I. What is causality?

II. How can we reason about causality mathematically?

From Bayesian Networks to Causal Bayesian Networks (causal DAGs)

III. Can we learn a causal DAG?

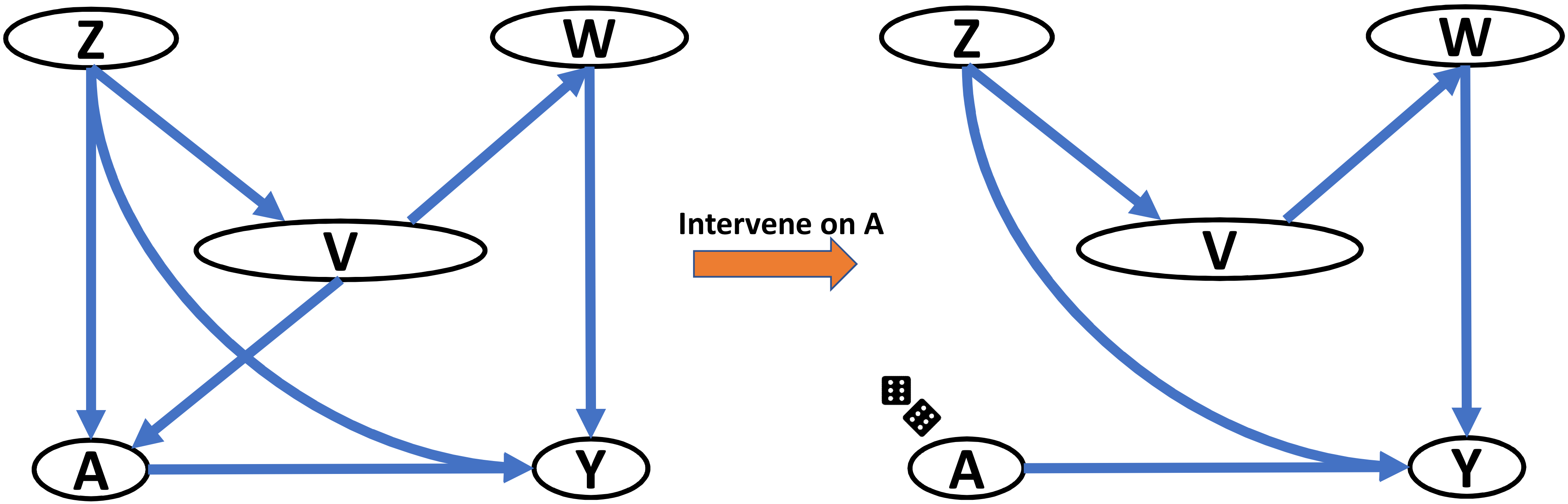
IV. Application 1: Estimating the effect of actions

V. Application 2: Building more generalizable prediction models

VI. Open questions

Estimating causal effect

- **Input:** Causal DAG, Action variable A, outcome variable Y
- **Output:** $P(Y|do(A))$, usually $E[Y|do(A)]$



Historical Data: $\boldsymbol{P}(.)$ **Randomized Experiment:** $\boldsymbol{P}^{*}(.)$

Causal Effect: $\mathbf{E_P}[\boldsymbol{Y}|\boldsymbol{do}(\boldsymbol{A})] = \boldsymbol{E_{P^*}}\ (\boldsymbol{Y}|\boldsymbol{A} = \boldsymbol{a})$

$\mathrm{E_P}[Y|do(A=1)] - \mathrm{E_P}[Y|do(A=0)] = E_{P^*}\ (Y|A=1) - E_{P^*}\ (Y|A=0)$

But what if randomized P^* distribution is not available?

Identification problem:

Can we $P(Y|do(A))$ purely as a probability expression computable over P?

Understanding **do-intervention** that leads to P^*:

The intervention makes A d-separated from Y after removing the A-> Y edge.

Q: Can we find how A can be d-separated from Y in the original distribution P?

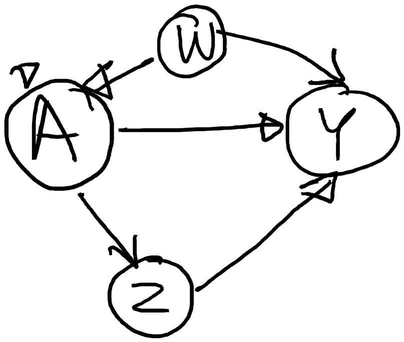
Identifiability

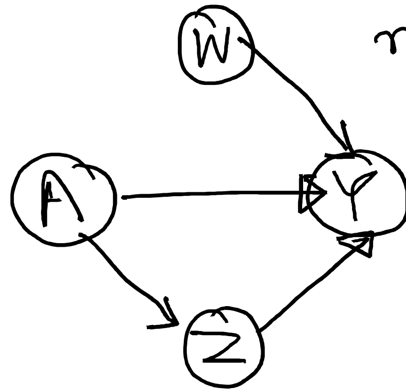
- Conditions under which observations of P() can be used to estimate effect on interventions on a variable A, let us call this $P^{A*} = P*$ for short. The graphical models for P and P* differ only on the CPD attached to A.
- When all variables are observed (no hidden confounders) then we can always estimate P* from observations from P. Let W denote the parents of A.
  - $P(Y|do(A) = a) = \sum_{w \in Dom(W)} P(Y|W = t, A = a)P(W = t)$

  Here is an intuitive proof. In the graph P and P*, when we fix the values of parents of A (W) to t and A to a, the graphs P and P* identical in the dependency structure with the only difference in P(A|W=t) and P*(A). But the conditional probability P(Y|A,W=t) is independent of the potential attached to A.

Example

to show that conditioning on parents of A makes $P(Y|A,W=t) = P^*(Y|A,W=t)$



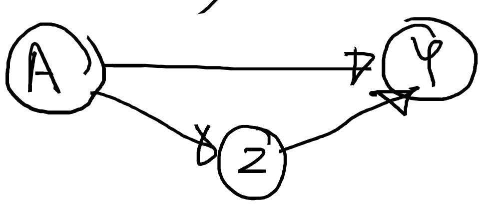


P

P^* ← where we intervene on A differs from P only on parents of P.

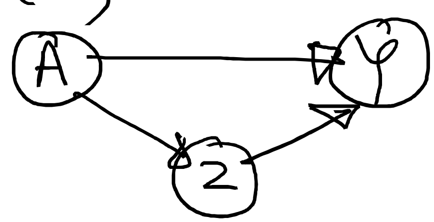
If we condition on parent $W=t$ we get

$P(A|W=t)$



$P(Y|A,Z,W=t)$

$P^*(A)$



$P(Y|A,Z,W=t)$

Now we can estimate

$$P(Y \mid A, W=t) = \sum_z P(Y \mid z, A, W=t)\, P(z \mid A, W=t)$$

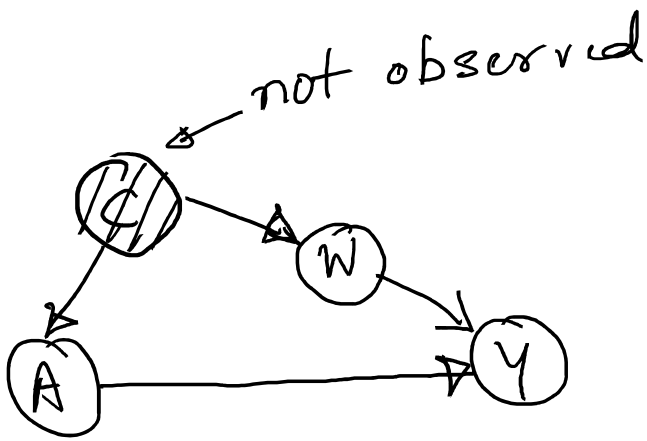
$$= \sum_z P(Y \mid z, A, W=t)\, P(z \mid A)$$

$$= \sum_z P^*(Y \mid z, A, W=t)\, P^*(z \mid A)$$

$$= P^*(Y \mid A, W=t)$$

When all variables are not observed

- Challenge is when subset of variables are not observed. Under what conditions can we still estimate $P^*(Y|A) = P(Y|do(A))$? In particular, if some parents are not observed, then can we find some other W instead of parents of A?



Yes, by conditioning on W.

The backdoor criterion

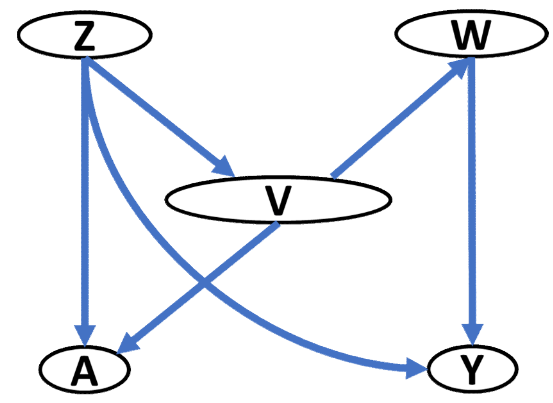
1.Remove any edges going out from A.

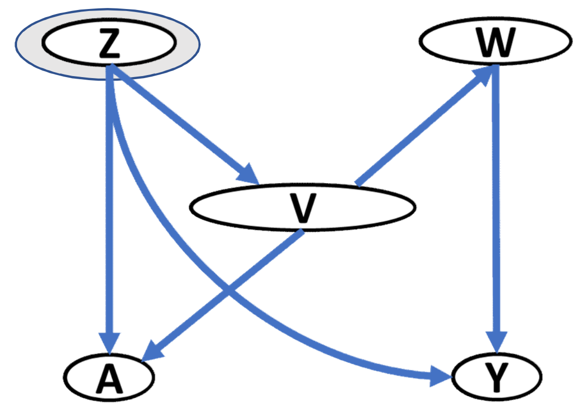
2. Find the set of variables such that A is conditionally d-separated from Y.

3. Condition on them.

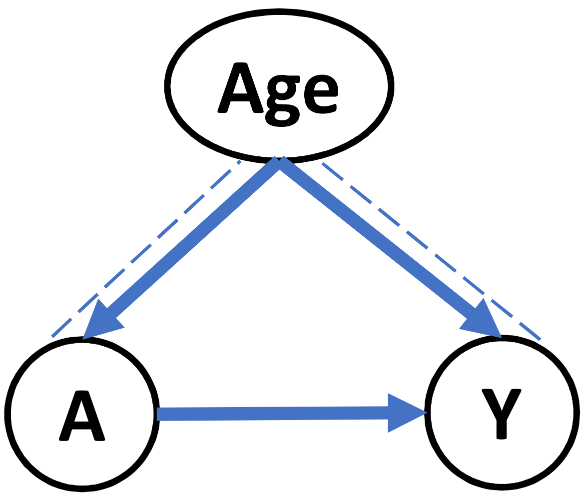
Backdoor set: $\{Z, W\}$ or $\{Z, V\}$

$$P(Y|do(A)) = \sum_{z,w} P(Y|A, Z = z, W = w)P(Z = z, W = w)$$

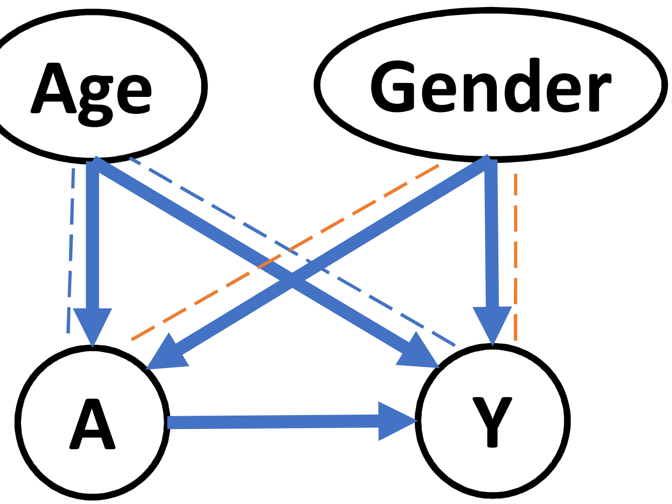




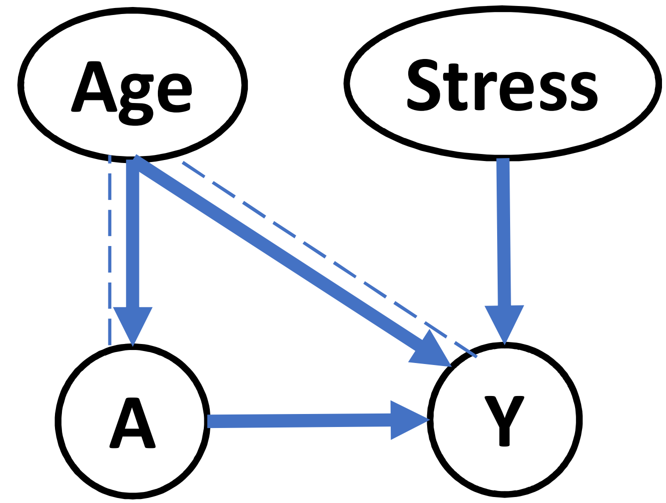
Find the backdoor set!



$\boldsymbol{B} = \{Age\}$

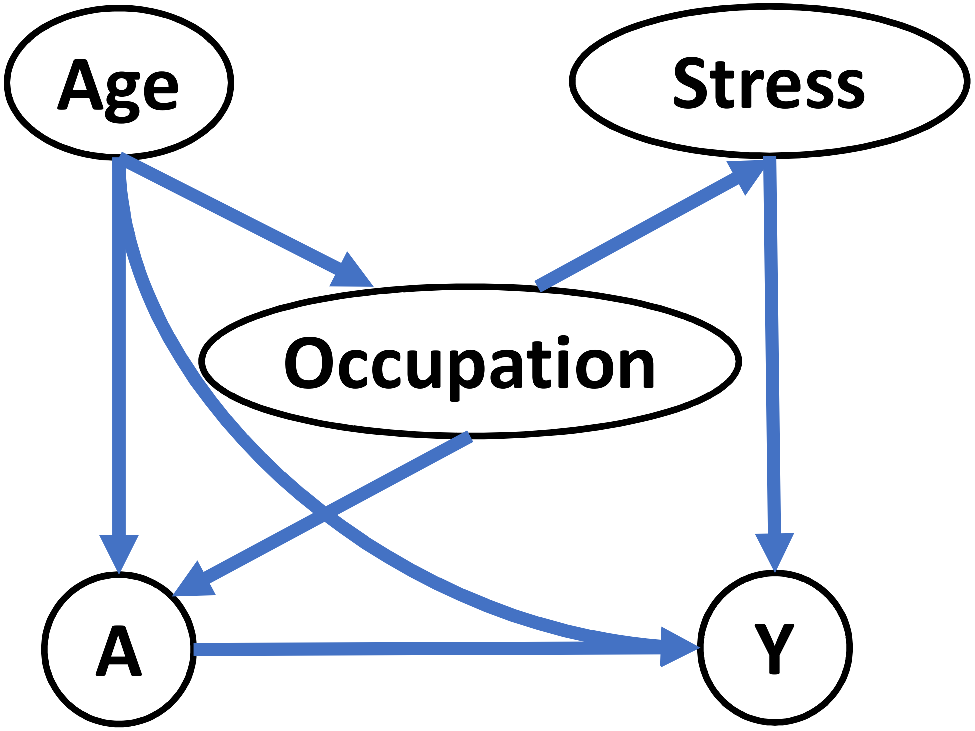


$\boldsymbol{B} = \{Age, Gender\}$

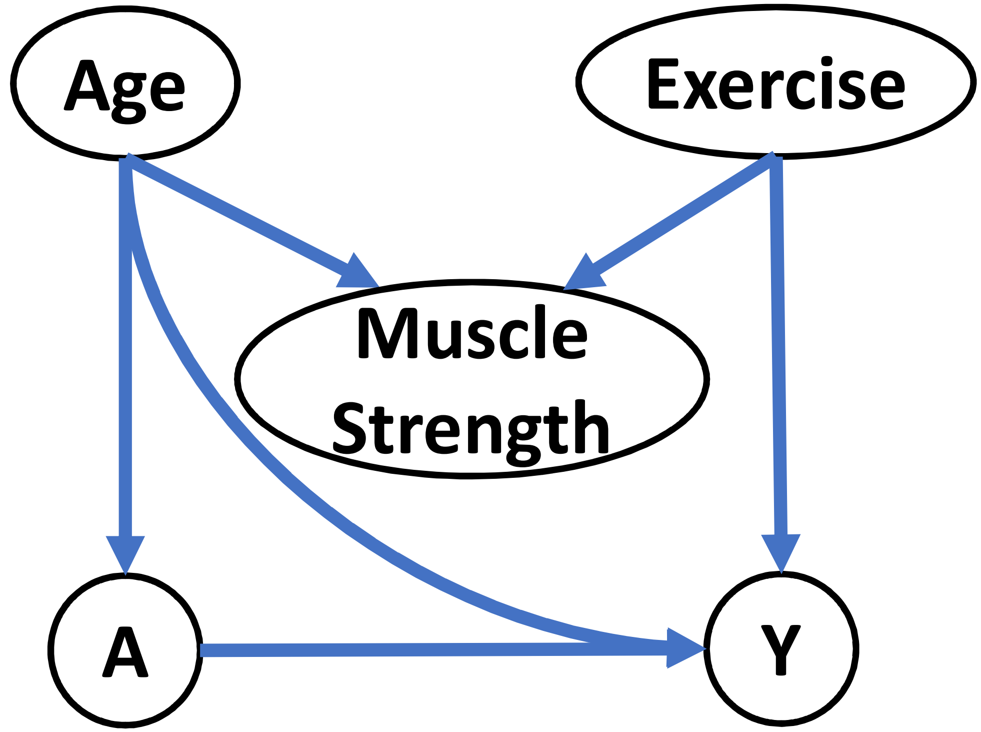


$\boldsymbol{B} = \{Age\}$

Find the backdoor set!



$$\boldsymbol{B} = \{Age, Stress\}$$
$$\boldsymbol{B} = \{Age, Occupation\}$$

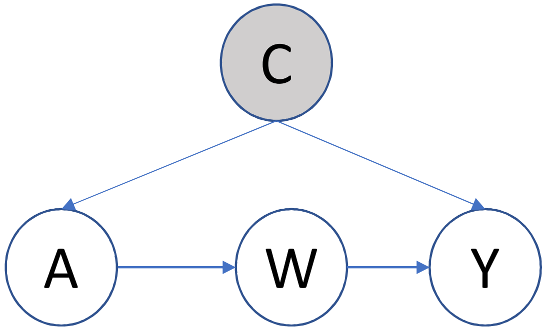


$$\boldsymbol{B} = \{Age, Exercise\}$$ ~~Correct~~?
$$\boldsymbol{B} \neq \{Age, MuscleStrength\}$$

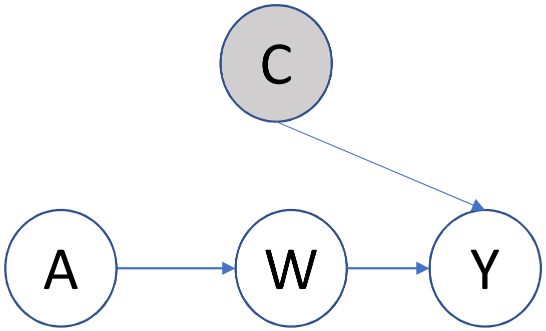
Backdoor is sufficient, but not a necessary criterion for identification

- There can be other ways to derive the do-expression (e.g., frontdoor criterion)
- Fortunately, there exists an algorithm that is both necessary and sufficient for an arbitrary causal DAG.
  - If it returns a probability expression, it is a valid identification.
  - If it fails to return an expression, then no valid non-parametric identification exists.
- Called ID algorithm [Shpitser and Pearl, 2006].
- Implemented in software libraries like DoWhy.

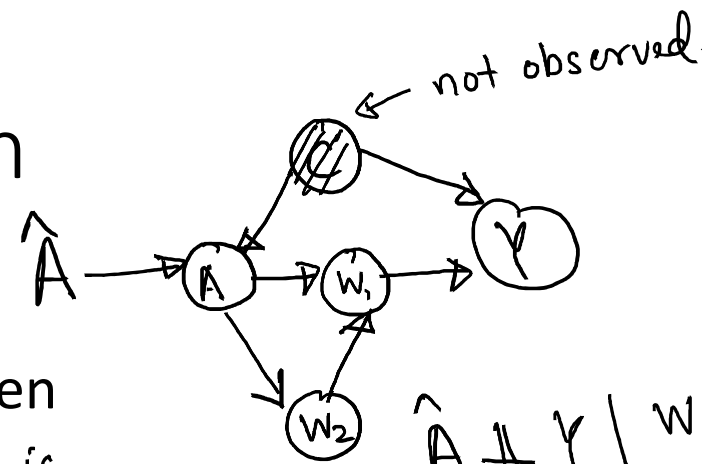
Frontdoor criteria for identifiability



- Consider a graph like the one on the side where C parent of A is not observed. Even in this case, we can compute P*(Y|A) by conditioning on W.
- Brief proof. Consider a fixed value w of W. Then, in P* we have that P*(Y|W=w,A) = P*(Y|W=w). It can be shown that this value will be equal to P(Y|W=w).



More general formulation



$\hat{A} \perp\!\!\!\perp Y \mid \underbrace{W_1, W_2}_{W}$

- Add an auxillary parent $\hat{A}$ to A
- If $\hat{A}$ is d-separated from Y given W then
  - $P(Y|do(A), W) = P(Y|W)$ where W is observed and $A \notin W$.
- Additionally if we have some criteria (e.g. backdoor) to estimate P(W|do(A)), then we
- $P(Y|do(A)) = \sum_w P(W = w|do(A))P(Y|w, A)$

More examples of identifiability

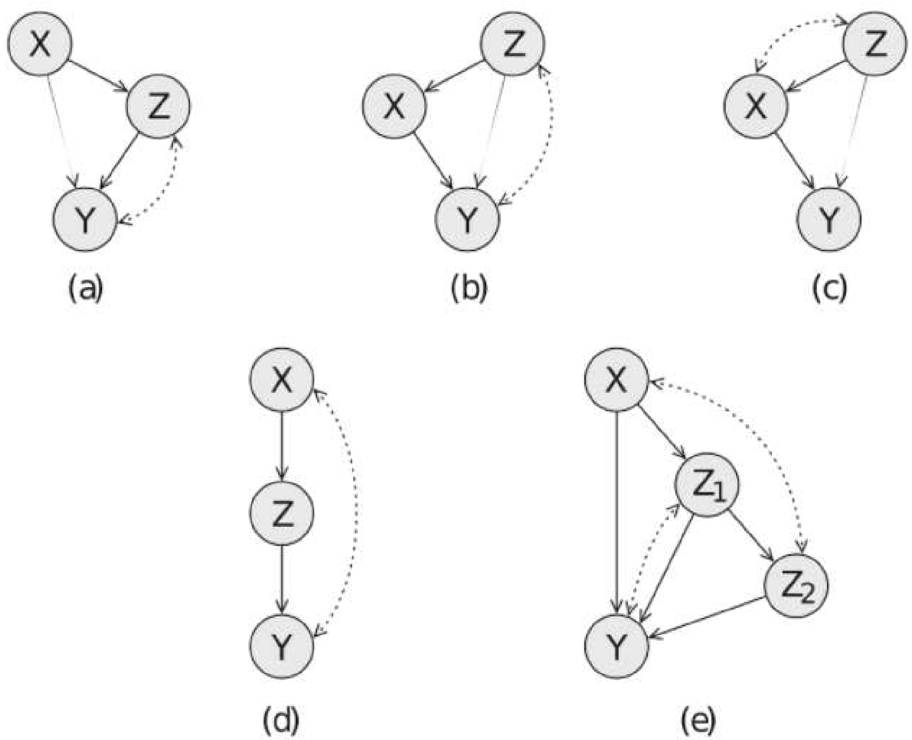


Figure 21.3 Examples of models where $P(Y \mid do(X))$ is identifiable. The bidirected dashed arrows denote cases where a latent variable affects both of the linked variables.

(a)

$P(Y|do(X))$

$= P(Y|X)$

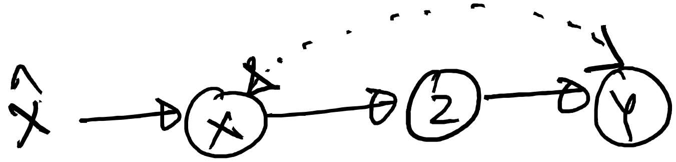
(by backdoor criteria $W = \phi$)

(b) $P(Y|do(X))$

$= \sum_{z} P(Y|X,z)P(z)$

(backdoor, $W = Z$)

& (c) Same as above.

(d) 

$\hat{X} \perp\!\!\!\perp Y \mid Z \Rightarrow$ Apply front-door criteria.

$$\left.\begin{aligned} P(Y \mid do(x)) &= \sum_z P(Y \mid z, x)\, P(z \mid do(x)) \\ &= \sum_z P(Y \mid z, x)\, P(z \mid x) \end{aligned}\right\} \to \text{back door.}$$

(e) Backdoor criteria on Z_2 separates X from Y after removing forward arrows.

$$P(Y \mid do(x)) = \sum_{z_2} P(Y \mid x, z_2)\, P(z_2 \mid do(x))$$

$$= \sum_{z_2} P(Y \mid x, z_2) \sum_{z_1} P(z_2 \mid z_1, x)\, P(z_1 \mid x)$$

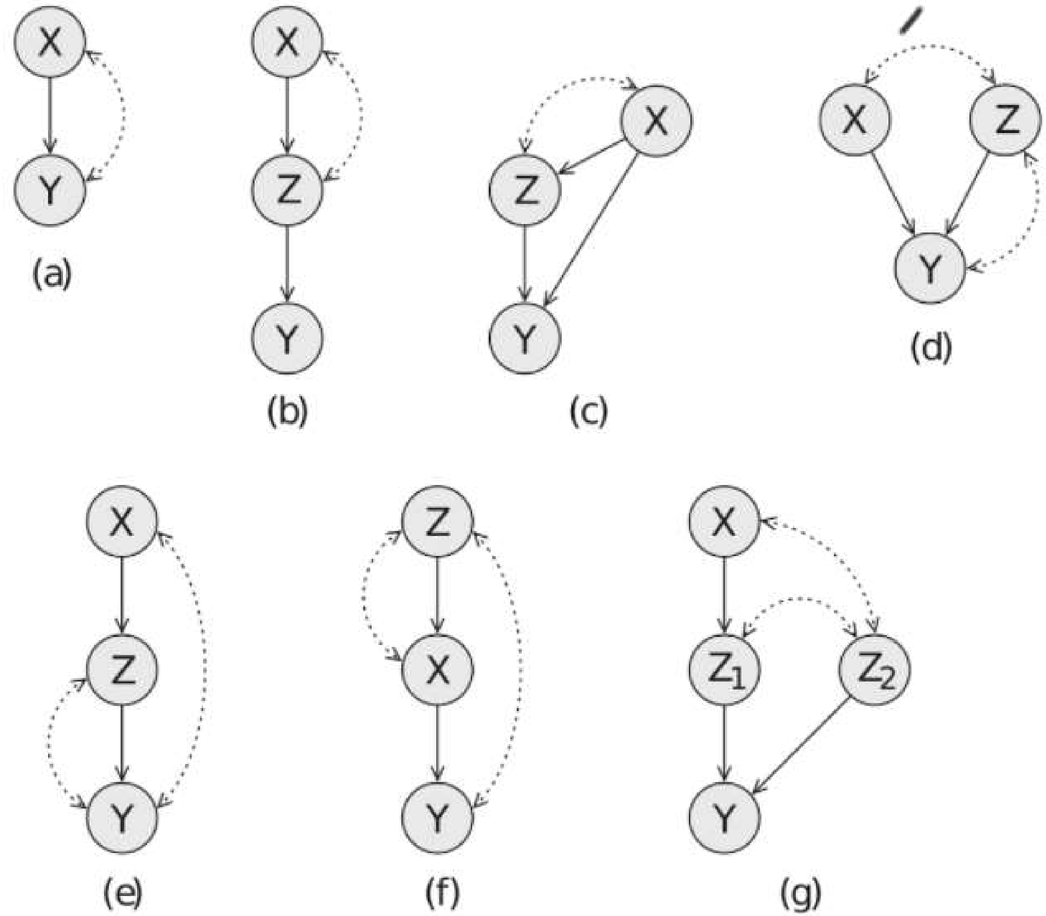


Figure 21.4 Examples of models where $P(Y \mid do(X))$ is not identifiable. The bidirected dashed arrows denote cases where a latent variable affects both of the linked variables.

End of syllabus for Spring 2022.