ETH zürich

Causal models and covariate adjustment I

Causality

Christina Heinze-Deml

Spring 2019

Last week

- Selection bias
- Causal effects and do-operator
- Causal graphical models
- Structural equation models

Do-operator and causal DAG models

- Mathematical notion of manipulation (see Pearl):
  - $do(X = x)$ (shorthand $do(x)$) represents a hypothetical intervention where X is set to the value x, uniformly over the entire population

- Let $G = (\boldsymbol{V}, E)$ be a DAG and p be the distribution of $X_{\boldsymbol{V}}$
- The pair (G, p) is a causal DAG model or a causal Bayesian network if for any $\boldsymbol{W} \subset \boldsymbol{V}$

$$p(x_{\boldsymbol{V}} \mid do(x_{\boldsymbol{W}} = x'_{\boldsymbol{W}})) = \prod_{i \in \boldsymbol{V} \setminus \boldsymbol{W}} p\left(x_i \middle| x_{\mathrm{pa}(i)}\right) 1\{x_{\boldsymbol{W}} = x'_{\boldsymbol{W}}\}$$

Interventions

- Surgical interventions
  - $do(X = x)$ (or shorthand $do(x)$) represents a hypothetical intervention where X is set to the value x, uniformly over the entire population

- Other intervention types possible
  - Imperfect interventions: $do(X_i = \widetilde{N}_i)$ with $\widetilde{N}_i \sim \tilde{F}$
  - Shift interventions: $X_i \leftarrow h_i\left(X_{\mathrm{pa}(i)}, \epsilon_i\right) + \widetilde{N}_i$
  - …

- Terminology not consistent in the literature

Example

- Suppose the distribution of (X, Y) is entailed by a SEM:

$$X \leftarrow N_X$$
$$Y \leftarrow 6X + N_Y$$

with $N_X, N_Y \sim \mathcal{N}(0,1)$ and DAG $X \rightarrow Y$

  - The marginal distribution of Y is $\mathcal{N}(0, 37)$.
  - The interventional distribution of $Y|do(X = 3)$ is $\mathcal{N}(18, 1)$.
  - The conditional distribution of $Y|X = 3$ is $\mathcal{N}(18, 1)$.

Intervening on X changes the distribution of Y

Here, intervening is the same as conditioning

Example

- Suppose the distribution of (X, Y) is entailed by a SEM:

$$X \leftarrow N_X$$
$$Y \leftarrow 6X + N_Y$$

with $N_X, N_Y \sim \mathcal{N}(0,1)$ and DAG $X \rightarrow Y$

  - The marginal distribution of X is $\mathcal{N}(0,1)$.
  - The interventional distribution of $X|do(Y=3)$ is $\mathcal{N}(0,1)$.
  - The conditional distribution of $X|Y=3$ is $\mathcal{N}\left(\frac{18}{37}, \frac{1}{37}\right)$.

Intervening on Y does not change the distribution of X

Intervening is <u>not</u> the same as conditioning.

Total causal effect

- The following statements are equivalent
  - There is a causal effect from X to Y.
  - There are x' and $\tilde{x}$ such that $p(y|do(X = x')) \neq p(y|do(X = \tilde{x}))$.
  - There is x' such that $p(y|do(X = x')) \neq p(y)$.
  - $X \not\perp\!\!\!\perp Y$ in $p\big(x, y \big| do\big(X = \tilde{N}_X\big)\big)$ if $\mathrm{Var}(\tilde{N}_X) > 0$.

Clicker question – Observational and interventional distributions

- Suppose the distribution of (X, Y) is entailed by a SEM:

$$X \leftarrow N_X$$
$$Y \leftarrow 1 + 2X + N_Y$$

with $N_X, N_Y \sim \mathcal{N}(0,1)$ and DAG $X \to Y$

  - The marginal distribution of Y is $\mathcal{N}(1,5)$. Yes
  - The interventional distribution of $Y|do(X = 2)$ is equal to the marginal distribution of Y. No
  - The interventional distribution of $X|do(Y = 1)$ is equal to the marginal distribution of X. Yes
  - The interventional distribution of $Y|\mathrm{do}(X = 2)$ is equal to the conditional distribution of $Y|X = 2$. Yes

Today

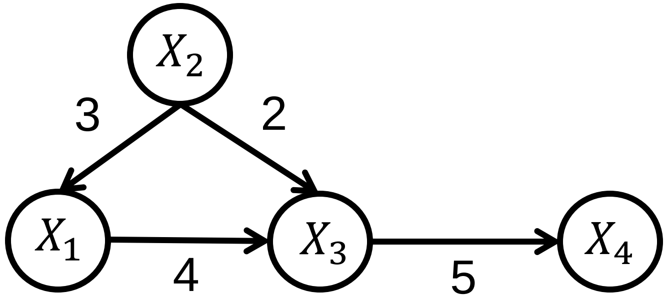
- Interventions
- Total causal effect definitions
- Path method
- Covariate adjustment – part 1

Course outline

- Background and framework
- Using the known causal graph structure to identify and estimate causal effects
- Causal structure learning

Linear structural equation models

- Linear SEMs: all structural equations are linear and the noise is additive
- Example:
  - $X_1 \leftarrow 3X_2 + \varepsilon_1$
  - $X_2 \leftarrow \varepsilon_2$
  - $X_3 \leftarrow 4X_1 + 3X_2 + \varepsilon_3$
  - $X_4 \leftarrow 5X_3 + \varepsilon_4$



- What is the total average causal effect of X_1 on X_4?
  - Increasing X_1 by 1 will on average increase X_3 by 4*1=4.
  - Increasing X_3 by 4 will on average increase X_4 by 4*5=20.

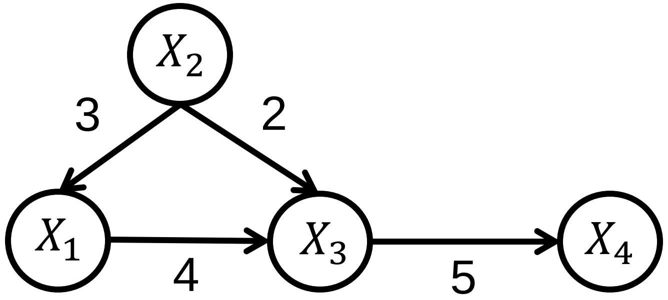
Causal effects in linear SEMs via the path method

- Path method to compute the total causal effect of X_i on X_j in a linear SEM:
  - For each directed path from X_i to X_j, multiply the edge weights along the path
  - Sum up the results over all paths

- See R and note that it matches up with simulating from interventional distributions

Linear structural equation models (linear SEMs)

- Linear SEMs: all structural equations are linear and the noise is additive
- Example:
  - $X_1 \leftarrow 3X_2 + \varepsilon_1$
  - $X_2 \leftarrow \varepsilon_2$
  - $X_3 \leftarrow 4X_1 + 3X_2 + \varepsilon_3$
  - $X_4 \leftarrow 5X_3 + \varepsilon_4$



- We saw in R that the total causal effect of X_1 on X_4 can be estimated by the coefficient of X_1 in $\mathrm{lm}(X_4 \sim X_1 + X_2)$
- In other words, we had to adjust for X_2

Determining adjustment sets

$$\begin{tikzcd} & \text{Gender } G \arrow[dl] \arrow[dr] & \\ \text{Treatment } T \arrow[rr] & & \text{Outcome } Y \end{tikzcd}$$

$$P(Y \mid do(T=1)) = \sum_g P(G=g)\, P(Y \mid G=g, T=1)$$

$\rightarrow$ have to adjust for G

Determining adjustment sets

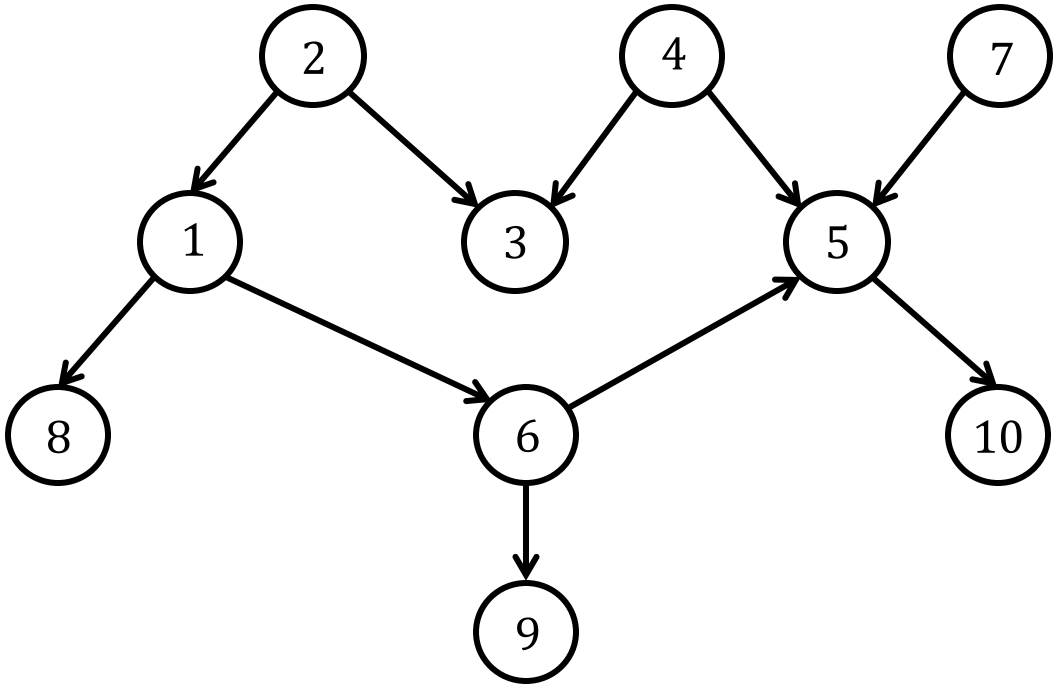
- Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a causal Bayesian network, $(i, k) \in \boldsymbol{V}, i \neq k$

- Can we find a graphical criterion for sets $\boldsymbol{Z} \subset \boldsymbol{V}$ that satisfy
$$p(x_k | do(x_i)) = \int_{x_{\boldsymbol{Z}}} p(x_k | x_i, x_{\boldsymbol{Z}}) p(x_{\boldsymbol{Z}}) dx_{\boldsymbol{Z}} \qquad (1)$$
for all $p(\cdot)$ such that (G, p) is a causal Bayesian network?

- Eq. (1) is called the adjustment formula
- Sets $\boldsymbol{Z}$ satisfying Eq. (1) are called valid adjustment sets

Determining adjustment sets

- How can we determine valid adjustment sets?
  - Should we adjust for as many variables as possible?
  - Is it always safe to adjust for "pre-treatment" variables?



No direct causes

- Let (G, p) be a causal Bayesian network
- Assume there are no edges into X_i, i.e., $\mathrm{pa}(i) = \emptyset$
- Then it follows from the truncated factorization formula that:

$$p\left(x_{\boldsymbol{V}\setminus\{i\}} \middle| do(x_i{}')\right) = \prod_{j \in \boldsymbol{V}\setminus\{i\}} p\left(x_j \middle| x_{\mathrm{pa}(j)}\right)\Big|_{x_i'} = \frac{p(x_{\boldsymbol{V}})}{p(x_i)}\Big|_{x_i'} = p\left(x_{\boldsymbol{V}\setminus\{i\}} \middle| x_i'\right)$$

- In this special case do-operator is the same as regular conditioning
- For any $k \neq i$, integrating out the other variables yields

$$p(x_k | do(x_i{}')) = p(x_k | x_i{}')$$

No direct causes

- Note: The situation with $\mathrm{pa}(i) = \emptyset$ arises in a randomized controlled trial where treatment $T = X_i$ is randomized. This shows why causal inference is straightforward in this setting.

Observational

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ T \arrow[rr] & & Y \end{tikzcd}$$

Randomized

$$\begin{tikzcd} & Z \arrow[dr] & \\ T \arrow[rr] & & Y \end{tikzcd}$$

↳ in this graph do(T) is the same as conditioning on T

In a randomized trial, treatment assignment T is determined by a coin toss - effectively erasing the arrow from Z to T

Reweighting

- Let (G, p) be a causal Bayesian network
- Then it follows from the truncated factorization formula that:

$$p\left(x_{\boldsymbol{V}\setminus\{i\}} \middle| do(x_i')\right) = \prod_{j\in\boldsymbol{V}\setminus\{i\}} p\left(x_j \middle| x_{\mathrm{pa}(j)}\right)\Big|_{x_i'} = \frac{p(x_{\boldsymbol{V}})}{p\left(x_i \middle| x_{\mathrm{pa}(i)}\right)}\Bigg|_{x_i'}$$

- Thus, the interventional distribution is a re-weighted version of the observational distribution, using weights $1/p(x_i|x_{\mathrm{pa}(i)})$
- This is used in inverse probability weighting (IPW) in marginal structural models (Robins, Hernan, Brumback).

Adjusting for direct causes

- Let (G, p) be a causal Bayesian network
- Rewriting the formula from the previous slide yields:

$$p\left(x_{\boldsymbol{V}\setminus\{i\}} \middle| do(x_i)\right) = \frac{p(x_{\boldsymbol{V}})}{p\left(x_i \middle| x_{\mathrm{pa}(i)}\right)} = p\left(x_{\boldsymbol{V}\setminus\{i,\mathrm{pa}(i)\}} \middle| x_i, x_{\mathrm{pa}(i)}\right) p\left(x_{\mathrm{pa}(i)}\right)$$

- Let $k \notin \{i, pa(i)\}$, then integrating out all variables other than X_i and X_k yields

$$p(x_k | do(x_i)) = \int_{x_{\mathrm{pa}(i)}} p\left(x_k \middle| x_i, x_{\mathrm{pa}(i)}\right) p\left(x_{\mathrm{pa}(i)}\right) dx_{\mathrm{pa}(i)}$$

- This is known as adjusting for $X_{\mathrm{pa}(i)}$

Recap

- Concepts to know:
  - Different intervention types
  - Total causal effect definitions
  - Path method
  - Connection between “no direct causes” and RCT
  - Parent adjustment

References and acknowledgments

- Slides adapted from M. Maathuis