Causal and Probabilistic Reasoning

Rina Dechter

The Identification Problem
The Front-Door Criterion,
The Do-calculus

Based on Elias Bareinboim slides

Primer, chapter 3, Causality 3.3, 3.4, 2.5, (Biometrica 1995)

Outline

The backdoor criterion and the adjustment formula
Computing bd: Inverse probability weighting
Conditional intervention
Front door condition
The do calculus

Outline

The backdoor criterion and the adjustment formula
Computing bd: Inverse probability weighting
Conditional intervention
Front door condition
The do calculus

How could adjustment help in real data analysis?

(The Problem of Confounding)

How could adjustment help in real data analysis?

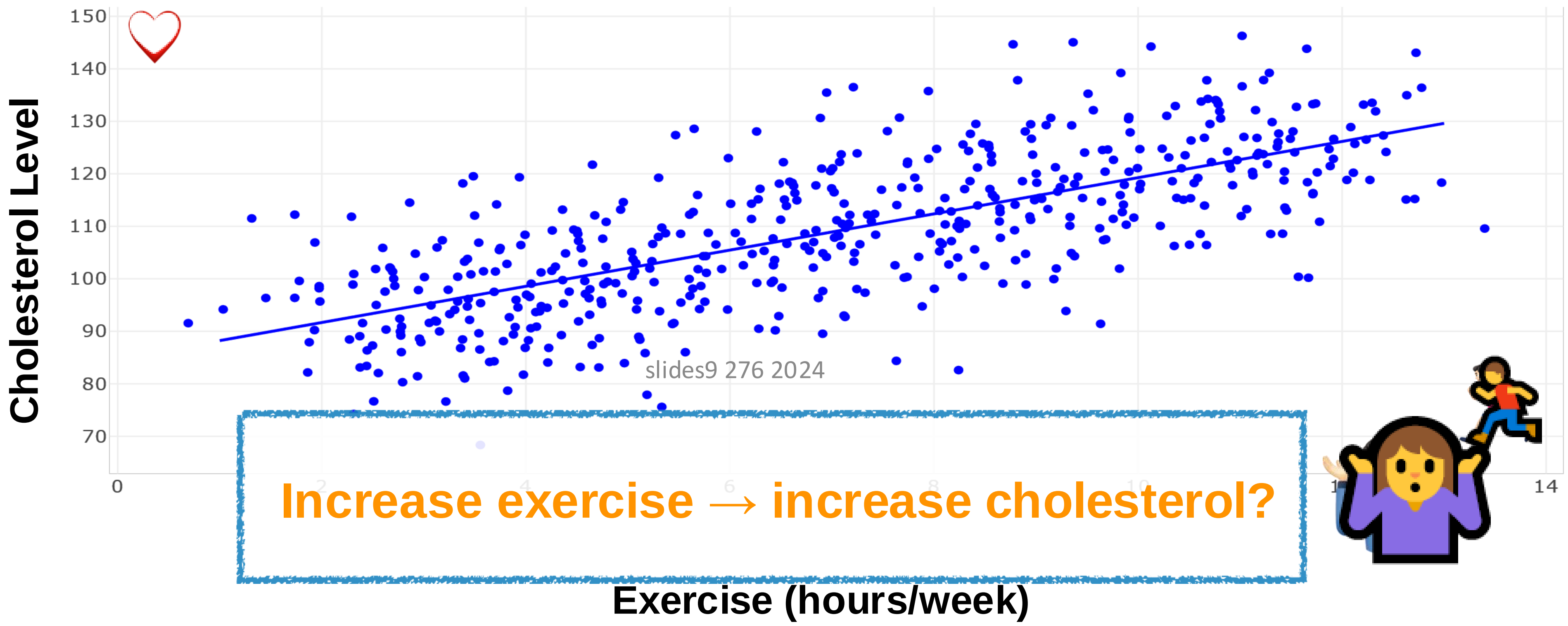
(The Problem of Confounding)

Confounding Bias

What's the causal effect of Exercise on Cholesterol?

What about $P(cholesterol \mid exercise)$?

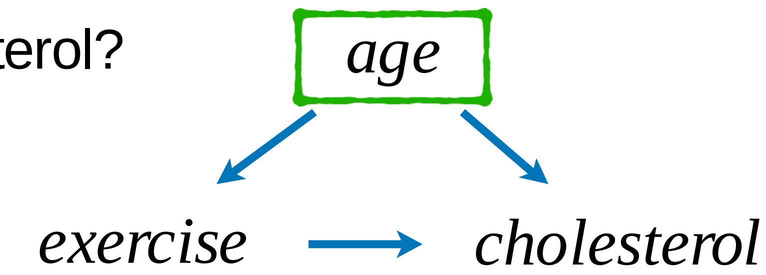
$$\begin{array}{ccc} & age & \\ \swarrow & & \searrow \\ exercise & \longrightarrow & cholesterol \end{array}$$

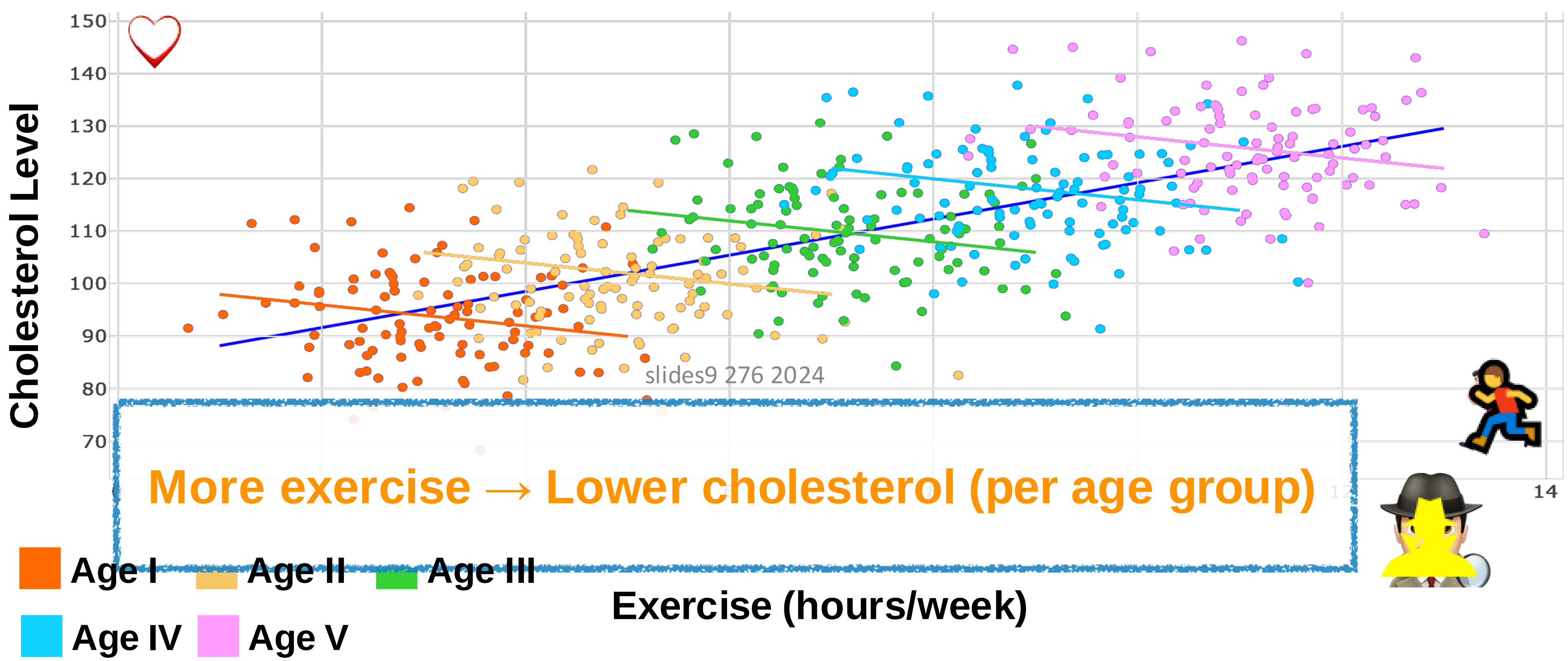


Confounding Bias

What's the causal effect of Exercise on Cholesterol?

What about *P(cholesterol | exercise)* ?





Adjustment by Direct Parents

Thm. Given a causal diagram G of any Markovian system, the causal quantity $Q = P(y \mid do(x))$ is identifial … hat is, wheneve … ariables $\boldsymbol{X}$ are mea … hen obtained …

What if the parents are not visible?

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{pa_{\mathbf{x}}} P\left(\mathbf{y} \mid \mathbf{x}, pa_{\mathbf{x}}\right) P\left(pa_{\mathbf{x}}\right)$$

Quiz: 1) derive from previous slide

2) derive for non-Markovian models

If Season is latent, is the effect still computable?

Queries:

$Q_2 = P(wet \mid do(Sprinkler = on))$

$$= \frac{\sum_{se,ra} P(we \mid Sp = on, ra) P(Sp = on) P(ra \mid se) P(se)}{P(Sp = on)}$$

equal to 1

$$= \sum_{se,ra} P(we \mid Sp = on, ra) P(ra \mid se) P(se)$$

$$= \sum_{se,ra} P(we \mid Sp = on, ra) P(ra, se)$$

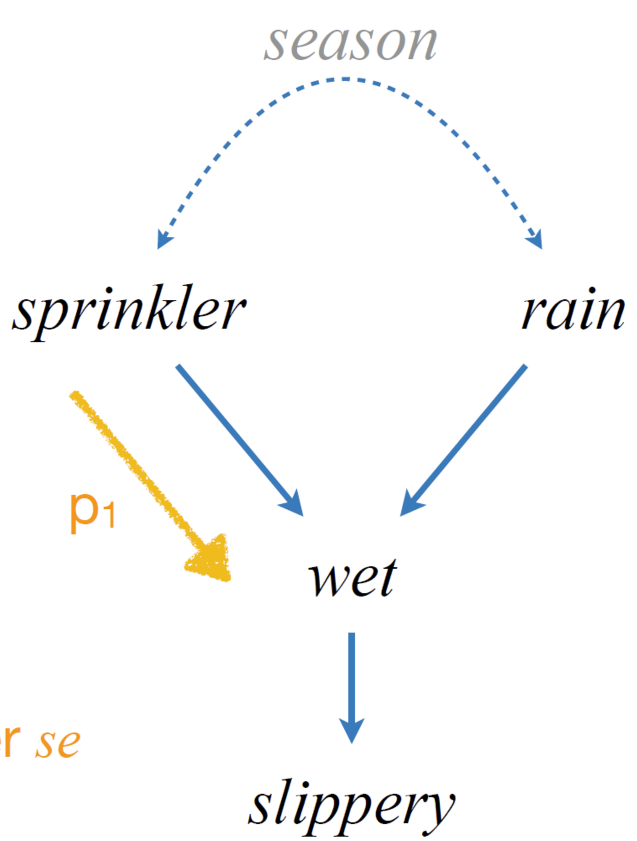
chain rule

$$= \sum_{ra} P(we \mid Sp = on, ra) \sum_{se} P(ra, se)$$

summing over *se*

$$\boxed{= \sum_{ra} P(we \mid Sp = on, ra) P(ra)}$$

Adjustment by Rain!



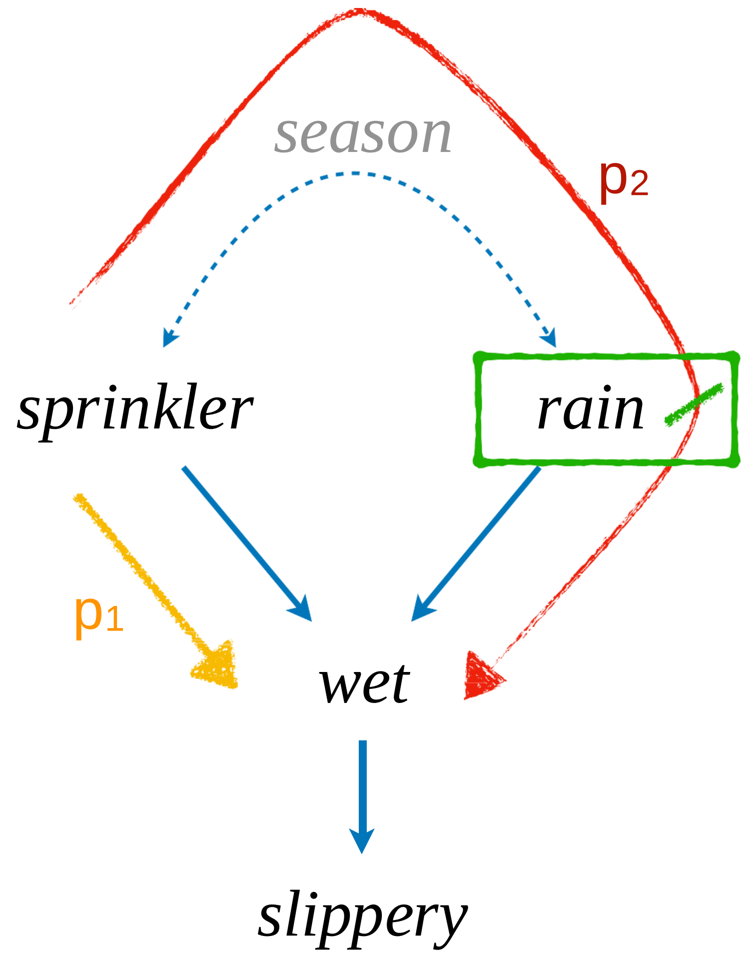
If Season is latent,
is the effect still computable?

Queries:

$Q_2 = P(wet \mid do(Sprinkler = on))$

$$\boxed{= \sum_{ra} P(we \mid Sp = on, ra)P(ra)}$$

By conditioning on rain,
- p_2 (the non-causal path) is blocked, and
- p_1 (the causal path) remains unaffected!



3.3 The Backdoor Criterion

Definition 3.3.1 (The Backdoor Criterion) *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*

If a set of variables Z satisfies the backdoor criterion for X and Y, then the causal effect of X on Y is given by the formula

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$$

Rationale:

1. We block all spurious paths between X and Y.
2. We leave all directed paths from X to Y unperturbed.
3. We create no new spurious paths.

The Back-door Adjustment

If a set $\boldsymbol{Z}$ satisfies the bdc w.r.t the pair $\boldsymbol{X},\boldsymbol{Y}$, the effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is identifiable and given by:

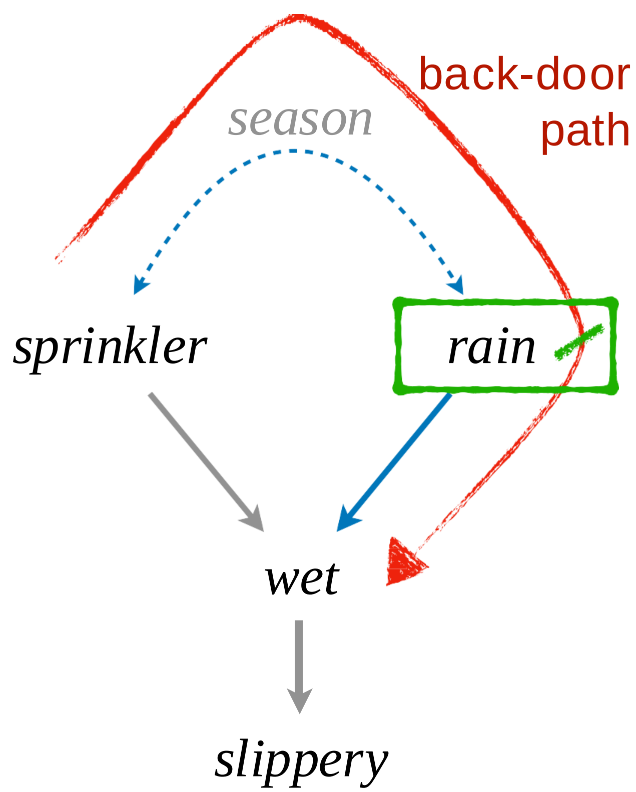
$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z})$$

Back-Door Sets as Substitutes of the Direct Parents of X

Rain satisfies the back-door criterion relative to *Sprinkler* and *Wet*:

(i) Rain is not a descendant of *Sprinkler*, and

(ii) Rain blocks the only back-door path from *Sprinkler* to *Wet*.



Adjusting for the direct parents of *Sprinkler*, we have:

Direct derivation, showing it works

$$
\begin{aligned}
P(we \mid do(sp)) &= \sum_{se} P(we \mid sp, se) P(se) \\
&= \sum_{se,ra} P(we \mid sp, se, ra) P(ra \mid sp, se) P(se) \\
&= \sum_{se,ra} P(we \mid sp, ra) P(ra \mid se) P(se) \\
&= \sum_{ra} P(we \mid sp, ra) \sum_{se} P(ra, se) = \boxed{\sum_{ra} P(we \mid sp, ra) P(ra)}
\end{aligned}
$$

(Sp⫫Ra|Se)

(We⫫Se|Ra,Sp)

Adjustment by *Rain*

Adjustment by Direct Parents → Back-door Adjustment

More Generally:

(i) no node in $\boldsymbol{Z}$ is a descendent of $\boldsymbol{X}$; and $\Longrightarrow$ $(\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Z} \mid Pa_x)$

(ii)Z blocks every path between $\boldsymbol{X}$ and $\boldsymbol{Y}$ that contains an arrow into $\boldsymbol{X}$. $\Longrightarrow$ $(\boldsymbol{Y} \perp\!\!\!\perp Pa_x \mid \boldsymbol{Z}, \boldsymbol{X})$

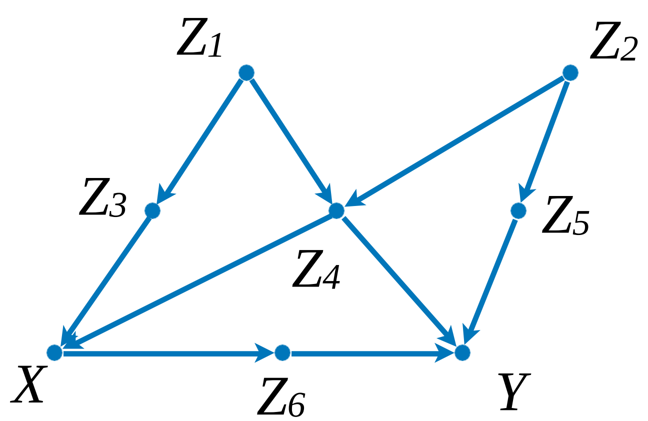
Then:

$$
\begin{aligned}
P(\mathbf{y} \mid do(\mathbf{x})) &= \sum_{pa_{\mathbf{x}}} P(\mathbf{y} \mid \mathbf{x}, pa_{\mathbf{x}}) P(pa_{\mathbf{x}}) \\
&= \sum_{\mathbf{z}, pa_{\mathbf{x}}} P(\mathbf{y} \mid \mathbf{x}, pa_{\mathbf{x}}, \mathbf{z}) P(\mathbf{z} \mid \mathbf{x}, pa_{\mathbf{x}}) P(pa_{\mathbf{x}}) \\
&= \sum_{\mathbf{z}, pa_{\mathbf{x}}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z} \mid pa_{\mathbf{x}}) P(pa_{\mathbf{x}}) \\
&= \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) \sum_{pa_{\mathbf{x}}} P(\mathbf{z}, pa_{\mathbf{x}}) = \boxed{\sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z})}
\end{aligned}
$$

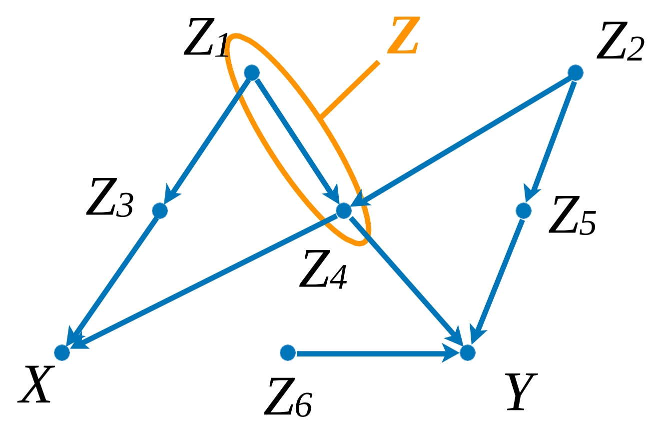
Adjustment by $\boldsymbol{Z}$ is equivalent to adjustment by direct parents whenever $\boldsymbol{Z}$ is bd-admissible!

How do we find these bd-sets? Graphical Condition

$P(y \mid do(x))$ is identifiable if there is a set $\boldsymbol{Z}$ that d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$ in $G_{\underline{X}}$ (the graph G where all arrows emanating from X are removed.)



G

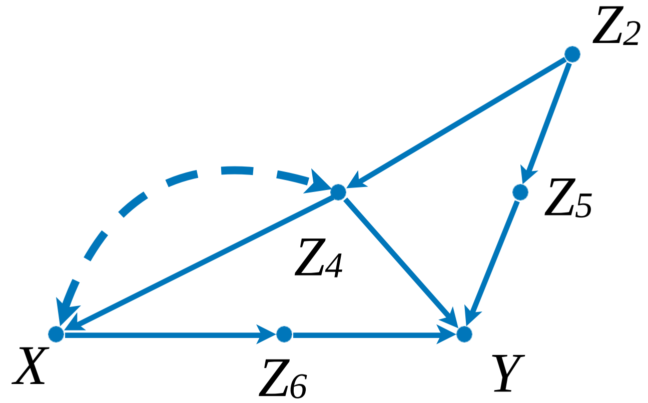


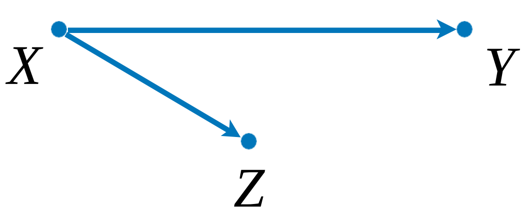
$G_{\underline{X}}$

$$P(y \mid do(x)) = \sum_{z_1, z_4} P(y \mid x, z_1, z_4) P(z_1, z_4)$$

Back-door Examples

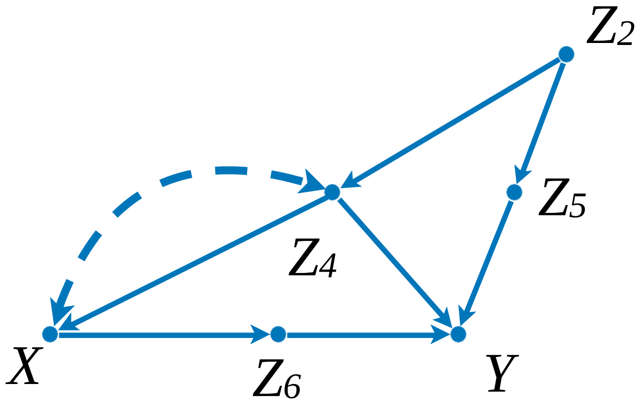
Are there admissible back-door sets (relative to $\boldsymbol{X},\boldsymbol{Y}$) for the following graphs?



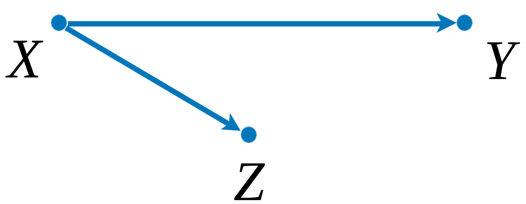


Back-door Examples

Are there admissible back-door sets (relative to $\boldsymbol{X},\boldsymbol{Y}$) for the following graphs?



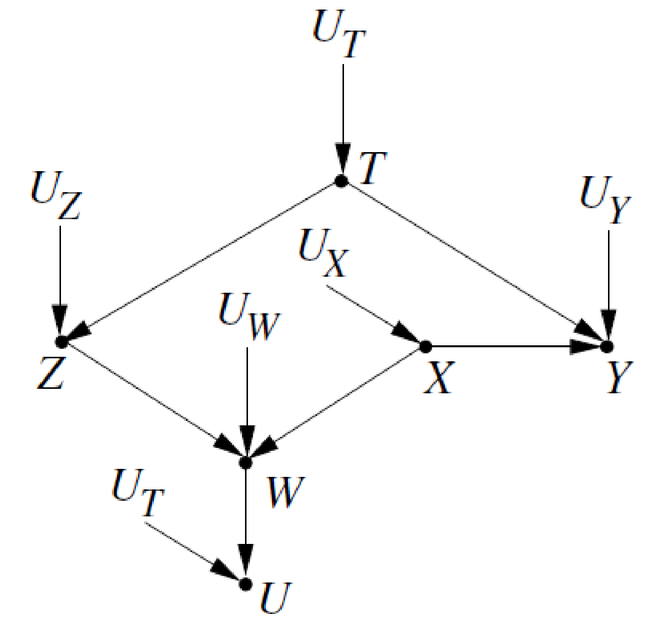
slides10 276 2024



$$\boldsymbol{Z} = \{Z_4, Z_2\}, \{Z_4, Z_5\}, \{Z_4, Z_2, Z_5\}$$

$$\boldsymbol{Z} = \emptyset$$

Examples



P(Y|do(X))?

No backdoors between X and Y and therefore: P(Y|do(X))= P(Y|X)

What if we adjust for W? … wrong!!!

But what if we want to determine P(Y|do(X),w)? What do we do with the spurious path $X \to W \leftarrow Z$ <- $T \to Y$?

if we condition on T, we would block the spurious path $X \to W \leftarrow Z$ <- $T \to Y$. We can compute:

$$P(Y=y|do(X=x),W=w) = \sum_t P(Y=y|X=x,W=w,T=t)P(T=t|W=w)$$

Example: W can be post-treatment pain

Adjusting for Colliders?

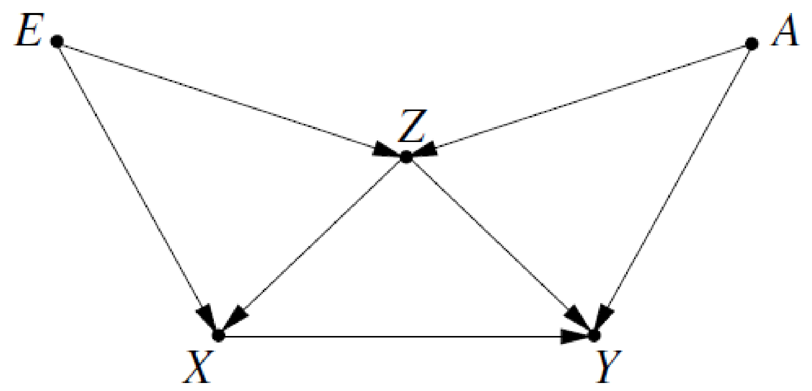
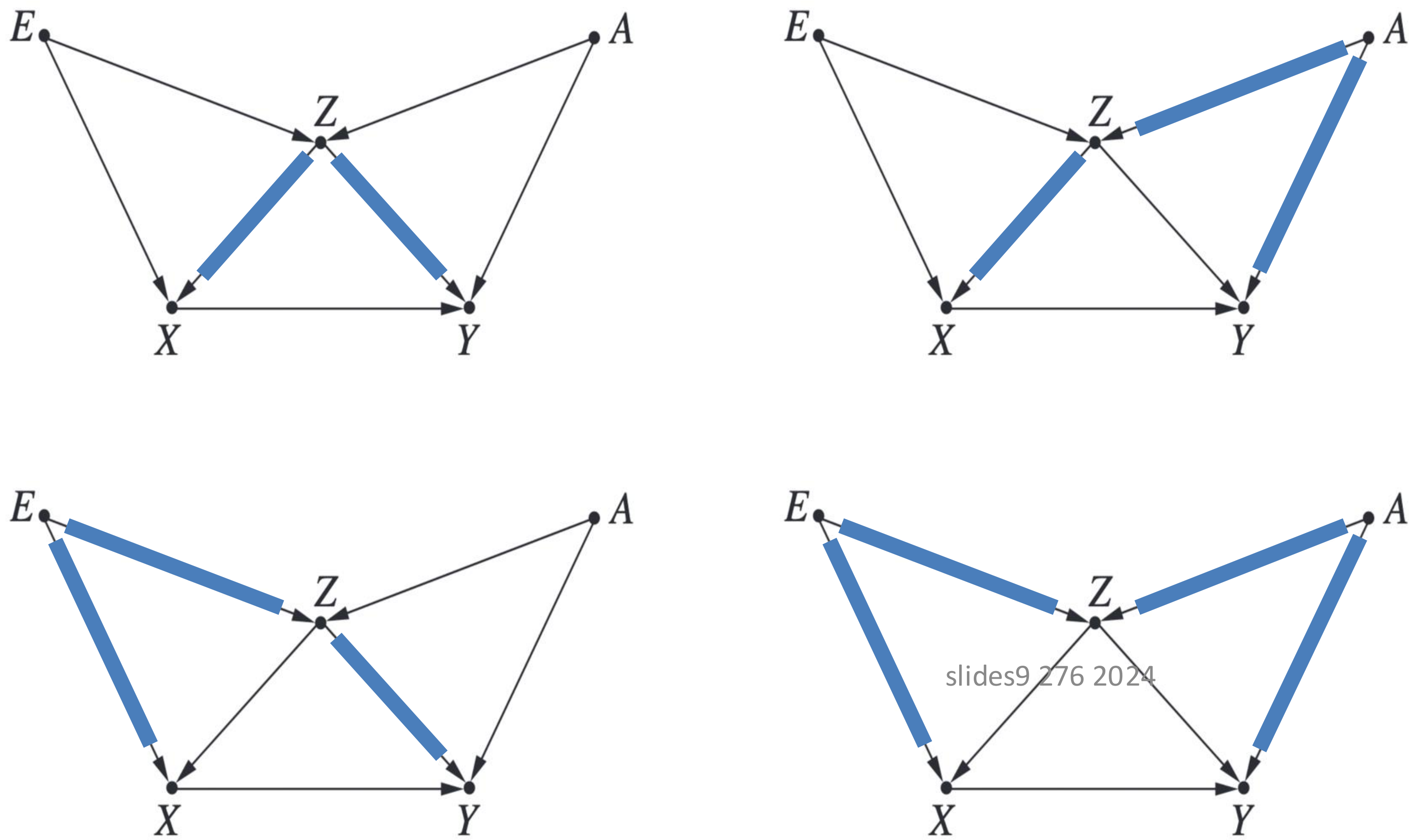


Figure 3.7: A graphical model in which the backdoor criterion requires that we condition on a collider (Z) in order to ascertain the effect of X on Y

There are 4 backdoor paths. We must adjust for Z, and one of E or A or both

Example: Backdoor

Backdoor for the effect of X on Y



enumerating backdoor paths

backdoor 1: *A, Z*

backdoor 2: *E, Z*

backdoor 3: *A, E, Z*

Outline

The backdoor criterion and the adjustment formula
Computing bd: Inverse probability weighting
Conditional intervention
Front door condition
The do calculus

Outline

Evaluating BD Adjustment

- The backdoor provides a criterion for deciding *when* a set of covariates $\boldsymbol{Z}$ is admissible for adjustment, i.e.,

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z})$$

- In practice, how should backdoor expressions be evaluated?
- There are sample & computational challenges entailed by the eval. of such expressions since one needs to
  - estimate the different distributions, and
  - evaluate them, summing over a possibly high-dimensional $\boldsymbol{Z}$ (i.e., time $O(exp(|\boldsymbol{Z}|))$).

Is it really exp in Z?

Inverse Probability Weighting (IPW)

- Let's rewrite the bd-expression,

$$P(\mathbf{y} \mid do(\mathbf{X}=\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) P(\mathbf{z})$$

$$= \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{x}, \mathbf{z})}{P(\mathbf{x}, \mathbf{z})} P(\mathbf{z})$$

$$= \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{x}, \mathbf{z})}{P(\mathbf{x} \mid \mathbf{z}) P(\mathbf{z})} P(\mathbf{z})$$

$$= \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{x}, \mathbf{z})}{P(\mathbf{x} \mid \mathbf{z})}$$

Entries of the joint distribution

Fit a function $g(z)$ that approximates this probability

Inverse Propensity score

Inverse Probability Weighting (IPW)

- Assume we have N samples, then

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \frac{P(\mathbf{y}, \mathbf{x}, \mathbf{z})}{P(\mathbf{x} \mid \mathbf{z})}$$

$$= \sum_{\mathbf{z}} \frac{\frac{1}{N} \sum_{i=1}^{N} 1_{\mathbf{Y}_i=\mathbf{y}, \mathbf{X}_i=\mathbf{x}, \mathbf{Z}_i=\mathbf{z}}}{g(\mathbf{z})}$$

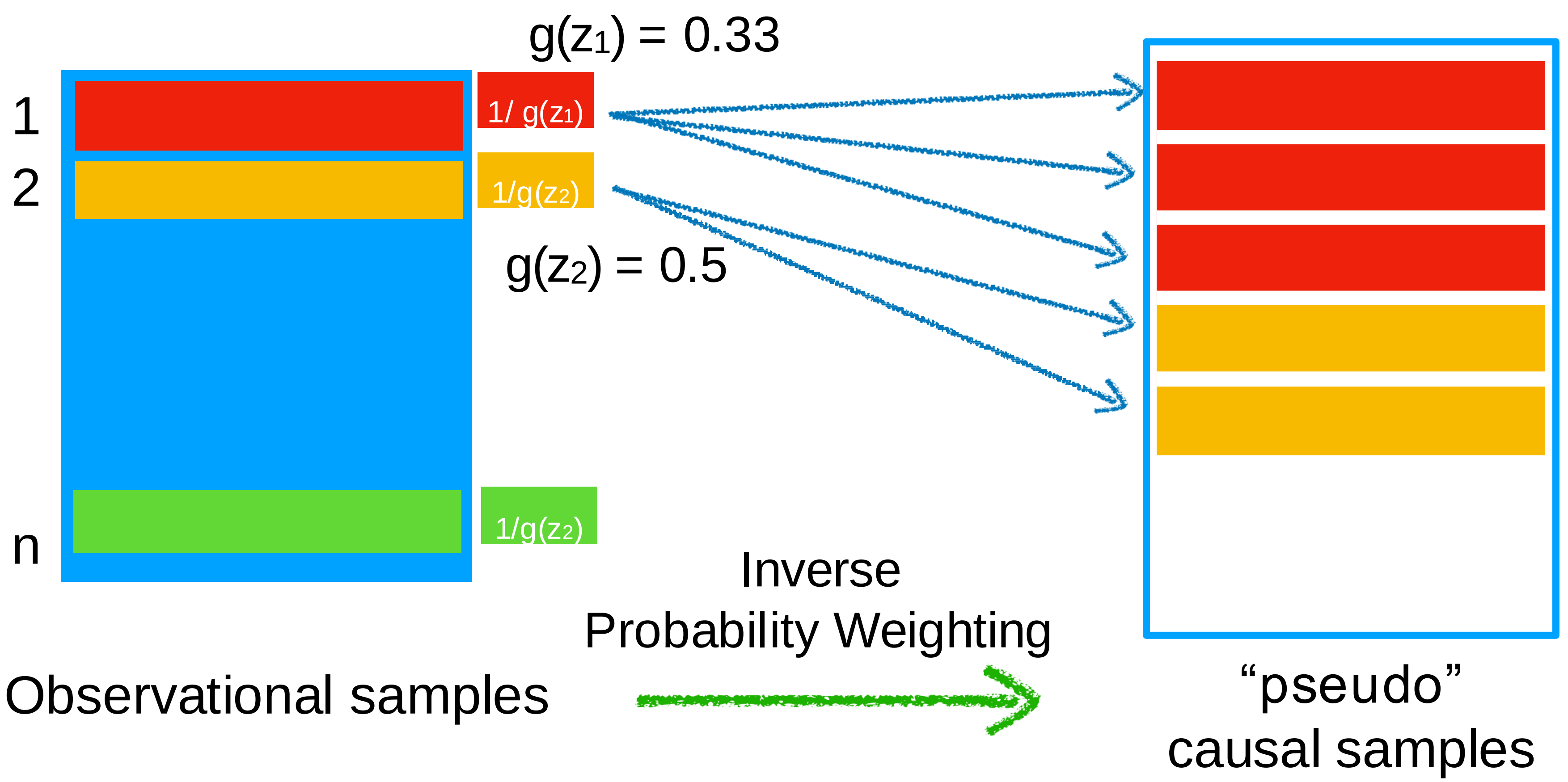
$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{\mathbf{z}} \frac{1_{\mathbf{Y}_i=\mathbf{y}, \mathbf{X}_i=\mathbf{x}, \mathbf{Z}_i=\mathbf{z}}}{g(\mathbf{z})}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \frac{1_{\mathbf{Y}_i=\mathbf{y}, \mathbf{X}_i=\mathbf{x}, \mathbf{Z}_i=\mathbf{z}}}{g(\mathbf{z})}$$

Requires time proportional to the number of samples N

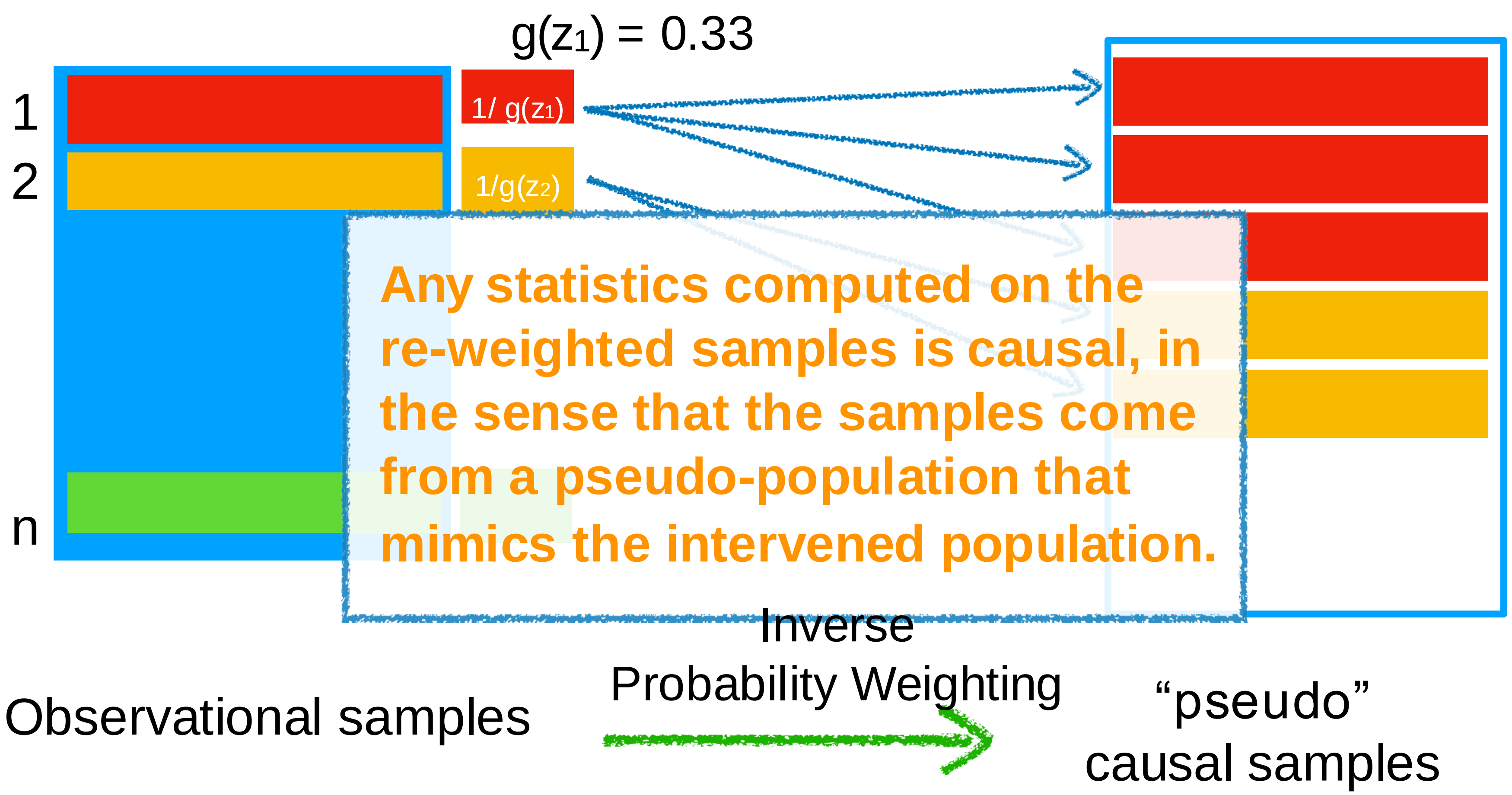
Inverse Probability Weighting (IPW)

- In practice, evaluating the expr. can be seen as: $\frac{1}{N}\sum_{i=1}^{N}\frac{1_{\mathbf{Y}_i=\mathbf{y},\mathbf{X}_i=\mathbf{x},\mathbf{Z}_i=\mathbf{z}}}{g(\mathbf{z})}$



Inverse Probability Weighting (IPW)

- In practice, evaluating the expr. can be seen as: $$\frac{1}{N}\sum_{i=1}^{N}\frac{1_{\mathbf{Y}_i=\mathbf{y},\mathbf{X}_i=\mathbf{x},\mathbf{Z}_i=\mathbf{z}}}{g(\mathbf{z})}$$



Inverse Probability Weighting (IPW)

This provides us with a simple procedure of estimating $P(Y = y|do(X = x))$ when we have finite samples. If we weigh each available sample by a factor = $1/P(X = x|Z = z)$, we can then treat the reweighted samples as if they were generated from P_m, not P, and proceed to estimate $P(Y = y|do(x))$ accordingly.

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug-gender-recovery story of Chapter 1 (Table 1.1)

X	Y	Z	% of population
Yes	Yes	Male	0.116
Yes	Yes	Female	0.274
Yes	No	Male	0.01
Yes	No	Female	0.101
No	Yes	Male	0.334
No	Yes	Female	0.079
No	No	Male	0.051
No	No	Female	0.036

X=yes, and normalizing (dividing by 0.49)

Table 3.4 Conditional probability distribution $P(Y, Z|X)$ for drug users ($X = yes$) in the population of Table 3.3

X	Y	Z	% of population
Yes	Yes	Male	0.232
Yes	Yes	Female	0.547
Yes	No	Male	0.02
Yes	No	Female	0.202

Rewighting by 1/P(x=yes|Z=male) =0.233
Or P(X=yes|Z=female)= 0.765

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $do(X = Yes)$, determined via the inverse probability method

X	Y	Z	% of population
Yes	Yes	Male	0.476
Yes	Yes	Female	0.357
Yes	No	Male	0.041
Yes	No	Female	0.132

This will provide saving if the number of samples is far smaller than domain of Z

Here
P(Y|DO(X=yes) =
0.476+0.357=0.833

Outline

Computing bd: Inverse probability weighting
Conditional intervention
Front door condition
The do calculus

Conditional Intervention

Suppose a policy maker contemplates an age-dependent policy whereby an amount x of drug is to be administered to patients, depending on their age Z. We write it as $do(X = g(Z))$.
To find out the distribution of outcome Y that results from this policy, we seek to estimate $P(Y = y|do(X = g(Z)))$.

We can often get it through z-specific effect of P(Y|do(X=x),Z=z)

Rule 2 *The z-specific effect $P(Y = y|do(X = x), Z = z)$ is identified whenever we can measure a set S of variables such that $S \cup Z$ satisfies the backdoor criterion. Moreover, the z-specific effect is given by the following adjustment formula*

$$P(Y = y|do(X = x), Z = z)$$
$$= \sum_s P(Y = y|X = x, S = s, Z = z)P(S = s)$$

Conditional Intervention

We now show that identifying the effect of such policies is equivalent to identifying the expression for the z-specific effect $P(Y = y|do(X = x), Z = z)$.

To compute $P(Y = y|do(X = g(Z)))$, we condition on $Z = z$ and write

$$
\begin{aligned}
& P(Y = y|do(X = g(Z))) \\
& = \sum_z P(Y = y|do(X = g(Z)), Z = z) P(Z = z|do(X = g(Z))) \\
& = \sum_z P(Y = y|do(X = g(z)), Z = z) P(Z = z) \qquad (3.17)
\end{aligned}
$$

The equality

$$P(Z = z|do(X = g(Z))) = P(Z = z)$$

stems, of course, from the fact that Z occurs before X; hence, any control exerted on X can have no effect on the distribution of Z. Equation (3.17) can also be written as

$$\sum_z P(Y = y|do(X = x), z)|_{x=g(z)} P(Z = z)$$

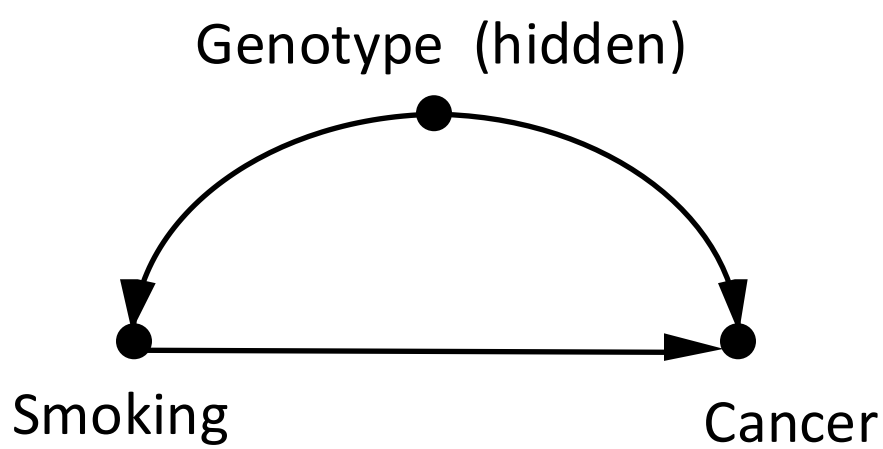
which tells us that the causal effect of a conditional policy $do(X = g(Z))$ can be evaluated directly from the expression of $P(Y = y|do(X = x), Z = z)$ simply by substituting $g(z)$ for x and taking the expectation over Z (using the observed distribution $P(Z = z)$).

So if we can compute conditional interventions we can compute conditional policies.

Outline

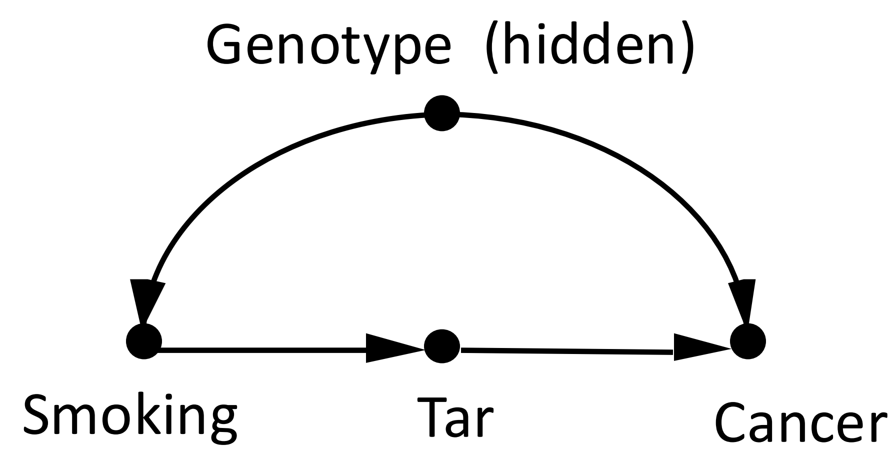
Computing bd: Inverse probability weighting
Conditional intervention
Front door condition
The do calculus

Incompleteness of Backdoor Criterion



no backdoor

causal effect **is not identifiable**



no backdoor

causal effect **is identifiable!**

causal effect of smoking on cancer $Pr(c|do(s))$

Reminder: Truncated Product in Semi-Markovian Models

The distribution generated by an intervention *do(**X**=**x**)* in a Semi-Markovian model M is given by the (generalized) truncated factorization product, namely,

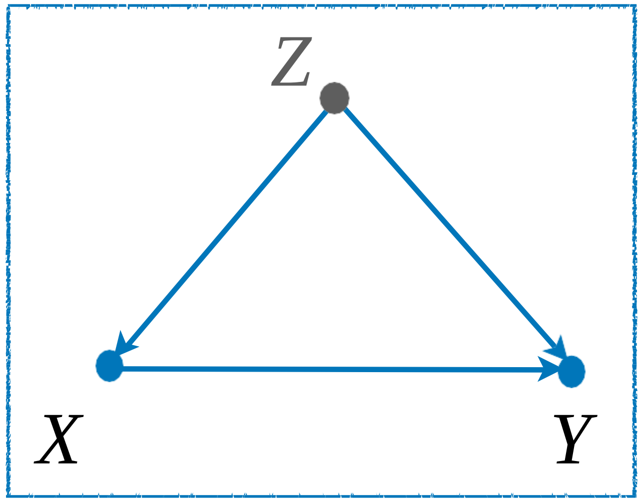
$$P(\mathbf{v} \mid do(\mathbf{x})) = \sum_{\mathbf{u}} \prod_{\{V_i \in \mathbf{V} \setminus \mathbf{X}\}} P\left(v_i \mid pa_i, u_i\right) P(\mathbf{u})$$

And the effect of such intervention on a set Y is

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{v} \setminus (\mathbf{y} \cup \mathbf{x})} \sum_{\mathbf{u}} \prod_{\{V_i \in \mathbf{V} \setminus \mathbf{X}\}} P\left(v_i \mid pa_i, u_i\right) P(\mathbf{u})$$

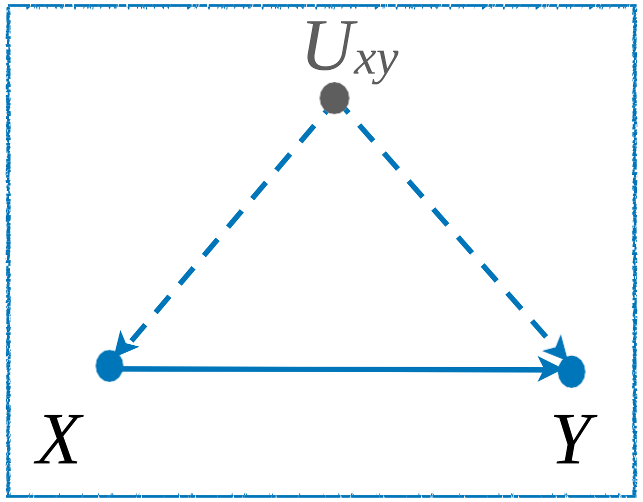
Interventions - Another Example

Real world



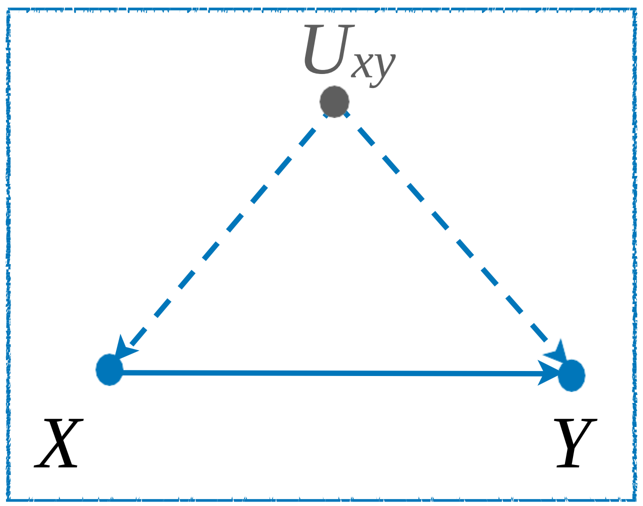
Interventions - Another Example

Real world



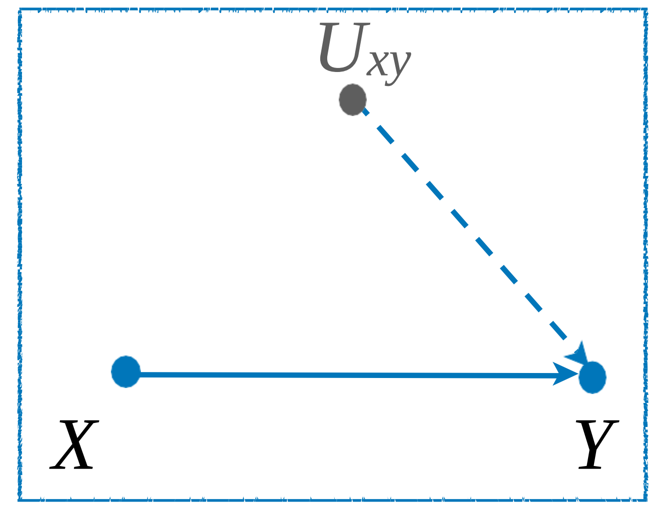
Interventions - Another Example

Real world



Intervention

Alternative world



$$M = \begin{cases} X \leftarrow f_X(u_{xy}, u_x) \\ Y \leftarrow f_Y(x, u_{xy}, u_y) \end{cases}$$

$do(X{=}x)$

$$M_x = \begin{cases} \cancel{X \leftarrow f_X(u_{xy}, u_x)} \quad X = x \\ Y \leftarrow f_Y(x, u_{xy}, u_y) \end{cases}$$

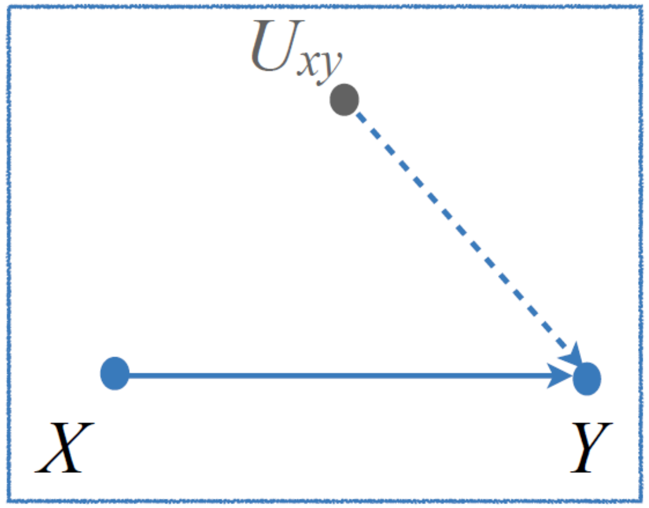
Interventions - Another Example

Re-writing the interventional distribution,

$$P(\mathbf{v} \mid do(x)) = \sum_{u_x, u_y, u_{xy}} \cancel{P(x \mid u_{xy}, u_x)} P(y \mid x, u_{xy}, u_y) P(u_x, u_y, u_{xy})$$

$$= \left(\sum_{u_{xy}} \left(\sum_{u_y} P(y \mid x, u_{xy}, u_y) P(u_y) \right) P(u_{xy}) \right) \left(\sum_{u_x} P(u_x) \right)$$

Alternative world



$$P(y \mid do(x)) = \sum_{u_{xy}} \underline{P(y \mid x, u_{xy}) P(u_{xy})}$$

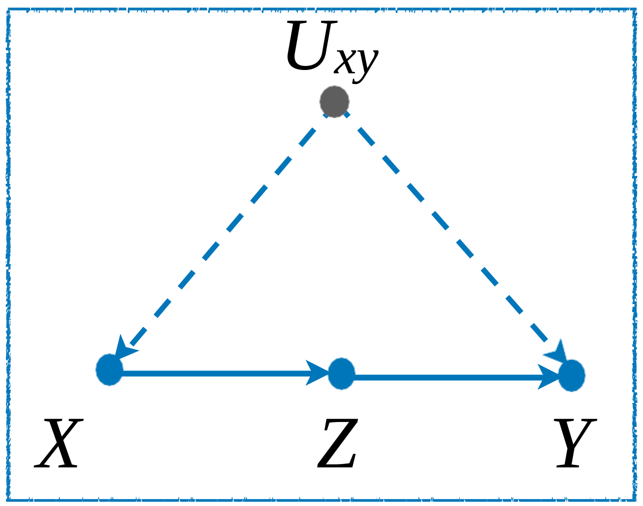
We can get rid of U_y
But not of U_xy

$$M_x = \begin{cases} \cancel{X \leftarrow f_X(u_{xy}, u_x)} \; X = x \\ Y \leftarrow f_Y(x, u_{xy}, u_y) \end{cases}$$

These distributions are not observed,
and nothing more can be removed.

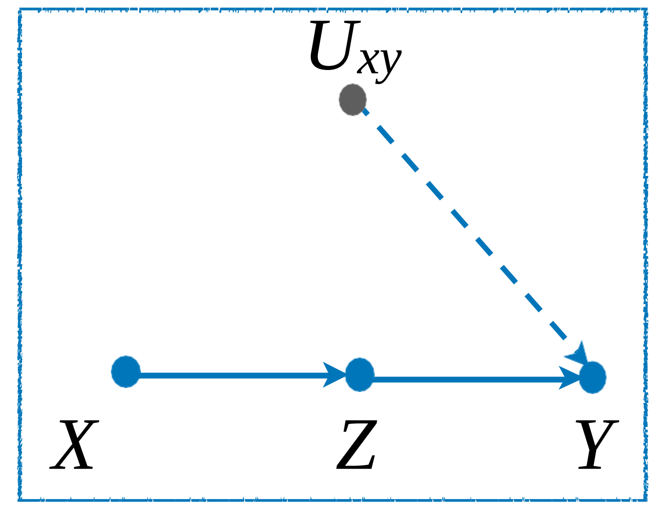
The Front-door Case

Real world



intervention

Alternative world



$$M = \begin{cases} X \leftarrow f_X(u_{xy}, u_x) \\ Z \leftarrow f_Z(x, u_z) \\ Y \leftarrow f_Y(z, u_{xy}, u_y) \end{cases}$$

$do(X{=}x)$

$$M_x = \begin{cases} \cancel{X \leftarrow f_X(u_{xy}, u_x)} \quad X = x \\ Z \leftarrow f_Z(x, u_z) \\ Y \leftarrow f_Y(z, u_{xy}, u_y) \end{cases}$$

$$P(\mathbf{v}) = \sum_{\mathbf{u}} P(x \mid u_{xy}, u_x) P(z \mid x, u_z) P(y \mid z, u_{xy}, u_y) P(\mathbf{u})$$

$$P(\mathbf{v} \mid do(x)) = \sum_{\mathbf{u}} \cancel{P(x \mid u_{xy}, u_x)} P(z \mid x, u_z) P(y \mid z, u_{xy}, u_y) P(\mathbf{u})$$

The Front-door Case

1. $(Y \perp\!\!\!\perp X \mid Z, U_{xy})$
2. $(Z \perp\!\!\!\perp U_{xy} \mid X)$

Re-writing the interventional distribution…

$$P(\mathbf{v} \mid do(x)) = \sum_{\mathbf{u}} \cancel{P(x \mid u_{xy}, u_x)} P(z \mid x, u_z) P(y \mid z, u_{xy}, u_y) P(\mathbf{u})$$

$$= \left(\sum_{u_z} P(z \mid x, u_z) P(u_z) \right) \left(\sum_{u_{xy}, u_y} P(y \mid z, u_{xy}, u_y) P(u_{xy}, u_y) \right) \left(\sum_{u_x} P(u_x) \right)$$

$$= P(z \mid x) \sum_{u_{xy}} P(y \mid z, u_{xy}) P(u_{xy})$$

Summing over X

$$= P(z \mid x) \sum_{x', u_{xy}} P(y \mid z, u_{xy}) P(u_{xy} \mid x') P(x')$$

$(Y \perp\!\!\!\perp X \mid Z, U_{xy})$

$$= P(z \mid x) \sum_{x', u_{xy}} P(y \mid z, x', u_{xy}) P(u_{xy} \mid x') P(x')$$

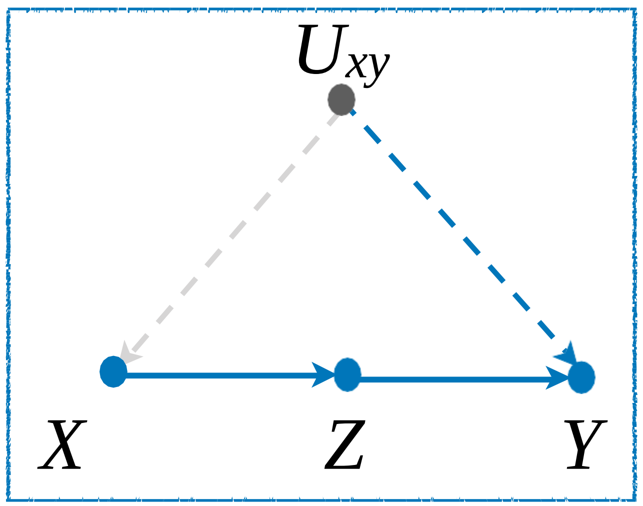
$(U_{xy} \perp\!\!\!\perp Z \mid X)$

$$= P(z \mid x) \sum_{x', u_{xy}} P(y \mid z, x', u_{xy}) P(u_{xy} \mid x', z) P(x')$$

Chain rule and sum out U_{xy}

$$= P(z \mid x) \sum_{x'} \sum_{u_{xy}} P(y, u_{xy} \mid z, x') P(x')$$

Alternative world



$$P(\mathbf{v} \mid do(x)) = P(z \mid x) \sum_{x'} P(y \mid z, x') P(x')$$

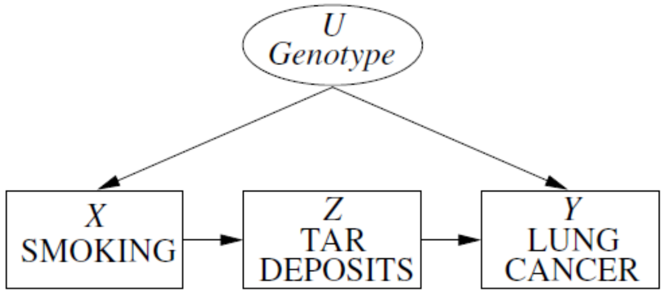
$$P(y \mid do(x)) = \sum_{z} \underline{P(z \mid x)} \sum_{x'} \underline{P(y \mid z, x') P(x')}$$

These factors can be computed from the observed distribution

Front-door Condition

Definition 3.4.1 (Front-Door)
A set of variables Z is said to satisfy the front-door *criterion relative to an ordered pair of variables (X, Y) if*

1. *Z intercepts all directed paths from X to Y.*
2. *There is no unblocked backdoor path from X to Z.*
3. *All backdoor paths from Z to Y are blocked by X.*



(b)

Theorem 3.4.1 (Front-Door Adjustment)
If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x'). \tag{3.16}$$

Example (Front-door)

Table 3.1: A hypothetical dataset of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar 400		No tar 400		All subjects 800	
	Smokers *380*	Nonsmokers *20*	Smokers *20*	Nonsmokers *380*	Smokers *400*	Nonsmokers *400*
No cancer	323 (85%)	1 (5%)	18 (90%)	38 (10%)	341 (85%)	39 (9.75%)
Cancer	57 (15%)	19 (95%)	2 (10%)	342 (90%)	59 (15%)	361 (90.25%)

Tobaco industry:
Only 15% of smoker developed cancer while 90% from the non-smoker

Antismoke lobbyist:
If you smoke you have 95% tar vs no smokers (380/400 vs 20/400)

Table 3.2 Reorganization of the dataset of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	SMOKERS 400		NON-SMOKERS 400		ALL SUBJECTS 800	
	Tar *380*	No tar *20*	Tar *20*	No tar *380*	Tar *400*	No tar *400*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (19%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (9%)	344 (81%)

If you have more tar, you increase the chance of cancer in both smoker (from 10% to 15%) and non-smokers (from 90% To 95%).

The Syntactical Goal on Identification of Causal Effects

- For both back- and front-door settings, the goal was to reduce the quantity $Q = P(\mathbf{y}|do(\mathbf{x}))$ into an expression with no $do(.)$, i.e., estimable from the observational distribution $P(\mathbf{v})$.

- We are interested in rules or a set of axioms that allow the systematic transformation of a $do(.)$ expression into a do-free expression while preserving the equivalence to the target effect.

Outline

Computing bd: Inverse probability weighting
Conditional intervention
Front door condition
The do calculus

Rules of Do-Calculus

Theorem 3.4.1. The following transformations are valid for any do-distribution induced by a causal model M:

Rule 1: Adding/removing Observations

$$P(y|do(x),z,w)=P(y|do(x),w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid W)_{G_{\overline{X}}}$$

Rule 2: Action/observation exchange

$$P(y|do(x),do(z),w)=P(y|do(x),z,w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid X, W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3: Adding/removing Actions

$$P(y|do(x),do(z),w)=P(y|do(x),w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$$

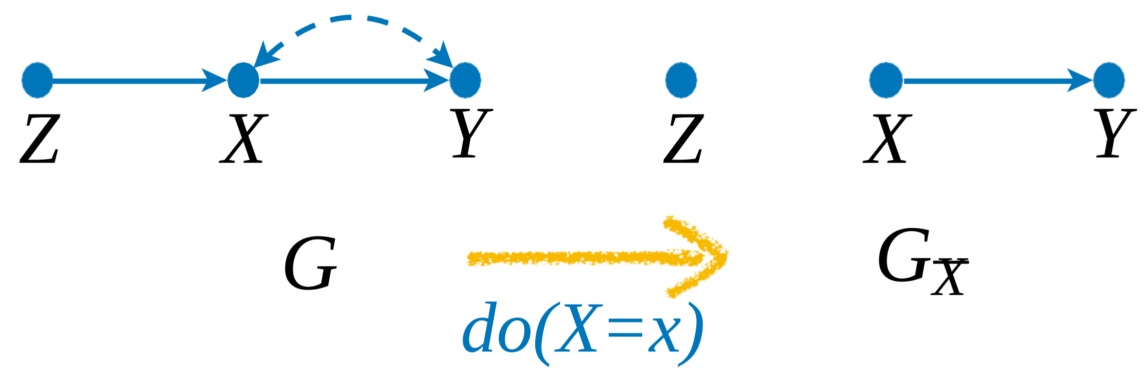
where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

Insight 1: Adding/removing Observations

- Adding/removing observations

In the original model, Z and Y may be not separable, e.g.:

$$(Z \not\perp\!\!\!\perp Y), (Z \not\perp\!\!\!\perp Y \mid X)$$



However, in the the do(X)-world (model M_x), Y and Z are d-separated, that is,

$$(Z \perp\!\!\!\perp Y)_{G_{\overline{X}}} \implies P(y|do(x),z)=P(y|do(x))$$

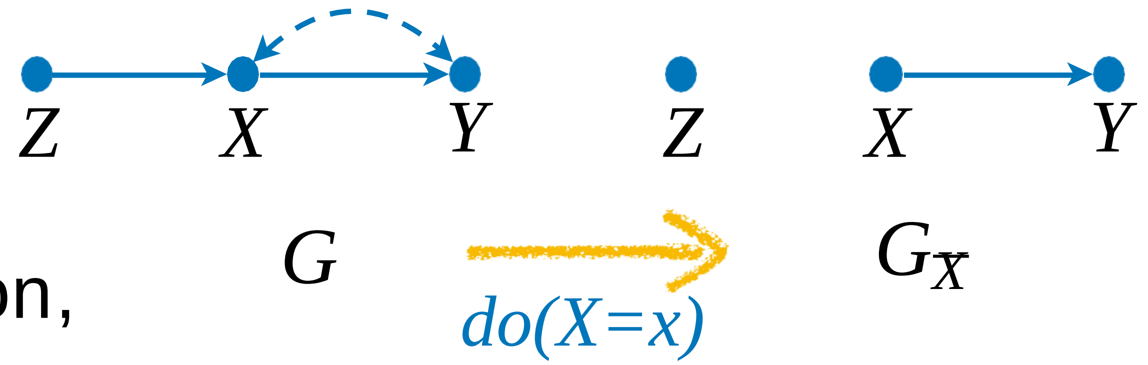
Let's verify this equality!

Try it yourself

Insight 1: Adding/removing Observations

- Adding/removing observations

$$P(y|do(x),z)=P(y|do(x))\ ?$$



First, let's write the interventional distribution,

$$P(\mathbf{v}\,|\,do(x))$$

$$= \sum_{\mathbf{u}} P(z\,|\,u_z)P(y\,|\,x,u_y,u_{xy})P(\mathbf{u})$$

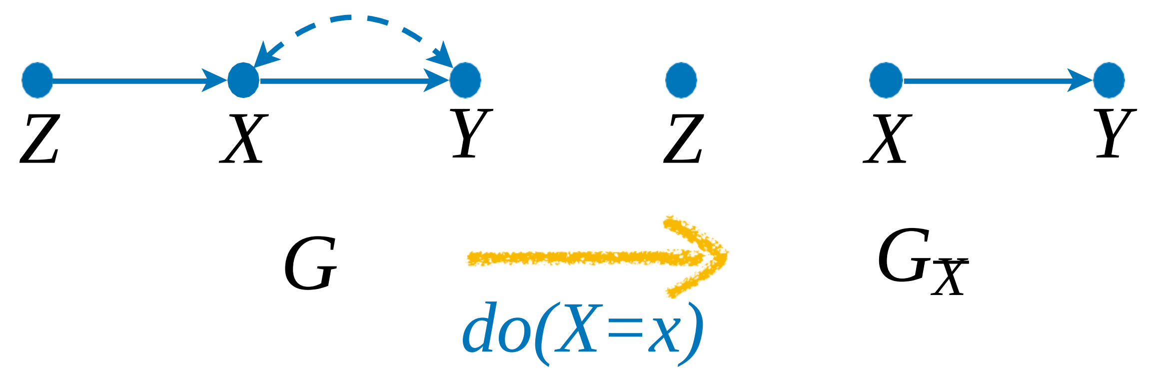
$$= P(z)\sum_{u_{xy}} P(y\,|\,x,u_{xy})P(u_{xy})$$

Let's keep the truncated in this form and …

Insight 1: Adding/removing Observations

- Adding/removing observations

$$P(y|do(x),z)=P(y|do(x))\ ?$$



And, let's rewrite the conditional effects,

$$P(y \mid do(x), z) = \frac{P(y, z \mid do(x))}{P(z \mid do(x))}$$

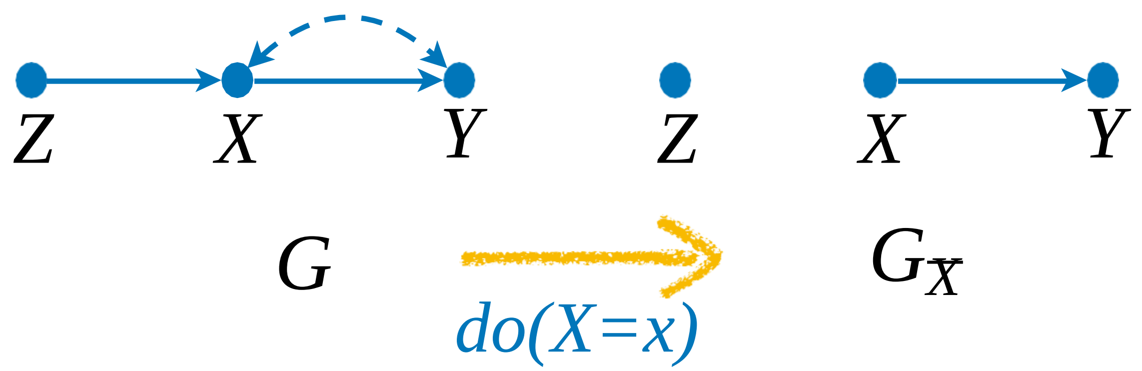
$$P(y, z \mid do(x)) = P(z) \sum_{u_{xy}} P(y \mid x, u_{xy}) P(u_{xy})$$

$$P(z \mid do(x)) = \sum_{y} P(z) \sum_{u_{xy}} P(y \mid x, u_{xy}) P(u_{xy}) = P(z)$$

Insight 1: Adding/removing Observations

- Adding/removing observations

$P(y|do(x),z)=P(y|do(x))$?



Substituting the factors back…

$$\boxed{P(y\,|\,do(x),z)} = \frac{P(z)\sum_{u_{xy}} P(y\,|\,x,u_{xy})P(u_{xy})}{P(z)}$$

$$= \sum_{u_{xy}} P(y\,|\,x,u_{xy})P(u_{xy})$$

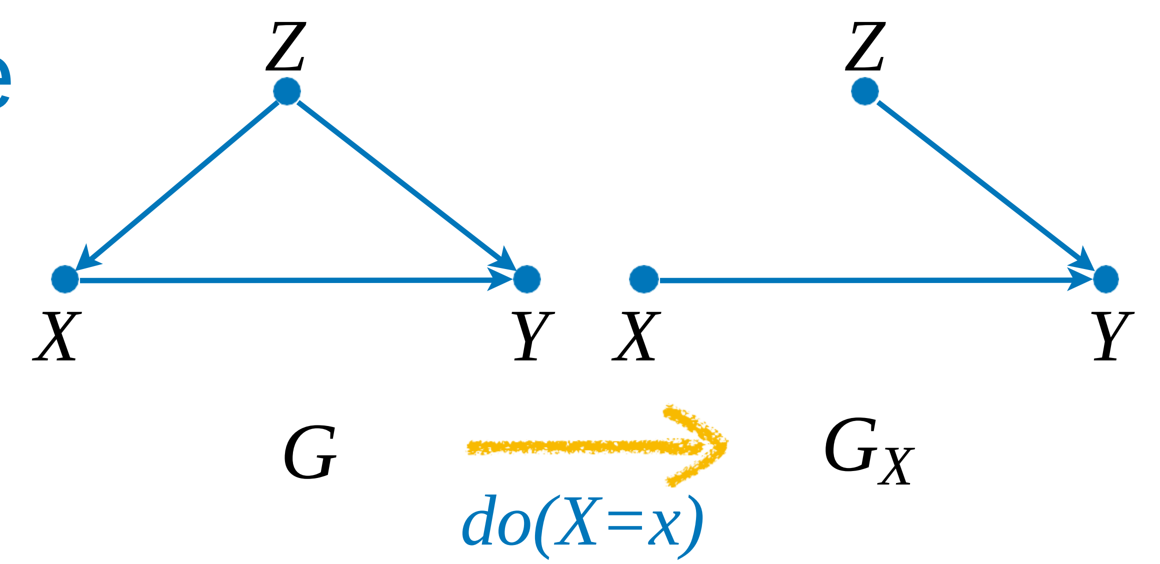
$$= \sum_{z} P(z) \sum_{u_{xy}} P(y\,|\,x,u_{xy})P(u_{xy})$$

$$= \sum_{z} P(\mathbf{v}\,|\,do(x)) \boxed{= P(y\,|\,do(x))}$$

$$(Z \perp\!\!\!\perp Y)_{G_{\overline{X}}}$$

Insight 2: Action/Observation Exchange

- Action/Observation Exchange



After observing Z, variable Y reacts to X in the same way, with and without intervention.

Note that given Z, Y is correlated with X only through causal paths, hence, $see(X=x)$ will be equiv. to $do(X=x)$.

Idea. If Z blocks all bd-paths w.r.t (X, Y), then cond. on Z, all the remaining association is equal to the causation.

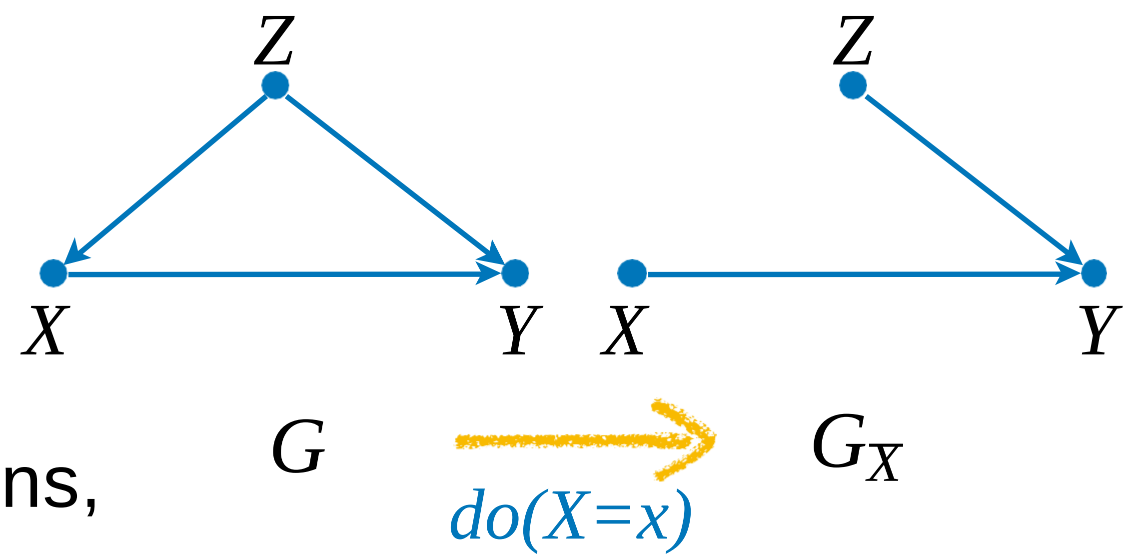
$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}} \quad \Longrightarrow \quad P(y|do(x),z)=P(y|x,z)$$

Let's verify this equality!

Insight 2: Action/Observation Exchange

- Action/Observation Exchange

$P(y|do(x),z)=P(y|x,z)$?



First, let's write the interventional distributions,

$$P(y,z|do(x)) = \sum_{\mathbf{u}} P(z|u_z)P(y|x,z,u_y)P(\mathbf{u}) = P(z)P(y|x,z)$$

$$P(z|do(x)) = \sum_{y} P(z)P(y|x,z) = P(z)$$

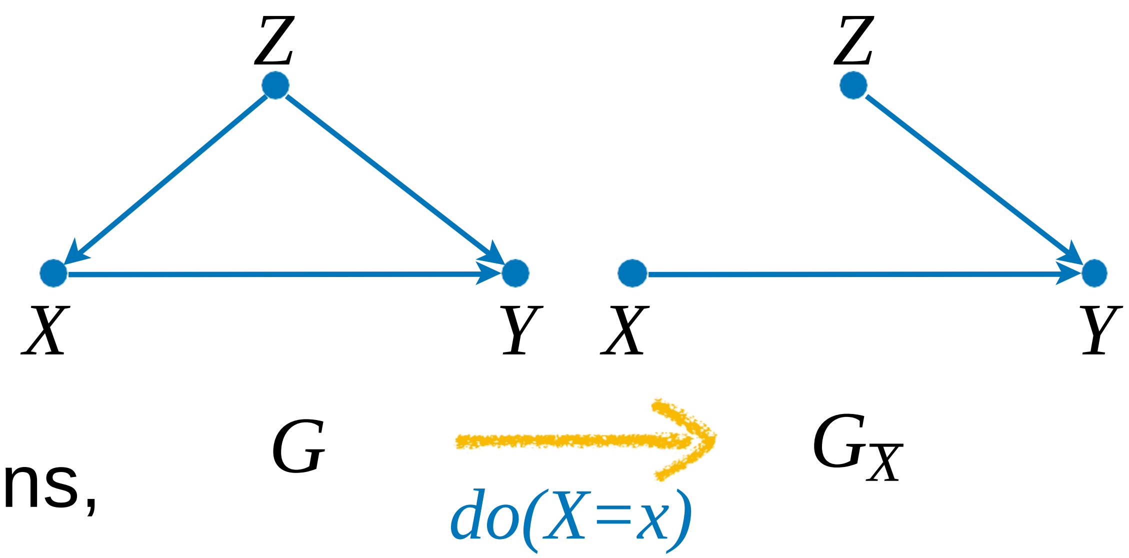
$$\boxed{P(y|do(x),z)} = \frac{P(z,y|do(x))}{P(z|do(x))} = \frac{P(z)P(y|x,z)}{P(z)} \boxed{= P(y|x,z)}$$

$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}}$$

Insight 2: Action/Observation Exchange

- Action/Observation Exchange

$P(y|do(x),z)=P(y|x,z)$?



First, let's write the interventional distributions,

$$P(y,z|do(x)) = \sum_{\mathbf{u}} P(z|u_z)P(y|x,z,u_y)P(\mathbf{u}) = P(z)P(y|x,z)$$

Looks familiar?

BD perhaps?

$$P(z|do(x)) = \sum_{y} P(z)P(y|x,z) = P(z)$$

$$\boxed{P(y|do(x),z)} = \frac{P(z,y|do(x))}{P(z|do(x))} = \frac{P(z)P(y|x,z)}{P(z)} \boxed{= P(y|x,z)}$$

$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}}$$

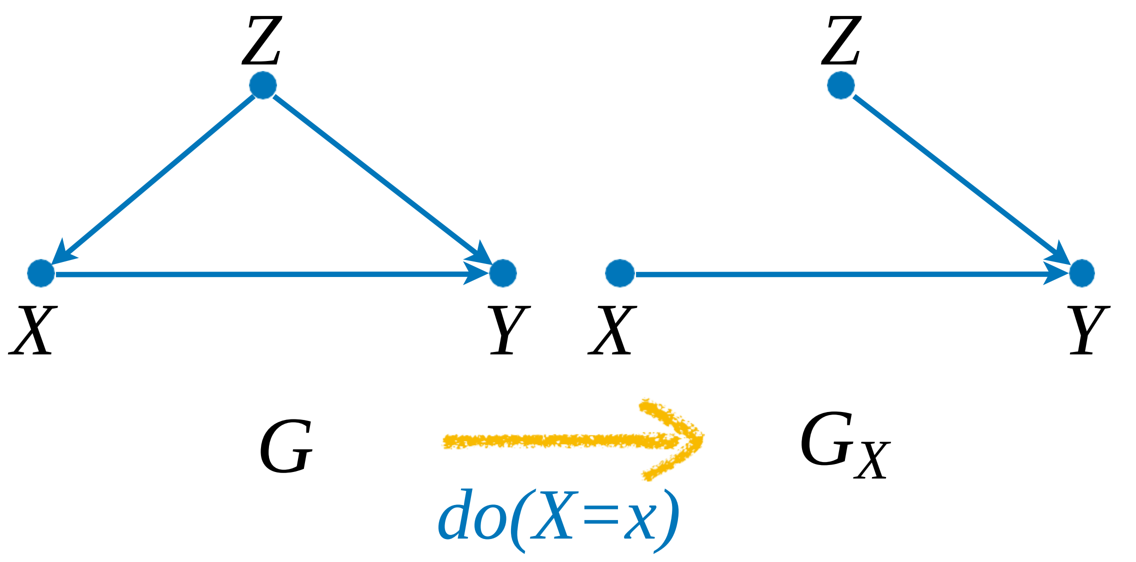
Insight 2: Action/Observation Exchange

- Action/Observation Exchange

Great, but what about the equality

$$P(y|do(x))=P(y|x)?$$

$$(Y \perp\!\!\!\perp X)_{G_{\underline{X}}}$$



Let's compare left and right-hand sides:

$$P(y \mid do(x)) = \sum_z \sum_{\mathbf{u}} P(y \mid x, z, u_y) P(z \mid u_z) P(\mathbf{u})$$

$$= \sum P(y \mid x, z) \boxed{P(z)}$$

$$P(y \mid x) = \sum_z P(y \mid x, z) \boxed{P(z \mid x)}$$

Almost any model compatible with this causal graph, $P(y|x)$ and $P(y \mid do(x))$ will **not** be equal since $P(z) \neq P(z \mid x)$ almost surely.

Rules of Do-Calculus

Theorem 3.4.1. The following transformations are valid for any do-distribution induced by a causal model M:

Rule 1: Adding/removing Observations

$$P(y|do(x),z,w)=P(y|do(x),w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid W)_{G_{\overline{X}}}$$

Rule 2: Action/observation exchange

$$P(y|do(x),do(z),w)=P(y|do(x),z,w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid X, W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3: Adding/removing Actions

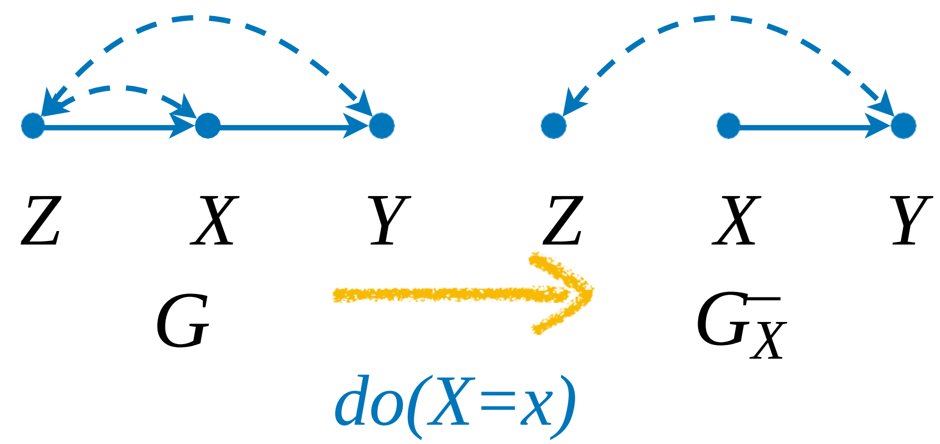
$$P(y|do(x),do(z),w)=P(y|do(x),w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$$

where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

Insight 3: Adding/Removing Actions

- Adding/Removing Actions

If there is no causal path from X to Z, then an intervention on X will have no effect on Z.



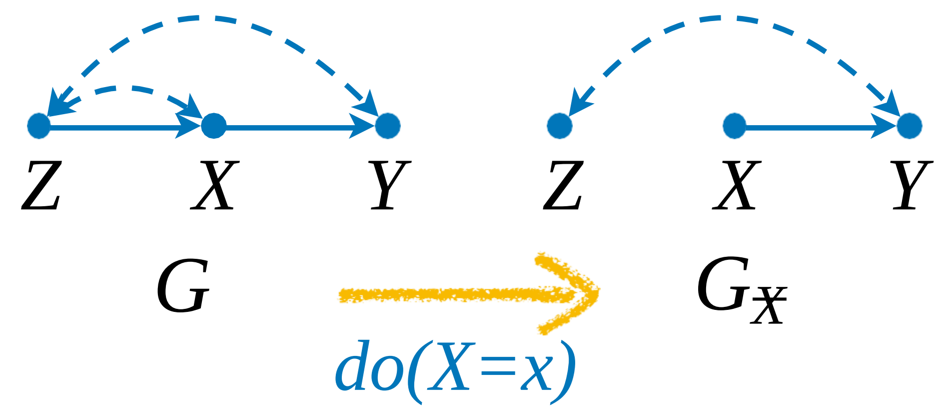
$$(Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \quad \Longrightarrow \quad P(z|do(x))=P(z)$$

Let's verify this equality!

Insight 3: Adding/Removing Actions

- Adding/Removing Actions

$$P(z|\textcolor{orange}{do(x)})=P(z)\ ?$$



$$\boxed{P(z\,|\,do(x))} = \sum_{y} P(\mathbf{v}\,|\,do(x))$$

$$= \sum_{y}\ \sum_{u_{zy},u_{zx}} P(z\,|\,u_{zy},u_{zx})P(y\,|\,x,u_{zy})P(u_{zy},u_{zx})$$

$$= \sum_{u_{zy},u_{zx}} P(z\,|\,u_{zy},u_{zx})P(u_{zy},u_{zx})$$

$$= \boxed{P(z)} \qquad (Z \perp\!\!\!\perp X)_{G_{\overline{X}}}$$

Rules of Do-Calculus

Theorem 3.4.1. The following transformations are valid for any do-distribution induced by a causal model M:

Rule 1: Adding/removing Observations

$$P(y|do(x),z,w)=P(y|do(x),w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid W)_{G_{\overline{X}}}$$

Rule 2: Action/observation exchange

$$P(y|do(x),do(z),w)=P(y|do(x),z,w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid X, W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3: Adding/removing Actions

$$P(y|do(x),do(z),w)=P(y|do(x),w) \quad \text{if} \quad (Z \perp\!\!\!\perp Y \mid X, W)_{G_{\overline{X}\overline{Z(W)}}}$$

where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

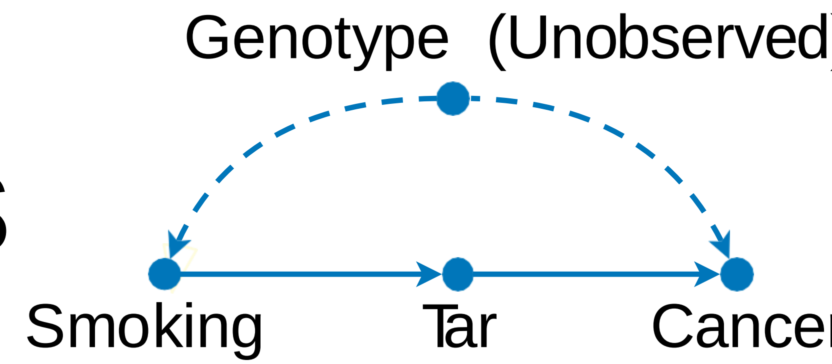
Properties of Do-Calculus

Theorem (soundness and completeness of do-calculus for causal identifiability from P(v)).

The causal quantity $Q = P(y|do(x))$ is identifiable from $P(v)$ and G if and only if there exists a sequence of application of the rules of do-calculus and the probability axioms that reduces Q into a do-free expression.

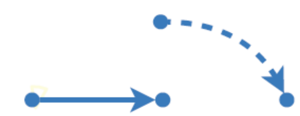
Syntactic goal: Re-express original Q without do()!

Derivation in Do-Calculus



$$P(c \mid do(s)) = \sum_t P(c \mid do(s), t) P(t \mid do(s))$$ Probability Axioms

$$= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s))$$ Rule 2 $(T \perp\!\!\!\perp C \mid S)_{G_{\underline{T}}}$

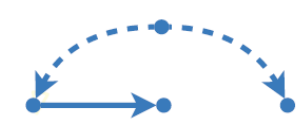


$$= \sum_t P(c \mid do(t)) P(t \mid do(s))$$ Rule 3 $(S \perp\!\!\!\perp C \mid T)_{G_{\overline{C,T}}}$

$$= \sum_t P(c \mid do(t)) P(t \mid s)$$ Rule 2 $(S \perp\!\!\!\perp T)_{G_{\underline{S}}}$

$$= \sum_t \sum_{s'} P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s)$$ Probability Axioms

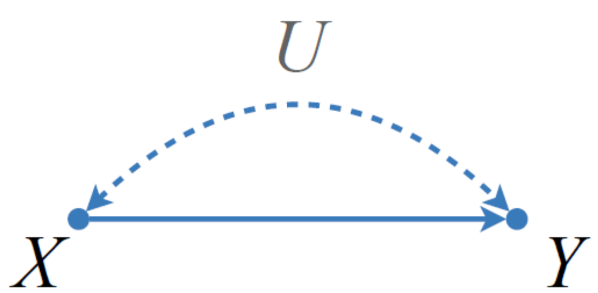
$$= \sum_t \sum_{s'} P(c \mid t, s') P(s' \mid do(t)) P(t \mid s)$$ Rule 2 $(T \perp\!\!\!\perp C \mid S)_{G_{\underline{T}}}$



$$= \sum_t \sum_{s'} P(c \mid t, s') P(s') P(t \mid s)$$ Rule 3 $(T \perp\!\!\!\perp S)_{G_{\overline{T}}}$

Example. Non-identifiable Effect

- Let M be a model compatible with G and inducing an observational distribution $P(\mathbf{v})$:



$$M = \begin{cases} X \leftarrow U \\ Y \leftarrow f_Y(X, U) \end{cases}$$

$U \sim P(u) \quad P(x, y) > 0$

$$M^{(1)} = \begin{cases} X \leftarrow U_{xy} \\ Y \leftarrow \begin{cases} f_Y(X, U) & \textbf{if } U = X \\ 0 & \textbf{otherwise} \end{cases} \end{cases}$$

$$M^{(2)} = \begin{cases} X \leftarrow U_{xy} \\ Y \leftarrow \begin{cases} f_Y(X, U) & \textbf{if } U = X \\ 1 & \textbf{otherwise} \end{cases} \end{cases}$$

$P^{(1)}(U) = P^{(2)}(U) = P(U)$

- Without intervention, U is always equal to X in both models, hence Y always outputs $f_Y(X, U)$ and $P^{(1)}(\mathbf{v}) = P^{(2)}(\mathbf{v}) = P(\mathbf{v})$.

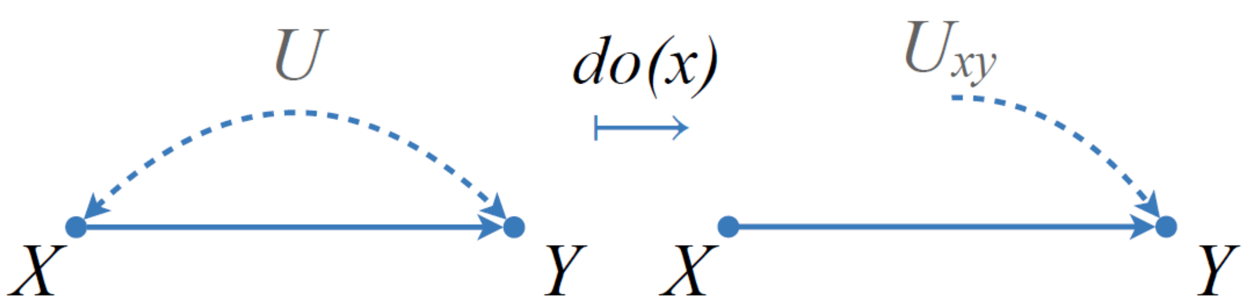
$1[x = u]$

$$\begin{aligned} &P^{(i)}(x, y) \\ &= \sum_u \boxed{P^{(i)}(x \mid u)} P^{(i)}(y \mid x, u) P(u) \\ &= P^{(i)}(y \mid x, U = x) P(U = x) \\ &= P(y \mid x) P(x) \\ &= \boxed{P(x, y)} \end{aligned}$$

Both models induce the same graph G and have the same $P(\mathbf{v})$

Example. Non-identifiable Effect

- Let M be a model compatible with G and inducing an observational distribution $P(\mathbf{v})$:



$$M_x^{(1)} = \begin{cases} X \leftarrow \cancel{U_{xy}}\ x \\ Y \leftarrow \begin{cases} f_Y(X, U) & \textbf{if } U = X \\ 0 & \textbf{otherwise} \end{cases} \end{cases}$$

$$M_x^{(2)} = \begin{cases} X \leftarrow \cancel{U_{xy}}\ x \\ Y \leftarrow \begin{cases} f_Y(X, U) & \textbf{if } U = X \\ 1 & \textbf{otherwise} \end{cases} \end{cases}$$

$P^{(1)}(U) = P^{(2)}(U) = P(U)$

- Under intervention $do(X=x)$, U and X do not need to match, hence $M_x^{(1)}$ and $M_x^{(2)}$ will output $Y=1$ with different probability:

$$P^{(i)}(y \mid do(x)) \qquad \textit{0 in } M_x^{(1)}, \textit{ 1 in } M_x^{(2)}$$

$$= \sum_u P^{(i)}(y \mid x, u) P(u)$$

$$= P^{(i)}(y \mid x, U = x) P(U = x)$$

$$+ \boxed{P^{(i)}(y \mid x, U \neq x)} P(U \neq x)$$

$$= P(y \mid x) P(x) + \boxed{1[i = 1](1 - P(x))}$$

Even though both models induce the same graph G and have the same $P(\mathbf{v})$, the causal effect $P^{(1)}(y|do(x)) \neq P^{(2)}(y|do(x))$!

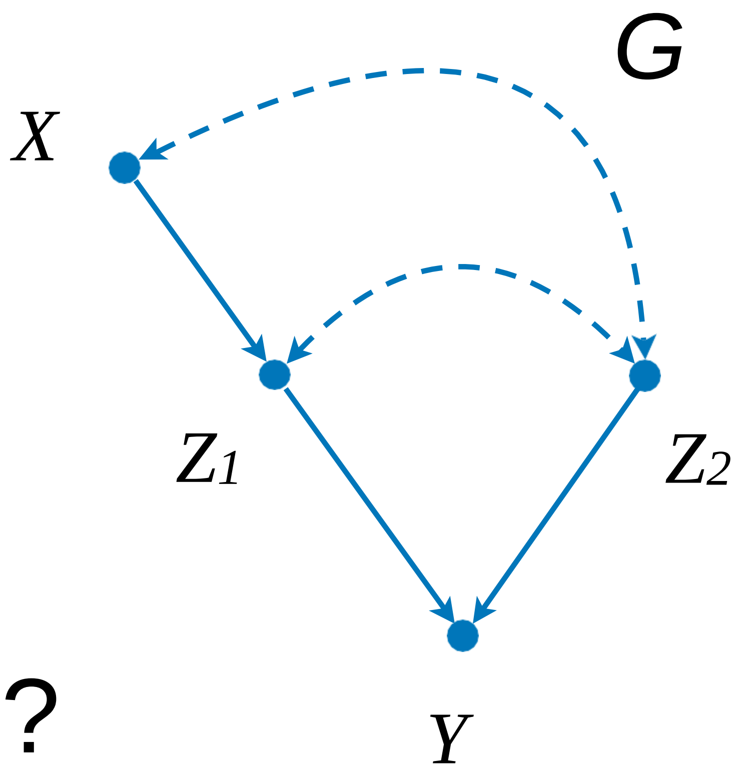
Non-identifiability Machinery

Lemma (Graph-subgraph ID (Tian and Pearl, 2002))

- If $Q = P(y \mid do(x))$ is not identifiable in G, then Q is not identifiable in the graph resulting from adding a directed or bidirected edge to G.
- Converse. If $Q = P(y|do(x))$ is identifiable in G, Q is still identifiable in the graph resulting from removing a directed or bidirected edge from G.

Non-identifiability Puzzle

- Is P(y | do(x)) identifiable from *G*?
  - Is *G* of bow-shape?



- Is P(y | do(x), z2) identifiable from *G*?
- Is P(y | do(x, z1)) identifiable from *G*?

Try it at home

P(Y|do(x) is not identifiable
But when conditioning on Z_1, or Z_2 they are.
So, computing the effect of a joint intervention can be easier than
Their individual interventions.

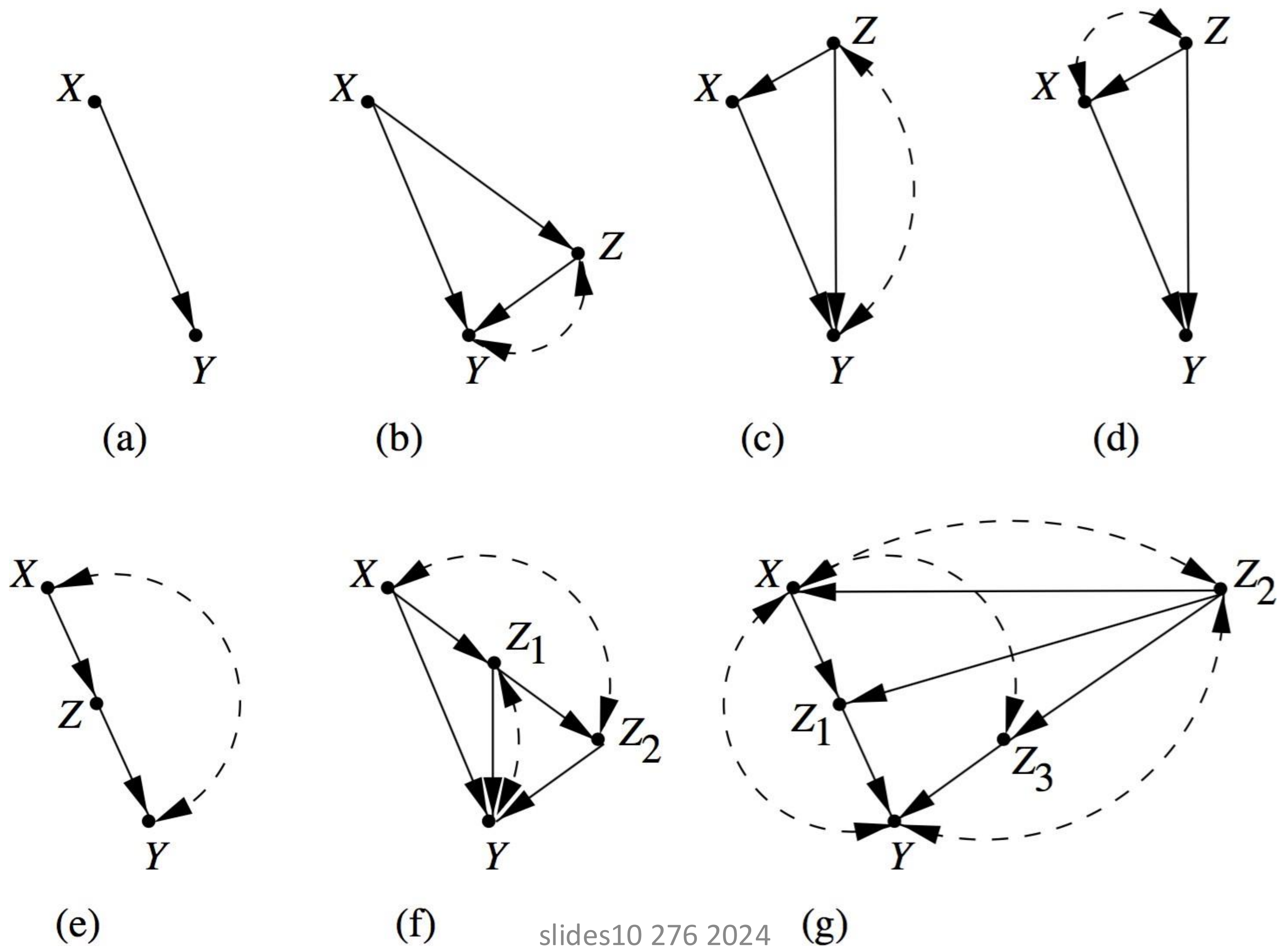
[C] sec 35.

Non-Identifiability Criterion

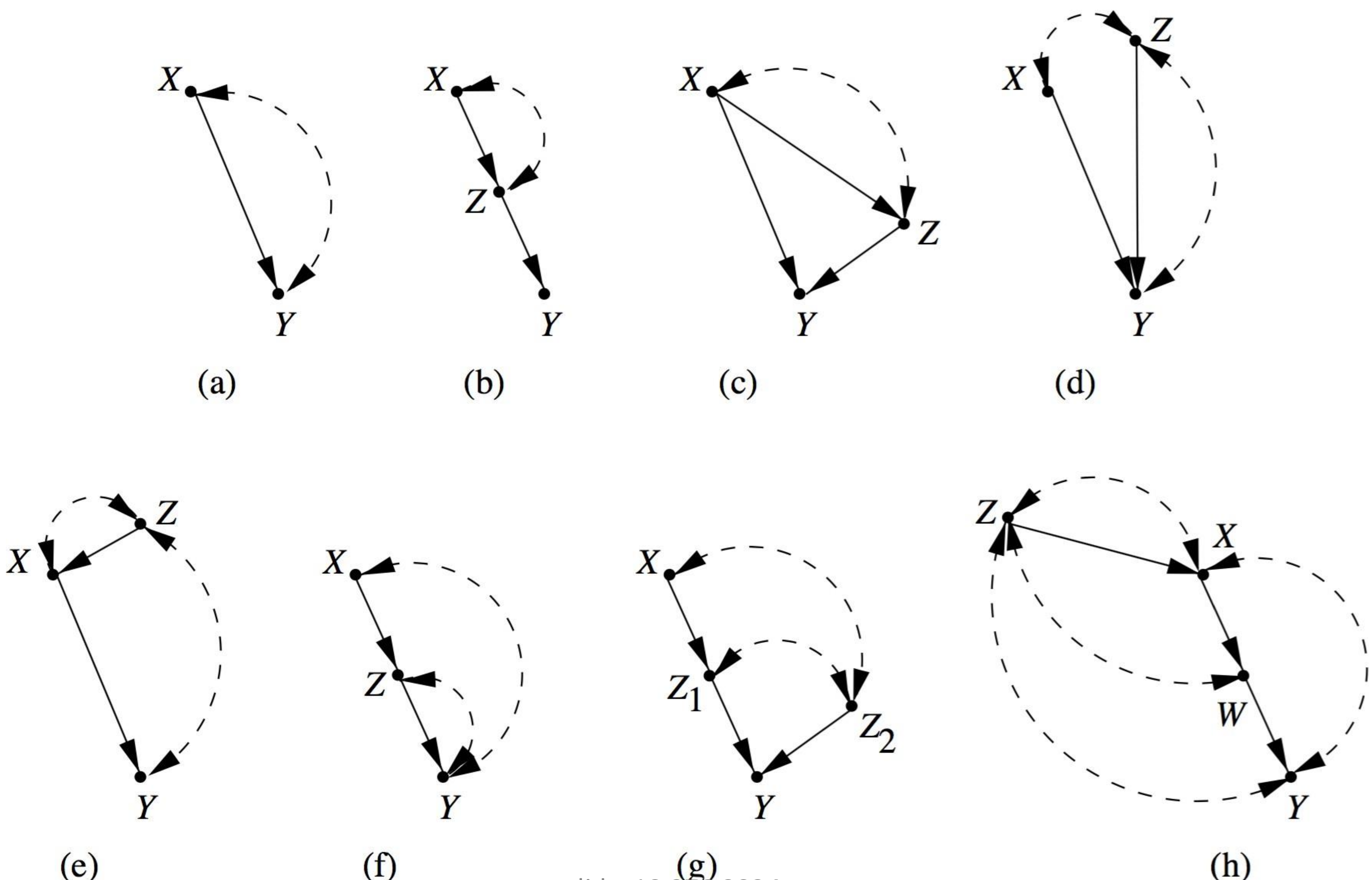
Theorem (Graphical criterion for non-identifiability of joint interventional distributions (Tian, 2002)).

If there is a bidirected path connecting X to any of its children in G, then $P(\mathbf{v}|do(x))$ is not identifiable from $P(\mathbf{v})$ and G.

Some Identifiable Graphs



Some Non-Identifiable Graphs



Summary

- The do-calculus provides a syntactical characterization to the problem of policy evaluation for atomic interventions.
- The problem of confounding and identification is essentially solved, non-parametrically.
- Simpson's Paradox is mathematized and dissolved.
- Applications are pervasive in the social and health sciences as well as in statistics, machine learning, and artificial intelligence.