Multivariate Causal Models

Structural Causal Models (SCM) and Interventions

Christoph Brause

Institute of Stochastics, TU Bergakademie Freiberg

22. Mai 2024

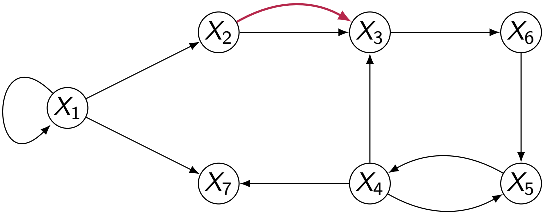
Content

Graph Terminology

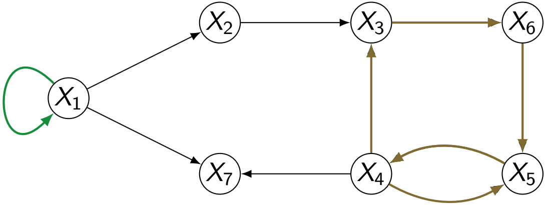
Graph theoretical terminology

Definition

A digraph $\mathcal{G} = (\boldsymbol{V}, \mathcal{E})$ consists of a set $\boldsymbol{V}$ of vertices (nodes) and a set of edges $\mathcal{E}$ with $\mathcal{E} \subseteq V \times V$.



- When working with random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, we assume $\boldsymbol{V} = \{X_1, X_2, \ldots, X_n\}$.



- A digraph without loops, i.e. edges (X_k, X_k), is a simple digraph, "graph" for short.
- A simple digraph without directed cycles is acyclic, a "DAG" for short.
- A simple digraph without directed cycles of length at least 3 is an partially directed acyclic graph, a "PDAG" for short.

Let $\mathcal{G} = (\boldsymbol{V}, \mathcal{E})$ be a digraph. The adjacency matrix $A_{\mathcal{G}} = (a_{i,j})_{i,j=1}^{d}$ is defined by

$$a_{i,j} = \begin{cases} 1 & \text{if } (X_i, X_j) \in \mathcal{E}, \\ 0 & \text{else.} \end{cases}$$

For some vertex $x_k \in \boldsymbol{V}$, let $\boldsymbol{PA}_k^{\mathcal{G}}$ and $\boldsymbol{CH}_k^{\mathcal{G}}$ be the set of parents and children of k, respectively, i.e.

$$\boldsymbol{PA}_k^{\mathcal{G}} = \{X_i \in \boldsymbol{V} : (X_i, X_k) \in \mathcal{E}\} \quad \text{and} \quad \boldsymbol{CH}_k^{\mathcal{G}} = \{X_i \in \boldsymbol{V} : (X_k, X_i) \in \mathcal{E}\}.$$

Furthermore, $\boldsymbol{AN}_k^{\mathcal{G}}$ and $\boldsymbol{DE}_k^{\mathcal{G}}$ denote the sets of ancestors and descendants of k, respectively, i.e.

$$\boldsymbol{AN}_k^{\mathcal{G}} = \{X_i \in \boldsymbol{V} : X_i = X_{j_1} \to X_{j_2} \to \ldots \to X_{j_\ell} = X_k\},$$

$$\boldsymbol{DE}_k^{\mathcal{G}} = \{X_i \in \boldsymbol{V} : X_i = X_{j_1} \leftarrow X_{j_2} \leftarrow \ldots \leftarrow X_{j_\ell} = X_k\}.$$

Theorem

If $\mathcal{G} = (\boldsymbol{V}, \mathcal{E})$ is a digraph, the following assertions are equivalent:

(a) *$\mathcal{G}$ is a DAG.*

(b) *There is a causal ordering (topological ordering), i.e. a permutation $\pi \colon [d] \to [d]$ such that $\pi(i) < \pi(j)$ for all $X_i \in \boldsymbol{V}$ and all $X_j \in \boldsymbol{DE}_i^{\mathcal{G}}$.*

(c) *For all $k \in [d]$, $\boldsymbol{AN}_k^{\mathcal{G}} \cap \boldsymbol{DE}_k^{\mathcal{G}} = \emptyset$ and $X_k \notin \boldsymbol{AN}_k^{\mathcal{G}} \cup \boldsymbol{DE}_k^{\mathcal{G}}$.*

(d) *The eigenvalues of $A_{\mathcal{G}} + \mathrm{Id}$ are real and positive.*

(e) *There is a permutation $\pi \colon [d] \to [d]$ such that $a_{\pi(i),\pi(j)} = 0$ if $i \geq j$.*

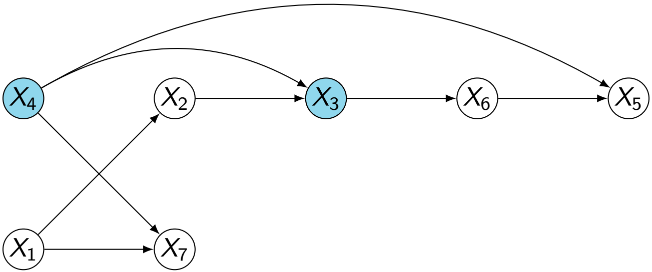
Definition

Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{S}$ be three pairwise disjoint vertex sets of $\mathcal{G}$. The sets $\boldsymbol{A}$ and $\boldsymbol{B}$ are d-separated by $\boldsymbol{S}$, denoted by

$$\boldsymbol{A} \perp\!\!\!\perp_{\mathcal{G}} \boldsymbol{B} \mid \boldsymbol{S},$$

if, for every undirected $\boldsymbol{A} - \boldsymbol{B}$-path $P : X_{i_1} X_{i_2} \ldots X_{i_k}$,

- there is a vertex $X_{i_j} \in \boldsymbol{S}$ such that
  (i) $X_{i_{j-1}} \to X_{i_j} \to X_{i_{j+1}}$ or
  (ii) $X_{i_{j-1}} \leftarrow X_{i_j} \leftarrow X_{i_{j+1}}$ or
  (iii) $X_{i_{j-1}} \leftarrow X_{i_j} \to X_{i_{j+1}}$ or
- there is a vertex $X_{i_j} \notin \boldsymbol{S}$ such that $X_{i_{j-1}} \to X_{i_j} \leftarrow X_{i_{j+1}}$ and there is no directed $X_{i_j} \leftarrow \boldsymbol{S}$-path.



- $X_1 \not\perp\!\!\!\perp_{\mathcal{G}} X_5 \mid X_3$
- $X_1 \perp\!\!\!\perp_{\mathcal{G}} X_5 \mid X_3, X_4$

Structural Causal Models

Definition

A structural causal model (SCM for short) $\mathfrak{C}$ is a pair $(\boldsymbol{S}, P_{\boldsymbol{N}})$ that consists of

- a set $\boldsymbol{S}$ of d structural assignments

$$X_j := f_j(\boldsymbol{PA}_j, N_j), \quad j \in [d]$$

  where

  - $\boldsymbol{PA}_j$ is an ℓ_j-tuple $(X_{i_1}, X_{i_2}, \dots, X_{i_{\ell_j}})$ of ℓ_j pairwise disjoint parents and
  - f_j denotes a measurable causal-effect mechanism, and

- a joint distribution $P_{\boldsymbol{N}} = P_{N_1} \times P_{N_2} \times \dots \times P_{N_d}$ with "noise" random variables $N_1, N_2 \dots, N_d$ on measurable spaces $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_d$, respectively.

Definition

The (causal) graph $\mathcal{G}$ of an SCM $\mathfrak{C}$ has vertex set $\{X_1, X_2, \ldots, X_d\}$ and edge set

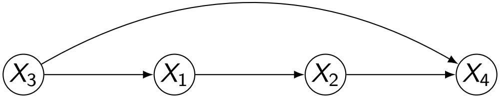
$$\{(X_i, X_j) : X_i \in \boldsymbol{PA}_j\}.$$

For $X_i \in \boldsymbol{PA}_j$, X_i is a direct cause of X_j and X_j is a direct effect of X_i.

- In what follows we assume that the causal graph of an SCM $\mathfrak{C}$ is a DAG.

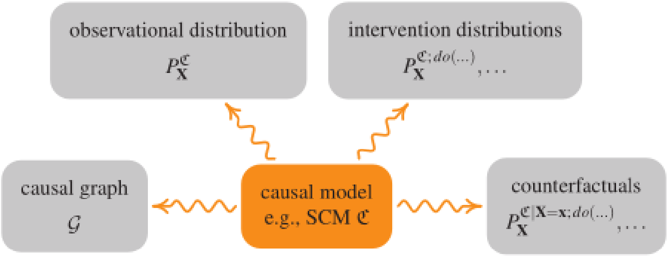
Example

$$\begin{aligned}\mathfrak{C} &:= (\{(1),(2),(3),(4)\}, P_{\mathbf{N}}) \text{ with}\\ X_1 &:= 5 \cdot X_3 + N_1 && (1)\\ X_2 &:= 3 \cdot X_1 + N_2 && (2)\\ X_3 &:= N_3 && (3)\\ X_4 &:= X_2 + X_3 + N_4 && (4)\\ P_{\mathbf{N}} &= P_{N_1} \times P_{N_2} \times P_{N_3} \times P_{N_4}\\ N_i &\sim \mathrm{N}(\mu_i, \sigma_i^2)\end{aligned}$$



Benefits

- SCMs are the key for formalizing causal reasoning and causal learning
- SCMs entail observational distribution and intervention distribution and counterfactuals.



J. Peters, D. Janzing, and B. Schölkopf, Elements of Causal Inference, MIT Press, 2017

Entailed Distribution

Proposition

An SCM $\mathfrak{C}$ yields a unique entailed distribution $P_{\mathbf{X}}^{\mathfrak{C}}$ ($P_{\mathbf{X}}$ for short).

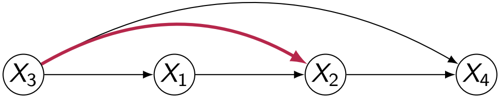
Sketch of a proof:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix} := \begin{pmatrix} f_1((f_2(N_2), f_5(N_5)), N_1) \\ f_2(N_2) \\ \vdots \\ f_d(f_2(N_2), N_1) \end{pmatrix}$$

$$P_{\mathbf{N}} = P_{N_1} \times P_{N_2} \times \ldots \times P_{N_d}$$

Structural minimality

$$
\begin{aligned}
\mathfrak{C}' &:= (\{(1),(2),(3),(4)\}, P_{\boldsymbol{N}}) \text{ with} \\
X_1 &:= 5 \cdot X_3 + N_1 && (1)\\
X_2 &:= 0 \cdot X_3 + 3 \cdot X_1 + N_2 && (2)\\
X_3 &:= N_3 && (3)\\
X_4 &:= X_2 + X_3 + N_4 && (4)\\
P_{\boldsymbol{N}} &= P_{N_1} \times P_{N_2} \times P_{N_3} \times P_{N_4} \\
N_i &\sim \mathrm{N}(\mu_i, \sigma_i^2)
\end{aligned}
$$



Definition

Let $\mathfrak{C}$ be an SCM with structural assignments $X_j := f_j(\boldsymbol{PA}_j, N_j)$, $j \in [d]$. If, for every $j \in [d]$ and every tuple $\boldsymbol{PA}_j^\star \subsetneq \boldsymbol{PA}_j$, there is no measurable function g such that

$$f_j(\boldsymbol{PA}_j, N_j) = g(\boldsymbol{PA}_j^\star, N_j) \quad \text{almost surely,}$$

then $\mathfrak{C}$ satisfies structural minimality.

Proposition

Given an SCM $\mathfrak{C}$, we can uniquely structural minimize $\mathfrak{C}$.

Convention

Given an SCM $\mathfrak{C}$, we assume that $\mathfrak{C}$ satisfies structural minimality.

- Causal minimality implies structural minimality but not (necessarily) vice versa ($\nearrow$ talk K. Bitterlich).

Sidenote – Linear SCMs whose causal graph is not a DAG

- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ and $\boldsymbol{N} = (N_1, N_2, \ldots, N_d)$. The set $\boldsymbol{S}$ of structural assignments for a linear SCM is described by

$$\boldsymbol{X} := B\boldsymbol{X} + \boldsymbol{N}$$

  for some $d \times d$-matrix B.
- If $\mathrm{Id} - B$ is invertible, then

$$\boldsymbol{X} := (\mathrm{Id} - B)^{-1}\boldsymbol{N} \tag{1}$$

  is a unique solution.
- One way to interpret (1) is to interpret it as a solution to the equilibration process

$$\boldsymbol{X}_t = B\boldsymbol{X}_{t-1} + \boldsymbol{N}$$

  with a sequence $(\boldsymbol{X}_t)$ of random variables $\boldsymbol{X}_t$, $t \geq 1$.
- The sequence $(\boldsymbol{X}_t)$ converges if $B^t \to (0)$ as $t \to \infty$.

Interventions

- An intervention is (usually) a change of (one of) the assignments in the SCM
- An intervention typically yields a different distribution – different from the unintervened distribution.

	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)

Charig et al., Comparison of treatment of renal calculi by [. . .]
British Medical Journal (Clin Res Ed), 292(6254):879–882, 1986.

What happens if the doctors force all patients to take treatment *a*?

Definition

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ be finitely random variables, and $\mathfrak{C} = (\boldsymbol{S}, P_{\boldsymbol{N}})$ and $\tilde{\mathfrak{C}} = (\tilde{\boldsymbol{S}}, \tilde{P}_{\boldsymbol{N}})$ be two SCMs on $\boldsymbol{X}$ with acyclic causal graphs.

- We say that the variables whose structural assignments differ in $\mathfrak{C}$ and $\tilde{\mathfrak{C}}$ have been intervened.
- If $X_{k_1}, \ldots, X_{k_\ell}$ denote the intervened variables, then

$$\mathfrak{C}; do(X_{k_1} := \tilde{f}_{k_1}(\widetilde{\boldsymbol{PA}}_{k_1}, \tilde{N}_{k_1}), \ldots, X_{k_\ell} := \tilde{f}_{k_\ell}(\widetilde{\boldsymbol{PA}}_{k_\ell}, \tilde{N}_{k_\ell})) := \tilde{\mathfrak{C}}.$$

- The distribution $P_{\boldsymbol{X}}^{\tilde{\mathfrak{C}}}$ is also known as intervention distribution.

- In what follows, we mainly consider $\widetilde{\boldsymbol{PA}}_{k_i} = \boldsymbol{PA}_{k_i}$ and $\widetilde{\boldsymbol{PA}}_{k_i} = ()$.

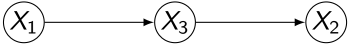
Definition

- If $\tilde{f}(\widetilde{\boldsymbol{PA}}_{k_i}, \tilde{N}_{k_i})$ sets X_{k_i} to a specific value x, then we write

$$\mathfrak{C}; do(\ldots, X_{k_i} := x, \ldots).$$

  The intervention is called atomic (hard, ideal, structural, surgical, independent, deterministic).
- If $\widetilde{\boldsymbol{PA}}_{k_i} = \boldsymbol{PA}_{k_i}$, then the intervention is called imperfect (soft, parametric, dependent, soft, mechanism change).

$$
\begin{aligned}
\mathfrak{C} &:= (\{(1),(2),(3)\}, P_{\boldsymbol{N}}) \text{ with} \\
X_1 &:= N_1 && (1) \\
X_2 &:= X_3 + N_2 && (2) \\
X_3 &:= X_1 + N_3 && (3) \\
N_1, N_3 &\sim \mathrm{N}(0,1),\ N_2 \sim \mathrm{N}(0, 0.1)
\end{aligned}
$$



- $P_{X_3}^{\mathfrak{C};do(X_2:=\tilde{N})} = \mathrm{N}(0,2) = P_{X_3}^{\mathfrak{C}}$,
  i.e. intervene on X_1 does not chance the distribution of X_3.
- $P_{X_3}^{\mathfrak{C};do(X_1:=\tilde{N})} = P_{\tilde{N}+N_3} \neq P_{N_1+N_3} = P_{X_3}^{\mathfrak{C}}$ assuming $P_{\tilde{N}} \neq P_{N_1}$,
  i.e. intervene on X_1 may chance the distribution of X_3.
- $P_{X_3}^{\mathfrak{C};do(X_2:=x)} = P_{X_3}^{\mathfrak{C}} = \mathrm{N}(0,2) \neq P_{X_3|X_2=x_2}^{\mathfrak{C}}$,
  i.e. intervention distribution may differ from conditional distribution.
- Intervening on a good predictor for a target variable may leave the target variable unaffected.

I: ice cream sales   H: heat strokes   T: temperature

$$\begin{aligned}
\mathfrak{C} &:= (\{(1), (2), (3)\}, P_{\boldsymbol{N}}) \text{ with} \\
T &:= N_T \quad (4)\\
H &:= f_H(T, N_H) \quad (5)\\
I &:= f_I(T, N_I) \quad (6)
\end{aligned}$$

- In $\mathfrak{C}; do(I := \tilde{N}_I)$, we have

$$H := f_H(N_T, N_H),$$

which indicates that there is no causal effect from I to H.

Definition

Let $\mathfrak{C}$ be an SCM. There is a total causal effect from X_1 to X_2 if and only if, for some random variable $\tilde{N}_1$ and the SCM $\mathfrak{C}' := \mathfrak{C}; do(X_1 := \tilde{N}_1)$, we have

$$X_1 \not\perp\!\!\!\perp X_2.$$

Total causal effect

Proposition

Let $\mathfrak{C}$ be an SCM with causal graph $\mathcal{G}$.

(i) *If there is no directed $X_1 \to X_2$-path in $\mathcal{G}$, then there is no total causal effect.*

(ii) *If there is a directed $X_1 \to X_2$-path in $\mathcal{G}$, then there might be no total causal effect.*

Sketch of the proof:

- Proof for (i): ↗ talk K. Bitterlich:

$$\boldsymbol{A} \perp\!\!\!\perp_{\mathcal{G}} \boldsymbol{B} \mid \boldsymbol{S} \quad \Rightarrow \quad \boldsymbol{A} \perp\!\!\!\perp \boldsymbol{B} \mid \boldsymbol{S}.$$

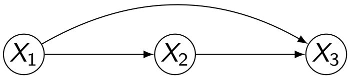
  Furthermore,

$$X_1 \perp\!\!\!\perp_{\mathcal{G}'} X_2 \mid \emptyset$$

  in the causal graph $\mathcal{G}'$ of $\mathfrak{C}; do(X_1 := \tilde{N}_1)$.

- An example for (ii):

$$
\begin{aligned}
\mathfrak{C} &:= (\{(1),(2),(3)\}, P_{\boldsymbol{N}}) \text{ with} \\
X_1 &:= N_1 && (1) \\
X_2 &:= a \cdot X_1 + N_2 && (2) \\
X_3 &:= -ab \cdot X_1 + b \cdot X_2 + N_3 && (3) \\
N_i &\sim \mathrm{N}(0, \sigma_i^2)
\end{aligned}
$$



$$X_1 \perp\!\!\!\perp X_3 \quad \text{but} \quad X_1 \not\perp\!\!\!\perp_{\mathcal{G}} X_3 \mid \emptyset$$

Alternative concepts of intervention

Proposition

Given an SCM $\mathfrak{C}$, the following assertions are equivalent:

(i) *There is a total causal effect from X_1 to X_2.*

(ii) *If $\tilde{N}_1$ is a random variable whose distribution has full support, then, for the SCM $\mathfrak{C}' := \mathfrak{C}; do(X_1 := \tilde{N}_1)$, we have*

$$X_1 \not\perp\!\!\!\perp X_2.$$

(iii) *There is an x_1 such that $P_{X_2}^{\mathfrak{C};do(X_1:=x_1)} \neq P_{X_2}^{\mathfrak{C}}$*

(iv) *There is are x_1, x_1' such that $P_{X_2}^{\mathfrak{C};do(X_1:=x_1)} \neq P_{X_2}^{\mathfrak{C};do(X_1:=x_1')}$*

In what follows, we describe an alternative approach to formalize intervention (atomic, on a single variable):

- Let $\mathfrak{C}$ be an SCM with causal graph $\mathcal{G}$.
- For each variable X_k, we insert a new variable I_k – a parentless variable with an edge to X_k only – where

$$\mathrm{Im}(I_k) = \mathrm{Im}(X_k) \cup \{\texttt{idle}\}.$$

  $I_k = \texttt{idle}$ means the variable has not been intervened and $I_k = x_k$ says that X_k is set to x_k.
- Replace $X_k := f_k(\boldsymbol{PA}_k, N_k)$ by

$$X_k := \begin{cases} f_k(\boldsymbol{PA}_k, N_k) & \text{if } I_k = \texttt{idle}, \\ I_j & \text{else} \end{cases}$$

- Add new noises $N'_1, \dots, N'_d$ such that $N_1, N'_1, \dots, N_d, N'_d$ are independent.
- For each variable I_k, add a structural assignments $I_j := f'_k(N'_k)$.

Remark

For the obtained SCM $\mathcal{C}^\star$, we have

$$P_{X_j}^{\mathfrak{C}, do(X_k := x_k)} = P_{X_j | I_k = x_k}^{\mathfrak{C}^\star}.$$