ETH zürich

Covariate adjustment II

Causality

Christina Heinze-Deml

Spring 2019

Announcements

- Series 2 is due today
- Series 3 will be uploaded later today

Last week

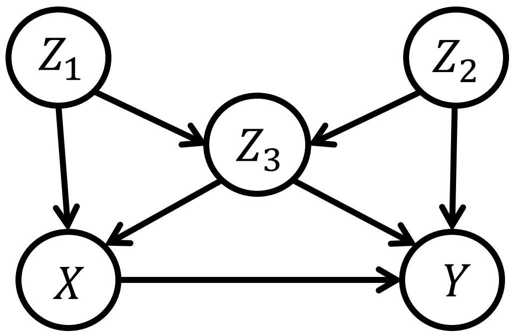
- Interventions
- Total causal effect definitions
- Path method
- Covariate adjustment – part 1

Today

- Covariate adjustment – part 2
- Frontdoor criterion

Example

- Interested in the causal effect of X on Y



Interested in $p(y \mid do(x))$

Could use

- truncated factorization
- parent adjustment

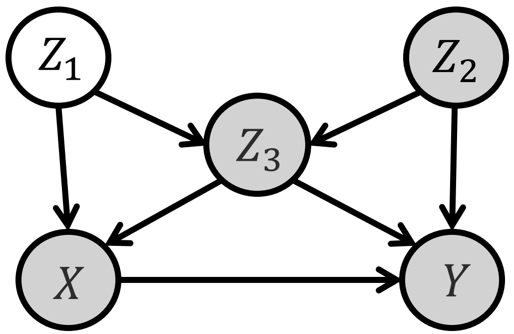
here: $Z = \{Z_1, Z_3\}$

Example

- Interested in the causal effect of X on Y

- Can we compute $p(y|do(x))$ if (only) Z_1 is not measured?
  - I.e., is $p(y|do(x))$ identifiable if (only) Z_1 is not measured?

Shaded nodes are observed



Identifiability

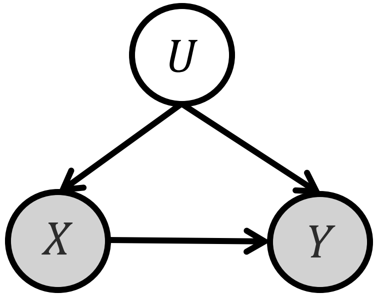
- An aspect of a statistical model is identifiable when it cannot be changed without there also being *some* change in the distribution of the observable variables.
- If we can alter part of a model with no observable consequences, that part of the model is unidentifiable.

- Identification is about the true distribution, not about finite data.

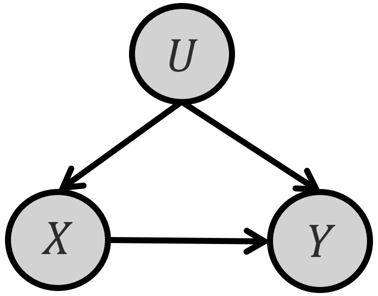
Identifiability

Shaded nodes are observed

- X and Y observable, U unobservable
  - $p(y|x)$ is identifiable
  - $p(y|do(x))$ not identifiable: can have different $p(y|do(x))$ with same distribution of observables $p(x,y)$ (compensating changes to other parts of the model)
  - Cannot estimate $p(y|do(x))$ from observational data



- X, Y and U observable
  - Can write $p(y|do(x))$ in terms of distribution of observables
  - Confounding can be removed by an identification strategy
  - $p(y|do(x))$ identifiable
  - Can estimate $p(y|do(x))$ from observational data



Identification strategies

- Interventional distribution is identifiable if it can be computed from the observational distribution and the graph structure
  - If there is a valid adjustment set for (X, Y), $p(y|do(x))$ is identifiable
  - Other means (discussed later):
    - Frontdoor criterion
    - Instrumental variables

Determining adjustment sets

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

- Adjustment formula

$$p(x_k | do(x_i)) = \int_{x_{\boldsymbol{Z}}} p(x_k | x_i, x_{\boldsymbol{Z}}) p(x_{\boldsymbol{Z}}) dx_{\boldsymbol{Z}} \qquad (1)$$

- Sets $\boldsymbol{Z}$ satisfying Eq. (1) are called valid adjustment sets
- If no proper subset of $\boldsymbol{Z}$ satisfies (1), $\boldsymbol{Z}$ is called a minimal adjustment set

Adjustment sets and confounding

- There is no confounding of the effect of x_i on x_k given covariates $x_{\boldsymbol{Z}}$ if

$$p(x_k | do(x_i), x_{\boldsymbol{Z}}) = p(x_k | x_i, x_{\boldsymbol{Z}}) \qquad (2)$$

- $\boldsymbol{Z}$ is then sufficient to adjust for confounding

ETH zürich

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?

Adjusting for direct causes

- Let (G, p) be a causal Bayesian network
- Rewriting the truncated factorization formula 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

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

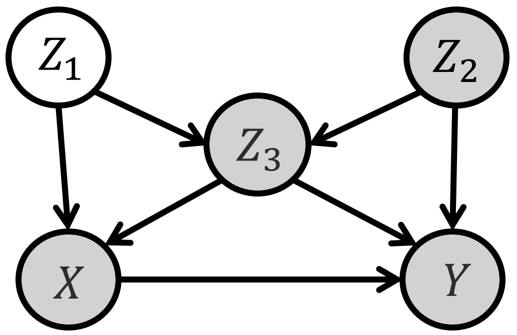
- This is known as adjusting for $X_{\mathrm{pa}(i)}$

Example

- Interested in the causal effect of X on Y
- Parent adjustment implies controlling for
$$\boldsymbol{Z} = \{Z_1, Z_3\}$$

- Can we compute $p(y|do(x))$ if (only) Z_1 is not measured?
  - I.e., is $p(y|do(x))$ identifiable if (only) Z_1 is not measured?

Shaded nodes are observed



Backdoor criterion (Pearl)

- Let $G = (\boldsymbol{V}, E)$ be a DAG and $i, k \in \boldsymbol{V}, i \neq k$. A set $\boldsymbol{Z} \subset \boldsymbol{V}$ (not containing i and k) satisfies the backdoor criterion relative to (i, k) in G if:
  i. $\boldsymbol{Z} \cap \mathrm{desc}(i) = \emptyset$, and
  ii. $\boldsymbol{Z}$ blocks all "backdoor paths" from i to k in G, i.e., all paths between i and k that start with an arrow into i $(i \leftarrow \cdots k)$

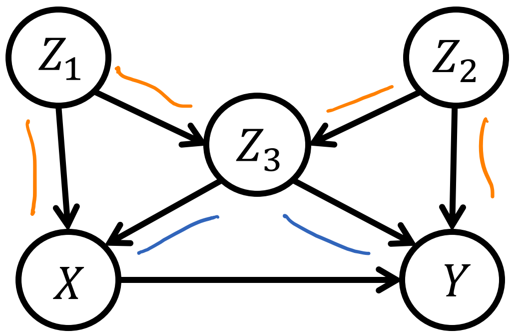
- If $\boldsymbol{Z} \subset \boldsymbol{V}$ satisfies the backdoor criterion relative to (i, k) in a DAG $G = (\boldsymbol{V}, E)$ then for all $p(\cdot)$ such that (G, p) is a causal Bayesian network, we have:

$$p(x_k | do(x_i)) = \int_{x_{\boldsymbol{Z}}} p(x_k | x_i, x_{\boldsymbol{Z}}) p(x_{\boldsymbol{Z}}) dx_{\boldsymbol{Z}}$$

Example

- Interested in the causal effect of X on Y

- Can we compute $p(y|do(x))$ if any of Z_1, Z_2, Z_3 is not measured?
- Valid adjustment sets:



- $Z \cap desc(X) = \emptyset$
- Z blocks all back-door paths

1st path: blocked by Z_3
2nd path: opened by Z_3

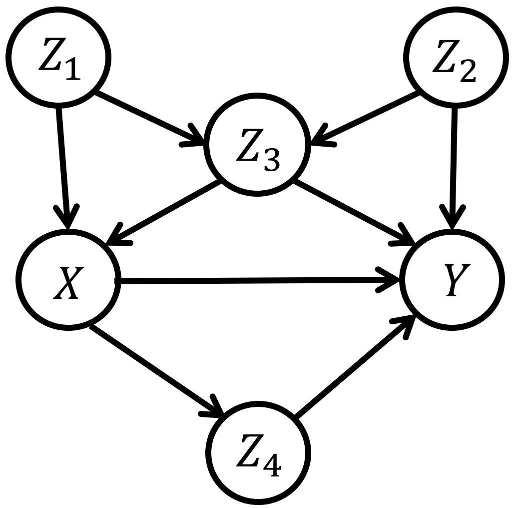
→ need to measure Z_3
→ need one of Z_1, Z_2

$\{Z_1, Z_3\}, \{Z_2, Z_3\}, \{Z_1, Z_2, Z_3\}$

Backdoor criterion

- Intuition behind backdoor criterion:
  - Backdoor paths carry spurious associations from X to Y
  - Paths directed along the arrows from X to Y carry causal associations
  - Blocking backdoor paths ensures that the measured association between X and Y is purely causal

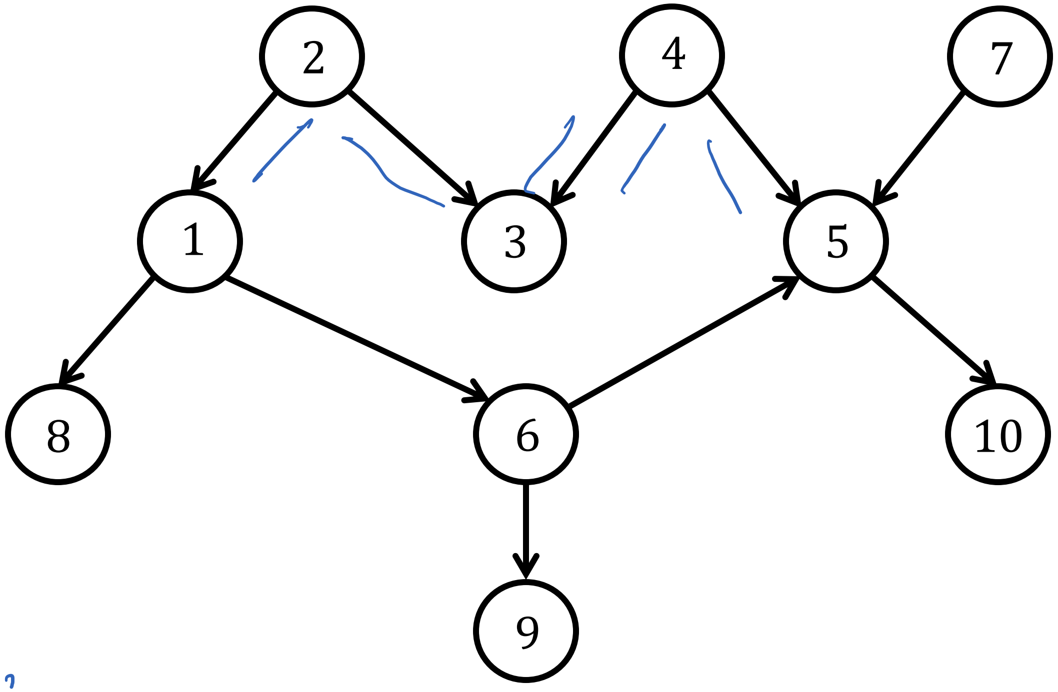
  - Don’t want to include descendants of X that are also ancestors of Y because this would block off a causal path
  - Don’t want to include descendants of X that are also descendants of Y because this would introduce collider bias



Backdoor criterion

Clicker question – Backdoor criterion

- Interested in the causal effect of X_1 on X_5
- Select all sets that satisfy the backdoor criterion:
  - $\{2\}$ ✓
  - $\{3\}$ ×
  - $\{2,7\}$ ✓
  - $\{\emptyset\}$ ϕ ✓
  - $\{7\}$ ✓
  - $\{2,3,4\}$ ✓
  - $\{8\}$ ×



- desc(1) = {1, 6, 5, 8, 9, 10}
  So we can only use subsets of {2, 3, 4, 7}
- Backdoor path:
  $1 \leftarrow 2 \rightarrow 3 \leftarrow 4 \rightarrow 5$
  blocked by {2}, {4}, {2, 4}, ϕ, {2, 3}, {3, 4}, {2, 3, 4}

we can add 7 to every set ⇒ 14 in total

Backdoor criterion

- The backdoor criterion is sufficient for adjustment
- Can show:
  If $\boldsymbol{Z}$ blocks all backdoor paths from i to k: $p(x_k | do(x_i), x_{\boldsymbol{Z}}) = p(x_k | x_i, x_{\boldsymbol{Z}})$

- If $G = (V, E)$ is a DAG, and $i, k \in V, i \neq k$, then the following hold:
  - If $k \notin \mathrm{pa}(i)$, then $\mathrm{pa}(i)$ satisfies the backdoor criterion relative to (i, k) in G
  - If $k \in \mathrm{pa}(i)$, then $p(x_k | do(x_i)) = p(x_k)$
  - [See Series 3.]

Positivity

- General requirement for identifiability:
  - Empirical basis for estimating the consequences of the contemplated interventions
  - Combinations of values under the interventional regime must also be possible under the observational regime

- Adjustment formula: $p(x_k|do(x_i)) = \int_{x_{\boldsymbol{Z}}} p(x_k|x_i, x_{\boldsymbol{Z}})p(x_{\boldsymbol{Z}})dx_{\boldsymbol{Z}}$
- In absence of further assumptions, positivity assumption requires:

$$p(x_i, x_{\boldsymbol{Z}}) > 0 \; \forall x_i \in \mathcal{X}_i, \, x_{\boldsymbol{Z}} \in \mathcal{X}_{\boldsymbol{Z}}$$

  - E.g. violation if we want to compare ”treatment” with “no treatment” in a patient group where some patients are so ill that they are never left untreated in practice

Simplification for multivariate Gaussian distributions

- Adjustment formula

$$p(x_k | do(x_i)) = \int_{x_{\boldsymbol{Z}}} p(x_k | x_i, x_{\boldsymbol{Z}}) p(x_{\boldsymbol{Z}}) dx_{\boldsymbol{Z}}$$

- May be hard to compute, especially in the case of continuous variables and high-dimensional $\boldsymbol{Z}$
- Simplification if the joint distribution p is Gaussian

Simplification for multivariate Gaussian distributions

- Let

$$p(x_k|do(x_i)) = \int_{x_{\boldsymbol{Z}}} p(x_k|x_i, x_{\boldsymbol{Z}})p(x_{\boldsymbol{Z}})dx_{\boldsymbol{Z}}$$

and let $p(x_{\boldsymbol{V}})$ be multivariate Gaussian. Then

$$E(X_k|do(x_i = x_i' + 1)) - E(X_k|do(x_i = x_i')) = \gamma$$

where γ is the coefficient of X_i in the linear regression of X_k on X_i and $X_{\boldsymbol{z}}$, i.e.

$$E(X_k|X_i, X_{\boldsymbol{Z}}) = \alpha + \gamma X_i + \beta^T X_{\boldsymbol{Z}}$$

for some α, β.

Simplification for multivariate Gaussian distributions

- Hence, we can then estimate the total effect of X_i on X_k in R by

```
coef(lm(xk ~ xi + xz))[2]
```

- See Jupyter notebook and R scripts

Adjustment criterion (Shpitser et al, Perkovic et al)

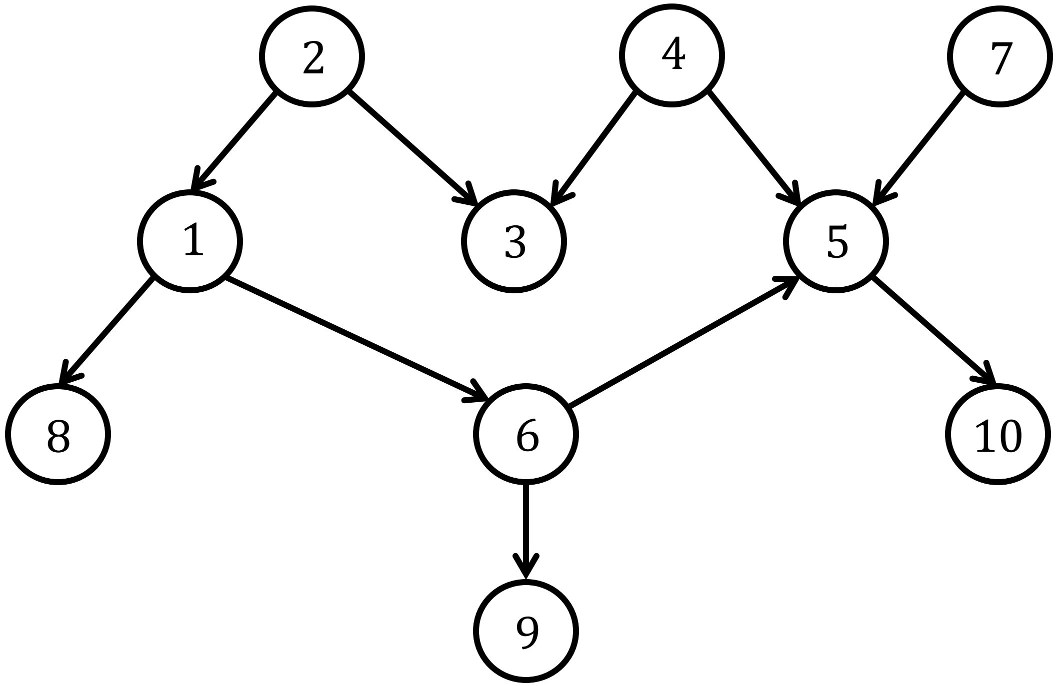
- Let $G = (\boldsymbol{V}, E)$ be a DAG and $i, k \in \boldsymbol{V}, i \neq k$. A set $\boldsymbol{Z} \subset \boldsymbol{V}$ (not containing i and k) satisfies the adjustment criterion relative to (i, k) in G if:
  - $\boldsymbol{Z}$ does not contain any descendants of nodes $r \neq i$ on a directed path from i to k in G
  - $\boldsymbol{Z}$ blocks all paths between i and k in G that are not directed from i to k

- A set $\boldsymbol{Z} \subset \boldsymbol{V}$ satisfies the adjustment criterion relative to (i, k) in a DAG $G = (\boldsymbol{V}, E)$ if and only if for all p such that (G, p) is a causal Bayesian network, we have:

$$p(x_k | do(x_i)) = \int_{x_{\boldsymbol{Z}}} p(x_k | x_i, x_{\boldsymbol{Z}}) p(x_{\boldsymbol{Z}}) dx_{\boldsymbol{Z}}$$

Example

- There are 28 sets satisfying the adjustment criterion.
  - [Exercise: Verify this.]



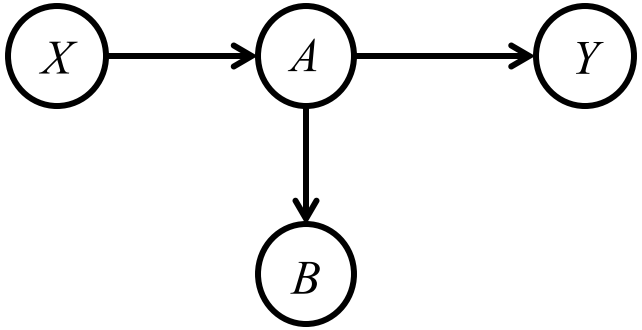
Adjustment criterion

- The adjustment criterion is necessary and sufficient for identifying total causal effects via adjustment.
- It is only sufficient for the identification of total causal effects.
  - Some effects are identified by other means, e.g., via the frontdoor criterion.

Determining adjustment sets

- Should we adjust for as many variables as possible?

  - X: Smoking
  - Y: Future miscarriages
  - A: Physiological abnormality induced by smoking
  - B: Previous miscarriages

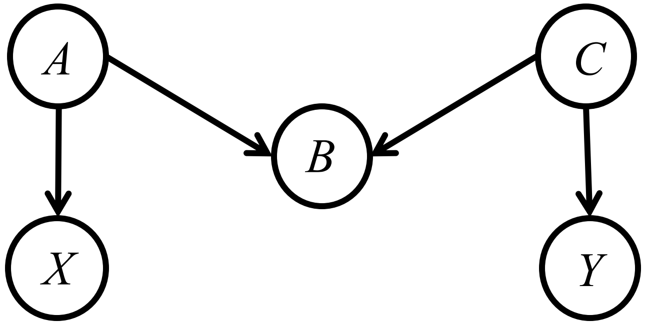


$\Rightarrow$ should not adjust for
A and/or B

Determining adjustment sets

- Is it always safe to adjust for “pre-treatment” variables?

  - X: Smoking
  - Y: Adult asthma
  - A: Parental smoking
  - B: Childhood asthma
  - C: Predisposition toward asthma



Control for : $\{\}$, $\{A, B\}$, $\{B, C\}$, $\{A, B, C\}$

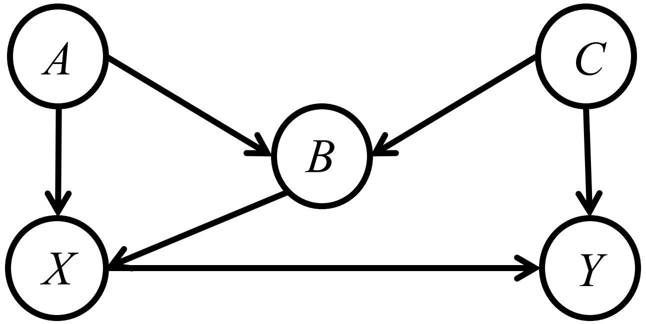
$\{A\}$, $\{C\}$, $\{A, C\}$

Only controlling for B would introduce bias – "M-bias"

Determining adjustment sets

- Is it always safe to adjust for “pre-treatment” variables?

  - X: Smoking
  - Y: Adult asthma
  - A: Parental smoking
  - B: Childhood asthma
  - C: Predisposition toward asthma



control for {A, B}, {C, B}, {A, B, C}, {C}

Still: only controlling for B would introduce M-bias

Summary: Determining adjustment sets

- Should we adjust for as many variables as possible?
  - No. Adjusting for certain variables can create bias.
- Is it always safe to adjust for “pre-treatment” variables?
  - No. This can create so-called M-bias.
- If we want the total effect of X_i on X_k in G $(k \notin pa(i))$ then:
  - $\mathrm{pa}(i)$ is a valid adjustment set (this includes not adjusting for anything if $\mathrm{pa}(i) = \emptyset$).
  - Any set $\boldsymbol{Z}$ satisfying the backdoor criterion relative to (i, k) in G is a valid adjustment set.
  - A set $\boldsymbol{Z}$ is a valid adjustment set if and only if it satisfies the adjustment criterion relative to (i, k) in G.

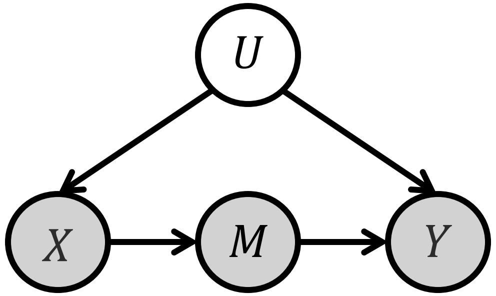
Statistical efficiency

- We focused so far on sets that provide asymptotically correct causal effects.
- We did not consider statistical efficiency.
- Rules of thumb for statistically efficient estimates in linear regression setting:
  - Try to avoid variables that are strongly correlated with X_i.
    - This blows up the standard error.
  - Try to use variables that help predict X_k.
    - This decreases the residual variance and hence decreases the standard error.
    - This may mean using optional variables that are not strictly needed.
- [See Series 3.]

Frontdoor criterion

- X, Y and M observable, U unobservable
- Cannot use the backdoor or the adjustment criterion since U is unobservable
- Idea:
  - Find set of variables $\boldsymbol{M}$ which mediate the causal influence of X on Y, i.e., all direct paths from X to Y pass through $\boldsymbol{M}$
  - If we can identify the effects of M on Y and of X on M, then we can combine them to get the effect of X on Y
  - “study the mechanisms by which X influences Y”

Shaded nodes are observed



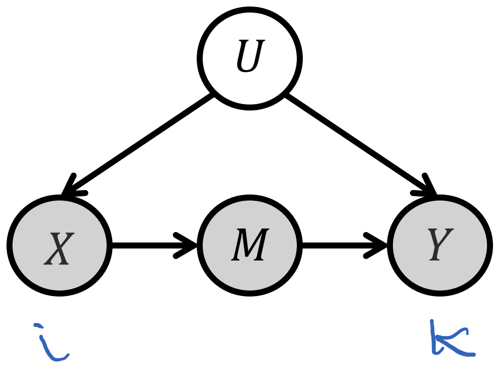
Frontdoor criterion

- Let $G = (\boldsymbol{V}, E)$ be a DAG and $i, k \in \boldsymbol{V}, i \neq k$.

  A set $\boldsymbol{M} \subset \boldsymbol{V}$ (not containing i and k) satisfies the frontdoor criterion relative to (i, k) in G if:

  i. $\boldsymbol{M}$ blocks all directed paths from i to k in G
  ii. There are no unblocked backdoor paths from i to $\boldsymbol{M}$ in G
  iii. i blocks all backdoor paths from $\boldsymbol{M}$ to k in G

Shaded nodes are observed



Frontdoor criterion

- If $\boldsymbol{M} \subset \boldsymbol{V}$ satisfies the frontdoor criterion relative to (i, k) in a DAG $G = (\boldsymbol{V}, E)$ then for all $p(\cdot)$ such that (G, p) is a causal Bayesian network, we have:

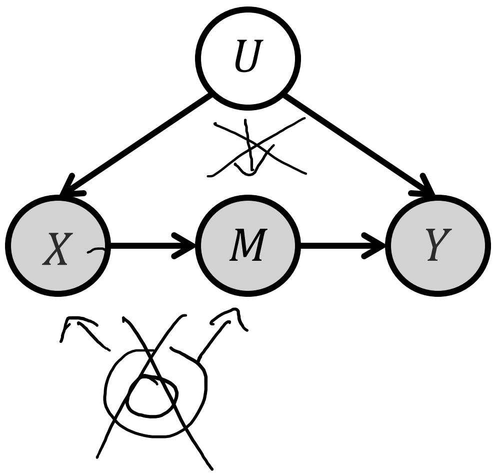
$$p(x_k | do(x_i')) = \int_{x_{\boldsymbol{M}}} p(x_{\boldsymbol{M}} | x_i') \int_{x_i} p(x_k | x_i, x_{\boldsymbol{M}}) p(x_i) dx_i \, dx_{\boldsymbol{M}}$$

ETH zürich

Example

- X: Smoking
- Y: Lung cancer
- U: Carcinogenic genotype
- M: Amount of tar in lungs

Shaded nodes are observed



- Assume that
  - smoking cigarettes has no effect on the production of lung cancer except as mediated through tar deposits (i.) → no direct edge from x to y
  - genotype has no direct effect on the amount of tar in the lungs (ii. + iii.)
  - no other factor that affects tar deposit has any influence on smoking (ii)

Recap

- Concepts to know:
  - Identifiability
  - Positivity
  - Backdoor criterion
  - Simplification in Gaussian setting
  - Adjustment criterion
  - Frontdoor criterion

References and acknowledgments

- Slides adapted from M. Maathuis
- Some examples from
  - Shalizi (2019). Chapter 22.
  - Pearl and Mackenzie (2018). The Book of Why.
  - Pearl (2009). Causality: Models, Reasoning and Inference.