Causal Graphical Models (Summer Short Course)

Qingyuan Zhao (Statistical Laboratory, University of Cambridge)

4-5 August, 2025 @ Peking University

Contents

0.1 Literature

0.1.1 Day 1 (Lecture 1-4)

- Zhao, Q. (2024). A matrix algebra for graphical statistical models. (arXiv:2407.15744)

0.1.2 Day 2 (Lecture 5-8)

- Lauritzen, S. L. (1996). *Graphical Models.* Clarendon Press.
- Zhao, Q. (2025). On statistical and causal models associated with acyclic directed mixed graphs. (arXiv:2501.03048)
- Guo, F. R., & Zhao, Q. (2023). Confounder selection via iterative graph expansion. (arXiv:2309.06053)

1 Lecture 1: Directed mixed graphs and linear systems

1.1 History

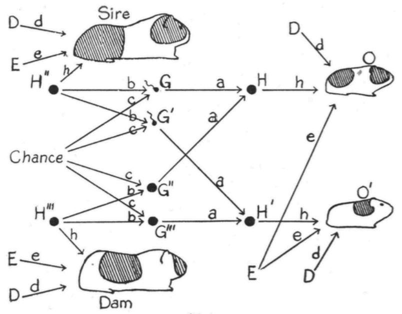


FIG. 5.

Diagram illustrating the casual relations between litter mates (O, O') and between each of them and their parents. H, H', H'', H,''' represent the genetic constitutions of the four individuals, G, G', G'', and G''' that of four germ cells. E represents such environmental factors as are common to litter mates. D represents other factors, largely ontogenetic irregularity. The small letters stand for the various path coefficients.

- **Sewall Green Wright** (December 21, 1889 – March 3, 1988) was an American geneticist known for his influential work on evolutionary theory and also for his work on **path analysis**. He was a founder of population genetics alongside Ronald Fisher and J. B. S. Haldane, which was a major step in the development of the modern synthesis combining genetics with evolution.

1.2 Directed mixed graphs

We will consider graphs with two types of edges: directed ($\longrightarrow$) and bidirected ($\longleftrightarrow$).

1.2.1 Definition

A **directed mixed graph (DMG)** $\mathrm{G} = (V, \mathcal{D}, \mathcal{B})$ consists of a finite vertex set V, a directed edge set $\mathcal{D} \subseteq V \times V$ that contains ordered pairs of vertices, and a bidirected edge set $\mathcal{B} \subseteq V \times V$ that contains unordered pairs of vertices (so $(j,k) \in \mathcal{B}$ implies $(k,j) \in \mathcal{B}$) such that

$$(j,k) \in \mathcal{B} \Longrightarrow (j,j) \in \mathcal{B}, (k,k) \in \mathcal{B}, \quad \text{for all } j,k \in V.$$

Let $\mathbb{G}(V)$ denote the collection of all directed mixed graphs with vertex set V.

We say the **directed edge** "$j \longrightarrow k$" is contained in G if $(j,k) \in \mathcal{D}$, and in this case we say this is an *incoming edge* for k, an **outgoing edge** for j, the vertex j is a **parent** of k, and k is a **child** of j in G. Likewise, we say the **bidirected edge** "$j \longleftrightarrow k$" is contained in G if $(j,k) \in \mathcal{B}$.

1.3 Causal interpretation

- Directed edges mean direct causal effects.
- Bidirected edges mean unspecified, residual/exogenous correlations.

1.3.1 Why directed edges?

- Causality is **transitive** (A causes B and B causes C $\Rightarrow$ A causes C). This defines a **pre-order**.
  - This can be described by the reachability relationship of a **directed graph**.
- Often we think causality is **irreflexive** (A does not cause itself) and **asymmetric** (A and B cannot be causes of each other). This defines a **partial order**.
  - This can be described by the reachability relationship of a **directed acyclic graph**.

1.3.2 Why bidirected edges?

- In statistics and causal inference we are often concerned with **latent variables**.

1.4 Canonical graphs

1.4.1 Definition

- We say the directed mixed graph is **canonical** if it contains all bidirected loops.
  - The full collection with vertex set V is denoted by $\mathbb{G}^*(V)$.
- We say the graph is **canonically directed** if it is canonical and contains no other bidirected edges.
  - The full collection with vertex set V is denoted by $\mathbb{G}^*_{\mathrm{D}}(V)$.

For such graphs, it is usually more convenient to use the trimmed graph obtained by

$$\text{trim} : (V, \mathcal{D}, \mathcal{B}) \mapsto (V, \mathcal{D}, \mathcal{B} \setminus \{(j, j) : j \in V\}).$$

1.5 Walk, path, cycle

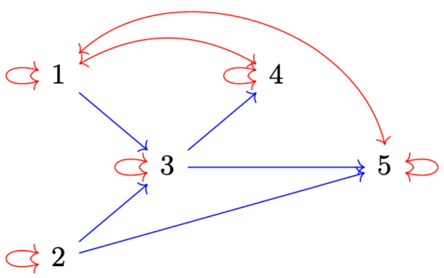
1.5.1 Definition

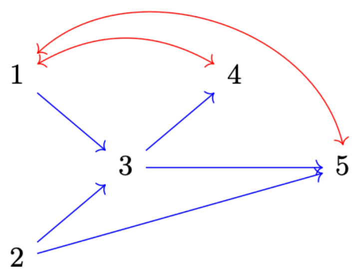
- A **walk** is an ordered sequence of connected edges ignoring edge direction.
- A **path** is a walk with no repeated vertices.
- A **cycle** is a walk with the same starting and ending vertices.
- Vertices at the two ends of a walk are called its **endpoints**, and the other vertices are called **non-endpoints**.
- A walk is **directed** if all its edges have the same direction, like $j \longrightarrow \cdots \longrightarrow k$.
- The graph G (or its direct subgraph) is **acyclic** if it contains no directed cycles.

1.5.2 Notation

- $\mathbb{G}^*_{\mathrm{A}}(V)$: the collection of acyclic, canonical DMGs with vertex set V.
- $\mathbb{G}^*_{\mathrm{DA}}(V)$: the collection acyclic, canonically directed DMGs (basically DAGs).

1.6 Demonstration of trimming





Is the directed mixed graph on the left

- canonical?
- canonically directed?
- acyclic?

1.7 Gaussian linear system on a graph

We will not distinguish a random vector $V = (V_1, \ldots, V_d)$ (probability theory) with a vertex set of random variables $V = \{V_1, \ldots, V_d\}$ (graph theory).

1.7.1 Definition

We say a random vector V follows a **Gaussian linear system** wrt $\mathrm{G} \in \mathbb{G}(V)$ if

$$V = \beta^T V + E \text{ for some } E \sim \mathrm{N}(0, \Lambda)$$

where

- $\beta \in \mathbb{R}^{d\times d}$ respects the directed subgraph: $V_j \not\to V_k$ **in** $\mathrm{G} \Rightarrow \beta_{jk} = 0$.
- $\Lambda \in \mathbb{R}^{d\times d}$ respects the bidirected subgraph: $V_j \not\leftrightarrow V_k$ **in** $\mathrm{G} \Rightarrow \Lambda_{jk} = 0$.

1.7.2 Gaussian model

- Let $\mathbb{N}(\mathrm{G})$ denote the collection of probability distributions of such V.
- Let $\mathbb{N}^+(\mathrm{G})$ denote the subclass where Λ is positive definite (so $\mathrm{G} \in \mathbb{G}^*(V)$).

1.8 Roadmap: Basic questions

Let V be a random vector and $J = V_{\mathcal{J}}$, $K = V_{\mathcal{K}}$, $L = V_{\mathcal{L}}$ be sub-vectors of V.

1. What is the probability distribution of J?
2. Is $J \perp\!\!\!\perp K$ true?
3. Is $J \perp\!\!\!\perp K \mid L$ true?

1.8.1 Answer

If $V \sim \mathrm{N}(0, \Sigma)$ and Σ is positive definite, then

1. $J \sim \mathrm{N}(0, \Sigma_{\mathcal{J},\mathcal{J}})$.
2. $J \perp\!\!\!\perp K$ if and only if $\Sigma_{\mathcal{J},\mathcal{K}} = 0$.
3. $V_j \perp\!\!\!\perp V_k \mid V_{[d]\setminus\{j,k\}}$ if and only if $(\Sigma^{-1})_{jk} = 0$.

1.8.2 Lecture 1-3

Let $\mathsf{P} \in \mathbb{N}(\mathrm{G})$ be the distribution of V. Can we answer these questions using just G?

1.9 Roadmap: From linear algebra to graphs

1.9.1 Basic combinatoric result

Let A be the adjacency matrix of a directed graph G. Then

$$(A^r)_{jk} = |\{\text{directed walks from } j \text{ to } k \text{ of length } r\}|,\ r \geq 1,$$

$$[(Id - A)^{-1}]_{jk} = [Id + A + A^2 + \ldots]_{jk} = \delta_{jk} + |\{\text{directed walks from } j \text{ to } k\}|.$$

So **matrix multiplication is similar to walking on a graph**.

- Thinking abstractly, edges in a graph encode certain local relations. By composing ("multiplying") those edges, we can obtain new global relations.

1.10 Matrices of walks

1.10.1 Basic matrices

Edges in $\mathrm{G} = (V, \mathcal{D}, \mathcal{B}) \in \mathbb{G}(V)$ can be rearranged into:

$$W[j \longrightarrow k \textbf{ in } \mathrm{G}] = \begin{cases} \{j \longrightarrow k\}, & \text{if } (j,k) \in \mathcal{D}, \\ \emptyset, & \text{otherwise,} \end{cases}$$
$$W[j \longleftrightarrow k \textbf{ in } \mathrm{G}] = \begin{cases} \{j \longleftrightarrow k\}, & \text{if } (j,k) \in \mathcal{B}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

- Examples (on blackboard).

1.11 Basic operations

1.11.1 On sets of walks

1. Addition + means set union.
   - Example: $\{V_2 \longrightarrow V_5\} + \{V_2 \longrightarrow V_3 \longrightarrow V_5\} =?$.
2. Multiplication $\cdot$ means concatenation.
   - Example: $\{V_2 \longleftrightarrow V_2\} \cdot \{V_2 \longrightarrow V_5,\ V_2 \longrightarrow V_3 \longrightarrow V_5\} =?$.
3. Transpose ^{T} means reversing direction.
   - Example: $\{V_2 \longrightarrow V_5,\ V_2 \longrightarrow V_3 \longrightarrow V_5\}^T =?$.

1.11.2 On matrices

Examples (on blackboard).

$$\begin{aligned} (W + W')[V_j, V_k] &= W[V_j, V_k] + W'[V_j, V_k], \\ (W \cdot W')[V_j, V_k] &= \sum_{V_l \in V} W[V_j, V_l] \cdot W'[V_l, V_k], \\ (W^T)[V_j, V_k] &= (W[V_k, V_j])^T = \{w^T : w \in W[V_k, V_j]\}. \end{aligned}$$

1.12 Further definitions

- **Right-directed walks**: $W[V \rightsquigarrow V] = \sum_{q=1}^{\infty} (W[V \longrightarrow V])^q$.
- **Left-directed walks**: $W[V \leftsquigarrow V] = (W[V \rightsquigarrow V])^T$.
- **Identity matrix** (for multiplication): $\mathrm{Id} = \mathrm{diag}(\mathrm{id}, \dots, \mathrm{id})$, where id is the trivial walk with length 0 such that
$$\mathrm{id} \cdot w = w \cdot \mathrm{id} = w \quad \text{and} \quad \mathrm{Id} \cdot W = W \cdot \mathrm{Id} = W$$
- Let $\emptyset$ also denote the matrix with empty sets of walks. This is the identity element for matrix addition (set union).

1.13 Weight function

- Recall that $\mathsf{P} \in \mathbb{N}(\mathrm{G})$ means $V = \beta^T V + E$ where $E \sim \mathrm{N}(0, \Lambda)$.
- So $V = (\mathrm{Id} - \beta)^{-T} E \sim \mathrm{N}(0, \Sigma)$ where
$$\Sigma = (Id - \beta)^{-T} \Lambda (Id - \beta)^{-1}.$$
How can we represent this graphically?
- Let σ be the **weight function** on all walks in G generated by
$$\beta = \sigma(W[V \longrightarrow V]) \quad \text{and} \quad \Lambda = \sigma(W[V \longleftrightarrow V]).$$
- Example: $\sigma(\{V_1 \longleftrightarrow V_4, V_1 \longleftrightarrow V_1 \longrightarrow V_3 \longrightarrow V_4\}) = \Lambda_{14} + \Lambda_{11}\beta_{13}\beta_{34}$.

1.13.1 Lemma

If β is stable (spectral radius < 1), then
$$(\mathrm{Id} - \beta)^{-1} = \mathrm{Id} + \sigma(W[V \rightsquigarrow V]).$$

2 Lecture 2: Path analysis and graph marginalization

2.1 Trek rule

This motivates us to define **treks** or **t-connected walks** in G as (expand on blackboard)
$$W[V \overset{\mathrm{t}}{\leftrightsquigarrow} V] = (\mathrm{Id} + W[V \leftsquigarrow V]) \cdot W[V \longleftrightarrow V] \cdot (\mathrm{Id} + W[V \rightsquigarrow V]).$$

2.1.1 Theorem

Suppose $\mathrm{G} \in \mathbb{G}(V)$ and $\mathsf{P} \in \mathbb{N}(\mathrm{G})$ with weight function σ, then
$$\mathsf{Cov}_{\mathsf{P}}(V) = \sigma(W[V \overset{\mathrm{t}}{\leftrightsquigarrow} V \textbf{ in } \mathrm{G}]).$$

2.1.2 Examples

- $\mathsf{Var}(V_3) =?$
- $\mathsf{Cov}(V_3, V_4) =?$

2.2 Arcs

2.2.1 Definitions

- We say a walk is an **arc** or **m-connected** if it has no collider (like $\longrightarrow j \longleftarrow$).
- We say a walk is **d-connected** if it is an arc and contains no bidirected edge, so

2.2.2 Proposition

An arc has exactly zero or one bidirected edge.

2.2.3 Notation

A squiggly line ($\rightsquigarrow$) means no collider, and we use no/half/full arrowheads at both ends.
$$W[V \overset{\mathrm{d}}{\leftrightsquigarrow} V] = W[V \leftsquigarrow V] + W[V \rightsquigarrow V] + W[V \leftsquigarrow V \rightsquigarrow V],$$
$$W[V \leftrightsquigarrow V] = W[V \overset{\mathrm{t}}{\leftrightsquigarrow} V] + W[V \overset{\mathrm{d}}{\leftrightsquigarrow} V].$$

2.3 From treks to paths

- Notation: $P[\cdots] = W[\cdots] \cap \mathcal{P}_{\mathrm{G}}$ where $\mathcal{P}_{\mathrm{G}}$ contains all paths on G and $\cap$ is applied entry-wise.

2.3.1 Lemma

For any $\mathrm{G} \in \mathbb{G}^*(V)$ and any $j, k \in V$, $j \neq k$, we have

$$j \overset{\mathrm{t}}{\leftrightsquigarrow} k \textbf{ in } \mathrm{G} \iff P[j \overset{}{\leftrightsquigarrow} k \textbf{ in } \mathrm{G}] \neq \emptyset \iff P[j \leftrightsquigarrow k \textbf{ in } \mathrm{trim}(\mathrm{G})] \neq \emptyset.$$

- Proof on blackboard (assuming G is acyclic).
- Key definition:

$$P[j \overset{\mathrm{d}}{\leftrightsquigarrow} k \text{ via root } r] = \begin{cases} P[j \rightsquigarrow k], & \text{if } r = j, \\ P[j \leftsquigarrow k], & \text{if } r = k, \\ (P[j \leftsquigarrow r] \cdot P[r \rightsquigarrow k]) \cap \mathcal{P}_{\mathrm{G}}, & \text{otherwise.} \end{cases}$$

2.4 Wright's path analysis

2.4.1 Theorem

Suppose $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}(V)$ and $\mathsf{P} \in \mathbb{N}(\mathrm{G})$ with weight function σ, then for any $V_j, V_k \in V$, $j \neq k$, we have

$$\begin{aligned} \mathsf{Cov}_{\mathsf{P}}(V_j, V_k) =& \sigma(P[V_j \overset{\mathrm{t}}{\leftrightsquigarrow} V_k \textbf{ in } \mathrm{G}]) \\ &+ \sum_{V_r \in V} \sigma(P[V_j \overset{\mathrm{d}}{\leftrightsquigarrow} V_k \text{ via root } V_r \textbf{ in } \mathrm{G}]) \cdot \mathsf{Var}_{\mathsf{P}}(V_r). \end{aligned}$$

- Proof on blackboard.
- Example: $\mathsf{Cov}(V_3, V_4) = ?$

2.5 Blocking arcs

2.5.1 Definition

We say an arc is **blocked** by $L \subseteq V$ if the arc has an non-endpoint in L.

2.5.2 Examples

1. Which of the following are blocked by 3?
   - $1 \longrightarrow 3 \longrightarrow 4$
   - $4 \longleftarrow 3 \longleftrightarrow 3 \longrightarrow 5$
   - $1 \longrightarrow 3 \longleftarrow 2$
2. Define $W[V \rightsquigarrow V \mid L]$ and $W[V \overset{\mathrm{t}}{\leftrightsquigarrow} V \mid L]$ using the matrix algebra.

2.5.3 Lemma

If $\beta = \sigma(W[V \longrightarrow V])$ is principally stable, then for any $L \subseteq V$, we have

$$(\mathrm{Id} - \beta_{L^c L^c})^{-1} = \mathrm{Id} + \sigma(W[L^c \rightsquigarrow L^c \mid L]).$$

2.6 Marginalization of graphs

- Notation: Write $W[\cdots \textbf{ in } \mathrm{G}] \neq \emptyset$ as $\cdots \textbf{ in } \mathrm{G}$.
- For $\mathrm{G} \in \mathbb{G}(V)$ and $\tilde{V} \subseteq V$, the **marginal graph** $\mathrm{margin}_{\tilde{V}}(\mathrm{G})$ is obtained by

$$
\begin{aligned}
j \longrightarrow k \textbf{ in } \tilde{\mathrm{G}} &\iff j \rightsquigarrow k \mid \tilde{V} \textbf{ in } \mathrm{G}, \quad j,k \in \tilde{V}, \\
j \longleftrightarrow k \textbf{ in } \tilde{\mathrm{G}} &\iff j \overset{\mathrm{t}}{\leftrightsquigarrow} k \mid \tilde{V} \textbf{ in } \mathrm{G}, \quad j,k \in \tilde{V}.
\end{aligned}
$$

- If $\tilde{V}$ is **ancestral** (meaning $\tilde{V}$ contains $\mathrm{an}(\tilde{V}) = \{V_j \in V : V_j \rightsquigarrow \tilde{V}\}$), then $\mathrm{margin}_{\tilde{V}}(\mathrm{G})$ is the subgraph of G restricted to $\tilde{V}$.
- It is often useful to view $\mathrm{margin}_{\tilde{V}}$ as a map of walks in G to walks in $\tilde{\mathrm{G}}$.

2.6.1 Proposition: Marginalization preserves unblocked directed walks and treks

For any $\mathrm{G} \in \mathbb{G}(V)$ and $L \subseteq \tilde{V} \subseteq V$,

$$
\mathrm{margin}_{\tilde{V}}\left(W \left[\tilde{V} \begin{Bmatrix} \rightsquigarrow \\ \leftsquigarrow \\ \overset{\mathrm{t}}{\leftrightsquigarrow} \end{Bmatrix} \tilde{V} \mid L \textbf{ in } \mathrm{G} \right] \right) = W \left[\tilde{V} \begin{Bmatrix} \rightsquigarrow \\ \leftsquigarrow \\ \overset{\mathrm{t}}{\leftrightsquigarrow} \end{Bmatrix} \tilde{V} \mid L \textbf{ in } \mathrm{margin}_{\tilde{V}}(\mathrm{G}) \right].
$$

2.7 Marginalization of linear systems

2.7.1 Theorem

For $\mathrm{G} \in \mathbb{G}(V)$ and $\mathsf{P} \in \mathbb{N}(\mathrm{G})$, if β is principally stable, then

$$
\mathrm{margin}_{\tilde{V}}(\mathsf{P}) \in \mathbb{N}(\mathrm{margin}_{\tilde{V}}(\mathrm{G}))
$$

with weight function generated by

$$
\begin{aligned}
\tilde{\sigma}(W[\tilde{V} \longrightarrow \tilde{V} \textbf{ in } \tilde{\mathrm{G}}]) &= \sigma(W[\tilde{V} \rightsquigarrow \tilde{V} \mid \tilde{V} \textbf{ in } \mathrm{G}]), \\
\tilde{\sigma}(W[\tilde{V} \longleftrightarrow \tilde{V} \textbf{ in } \tilde{\mathrm{G}}]) &= \sigma(W[\tilde{V} \overset{\mathrm{t}}{\leftrightsquigarrow} \tilde{V} \mid \tilde{V} \textbf{ in } \mathrm{G}]).
\end{aligned}
$$

- Proof of blackboard.

2.7.2 Example

- Find the marginal graph and linear system with $\tilde{V} = \{V_1, V_2, V_4, V_5\}$.

2.8 Marginalization of canonical graphs

2.8.1 Definition (marginalization of trimmed graphs)

Consider $\mathrm{G} \in \mathbb{G}^*(V)$ and $\mathrm{G}^* = \mathrm{trim}(\mathrm{G})$. For $\tilde{V} \subseteq V$, the marginal trimmed graph $\tilde{\mathrm{G}}^* = \mathrm{margin}^*_{\tilde{V}}(\mathrm{G}^*)$ is obtained by

$$
\begin{aligned}
j \longrightarrow k \textbf{ in } \tilde{\mathrm{G}}^* &\iff P[j \rightsquigarrow k \mid \tilde{V} \textbf{ in } \mathrm{G}^*] \neq \emptyset, \\
j \longleftrightarrow k \textbf{ in } \tilde{\mathrm{G}}^* &\iff P[j \leftrightsquigarrow k \mid \tilde{V} \textbf{ in } \mathrm{G}^*] \neq \emptyset.
\end{aligned}
$$

- This is the usual definition in the literature and is justified by the following commutative diagram (proof omitted; example on blackboard).

$$
\begin{tikzcd}
\mathbb{G}^*(V) \arrow[r, "\mathrm{trim}"] \arrow[d, "\mathrm{margin}_{\tilde{V}}"'] & \mathrm{trim}(\mathbb{G}^*(V)) \arrow[d, "\mathrm{margin}^*_{\tilde{V}}"] \\
\mathbb{G}^*(\tilde{V}) \arrow[r, mapsto, "\mathrm{trim}"] & \mathrm{trim}(\mathbb{G}^*(\tilde{V}))
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathrm{G} \arrow[r, mapsto, "\mathrm{trim}"] \arrow[d, mapsto, "\mathrm{margin}_{\tilde{V}}"'] & \mathrm{G}^* \arrow[d, mapsto, "\mathrm{margin}^*_{\tilde{V}}"] \\
\tilde{\mathrm{G}} \arrow[r, mapsto, "\mathrm{trim}"] & \tilde{\mathrm{G}}^*
\end{tikzcd}
$$

3 Lecture 3: m-separation and conditional independence

3.1 Conditional independence in multivariate normal

Suppose $V \sim \mathrm{N}(\mu, \Sigma)$ and Σ is positive definite.

3.1.1 Proposition

1. For $\mathcal{J} \subset [d]$ and $\mathcal{L} = [d] \setminus \mathcal{J}$, we have

$$V_{\mathcal{J}} \mid V_{\mathcal{L}} = v_{\mathcal{L}} \sim \mathrm{N}\left(\mu_{\mathcal{J}} + \Sigma_{\mathcal{J}\mathcal{L}}\Sigma_{\mathcal{L}\mathcal{L}}^{-1}(v_{\mathcal{L}} - \mu_{\mathcal{L}}),\, \Sigma_{\mathcal{J}\mathcal{J}} - \Sigma_{\mathcal{J}\mathcal{L}}\Sigma_{\mathcal{L}\mathcal{L}}^{-1}\Sigma_{\mathcal{L}\mathcal{J}}\right)$$

2. For any $j, k \in [d]$, $j \neq k$, we have

$$V_j \perp\!\!\!\perp V_k \mid V_{[d]\setminus\{j,k\}} \iff \mathsf{Cov}(V_j, V_k \mid V_{[d]\setminus\{j,k\}}) = 0 \iff (\Sigma^{-1})_{jk} = 0.$$

3.1.2 Proof sketch

1. $V_{\mathcal{J}} - \Sigma_{\mathcal{J}\mathcal{L}}\Sigma_{\mathcal{L}\mathcal{L}}^{-1}V_{\mathcal{L}}$ is independent of $V_{\mathcal{L}}$.
2. Use $\mathcal{J} = \{j, k\}$, then use the block matrix inverse formula.

3.2 Collider-connected walks

Suppose $\mathsf{P} \in \mathbb{N}^{+}(\mathrm{G})$ and $(\mathrm{Id} - \beta)$ is non-singular.

- From Lecture 2: $\Sigma = (\mathrm{Id} - \beta)^{-T}\Lambda(\mathrm{Id} - \beta)^{-1}$.
- So $\Sigma^{-1} = (\mathrm{Id} - \beta)\Lambda^{-1}(\mathrm{Id} - \beta)^{T}$. More specifically,

$$(\Sigma^{-1})_{jk} = (\Lambda^{-1})_{jk} - \sum_{m\in[d]} \beta_{jm}(\Lambda^{-1})_{mk} - \sum_{l\in[d]}(\Lambda^{-1})_{jl}\beta_{kl} + \sum_{m,l\in[d]} \beta_{jm}(\Lambda^{-1})_{ml}\beta_{kl}$$

- A sufficient condition for $(\Sigma^{-1})_{jk} = 0$ is that every RHS term vanishes.

3.2.1 Lemma

For any $V_j \neq V_k$, we have

$$W[V_j \hookrightarrow * \hookleftarrow V_k \textbf{ in } \mathrm{G}] = \emptyset \Longrightarrow V_j \perp\!\!\!\perp V_k \mid V \setminus \{V_j, V_k\} \textbf{ under } \mathsf{P}.$$

- $V_j \hookrightarrow * \hookleftarrow V_k$ means a walk from V_j to V_k where **every non-endpoint is a collider**.
- Proof on blackboard. (*Hint:* First assume $\beta = 0$.)

3.3 Towards the general case

- Notation: Write $W[\cdots] = \emptyset$ as **not** $\cdots$.

3.3.1 Goal

We would like to establish a graphical condition for $J \perp\!\!\!\perp K \mid L$.

- Because V is Gaussian, it suffices to consider $V_j \perp\!\!\!\perp V_k \mid L$ for all $V_j \in J, V_k \in K$.
- Using the last Lemma, we have (let $\tilde{V} = \{V_j, V_k\} \cup L$)

$$\textbf{not } V_j \hookrightarrow * \hookleftarrow V_k \textbf{ in } \mathrm{margin}_{\tilde{V}}(\mathrm{G}) \Longrightarrow V_j \perp\!\!\!\perp V_k \mid L \textbf{ under } \mathsf{P}.$$

3.3.2 Main problem

Can we conclude **not** $V_j \hookrightarrow * \hookleftarrow V_k$ **in** $\mathrm{margin}_{\tilde{V}}(\mathrm{G})$ **using** G **directly**?

3.4 General blocking

3.4.1 Definition

- We say a walk in $\mathrm{G} \in \mathbb{G}(V)$ is <u>(ancestrally)</u> **blocked** by $L \subseteq V$, if
  1. the walk contains a collider m such that $m \notin L$ <u>(and $m \not\rightsquigarrow L$)</u>, **or**
  2. the walk contains a non-colliding non-endpoint m such that $m \in L$.
- Let $W[V \leftrightsquigarrow * \leftrightsquigarrow V \mid L$ **in** G] collect all walks that are not blocked by L.
- We say $J, K \subseteq V$ are **m-separated** given L if **not** $J \leftrightsquigarrow * \leftrightsquigarrow K \mid L$ **in** G.

3.4.2 Remarks

- All walks are separated by 0, 1, or more colliders.
- $W[V \leftrightsquigarrow * \leftrightsquigarrow V] \neq W[V \leftrightsquigarrow * \leftrightsquigarrow V \mid \emptyset] = W[V \leftrightsquigarrow V]$.
- The literature usually uses ancestral blocking with paths (see next Proposition).

3.4.3 Example

- Find all m-separations in the running example. (*Hint:* There are 2 in total.)

3.5 Graph separation

- **t-separation**: only consider $\overset{\mathrm{t}}{\leftrightsquigarrow} * \overset{\mathrm{t}}{\leftrightsquigarrow}$ (one or more treks).
- **d-separation**: only consider walks consisting of directed edges only.

3.5.1 Lemma

For any $\mathrm{G} \in \mathbb{G}(V)$ and $J, K, L \subseteq \tilde{V} \subseteq V$, we have

$$J \overset{\mathrm{t}}{\leftrightsquigarrow} * \overset{\mathrm{t}}{\leftrightsquigarrow} K \mid L \textbf{ in } \mathrm{margin}_{\tilde{V}}(\mathrm{G}) \iff J \overset{\mathrm{t}}{\leftrightsquigarrow} * \overset{\mathrm{t}}{\leftrightsquigarrow} K \mid L \textbf{ in } \mathrm{G}\,.$$

3.5.2 Lemma

Consider $\mathrm{G} \in \mathbb{G}(V)$ and $\{j\}, \{k\}, L \subseteq V$. If G is canonical, then

$$\text{(i) } j \overset{\mathrm{t}}{\leftrightsquigarrow} * \overset{\mathrm{t}}{\leftrightsquigarrow} k \mid L \iff \text{(ii) } j \leftrightsquigarrow * \leftrightsquigarrow k \mid L \iff \text{(iii) } P[j \leftrightsquigarrow * \leftrightsquigarrow k \mid_a L] \neq \emptyset.$$

Furthermore, if G is canonically directed, then

$$\text{(i), (ii), (iii)} \iff \text{(iv) } j \overset{\mathrm{d}}{\leftrightsquigarrow} * \overset{\mathrm{d}}{\leftrightsquigarrow} k \mid L \iff \text{(v) } P[j \overset{\mathrm{d}}{\leftrightsquigarrow} * \overset{\mathrm{d}}{\leftrightsquigarrow} k \mid_a L] \neq \emptyset.$$

3.6 General result

3.6.1 Theorem

Suppose $\mathsf{P} \in \mathbb{N}^+(\mathrm{G})$ for some $\mathrm{G} \in \mathbb{G}^*(V)$ and $(\mathrm{Id} - \beta)$ is principally non-singular. Then for all disjoint $J, K, L \subseteq V$, we have

$$\textbf{not } J \leftrightsquigarrow * \leftrightsquigarrow K \mid L \textbf{ in } \mathrm{G} \Longrightarrow J \perp\!\!\!\perp K \mid L \textbf{ under } \mathsf{P}\,.$$

- Proof on blackboard using the results above.

4 Lecture 4: Linear structural equation model and identifiability

4.1 Potential outcome

4.1.1 Motivation

We can represent Hooke's law by (X force, Y compression distance, β elasticity):

$$X \longrightarrow Y \quad \text{and} \quad Y = \beta X.$$

- But equations have no direction: $Y = \beta X$ is equivalent to $X = Y/\beta$.
- To emphasize "force causes compression", we can write $Y(x) = \beta x$ for all x.

4.1.2 Definition

- Let $V(V_{\mathcal{I}} = v_{\mathcal{I}})$ (often abbreviated as $V(v_{\mathcal{I}})$) denote the **potential outcome** of the entire system under an intervention that sets $V_{\mathcal{I}}$ to $v_{\mathcal{I}}$.
- A **causal model** is a collection of probability distributions P on the **potential outcomes schedule** $V(\cdot) = (V(v_{\mathcal{I}}) : \mathcal{I} \in [d])$ such that

$$\mathsf{P}(V(v_{\mathcal{I}}, v_{\mathcal{J}}) = V(v_{\mathcal{I}}) \mid V_{\mathcal{J}}(v_{\mathcal{I}}) = v_{\mathcal{J}}) = 1, \text{ for all disjoint } \mathcal{I}, \mathcal{J} \subseteq [d],\ v \in \mathbb{V}.$$

4.2 Linear structural equation model (Linear SEM)

4.2.1 Definition

We say $V(\cdot)$ follows a **linear SEM** with respect to $\mathrm{G} \in \mathbb{G}(V)$, if there exist β and Λ compatible with G such that

$$V_j(v_{\mathcal{I}}) = \sum_{k \in \mathrm{pa}(j) \cap \mathcal{I}} \beta_{kj} v_k + \sum_{k \in \mathrm{pa}(j) \setminus \mathcal{I}} \beta_{kj} V_k(v_{\mathcal{I}}) + E_j, \text{ for all } j \in [d] \text{ and } \mathcal{I} \subseteq [d],$$

for some $E = (E_1, \ldots, E_d)$ with $\mathsf{Cov}(E) = \Lambda$ and $\mathrm{pa}_{\mathrm{G}}(j) = \{l \in [d] : V_l \longrightarrow V_j \textbf{ in } \mathrm{G}\}$.

- In words, every equation still "holds" (thus is "structural") under any intervention.

4.2.2 Example

In the running example, how do the structural equations look like under the intervention $(V_2, V_3) = (v_2, v_3)$?

4.3 Single-world intervention graphs

We can rewrite the structural equations for $V(v_{\mathcal{I}})$ in matrix form:

$$\begin{pmatrix} v_{\mathcal{I}} \\ V_{\mathcal{I}}(v_{\mathcal{I}}) \\ V_{\mathcal{I}^c}(v_{\mathcal{I}}) \end{pmatrix} = \begin{pmatrix} v_{\mathcal{I}} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ \beta^T_{\mathcal{I},\mathcal{I}} & 0 & \beta^T_{\mathcal{I}^c,\mathcal{I}} \\ \beta^T_{\mathcal{I},\mathcal{I}^c} & 0 & \beta^T_{\mathcal{I}^c,\mathcal{I}^c} \end{pmatrix} \begin{pmatrix} v_{\mathcal{I}} \\ V_{\mathcal{I}}(v_{\mathcal{I}}) \\ V_{\mathcal{I}^c}(v_{\mathcal{I}}) \end{pmatrix} + \begin{pmatrix} 0 \\ E_{\mathcal{I}} \\ E_{\mathcal{I}^c} \end{pmatrix}.$$

So $(v_{\mathcal{I}}, V(v_{\mathcal{I}}))$ follows a linear system with respect to the **single-world intervention graph (SWIG)** $\mathrm{G}(v_{\mathcal{I}})$ obtained by modifying G as follows:

1. each intervened vertex $i \in \mathcal{I}$ is split into two vertices, $V_i(v_{\mathcal{I}})$ and v_i, and each non-intervened vertex $j \notin \mathcal{I}$ is relabeled as $V_j(v_{\mathcal{I}})$.
2. the "random" vertex $V_i(v_{\mathcal{I}})$ inherits all "incoming" edges of V_i (edges like $* \longrightarrow V_i$ or $* \longleftrightarrow V_i$) in G;
3. the "fixed" vertex v_i inherits all "outgoing" edges of V_i (edges like $V_i \longrightarrow *$) in G.

(Example on blackboard.)

4.4 Causal effects in linear SEM

Due to linearity, we can define the **(joint) causal effect** of $V_{\mathcal{I}}$ on V as the matrix

$$D_{\mathcal{I}} = \{\nabla_{v_{\mathcal{I}}} V(v_{\mathcal{I}})\}^T \text{ with entries } D_{ij} = \frac{\partial}{\partial v_i} V_j(v_{\mathcal{I}}),\ i \in \mathcal{I}, j \in [d].$$

4.4.1 Theorem

Suppose P is a linear SEM with respect to $\mathrm{G} \in \mathbb{G}(V)$ and β is principally stable. Then

$$\{\nabla_{v_{\mathcal{I}}} V(v_{\mathcal{I}})\}^T = \sigma(W[V_{\mathcal{I}} \rightsquigarrow V \mid V_{\mathcal{I}} \textbf{ in } \mathrm{G}]).$$

Thus if G is acyclic, the total causal effect of V_i on V_j is

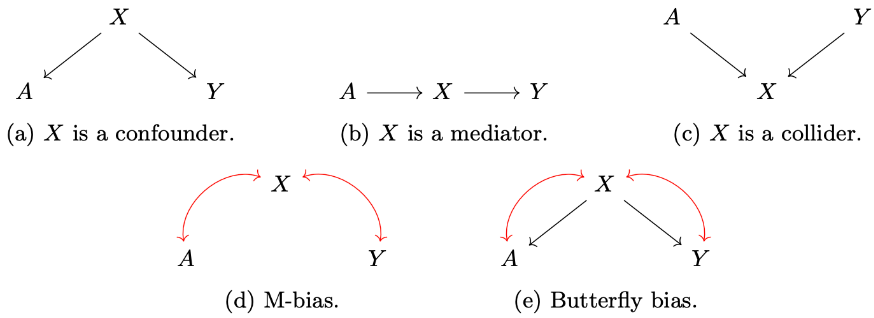
$$\frac{\mathrm{d}V_j(v_i)}{\mathrm{d}v_i} = \sigma(P[V_i \rightsquigarrow V_j]).$$

- Proof on blackboard.

4.5 Correlation is not causation

- Statistical dependence: $\longleftrightarrow$ (local), $\overset{\mathrm{t}}{\leftrightsquigarrow}$ (global), t/m-connection (conditional).
- Causal dependence: $\longrightarrow$ (local), $\rightsquigarrow$ (global).

4.5.1 Examples



(a) X is a confounder. (b) X is a mediator. (c) X is a collider. (d) M-bias. (e) Butterfly bias.

Assuming all variables are Gaussian and have unit variance, find $\mathsf{Cov}(A, Y)$, $\mathsf{Cov}(A, Y \mid X)$ and the causal effect of A on Y.

4.6 Identifiability

A central question of causal inference is identifiability. In linear models, this is asking whether the following map is injective:

$$\Sigma : (\beta, \Lambda) \mapsto (\mathrm{Id} - \beta)^{-T} \Lambda (\mathrm{Id} - \beta)^{-1}.$$

We say $\mathbb{N}^+(\mathrm{G})$ is

- **globally identifiable** if $\Sigma^{-1}(\Sigma(\beta, \Lambda))$ is a singleton for all (β, Λ);
- **generically identifiable** if $\Sigma^{-1}(\Sigma(\beta, \Lambda))$ is a singleton for almost all (β, Λ);
- **locally identifiable** if $\Sigma^{-1}(\Sigma(\beta, \Lambda))$ does not contain an open neighborhood of (β, Λ) for all (β, Λ).

4.7 Examples

1. Instrumental variables.

$$\begin{tikzcd} Z \arrow[r] & A \arrow[r] \arrow[r, bend left, leftrightarrow, red] & Y \end{tikzcd}$$

2. Factor analysis (U is not observed).

$$\begin{tikzcd} & & U \arrow[dll] \arrow[dl] \arrow[dr] & \\ V_1 & V_2 & \cdots & V_d \end{tikzcd}$$

3. Double negative controls (U is not observed).

$$\begin{tikzcd} & & U \arrow[dll] \arrow[dl] \arrow[dr] \arrow[drr] & & \\ Z \arrow[r] & A \arrow[rr] & & Y & W \arrow[l] \end{tikzcd}$$

5 Lecture 5: Conditional independence and undirected graphical models

5.1 Conditional independence

Let $\mathbb{P}(\mathbb{V})$ denote all absolutely continuous probability distributions wrt $\mathbb{V}$ (usually $\mathbb{R}^d$). Let p denote the density function of $\mathsf{P} \in \mathbb{P}(\mathbb{V})$. Consider disjoint $V_\mathcal{J}, V_\mathcal{K}, V_\mathcal{L} \subseteq V$.

5.1.1 Definition

- The **conditional density function** of $V_\mathcal{J}$ given $V_\mathcal{K}$ is given by

$$\mathsf{p}(V_\mathcal{J} = v_\mathcal{J} \mid V_\mathcal{K} = v_\mathcal{K}) = \frac{\mathsf{p}(V_\mathcal{J} = v_\mathcal{J}, V_\mathcal{K} = v_\mathcal{K})}{\mathsf{p}(V_\mathcal{K} = v_\mathcal{K})},$$

  which is well defined at any value $v_\mathcal{K}$ such that $\mathsf{p}(V_\mathcal{K} = v_\mathcal{K}) > 0$. We often abbreviate this as $\mathsf{p}(v_\mathcal{J} \mid v_\mathcal{K})$.

- We write $V_\mathcal{J} \perp\!\!\!\perp V_\mathcal{K} \mid V_\mathcal{L}$ **under** P if one of the next equivalent conditions hold:

  1. $\mathsf{p}(v_\mathcal{J}, v_\mathcal{K} \mid v_\mathcal{L}) = \mathsf{p}(v_\mathcal{J} \mid v_\mathcal{L})\, \mathsf{p}(v_\mathcal{K} \mid v_\mathcal{L})$.
  2. $\mathsf{p}(v_\mathcal{J} \mid v_\mathcal{L}, v_\mathcal{K}) = \mathsf{p}(v_\mathcal{J} \mid v_\mathcal{L})$.
  3. $\log \mathsf{p}(v_\mathcal{J}, v_\mathcal{K}, v_\mathcal{L}) = g_{\mathcal{J},\mathcal{K}}(v_\mathcal{J}, v_\mathcal{K}) + g_{\mathcal{K},\mathcal{L}}(v_\mathcal{K}, v_\mathcal{L})$ for some $g_{\mathcal{J},\mathcal{K}}$ and $g_{\mathcal{K},\mathcal{L}}$.

5.2 Graphoid axioms

5.2.1 Proposition

Consider $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ and disjoint subvectors $J, K, L, M \subseteq V$. We have

Symmetry $(J \perp\!\!\!\perp K \mid L) \iff (K \perp\!\!\!\perp J \mid L)$;

Chain rule $(J \perp\!\!\!\perp K \mid L, M)$ and $(J \perp\!\!\!\perp M \mid L) \iff (J \perp\!\!\!\perp K, M \mid L)$.

If $\mathsf{p}(v) > 0$ for all v, then we also have

Intersection $(J \perp\!\!\!\perp K \mid L, M)$ and $(J \perp\!\!\!\perp M \mid K, L) \implies (J \perp\!\!\!\perp K, M \mid L)$.

(Proof is left as an exercise.)

- A ternary relation that satisfy these axioms is called a **graphoid**. The terminology is justified by the following visualization.

<table>
<tr><td rowspan="2">J</td><td rowspan="2">L</td><td>K</td></tr>
<tr><td>M</td></tr>
</table>

5.3 Separation in undirected graphs

5.3.1 Definition

- Let $\mathbb{UG}(V)$ denote the collection of all simple undirected graphs with vertex set V.
- Given $\mathrm{G} \in \mathbb{UG}(V)$ and disjoint subsets $J, K, L \subset V$, we say L **separate** J and K in G and write

$$\textbf{not } J \longrightarrow * \longrightarrow K \mid L \textbf{ in } \mathrm{G}$$

  if every path from a vertex in J to a vertex in K in G contains a non-endpoint in L.

5.3.2 Interpretation

This is "dual" to separation in bidirected graphs:

- $J \longrightarrow L \longrightarrow K,\ J \longleftrightarrow M \longleftrightarrow K$ are blocked given L;
- $J \longrightarrow M \longrightarrow K,\ J \longleftrightarrow L \longleftrightarrow K$ are not blocked given L.

5.4 Undirected graphical models

- Let $\mathbb{P}_{\mathrm{GM}}(G)$ collects all $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ that satisfies the **global Markov property** wrt $\mathrm{G} \in \mathbb{UG}(V)$: for all disjoint $J, K, L \subset V$,

$$\textbf{not } J \longrightarrow * \longrightarrow K \mid L \textbf{ in } \mathrm{G} \Longrightarrow J \perp\!\!\!\perp K \mid L \textbf{ under } \mathsf{P}\,.$$

- Let $\mathbb{P}_{\mathrm{F}}(G)$ collects all $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ that **factorizes** wrt $\mathrm{G} \in \mathbb{UG}(V)$:

$$\mathsf{p}(v) = \prod_{V_{\mathcal{J}} \in \mathcal{C}(\mathrm{G})} f_{\mathcal{J}}(v_{\mathcal{J}}),$$

  for some $f_{\mathcal{J}}$, $\mathcal{J} \subseteq [d]$, where $\mathcal{C}(\mathrm{G})$ collects all "cliques" (complete subgraphs) of G.

5.4.1 Theorem (Hammersley-Clifford)

For any $\mathrm{G} \in \mathbb{UG}(V)$, we have $\mathbb{P}_{\mathrm{F}}(\mathrm{G}) \subseteq \mathbb{P}_{\mathrm{GM}}(\mathrm{G})$ and $\mathbb{P}^{+}_{\mathrm{F}}(\mathrm{G}) = \mathbb{P}^{+}_{\mathrm{GM}}(\mathrm{G})$.

- Proof of the first part. ($\mathbb{P}^{+}$ means positive density functions.)

5.5 Graph augmentation

Graph separations in undirected and directed graphs are closely related via the augmentation map $\mathrm{aug} : \mathbb{G}^*(V) \to \mathbb{UG}(V)$ defined by

$$V_j \longrightarrow V_k \textbf{ in } \mathrm{aug}(\mathrm{G}) \Longleftrightarrow V_j \hookrightarrow * \hookleftarrow V_k \textbf{ in } \mathrm{G},\ \text{for all } V_j \neq V_k.$$

- When restricted to $\mathbb{G}^*_{\mathrm{DA}}(V)$, this is called **moralization** in the literature because it connects any two parents of the same child.
- For $J \subseteq V$, define $\mathrm{an}(J) = \{V_k \in V : V_k \rightsquigarrow J \textbf{ in } \mathrm{G}\}$ and $\overline{\mathrm{an}}(J) = \mathrm{an}(J) \cup J$ (the smallest ancestral set containing J).

5.5.1 Proposition

For any $\mathrm{G} \in \mathbb{G}^*(V)$ and disjoint $J, K, L \subset V$, we have, with $\tilde{V} = \overline{\mathrm{an}}(J \cup K \cup L)$,

$$J \rightsquigarrow * \leftsquigarrow K \mid L \textbf{ in } \mathrm{G} \Longleftrightarrow J \longrightarrow * \longrightarrow K \mid L \textbf{ in } \mathrm{aug} \circ \mathrm{margin}_{\tilde{V}}(\mathrm{G}).$$

- Proof on blackboard.

6 Lecture 6: DAG models and ADMG models

6.1 DAG models

6.1.1 Definition

- Let $\mathbb{P}_{\mathrm{F}}(G)$ collects all $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ that **factorizes** wrt $\mathrm{G} \in \mathbb{G}^*_{\mathrm{DA}}(V)$:

$$\mathsf{p}(v) = \prod_{j=1}^{p} \mathsf{p}(v_j \mid v_{\mathrm{pa}_{\mathrm{G}}(j)}).$$

- Let $\mathbb{P}_{\mathrm{GM}}(G)$ collects all $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ that satisfies the **global Markov property** wrt $\mathrm{G} \in \mathbb{G}^*_{\mathrm{DA}}(V)$: for any disjoint subsets $J, K, L \subset V$,

$$\textbf{not } J \overset{\mathrm{d}}{\rightsquigarrow} * \overset{\mathrm{d}}{\leftsquigarrow} K \mid L \textbf{ in } \mathrm{G} \Longrightarrow J \perp\!\!\!\perp K \mid L \textbf{ under } \mathsf{P}.$$

6.1.2 Theorem

For any $\mathrm{G} \in \mathbb{G}^*_{\mathrm{DA}}(V)$, we have $\mathbb{P}_{\mathrm{F}}(\mathrm{G}) = \mathbb{P}_{\mathrm{GM}}(\mathrm{G})$.

- Proof on blackboard.

6.2 ADMG models

- Let $\mathbb{P}_{\mathrm{GM}}(G)$ collects all $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ that satisfies the **global Markov property** wrt $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$: for any disjoint subsets $J, K, L \subset V$,

$$\textbf{not } J \rightsquigarrow * \leftsquigarrow K \mid L \textbf{ in } \mathrm{G} \Longrightarrow J \perp\!\!\!\perp K \mid L \textbf{ under } \mathsf{P}.$$

- Alternatively, we can define ADMG models by using simpler expanded graphs.

6.2.1 Graph expansion

- We say G' is an **expansion** of $\mathrm{G} \in \mathrm{G}^*(V)$ if it is in

$$\mathrm{expand}(\mathrm{G}) = \mathrm{margin}_V^{-1}(\mathrm{G}) = \bigcup_{V' \supseteq V} \{\mathrm{G}' \in \mathbb{G}^*(V') : \mathrm{margin}_V(\mathrm{G}') = \mathrm{G}\}.$$

- Often, bidirected edges in G correspond to certain latent variables in G'.
- If we are satisfied with an expansion G' of G, we can use $\mathrm{margin}_V(\mathbb{P}_{\mathrm{GM}}(G'))$ as the model for G.

6.3 ADMG models

- **Pairwise expansion**: replace every $V_j \longleftrightarrow V_k$ with $V_j \longleftarrow U_{jk} \longrightarrow V_k$.
- **Clique expansion**: replace every bidirected clique $V_{\mathcal{J}}$ (complete bidirected subgraph) with directed edges $U_{\mathcal{J}} \longrightarrow V_j$, $j \in \mathcal{J}$.
- **Noise expansion**: add $U_j \longrightarrow V_j$ such that U_j inherits all bidirected edges of V_j.
- Example on blackboard.

Let the corresponding models be denoted as $\mathbb{P}_{\mathrm{PE}}(\mathrm{G})$, $\mathbb{P}_{\mathrm{CE}}(\mathrm{G})$, $\mathbb{P}_{\mathrm{NE}}(\mathrm{G})$.

6.3.1 Proposition

For any $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$, we have

$$\mathbb{P}_{\mathrm{PE}}(\mathrm{G}) \subseteq \mathbb{P}_{\mathrm{CE}}(\mathrm{G}) \subseteq \mathbb{P}_{\mathrm{NE}}(\mathrm{G}) \subseteq \mathbb{P}_{\mathrm{GM}}(\mathrm{G}),$$

and $\subseteq$ in above cannot be replaced by $=$ in general.

- The latent variable models (PE/CE/NE) have additional equality and inequality constraints.

6.4 Additional equality constraints

Consider the following (trimmed) ADMG and its expansion (U is latent).

$$\begin{tikzcd} A_1 \arrow[r] \arrow[rr, bend right] & L \arrow[r] \arrow[rr, leftrightarrow, bend left] & A_2 \arrow[r] & Y \end{tikzcd} \qquad \begin{tikzcd} & & U \arrow[dl] \arrow[dr] & \\ A_1 \arrow[r] \arrow[rr, bend right] & L \arrow[r] & A_2 \arrow[r] & Y \end{tikzcd}$$

Suppose $\mathsf{P} \in \mathbb{P}_{\mathrm{PE}}(\mathrm{G})$ (PE,CE,NE are actually equivalent in this example). Then

$$\int \mathsf{p}(y \mid a_1, l, a_2)\, \mathsf{p}(l \mid a_1)\, \mathrm{d}l \text{ does not depend on } a_1.$$

- Proof on blackboard.
- This can be understood as a "hidden" independence $Y \perp\!\!\!\perp A_1$ in the kernel

$$\mathsf{p}(a_1, l, y \mid \mathrm{do}(a_2)) = \frac{\mathsf{p}(a_1, l, a_2, y)}{\mathsf{p}(a_2 \mid a_1, l)} = \mathsf{p}(a_1)\, \mathsf{p}(l \mid a_1)\, \mathsf{p}(y \mid a_1, l, a_2).$$

6.5 Additional inequality constraints

Consider the following bidirected clique and its pairwise expansion.

$$\begin{tikzcd} & V_1 \arrow[dl, leftrightarrow] \arrow[dr, leftrightarrow] & \\ V_2 \arrow[rr, leftrightarrow] & & V_3 \end{tikzcd} \qquad \begin{tikzcd} U_{12} \arrow[r] \arrow[d] & V_1 & U_{13} \arrow[l] \arrow[d] \\ V_2 & U_{23} \arrow[l] \arrow[r] & V_3 \end{tikzcd}$$

If $\mathsf{P} \in \mathbb{P}_{\mathrm{PE}}(\mathrm{G})$ and $\mathbb{V} = \{-1, 1\}^3$, the following "perfect correlation" is impossible:

$$\mathsf{P}(V_1 = V_2 = V_3 = 1) = \mathsf{P}(V_1 = V_2 = V_3 = -1) = \frac{1}{2}.$$

- Heuristically, if $V_1 = V_2$, then V_1 cannot depend on U_{13}.
- $\mathbb{P}_{\mathrm{CE}}(\mathrm{G})$ or $\mathbb{P}_{\mathrm{NE}}(\mathrm{G})$ have no such constraints.
- Other related examples: Bell's inequality in quantum mechanics; Balke-Pearl bound for instrumental variable graph.

6.6 Advantages of the noise expansion model

1. It is equivalent to a **natural nonparametric generalization** of the linear SEM: if $\mathsf{P} \in \mathbb{P}_{\mathrm{NE}}(\mathrm{G})$, then V satisfies
$$V_j = f_j(V_{\mathrm{pa_G}(j)}, E_j)$$
for some functions $f_1, \ldots, f_d$ and noise variables $E_1, \ldots, E_d$ that satisfy
$$V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathrm{G} \Longrightarrow E_{\mathcal{J}} \perp\!\!\!\perp E_{\mathcal{K}}.$$

2. Let $\mathbb{G}^*_{\mathrm{UA}}(V)$ collects all **unconfounded ADMGs** ($V_j \longleftrightarrow V_k$ **in** G, $V_j \neq V_k$ implies $\mathrm{pa_G}(j) = \emptyset$) with vertex set V. Then
   - For all $\mathrm{G} \in \mathbb{G}^*_{\mathrm{UA}}(V)$, we have $\mathbb{P}_{\mathrm{NE}}(\mathrm{G}) = \mathbb{P}_{\mathrm{GM}}(\mathrm{G})$.
   - For all $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$, we have
   $$\mathbb{P}_{\mathrm{NE}}(\mathrm{G}) = \bigcup_{V' \supseteq V} \bigcup_{\mathrm{G}'} \mathrm{margin}_V(\mathbb{P}_{\mathrm{NE}}(\mathrm{G}')).$$
   (The second union is over $\mathrm{G}' \in \mathrm{expand}(\mathrm{G}) \cap \mathbb{G}^*_{\mathrm{UA}}(V')$.)

7 Lecture 7: Causal Markov model

7.1 Causal Markov model

Recall that a causal model is a collection of consistent probability distributions on the potential outcomes schedule $V(\cdot)$.

7.1.1 Definition

Let $\mathbb{CP}(\mathrm{G})$ collect all distributions P on $V(\cdot)$ that is **causal Markov** wrt $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$:

1. Recursive substitution: With P-probability 1, we have
$$V_j(v_{\mathcal{I}}) = V_j(v_{\mathrm{pa_G}(j)\cap\mathcal{I}}, V_{\mathrm{pa_G}(j)\setminus\mathcal{I}}(v_{\mathcal{I}})) \text{ for all } j \in [d], \mathcal{I} \subseteq [d], v \in \mathbb{V}.$$

2. Basic potential outcomes are Markov wrt bidirected subgraph:
$$V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathrm{G} \Longrightarrow V_{\mathcal{J}}(v) \perp\!\!\!\perp V_{\mathcal{K}}(v) \textbf{ under } \mathsf{P} \text{ for all disjoint } \mathcal{J}, \mathcal{K} \subset [d] \text{ and } v \in \mathbb{V}.$$

Example: What is $Y(a_1, a_2)$? What is $Y(a_1)$?

$$\begin{tikzcd} A_1 \arrow[r] \arrow[rr, bend right] & L \arrow[r] \arrow[rr, leftrightarrow, bend left] & A_2 \arrow[r] & Y \end{tikzcd}$$

7.2 Properties

Suppose $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ and $\mathsf{P} \in \mathbb{CP}(\mathrm{G})$.

7.2.1 Property 1 (Consistency of potential outcomes)

$\mathsf{P}(V(v_{\mathcal{I}}, v_{\mathcal{J}}) = V(v_{\mathcal{I}}) \mid V_{\mathcal{J}}(v_{\mathcal{I}}) = v_{\mathcal{J}}) = 1$, for all disjoint $\mathcal{I}, \mathcal{J} \subseteq [d]$, $v \in \mathbb{V}$.

7.2.2 Property 2 (Simplifying potential outcomes)

For any $V_{\mathcal{J}}, V_{\mathcal{K}}, V_{\mathcal{L}} \subseteq V$, $V_{\mathcal{K}} \cap V_{\mathcal{L}} = \emptyset$, we have

$$\textbf{not } V_{\mathcal{L}} \rightsquigarrow V_{\mathcal{J}} \mid V_{\mathcal{K}} \textbf{ in } \mathrm{G} \Longrightarrow \mathsf{P}(V_{\mathcal{J}}(v_{\mathcal{K}}, v_{\mathcal{L}}) = V_{\mathcal{J}}(v_{\mathcal{K}})) = 1, \text{ for all } v_{\mathcal{K}} \in \mathbb{V}_{\mathcal{K}}, v_{\mathcal{L}} \in \mathbb{V}_{\mathcal{L}}.$$

7.2.3 Property 3 (SWIG Markov property)

We have $\mathrm{margin}_{V(v_{\mathcal{I}})}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{G}(v_{\mathcal{I}}))$ for all $V_{\mathcal{I}} \subseteq V$ and $v \in \mathbb{V}$.

- Proof on blackboard.

8 Lecture 8: Causal identification and confounder selection

8.1 Identification by fixing

Consider $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$.

- We say $V_j \in V$ is **fixable** in G if there exists no $V_k \in V$ such that $V_j \rightsquigarrow V_k$ and $V_j \longleftrightarrow * \longleftrightarrow V_k$ **in** G.
- For $V_j \in V$, its **Markov background** in G is defined as

$$\mathrm{mbg}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \longleftrightarrow * \longleftrightarrow V_j \textbf{ in } \mathrm{G}\}.$$

8.1.1 Proposition

Consider $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ and $\mathsf{P} \in \mathbb{CP}(\mathrm{G})$. If $V_j \in V$ is fixable in G, then

$$\frac{\mathsf{p}(V_j(v_j) = \tilde{v}_j, V_{-j}(v_j) = v_{-j})}{\mathsf{p}(V_j = v_j, V_{-j} = v_{-j})} = \frac{\mathsf{p}(V_j = \tilde{v}_j \mid V_{\mathrm{mbg}(j)} = v_{\mathrm{mbg}(j)})}{\mathsf{p}(V_j = v_j \mid V_{\mathrm{mbg}(j)} = v_{\mathrm{mbg}(j)})}, \text{ for all } v \in \mathbb{V} \text{ and } v^*_j \in \mathbb{V}_j,$$

whenever $\mathsf{p}(V_j = v_j \mid V_{\mathrm{mbg}(j)} = v_{\mathrm{mbg}(j)}) > 0$.

- Proof on blackboard.

8.2 Example

$$\begin{tikzcd} A_1 \arrow[r] \arrow[rr, bend right] & L \arrow[r] \arrow[rr, leftrightarrow, bend left] & A_2 \arrow[r] & Y \end{tikzcd}$$

Show that the equality constraint

$$\int \mathsf{p}(y \mid a_1, l, a_2)\, \mathsf{p}(l \mid a_1)\, \mathrm{d}l \text{ does not depend on } a_1.$$

corresponds to

- the independence $Y(a_2) \perp\!\!\!\perp A_1$; or
- no direct $A_1 \longrightarrow Y$ effect: $Y(a_1, a_2) = Y(a_2)$.

8.3 Back-door criterion

Consider $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$, $\mathsf{P} \in \mathbb{CP}(\mathrm{G})$, $A, Y \in V$. Interested in the causal effect of A on Y.

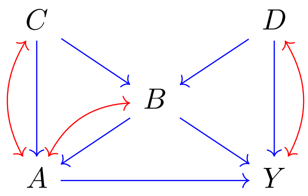
8.3.1 Theorem

Suppose $X \subset V$, $X \cap \{A, Y\} = \emptyset$ satisfies

1. $A \not\rightsquigarrow X$ **in** G;
2. $P[A \leftrightsquigarrow * \leftrightsquigarrow Y \mid_a X] = \emptyset$.

Then $\mathsf{p}(Y(a) = y \mid X = x) = \mathsf{p}(Y = y \mid A = a, X = x)$.

- Proof on blackboard.
- Example: Which $X \subseteq \{B, C, D\}$ meet the back-door criterion?



8.4 Confounder selection

Can we select a set of confounders X without knowing the full graph G?

8.4.1 Definition (symmetric back-door criterion)

- $X \subseteq V \setminus \{A, Y\}$ is an **adjustment set** for $A, Y \in V$ if $A \not\rightsquigarrow X$ and $Y \not\rightsquigarrow X$.
- An adjustment set X is **sufficient** if $P[A \leftrightsquigarrow * \leftrightsquigarrow Y \mid_a X] = \emptyset$.
- An adjustment set X is **primary** if $P[A \leftrightsquigarrow Y \mid X] = \emptyset$.

8.4.2 Heuristics

Directly blocking all confounding paths is difficult, because

$$P[A \leftrightsquigarrow * \leftrightsquigarrow Y \mid_a X] = \emptyset \not\Rightarrow P[A \leftrightsquigarrow * \leftrightsquigarrow Y \mid_a \tilde{X}] = \emptyset \text{ for } X \subset \tilde{X}.$$

But we can block confounding arcs recursively, because

$$P[A \leftrightsquigarrow Y \mid X] = \emptyset \Rightarrow P[A \leftrightsquigarrow Y \mid X'] = \emptyset \text{ for } X \subset X'.$$

8.5 District criterion

8.5.1 Theorem (marginalization preserves confounding arcs and paths)

Consider $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$, distinct $A, Y \in V$, $X \subseteq V \setminus \{A, Y\}$. For any vertex set $\tilde{V}$ such that $\{A, B\} \cup C \subseteq \tilde{V} \subseteq V$, we have

$$\begin{aligned} P[A \leftrightsquigarrow Y \mid X \textbf{ in } \mathrm{G}] = \emptyset &\iff P[A \leftrightsquigarrow Y \mid X \textbf{ in } \mathrm{margin}_{\tilde{V}}(\mathrm{G})] = \emptyset, \\ P[A \leftrightsquigarrow * \leftrightsquigarrow Y \mid_a X \textbf{ in } \mathrm{G}] = \emptyset &\iff P[A \leftrightsquigarrow * \leftrightsquigarrow Y \mid_a X \textbf{ in } \mathrm{margin}_{\tilde{V}}(\mathrm{G})] = \emptyset. \end{aligned}$$

As a corollary, we have

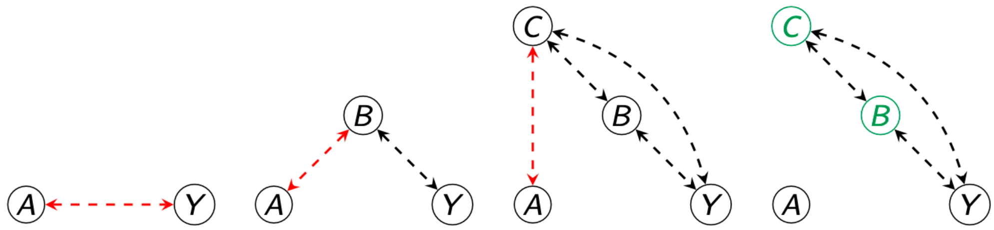
$$\begin{aligned} P[A \leftrightsquigarrow Y \mid X \textbf{ in } \mathrm{G}] = \emptyset &\iff \textbf{not } A \longleftrightarrow Y \textbf{ in } \mathrm{margin}_{\{A,Y\} \cup X}(\mathrm{G}), \\ P[A \leftrightsquigarrow * \leftrightsquigarrow Y \mid_a X \textbf{ in } \mathrm{G}] = \emptyset &\iff \textbf{not } A \longleftrightarrow * \longleftrightarrow Y \textbf{ in } \mathrm{margin}_{\{A,Y\} \cup X}(\mathrm{G}). \end{aligned}$$

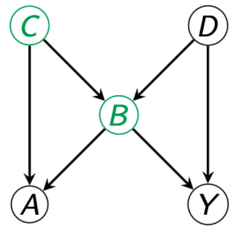
8.6 Iterative graph expansion

procedure ConfounderSelect(A, Y)
  $\mathcal{R} = \emptyset$
  procedure GraphExpand(X, $\mathcal{B}_y$, $\mathcal{B}_n$)
    if $A \longleftrightarrow * \longleftrightarrow Y$ by edges in $\mathcal{B}_y$ **then**
      return
    else if not $A \longleftrightarrow * \longleftrightarrow Y$ by edges in $(X \cup \{A, Y\}) \times (X \cup \{A, Y\}) \setminus \mathcal{B}_n$ **then**
      $\mathcal{R} = \mathcal{R} \cup \{X\}$
      return
    end if
    $(C \leftrightarrow D) =$ SelectEdge(A, Y, X, $\mathcal{B}_y$, $\mathcal{B}_n$)
    for X' in FindPrimary($C \leftrightarrow D$, X) **do**

```
            GraphExpand(X ∪ X′, B_y, B_n ∪ {C ↔ D})
        end for
        GraphExpand(X, B_y ∪ {C ↔ D}, B_n)
    end procedure
    GraphExpand(∅, ∅, ∅)
    return R
end procedure
```

8.7 Illustration





Shiny app: `https://ricguo.shinyapps.io/InteractiveConfSel/`