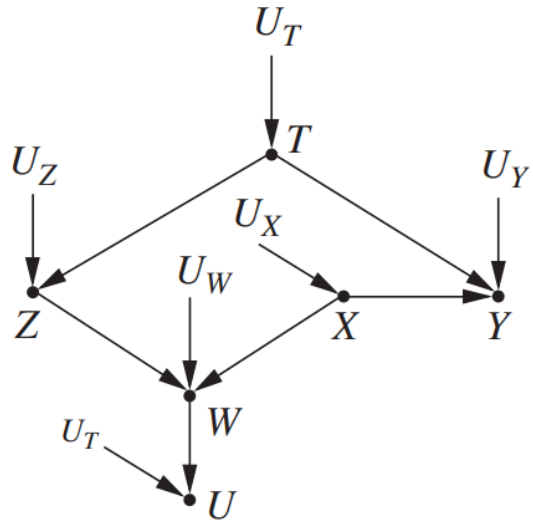
*(ii) neither $i_k$ nor any of its descendants is in* $\mathbf{S}$ *and*

$$i_{k-1} \to i_k \leftarrow i_{k+1}.$$

*Furthermore, in a DAG $\mathcal{G}$, we say that two disjoint subsets of vertices* $\mathbf{A}$ *and* $\mathbf{B}$ *are **$d$-separated** by a third (also disjoint) subset* $\mathbf{S}$ *if every path between nodes in* $\mathbf{A}$ *and* $\mathbf{B}$ *is blocked by* $\mathbf{S}$*. We then write*

$$\mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \,|\, \mathbf{S}.$$

# Example




- $Z$ and $Y$ are unconditionally dependent
- Condition on $T$, $Z$ and $Y$ become independent
- Condition on $\{T, W\}$, $Z$ and $Y$ become dependent
- Condition on $\{T, W, X\}$, $Z$ and $Y$ become independent again

# Contents

Graph Terminology

Structural Causal Models

Interventions

Counterfactuals

Markov Property, Faithfulness, and Causal Minimality

# Definition

**Definition 6.2 (Structural causal models)** *A **structural causal model (SCM)** $\mathfrak{C} := (\mathbf{S}, P_{\mathbf{N}})$ consists of a collection* $\mathbf{S}$ *of* $d$ ***(structural) assignments***

$$X_j := f_j(\mathbf{PA}_j, N_j), \qquad j = 1, \ldots, d, \tag{6.1}$$

*where* $\mathbf{PA}_j \subseteq \{X_1, \ldots, X_d\} \setminus \{X_j\}$ *are called **parents of** $X_j$; and a joint distribution* $P_{\mathbf{N}} = P_{N_1,\ldots,N_d}$ *over the noise variables, which we require to be jointly independent; that is,* $P_{\mathbf{N}}$ *is a product distribution.*

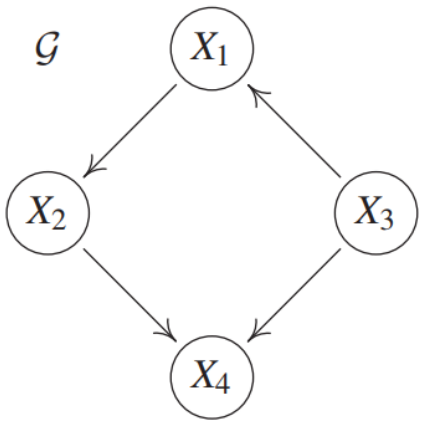
*The graph* $\mathcal{G}$ *of an SCM is obtained by creating one vertex for each* $X_j$ *and drawing directed edges from each parent in* $\mathbf{PA}_j$ *to* $X_j$*, that is, from each variable* $X_k$ *occurring on the right-hand side of equation* (6.1) *to* $X_j$ *(see Figure 6.1). We henceforth assume this graph to be acyclic.*

*We sometimes call the elements of* $\mathbf{PA}_j$ *not only parents but also **direct causes** of* $X_j$*, and we call* $X_j$ *a **direct effect** of each of its direct causes. SCMs are also called (nonlinear) SEMs.*

# Example

$$X_1 := f_1(X_3, N_1)$$
$$X_2 := f_2(X_1, N_2)$$
$$X_3 := f_3(N_3)$$
$$X_4 := f_4(X_2, X_3, N_4)$$

- $N_1, \ldots, N_4$ jointly independent
- $\mathcal{G}$ is acyclic




In this book we focus mainly on acyclic structures.

# Entailed distributions

**Proposition 6.3 (Entailed distributions)** *An SCM $\mathfrak{C}$ defines a unique distribution over the variables $\mathbf{X} = (X_1, \ldots, X_d)$ such that $X_j = f_j(\mathbf{PA}_j, N_j)$, in distribution, for $j = 1, \ldots, d$. We refer to it as the entailed distribution $P_{\mathbf{X}}^{\mathfrak{C}}$ and sometimes write $P_{\mathbf{X}}$.*

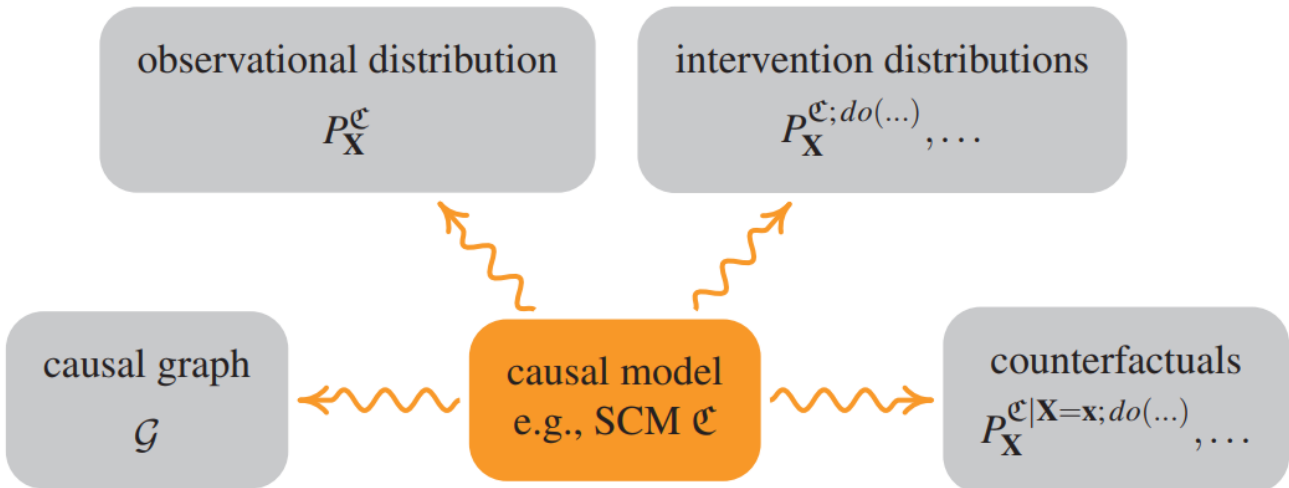



Figure 6.2: Causal models as SCMs do not only model an observational distribution $P$ (Proposition 6.3) but also intervention distributions (Section 6.3) and counterfactuals (Section 6.4).

# Structural minimality of SCMs

**Remark 6.6 (Structural minimality of SCMs)** Definition 6.2 can be read such that one distinguishes between the two SCMs

$$\mathbf{S}_1 : X := N_X,\ Y := 0 \cdot X + N_Y \quad \text{and}$$
$$\mathbf{S}_2 : X := N_X,\ Y := N_Y,$$

even though clearly $0 \cdot X = 0$. This contradicts our intuition. We therefore add the requirement that the functions $f_j$ depend on all of their input arguments. Mathematically speaking, whenever there is a $k \in \{1, \ldots, d\}$ and a function $g$ such that

$$f_k(\mathbf{pa}_k, n_k) = g(\mathbf{pa}_k^*, n_k), \qquad \forall \mathbf{pa}_k, \forall n_k \text{ with } p(n_k) > 0, \tag{6.5}$$

where $\mathbf{PA}_k^* \subsetneq \mathbf{PA}_k$, we choose the latter representation. In the preceding example, we would therefore choose the representation $\mathbf{S}_2$ over $\mathbf{S}_1$. We will see later that these two SCMs can indeed be identified in that they entail the same observational distribution, intervention distribution,[2] and counterfactuals (see Section 6.8).

# Contents

Graph Terminology

Structural Causal Models

Interventions

Counterfactuals

Markov Property, Faithfulness, and Causal Minimality

# Definition

**Definition 6.8 (Intervention distribution)** *Consider an SCM $\mathfrak{C} := (\mathbf{S}, P_{\mathbf{N}})$ and its entailed distribution $P_{\mathbf{X}}^{\mathfrak{C}}$. We replace one (or several) of the structural assignments to obtain a new SCM $\tilde{\mathfrak{C}}$. Assume that we replace the assignment for $X_k$ by*

$$X_k := \tilde{f}(\widetilde{\mathbf{PA}}_k, \tilde{N}_k).$$

*We then call the entailed distribution of the new SCM an intervention distribution and say that the variables whose structural assignment we have replaced have been **intervened** on. We denote the new distribution by*[4]

$$P_{\mathbf{X}}^{\tilde{\mathfrak{C}}} =: P_{\mathbf{X}}^{\mathfrak{C};do\left(X_k := \tilde{f}(\widetilde{\mathbf{PA}}_k, \tilde{N}_k)\right)}.$$

- The causal parents of the intervened variable have changed.
- Intervention distributions differ from the observational distribution.

# Example

Consider the SCM

$$
\begin{aligned}
X_1 &:= N_{X_1} \\
Y &:= X_1 + N_Y \\
X_2 &:= Y + N_{X_2}
\end{aligned}
$$


with $N_{X1}, N_Y \sim N(0,1)$ and $N_{X2} \sim N(0, 0.1)$.
Interventions on $X_2$ are useless

$$P_Y^{\mathfrak{C};do(X_2:=\tilde{N})} = P_Y^{\mathfrak{C}} \quad \text{for all variables } \tilde{N};$$

Intervention on $X_1$, however, does change the distribution of $Y$

$$P_Y^{\mathfrak{C};do(X_1:=\tilde{N})} = \mathcal{N}\left(\mathbb{E}[N_Y] + \mathbb{E}[\tilde{N}], \mathrm{var}[N_Y] + \mathrm{var}[\tilde{N}]\right) \neq P_Y^{\mathfrak{C}}$$

Intervening is usually different from conditioning

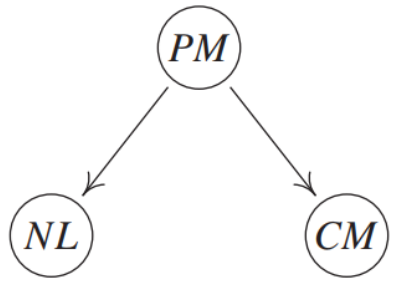
$$p_Y^{\mathfrak{C};do(X_2:=x)}(y) = p_Y^{\mathfrak{C}}(y) \neq p_Y^{\mathfrak{C}}(y|X_2 = x).$$

# Example: Myopia

Assume that the underlying SCM is of the form

$$\mathbf{S}: \quad \begin{aligned} PM &:= N_{PM} \\ NL &:= f(PM, N_{NL}) \\ CM &:= g(PM, N_{CM}) \end{aligned}$$

where PM stands for parent myopia, NL for night light, and CM for child myopia.




Replacing the structural assignment of NL with $NL := \tilde{N}_{NL}$, where $\tilde{N}_{NL}$ could randomly assign one out of the three night light conditions(darkness, night light, room light) with equal probability, we would find $NL \perp\!\!\!\perp CM$.

# Total causal effects

**Definition 6.12 (Total causal effect)** *Given an SCM $\mathfrak{C}$, there is a total causal effect from $X$ to $Y$ if and only if*

$$X \not\perp\!\!\!\perp Y \quad \text{in } P_{\mathbf{X}}^{\mathfrak{C};do(X:=\tilde{N}_X)}$$

*for some random variable $\tilde{N}_X$.*

**Proposition 6.13 (Total causal effects)** *Given an SCM $\mathfrak{C}$, the following statements are equivalent:*

*(i) There is a total causal effect from $X$ to $Y$.*

*(ii) There are $x^{\triangle}$ and $x^{\square}$ such that $P_Y^{\mathfrak{C};do(X:=x^{\triangle})} \neq P_Y^{\mathfrak{C};do(X:=x^{\square})}$.*

*(iii) There is $x^{\triangle}$ such that $P_Y^{\mathfrak{C};do(X:=x^{\triangle})} \neq P_Y^{\mathfrak{C}}$.*

*(iv) $X \not\perp\!\!\!\perp Y$ in $P_{X,Y}^{\mathfrak{C};do(X:=\tilde{N}_X)}$ for any $\tilde{N}_X$ whose distribution has full support.*

# Total causal effects

**Proposition 6.14 (Graphical criteria for total causal effects)** *Assume we are given an SCM $\mathfrak{C}$ with corresponding graph $\mathcal{G}$.*

*(i) If there is no directed path from $X$ to $Y$, then there is no total causal effect.*

*(ii) Sometimes there is a directed path but no total causal effect.*

# Example: Randomized trials

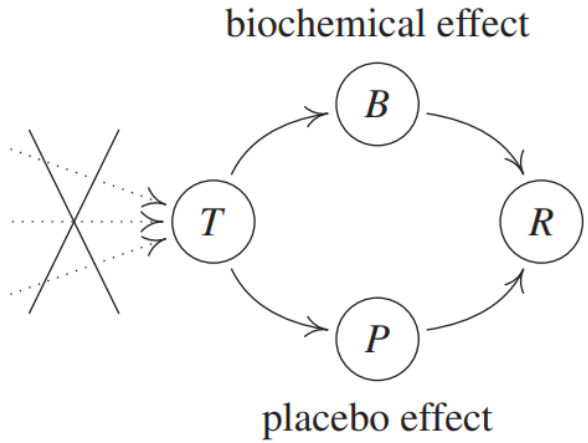



Figure 6.3: Simplified description of randomized studies. $T$ denotes the treatment, $P$ and $B$ the patient's psychology and some biochemical state, and $R$ indicates whether the patient recovers. The randomization over $T$ removes the influence of any other variable on $T$, and thus there cannot be any hidden common cause between $T$ and $R$. We distinguish between two different effects: the placebo effect via $P$ and the biochemical effect via $B$.

$T = 0$: no medication, $T = 1$: placebo, $T = 2$: drug of interest

# Example: Kidney stones

| | Overall | Patients with small stones | Patients with large stones |
|---|---|---|---|
| Treatment $a$: Open surgery | 78% (273/350) | **93%** (81/87) | **73%** (192/263) |
| Treatment $b$: Percutaneous nephrolithotomy | **83%** (289/350) | 87% (234/270) | 69% (55/80) |

Table 6.1: A classic example of Simpson's paradox. The table reports the success rates of two treatments for kidney stones [Bottou et al., 2013, Charig et al., 1986, tables I and II]. Although the overall success rate of treatment $b$ seems better (any bold number is largest in its column), treatment $b$ performs worse than treatment $a$ on both patients with small kidney stones and patients with large kidney stones (see Examples 6.37 and Section 9.2).

# Contents

Graph Terminology

Structural Causal Models

Interventions

Counterfactuals

Markov Property, Faithfulness, and Causal Minimality

# Definition

**Definition 6.17 (Counterfactuals)** *Consider an SCM* $\mathfrak{C} := (\mathbf{S}, P_{\mathbf{N}})$ *over nodes* $\mathbf{X}$. *Given some observations* $\mathbf{x}$, *we define a* counterfactual SCM *by replacing the distribution of noise variables:*

$$\mathfrak{C}_{\mathbf{X}=\mathbf{x}} := \left(\mathbf{S}, P_{\mathbf{N}}^{\mathfrak{C}|\mathbf{X}=\mathbf{x}}\right),$$

*where* $P_{\mathbf{N}}^{\mathfrak{C}|\mathbf{X}=\mathbf{x}} := P_{\mathbf{N}|\mathbf{X}=\mathbf{x}}$.[7] *The new set of noise variables need not be jointly independent anymore. Counterfactual statements can now be seen as do-statements in the new counterfactual SCM.*

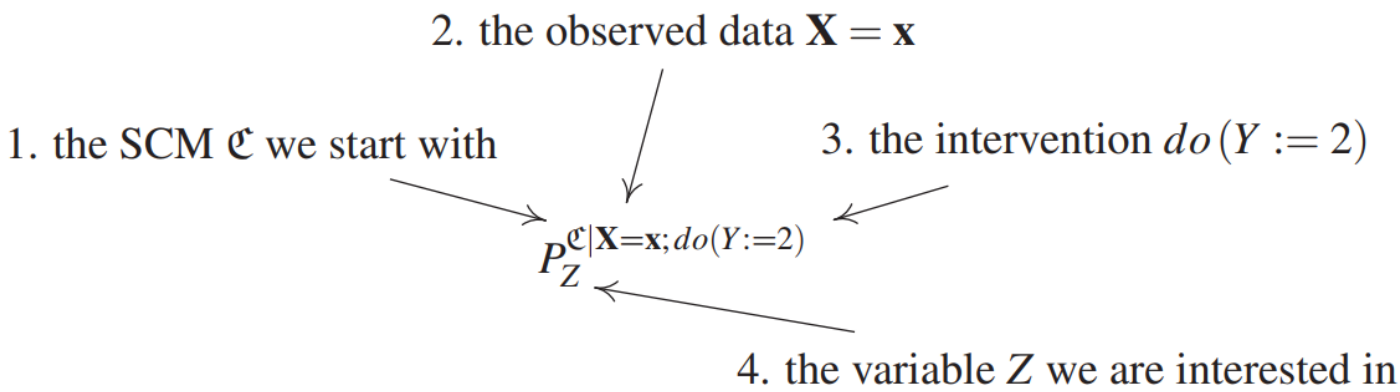
Counterfactual corresponds to updating the noise distributions of an SCM (by conditioning) and then performing an intervention.

# Example

Consider the SCM

$$
\begin{aligned}
X &:= N_X \\
Y &:= X^2 + N_Y \\
Z &:= 2 \cdot Y + X + N_Z
\end{aligned}
$$

with $N_x, N_y, N_z \overset{iid}{\sim} U(\{-5, -4, \cdots, 4, 5\})$. Assume that we observe $(X, Y, Z) = (1, 2, 4)$, then $Z$ would have been 11 had $X$ been set to 2.



# Example

This example shows two SCMs that induce the same graph, observational distributions, and intervention distributions but entail different counterfactual statements.

**Example 6.19** Let $N_1, N_2 \sim \text{Ber}(0.5)$, and $N_3 \sim \text{U}(\{0,1,2\})$, such that the three variables are jointly independent. That is, $N_1, N_2$ have a Bernoulli distribution with parameter 0.5 and $N_3$ is uniformly distributed on $\{0,1,2\}$. We define two different SCMs. First consider $\mathfrak{C}_A$:

$$
\begin{aligned}
X_1 &:= N_1 \\
X_2 &:= N_2 \\
X_3 &:= (1_{N_3>0} \cdot X_1 + 1_{N_3=0} \cdot X_2) \cdot 1_{X_1 \neq X_2} + N_3 \cdot 1_{X_1 = X_2}.
\end{aligned}
$$

If $X_1$ and $X_2$ have different values, depending on $N_3$ we either choose $X_3 = X_1$ or $X_3 = X_2$. Otherwise $X_3 = N_3$. Now, $\mathfrak{C}_B$ differs from $\mathfrak{C}_A$ only in the latter case:

$$
\begin{aligned}
X_1 &:= N_1 \\
X_2 &:= N_2 \\
X_3 &:= (1_{N_3>0} \cdot X_1 + 1_{N_3=0} \cdot X_2) \cdot 1_{X_1 \neq X_2} + (2 - N_3) \cdot 1_{X_1 = X_2}.
\end{aligned}
$$

# Contents

Graph Terminology

Structural Causal Models

Interventions

Counterfactuals

Markov Property, Faithfulness, and Causal Minimality

# Markov property

**Definition 6.21 (Markov property)** *Given a DAG $\mathcal{G}$ and a joint distribution $P_{\mathbf{X}}$, this distribution is said to satisfy*

*(i) the **global Markov property** with respect to the DAG $\mathcal{G}$ if*

$$\mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \,|\, \mathbf{C} \;\Rightarrow\; \mathbf{A} \perp\!\!\!\perp \mathbf{B} \,|\, \mathbf{C}$$

*for all disjoint vertex sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$ (the symbol $\perp\!\!\!\perp_{\mathcal{G}}$ denotes d-separation — see Definition 6.1),*

*(ii) the **local Markov property** with respect to the DAG $\mathcal{G}$ if each variable is independent of its non-descendants given its parents, and*

*(iii) the **Markov factorization property** with respect to the DAG $\mathcal{G}$ if*

$$p(\mathbf{x}) = p(x_1, \ldots, x_d) = \prod_{j=1}^{d} p(x_j \,|\, \mathbf{pa}_j^{\mathcal{G}}).$$

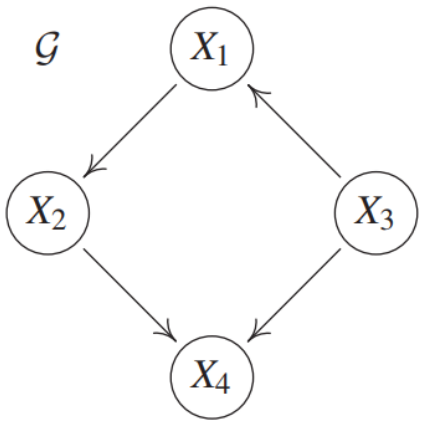
*For this last property, we have to assume that $P_{\mathbf{X}}$ has a density $p$; the factors in the product are referred to as **causal Markov kernels** describing the conditional distributions $P_{X_j|\mathbf{PA}_j^{\mathcal{G}}}$.*

**Theorem 6.22 (Equivalence of Markov properties)** *If $P_{\mathbf{X}}$ has a density $p$, then all Markov properties in Definition 6.21 are equivalent.*

# Example

$$X_1 := f_1(X_3, N_1)$$
$$X_2 := f_2(X_1, N_2)$$
$$X_3 := f_3(N_3)$$
$$X_4 := f_4(X_2, X_3, N_4)$$

- $N_1, \ldots, N_4$ jointly independent
- $\mathcal{G}$ is acyclic




The Markov condition relates statements about graph separation to conditional independences.

**Proposition 6.31 (SCMs imply Markov property)** *Assume that $P_{\mathbf{X}}$ is induced by an SCM with graph $\mathcal{G}$. Then, $P_{\mathbf{X}}$ is Markovian with respect to $\mathcal{G}$.*

# Markov equivalence of graphs

Different graphs may encode the exact same set of conditional independences.

**Definition 6.24 (Markov equivalence of graphs)** *We denote by $\mathcal{M}(\mathcal{G})$ the set of distributions that are Markovian with respect to $\mathcal{G}$:*

$$\mathcal{M}(\mathcal{G}) := \{P : P \text{ satisfies the global (or local) Markov property with respect to } \mathcal{G}\}.$$

*Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if $\mathcal{M}(\mathcal{G}_1) = \mathcal{M}(\mathcal{G}_2)$. This is the case if and only if $\mathcal{G}_1$ and $\mathcal{G}_2$ satisfy the same set of d-separations, which means the Markov condition entails the same set of (conditional) independence conditions.*

*The set of all DAGs that are Markov equivalent to some DAG is called* ***Markov equivalence class*** *of $\mathcal{G}$. It can be represented by a completed PDAG that is denoted by $CPDAG(\mathcal{G}) = (V, \mathcal{E})$; it contains the (directed) edge $(i, j) \in \mathcal{E}$ if and only if one member of the Markov equivalence class does; see Figure 6.4.*

# Markov equivalence of graphs

**Lemma 6.25 (Graphical criteria for Markov equivalence)** *Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if and only if they have the same skeleton and the same immoralities.*

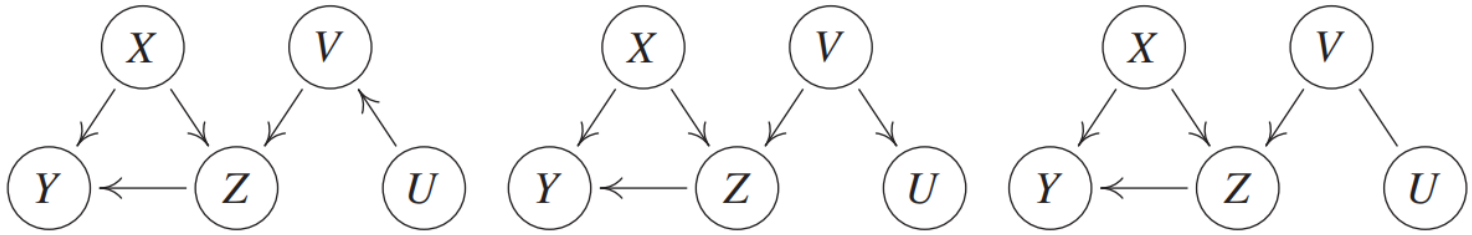



Figure 6.4: Two Markov equivalent DAGs (left and center); these are the only two DAGs in the corresponding Markov equivalence class that can be represented by the CPDAG on the right-hand side.

# Markov blanket

**Definition 6.26 (Markov blanket)** *Consider a DAG $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ and a target node $Y$. The Markov blanket of $Y$ is the smallest set $M$ such that*

$$Y \perp\!\!\!\perp_{\mathcal{G}} \mathbf{V} \setminus (\{Y\} \cup M) \text{ given } M.$$

If $P_{\mathbf{X}}$ is Markovian with respect to $\mathcal{G}$, then

$$Y \perp\!\!\!\perp \mathbf{V} \setminus (\{Y\} \cup M) \text{ given } M.$$

In other words, given $M$, the other variables do not provide any further information about $Y$.

**Proposition 6.27 (Markov blanket)** *Consider a DAG $\mathcal{G}$ and a target node $Y$. Then, the Markov blanket $M$ of $Y$ includes its parents, its children, and the parents of its children*

$$M = \mathbf{PA}_Y \cup \mathbf{CH}_Y \cup \mathbf{PA}_{\mathbf{CH}_Y}.$$

# Reichenbach's common cause principle

Markov property relates distributions and graphs, then it can be used to justify Reichenbach's common cause principle.

**Proposition 6.28 (Reichenbach's common cause principle)** *Assume that any pair of variables $X$ and $Y$ can be embedded into a larger system in the following sense. There exists a correct SCM over the collection $\mathbf{X}$ of random variables that contains $X$ and $Y$ with graph $\mathcal{G}$. Then Reichenbach's common cause principle follows from the Markov property. If $X$ and $Y$ are (unconditionally) dependent, then there is*

*(i) either a directed path from $X$ to $Y$, or*

*(ii) from $Y$ to $X$, or*

*(iii) there is a node $Z$ with a directed path from $Z$ to $X$ and from $Z$ to $Y$.*

# Causal graphical model

**Definition 6.32 (Causal graphical model)** *A causal graphical model over random variables* $\mathbf{X} = (X_1, \ldots, X_d)$ *contains a graph* $\mathcal{G}$ *and a collection of functions* $f_j(x_j, x_{\mathbf{PA}_j^{\mathcal{G}}})$ *that integrate to* 1*:*

$$\int f_j(x_j, x_{\mathbf{PA}_j^{\mathcal{G}}})\, dx_j = 1.$$

*These functions induce a distribution* $P_{\mathbf{X}}$ *over* $\mathbf{X}$ *via*

$$p(x_1, \ldots, x_d) = \prod_{j=1}^{d} f_j(x_j, x_{\mathbf{PA}_j^{\mathcal{G}}}),$$

*and thus play the role of conditionals:* $f_j(x_j, x_{\mathbf{PA}_j^{\mathcal{G}}}) = p(x_j | x_{\mathbf{PA}_j^{\mathcal{G}}})$*. A causal graphical model induces intervention distribution according to Equations* (6.8) *and* (6.9) *in Section 6.6. In the most general form, we can define*

$$p^{do\left(X_k := q(\cdot | x_{\widetilde{\mathbf{PA}}_k})\right)}(x_1, \ldots, x_d) = \prod_{j \neq k} f_j(x_j, x_{\mathbf{PA}_j^{\mathcal{G}}})\; q(\cdot | x_{\widetilde{\mathbf{PA}}_k}),$$

*with* $q(\cdot | x_{\widetilde{\mathbf{PA}}_k})$ *integrating to* 1 *and the new parents not leading to a cycle.*

# Faithfulness and Causal Minimality

Markov assumption enables us to read off independences from the graph structure. Faithfulness however, allows us to infer dependences from the graph structure.

**Definition 6.33 (Faithfulness and causal minimality)** *Consider a distribution $P_{\mathbf{X}}$ and a DAG $\mathcal{G}$.*

*(i) $P_{\mathbf{X}}$ is faithful to the DAG $\mathcal{G}$ if*

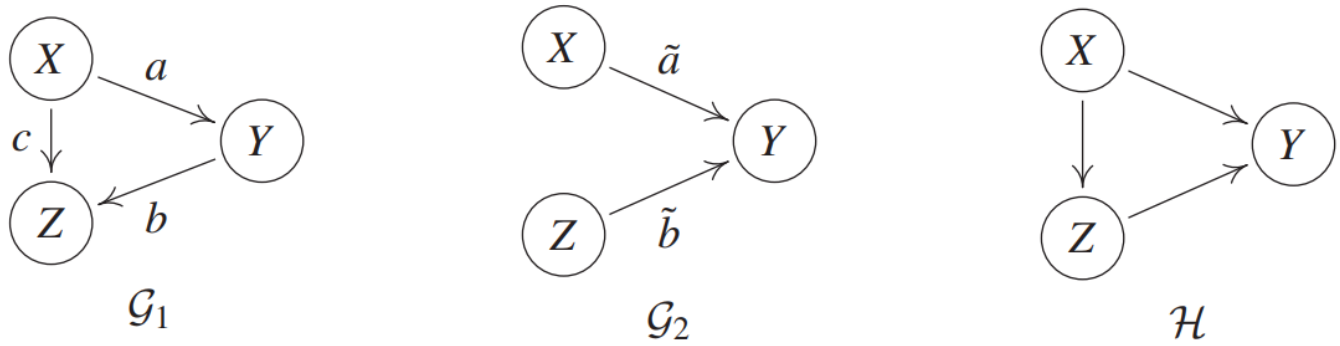
$$\mathbf{A} \perp\!\!\!\perp \mathbf{B} \,|\, \mathbf{C} \;\Rightarrow\; \mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \,|\, \mathbf{C}$$

*for all disjoint vertex sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$.*

*(ii) A distribution satisfies causal minimality with respect to $\mathcal{G}$ if it is Markovian with respect to $\mathcal{G}$, but not to any proper subgraph of $\mathcal{G}$.*

# Example

**Example 6.34 (Violation of faithfulness)** Consider the following figure.




We first look at a linear Gaussian SCM that corresponds to the left graph $\mathcal{G}_1$.

$$
\begin{aligned}
X &:= N_X, \\
Y &:= aX + N_Y, \\
Z &:= bY + cX + N_Z,
\end{aligned}
$$

If $ab + c = 0$, the distribution is not faithful with respect to $\mathcal{G}_1$ since we obtain $X \perp\!\!\!\perp Z$, which is not implied by the graph structure.

# Faithfulness and causal minimality

**Proposition 6.35 (Faithfulness implies causal minimality)** *If $P_{\mathbf{X}}$ is faithful and Markovian with respect to $\mathcal{G}$, then causal minimality is satisfied.*

A distribution is minimal with respect to $\mathcal{G}$ if and only if there is no node that is conditionally independent of any of its parents, given the remaining parents.

**Proposition 6.36 (Equivalence of causal minimality)** *Consider the random vector $\mathbf{X} = (X_1, \ldots, X_d)$ and assume that the joint distribution has a density with respect to a product measure. Suppose that $P_{\mathbf{X}}$ is Markovian with respect to $\mathcal{G}$. Then $P_{\mathbf{X}}$ satisfies causal minimality with respect to $\mathcal{G}$ if and only if $\forall X_j \, \forall Y \in \mathbf{PA}_j^{\mathcal{G}}$ we have that $X_j \not\perp\!\!\!\perp Y \,|\, \mathbf{PA}_j^{\mathcal{G}} \setminus \{Y\}$.*

# Reference

Pearl J, Glymour M, Jewell N P. Causal inference in statistics: A primer[M]. John Wiley & Sons, 2016.
Murphy K P. Machine learning: a probabilistic perspective[M]. MIT press, 2012.