CS 295: Causal 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)

Outline

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

Outline

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

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 $O(exp(|\mathbf{Z}|))$).

Inverse Probability Weighting (IPW)

- Let's rewrite the bd-expression,

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

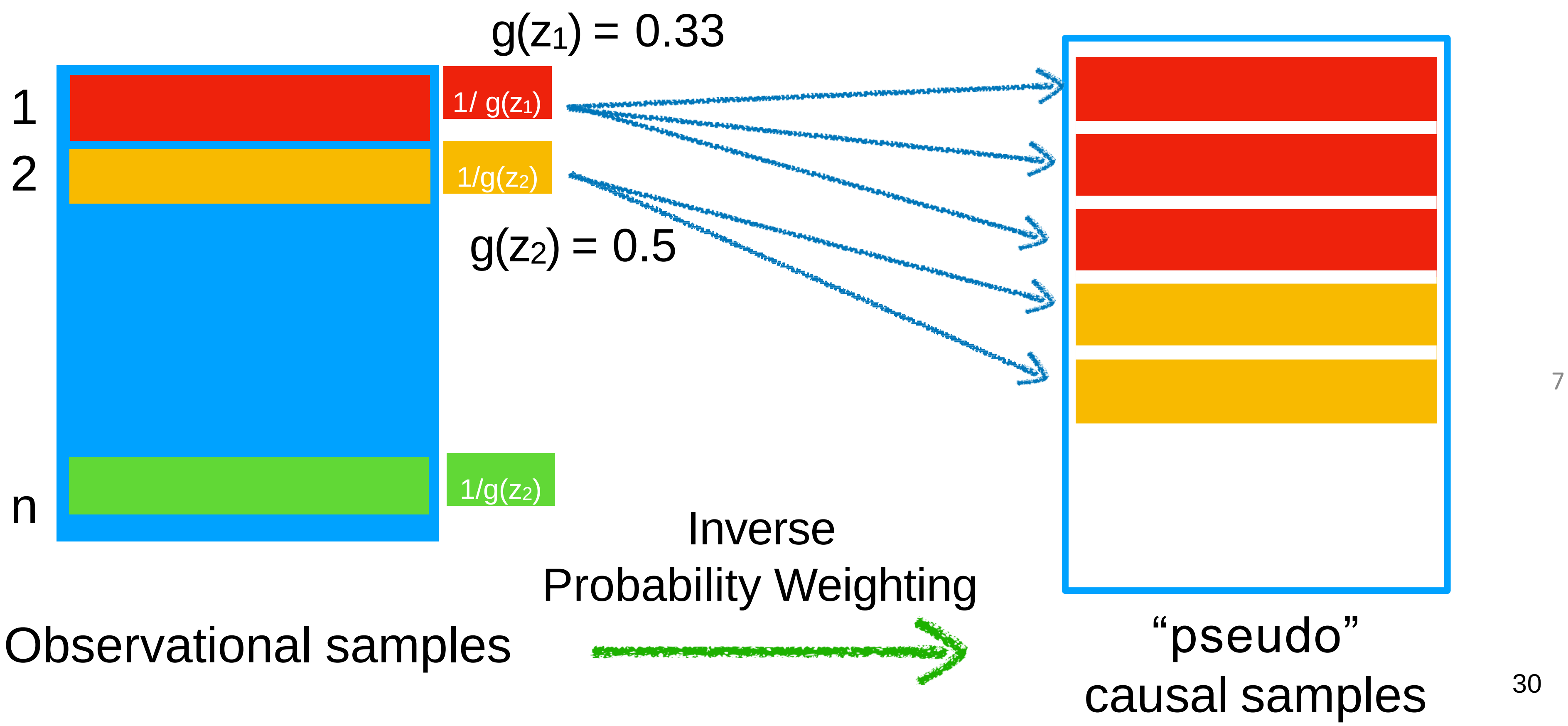
Entries of the joint distribution

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

Inverse Propensity score

Inverse Probability Weighting (IPW)

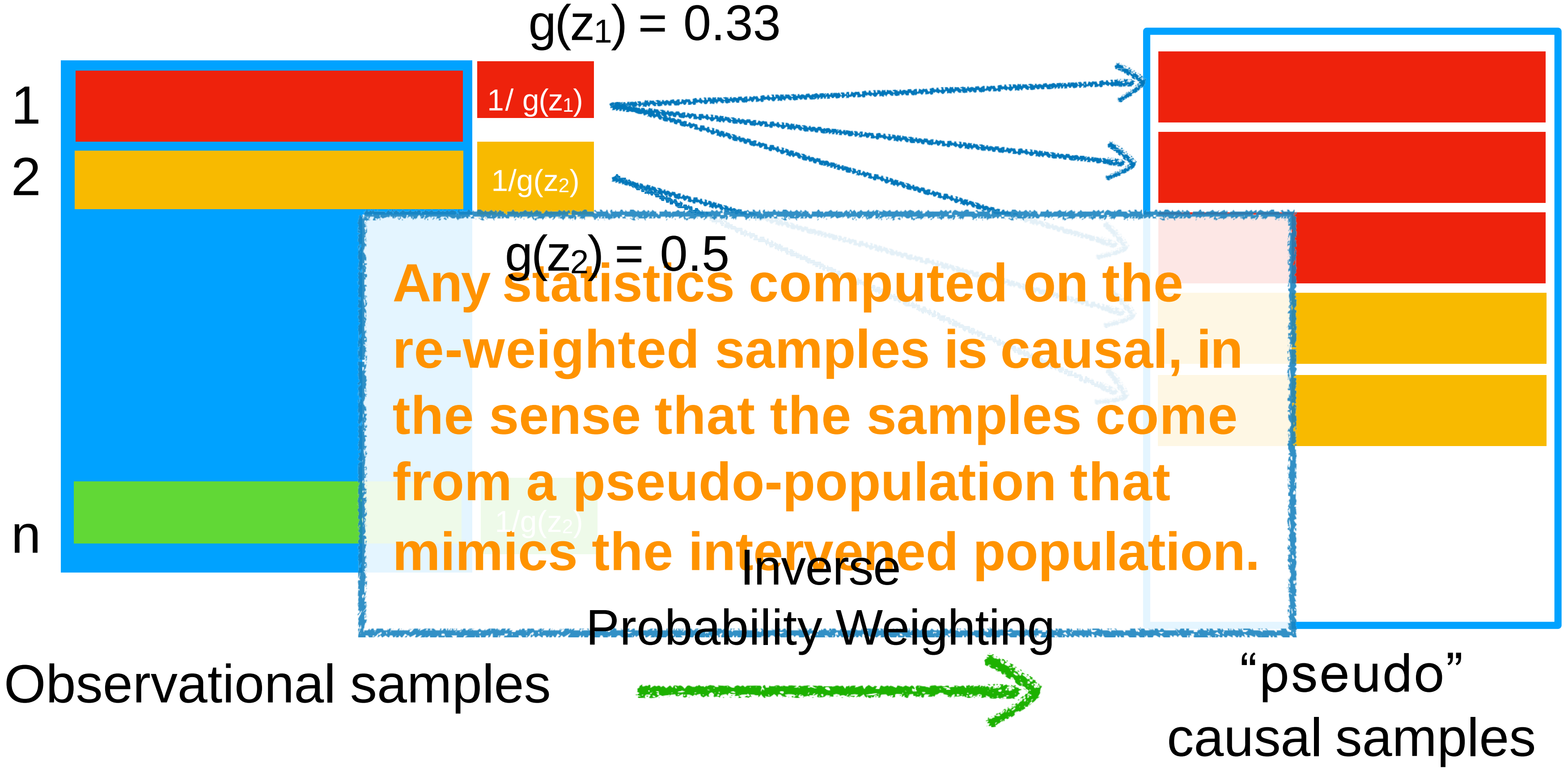
- In practice, evaluating the expr. can be seen as: $\frac{1}{N}\sum_{i=1}^{N}\frac{1_{\mathbf{Y}_i=\mathbf{y},\mathbf{X}_i=\mathbf{x},\mathbf{Z}_i=\mathbf{z}}}{g(\mathbf{z})}$



Inverse Probability Weighting (IPW)

- In practice, evaluating the expr. can be seen as:

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



Any statistics computed on the re-weighted samples is causal, in the sense that the samples come from a pseudo-population that mimics the intervened population.

8

31

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

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)
Or P(X=yes|Z=female)

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

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 $P(Z = z)$).

Outline

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

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,

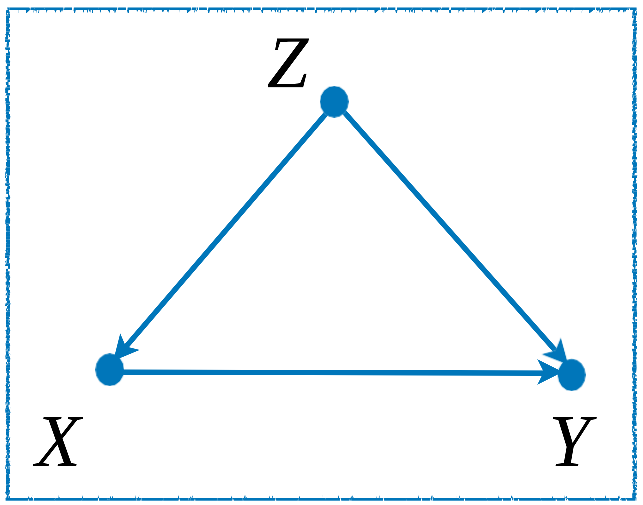
$$P(\mathbf{v} \mid do(\mathbf{x})) = \sum_{\mathbf{u}} \prod_{\{V_i \in \mathbf{V} \setminus \mathbf{X}\}} P\left(v_i \mid pa_i, u_i\right) P(\mathbf{u})$$

And the effect of such intervention on a set Y is

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

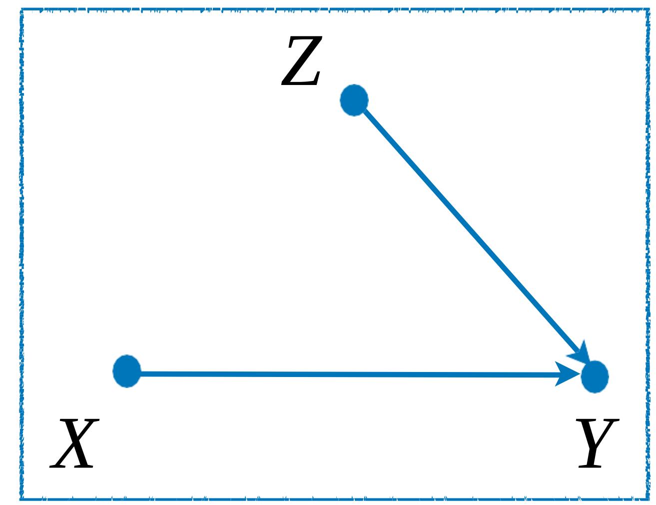
Interventions in Semi-Markovian

Real world



Intervention

Alternative world



$$M = \begin{cases} Z \leftarrow f_Z(u_z) \\ X \leftarrow f_X(z, u_x) \\ Y \leftarrow f_Y(x, z, u_y) \end{cases} \xrightarrow[do(X=x)]{} M_x = \begin{cases} Z \leftarrow f_Z(u_z) \\ \cancel{X \leftarrow f_X(z, u_x)} \quad X = x \\ Y \leftarrow f_Y(x, z, u_y) \end{cases}$$

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

Interventions in Semi-Markovian

Re-writing the interventional distribution…

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

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

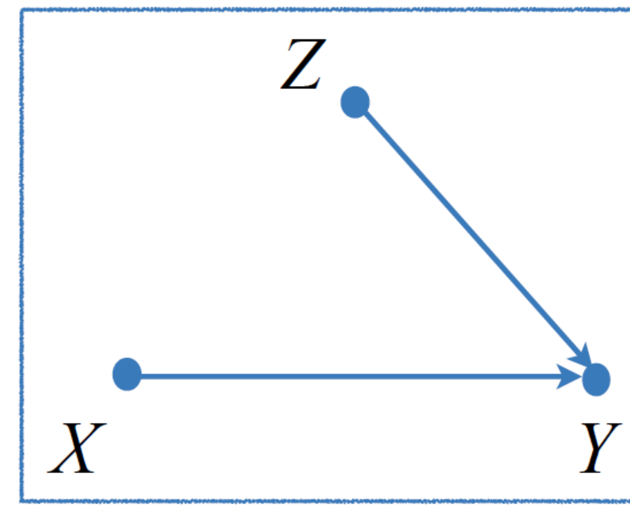
$$= P(z)\left(\sum_{u_y} P(y \mid x, z, u_y)P(u_y)\right)$$

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

$$= P(z)P(y \mid x, z)$$

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

Alternative world



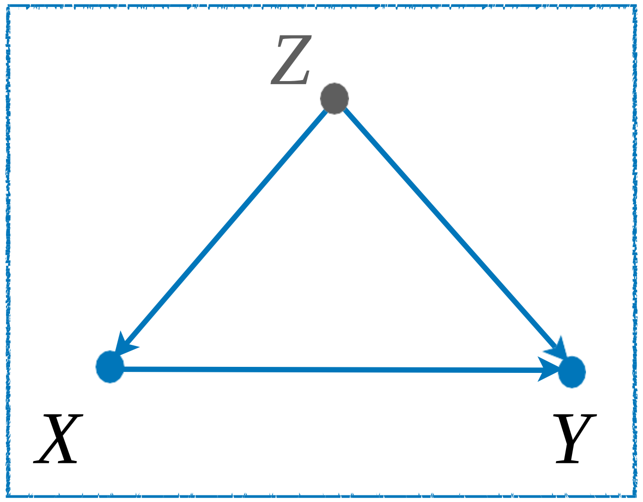
$$M_x = \begin{cases} Z \leftarrow f_Z(u_z) \\ \cancel{X \leftarrow f_X(z, u_x)} \quad X = x \\ Y \leftarrow f_Y(x, z, u_y) \end{cases}$$

These distributions can be computed from the obs. distribution *P(z,x,y)*.

(Alternative 'proof' for the backdoor!)

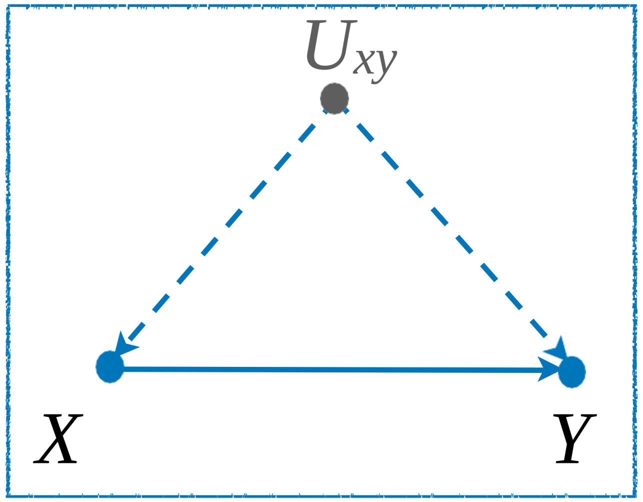
Interventions - Another Example

Real world



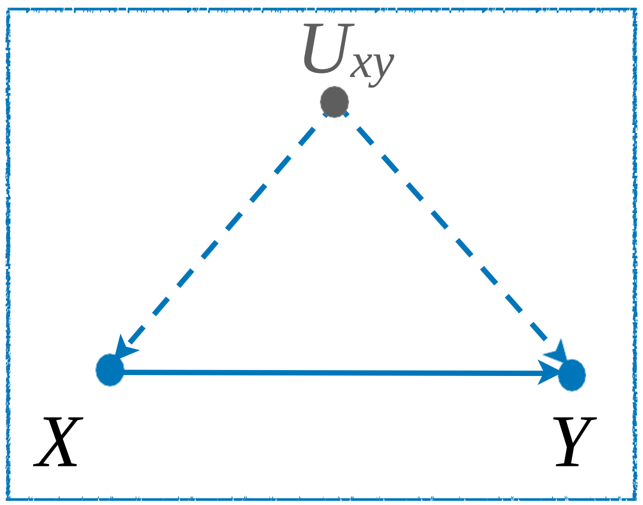
Interventions - Another Example

Real world



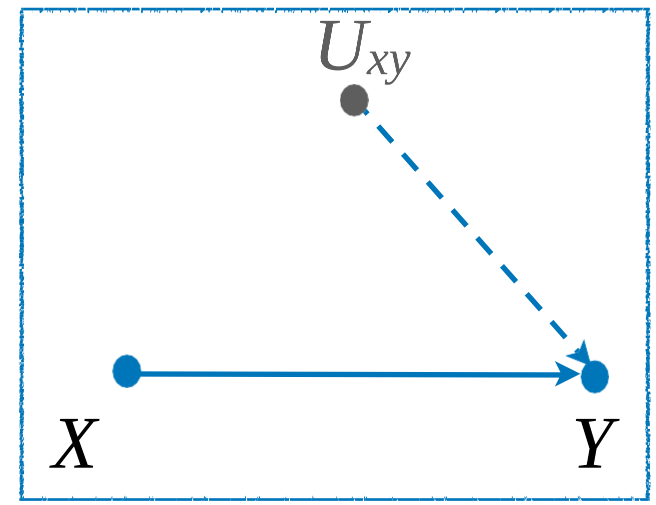
Interventions - Another Example

Real world



Intervention

Alternative world



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

$do(X{=}x)$

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

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

$$P(y \mid do(x)) = \sum_{u_x, u_y, u_z} \cancel{P(x \mid u_z, u_x)} P(y \mid x, u_z, u_y) P(u)$$

U_xy

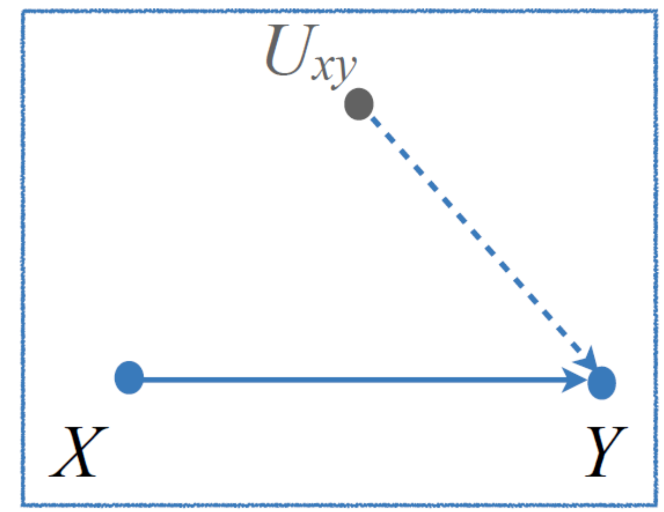
Interventions - Another Example

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



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

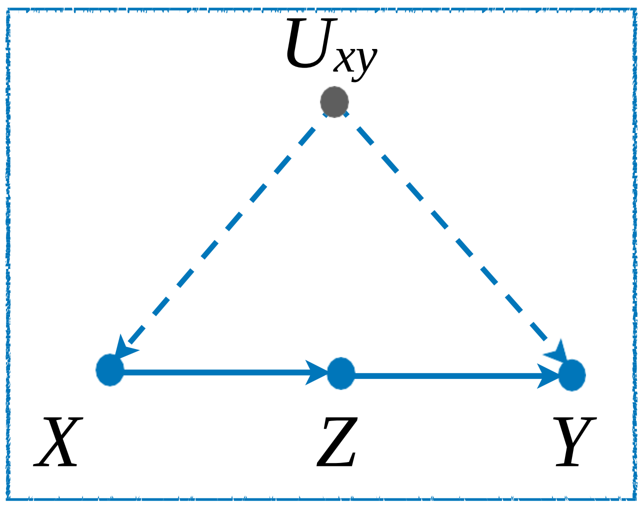
We can get rid of U_y
But not of U_xy

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

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

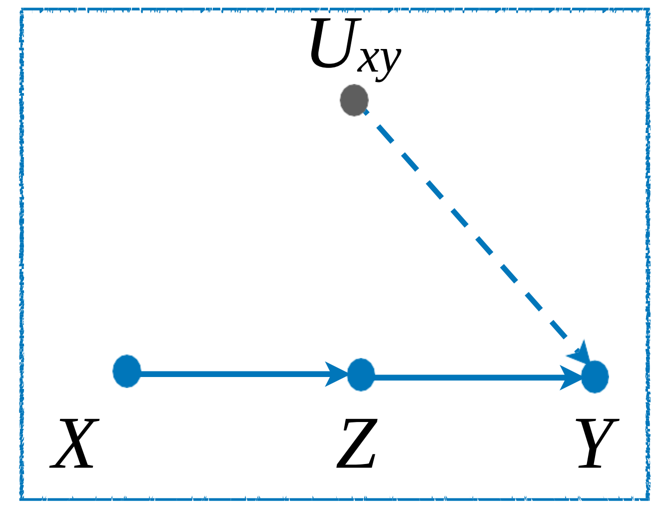
The Front-door Case

Real world



intervention

Alternative world



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

$$\xrightarrow{\quad} \quad do(X=x)$$

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

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

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

The Front-door Case

Re-writing the interventional distribution…

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

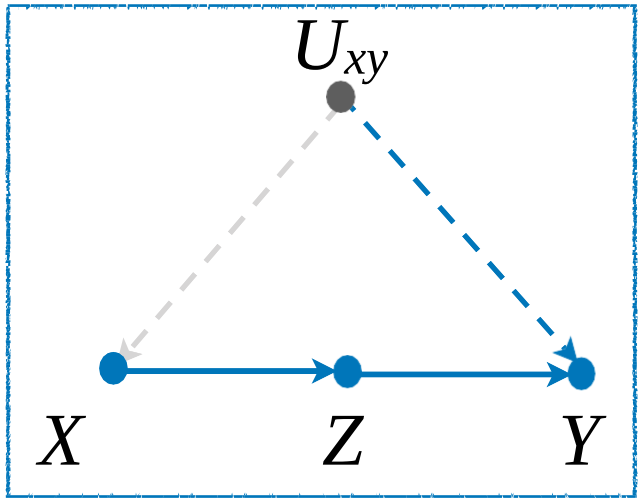
$$= \left(\sum_{u_z} P(z \mid x, u_z) P(u_z) \right) \left(\sum_{u_{xy}, u_y} P(y \mid z, u_{xy}, u_y) P(u_{xy}, u_y) \right) \left(\sum_{u_x} P(u_x) \right)$$

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

Note this seems similar to the previous example since ¬ ($U_{xy} \perp\!\!\!\perp Z$)!

Is this instance non-ID as well?

Alternative world



Note that the model entails different invariances, i.e.:

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

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

The Front-door Case

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

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

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

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

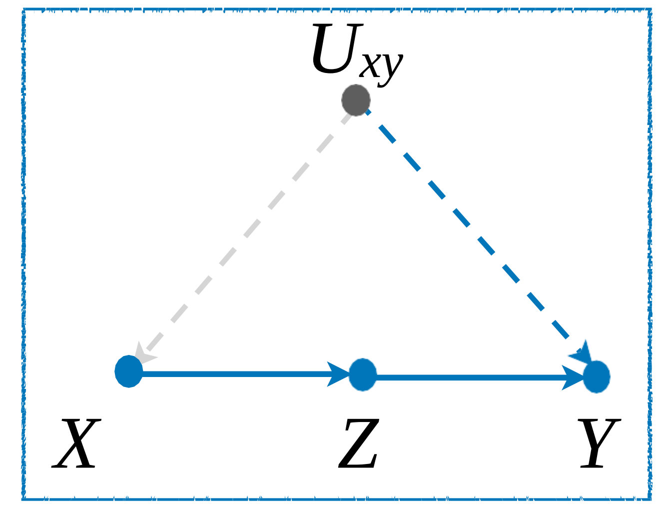
$(U_{xy} \perp\!\!\!\perp Z \mid X)$

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

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

The Front Door Criterion

When we cannot block a backdoor path, we may still have a front door path

Consider the century-old debate on the relation between smoking and lung cancer. In the years preceding 1970, the tobacco industry has managed to prevent antismoking legislation by promoting the theory that the observed correlation between smoking and lung cancer could be explained by some sort of carcinogenic genotype that also induces an inborn craving for nicotine.

A graph depicting this example is shown in Figure 3.10(a) This graph does not satisfy

Causal effect not identifiable here

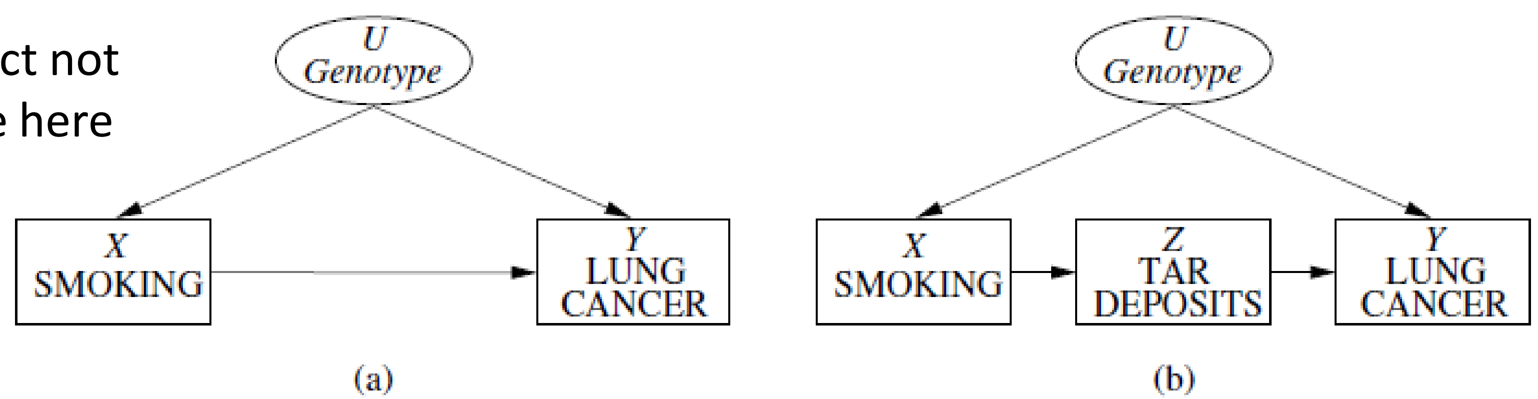


Figure 3.10: A graphical model representing the relationships between smoking (X) and lung cancer (Y), with unobserved confounder (U) and a mediating variable Z

Front Door Condition

We cannot satisfy the backdoor criterion since we cannot measure U. But consider the model in (b).
It does not satisfy the backdoor criterion, but we can measure the tar level, Z, which will allow identifiability of P(Y|do(X)),

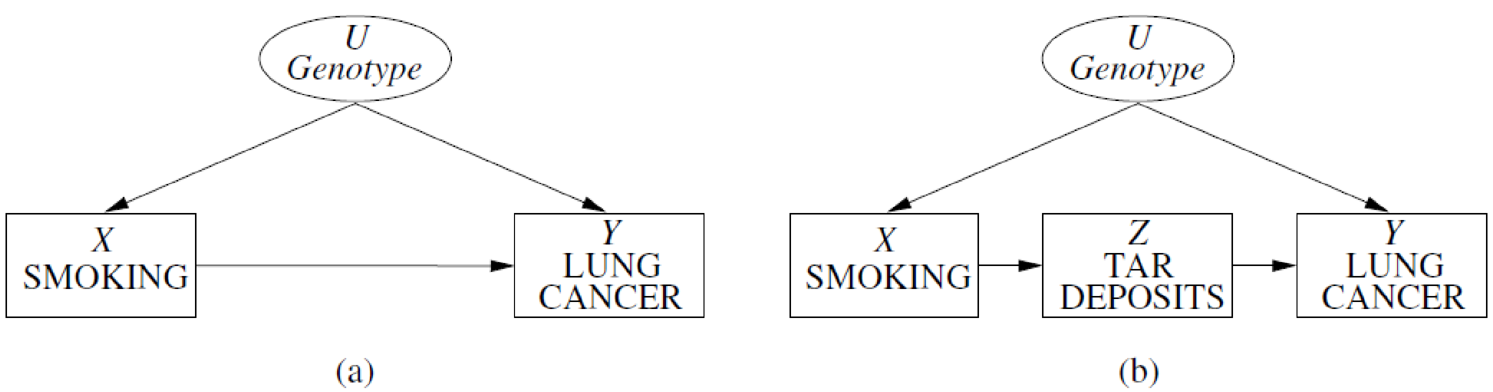


Figure 3.10: A graphical model representing the relationships between smoking (X) and lung cancer (Y), with unobserved confounder (U) and a mediating variable Z

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%).

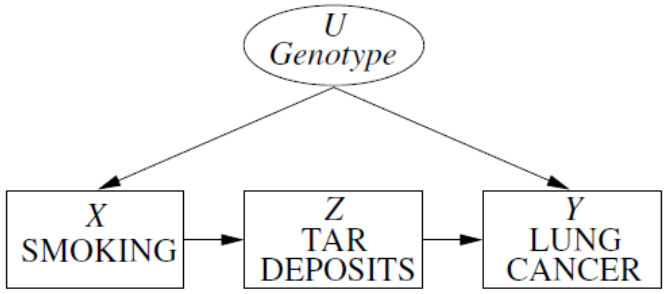
Front-door Condition

The graph of Figure 3.10(b) enables us to decide between these two groups of statisticians. First, we note that the effect of X on Z is identifiable, since there is no backdoor path from X to Z. Thus, we can immediately write

$$P(Z = z|do(X = x)) = P(Z = z|X = x) \tag{3.12}$$

Next we note that the effect of Z on Y is also identifiable, since the backdoor path from Z to Y, namely $Z \leftarrow X \leftarrow U \rightarrow Y$, can be blocked by conditioning on X. Thus we can write

$$P(Y = y|do(Z = z)) = \sum_x P(Y = y|Z = z, X = x)\ \mathrm{P(x)} \tag{3.13}$$



(b)

We are now going to chain together the two partial effects to obtain the overall effect of X on Y. The reasoning goes as follows: If nature chooses to assign Z the value z, then the probability of Y would be $P(Y = y|do(Z = z))$. But the probability that nature would choose to do that, given that we choose to set X at x, is $P(Z = z|do(X = x))$. Therefore, summing over all states z of Z we have

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|do(Z = z))P(Z = z|do(X = x)) \tag{3.14}$$

The terms on the right hand side of (3.14) were evaluated in (3.12) and (3.13), and we can substitute them to obtain a *do*-free expression for $P(Y = y|do(X = x))$. We also distinguish between the x that appears in (3.12) and the one that appears in (3.13), the latter of which is merely an index of summation and might as well be denoted x'. The final expression we have is

$$P(Y = y|do(X = x)) =$$
$$\textstyle\sum_z \sum_{x'} P(Y = y|Z = z, X = x')P(X = x')P(Z = z|X = x) \tag{3.15}$$

Equation (3.15) is known as the *front-door formula*.

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.*

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

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

Front-door Condition

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

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

The conditions stated in Definition 3.4.1 are overly conservative; some of the backdoor paths excluded by conditions (ii) and (iii) can actually be allowed provided they are blocked by some variables. There is a powerful symbolic machinery, called the *do-calculus*, that allows analysis of such intricate structures. In fact, the *do*-calculus uncovers *all* causal effects that can be identified from a given graph. Unfortunately, it is beyond the scope of this book (see Pearl 2009 and Shpitser and Pearl 2008 for details). But the combination of the adjustment formula, the backdoor criterion, and the front-door criterion covers numerous scenarios. It proves the enormous, even revelatory, power that causal graphs have in not merely representing, but actually discovering causal information.

Outline

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

Sleeping Too Little in Middle Age May Increase Dementia Risk, Study Finds

The research, tracking thousands of people from age 50 on, suggests those who sleep six hours or less a night are more likely (NYT, 2021)

The research, published Tuesday in the journal Nature Communications, has limitations but also several strengths. It followed nearly 8,000 people in Britain for about 25 years, beginning when they were 50 years old. It found that those who consistently reported sleeping six hours or less on an average weeknight were about 30 percent more likely than people who regularly got seven hours sleep (defined as "normal" sleep in the study) to be diagnosed with dementia nearly three decades later.

The team was able to adjust for several behaviors and characteristics that might influence people's sleep patterns or dementia risk, said an author of the study, Séverine Sabia, an epidemiologist at Inserm, the French public-health research center. Those included smoking, alcohol consumption, how physically active people were, body mass index, fruit and vegetable consumption, education level, marital status and conditions like hypertension, diabetes and cardiovascular disease.

In addition, most participants were white and better educated and healthier than the overall British population. And in relying on electronic medical records for dementia diagnoses, researchers might have missed some cases. They also could not identify exact types of dementia.

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 $P(\mathbf{v})$.

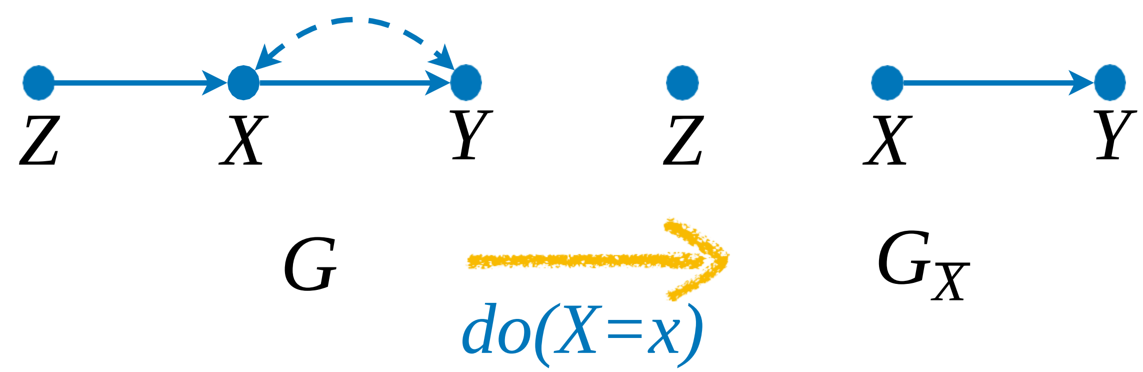
- We are interested in rules or a set of axioms that allow the systematic transformation of a $do(.)$ expression into a do-free expression while preserving the equivalence to the target effect.

Causal Calculus:
A systematic approach for identification

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,

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

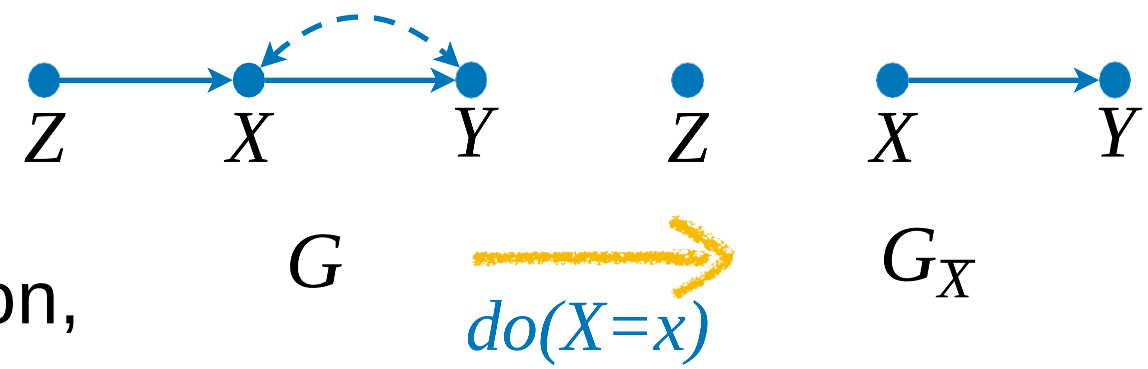
Let's verify this equality!

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,

$$P(\mathbf{v} \mid do(x))$$

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

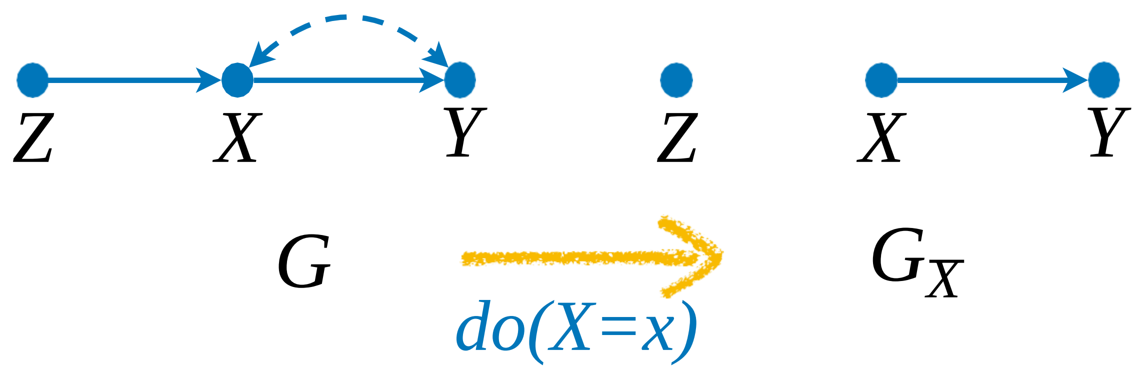
$$= P(z) \sum_{u_{xy}} P(y \mid x, u_{xy}) P(u_{xy})$$

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,

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

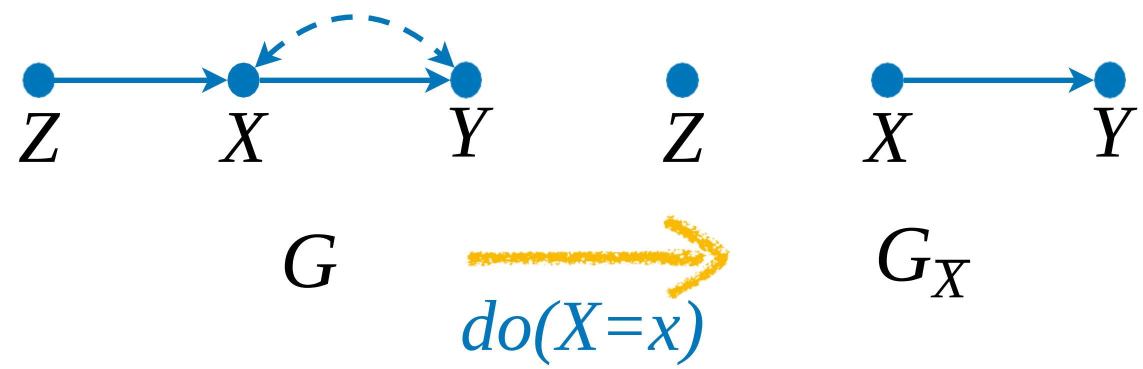
$$P(y, z \mid do(x)) = P(z) \sum_{u_{xy}} P(y \mid x, u_{xy}) P(u_{xy})$$

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

Insight 1: Adding/removing Observations

- Adding/removing observations

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



Substituting the factors back…

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

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

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

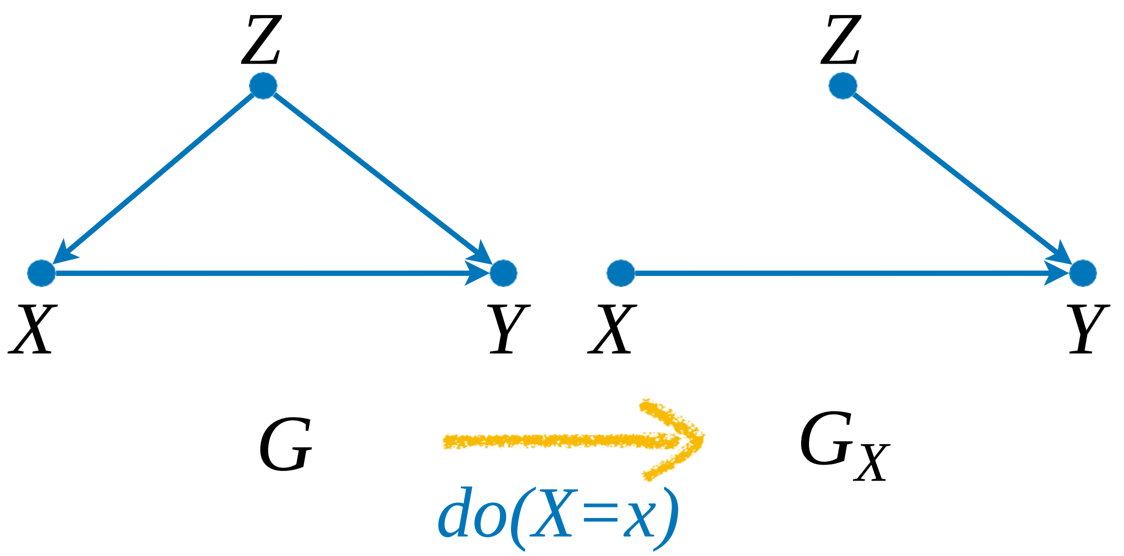
$$= \sum_{z} P(\mathbf{v}\,|\,do(x)) \boxed{= P(y\,|\,do(x))}$$

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

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!

Sleeping Too Little in Middle Age May Increase Dementia Risk, Study Finds

The research, tracking thousands of people from age 50 on, suggests those who sleep six hours or less a night are more likely (NYT, 2021)

The research, published Tuesday in the journal Nature Communications, has limitations but also several strengths. It followed nearly 8,000 people in Britain for about 25 years, beginning when they were 50 years old. It found that those who consistently reported sleeping six hours or less on an average weeknight were about 30 percent more likely than people who regularly got seven hours sleep (defined as “normal” sleep in the study) to be diagnosed with dementia nearly three decades later.

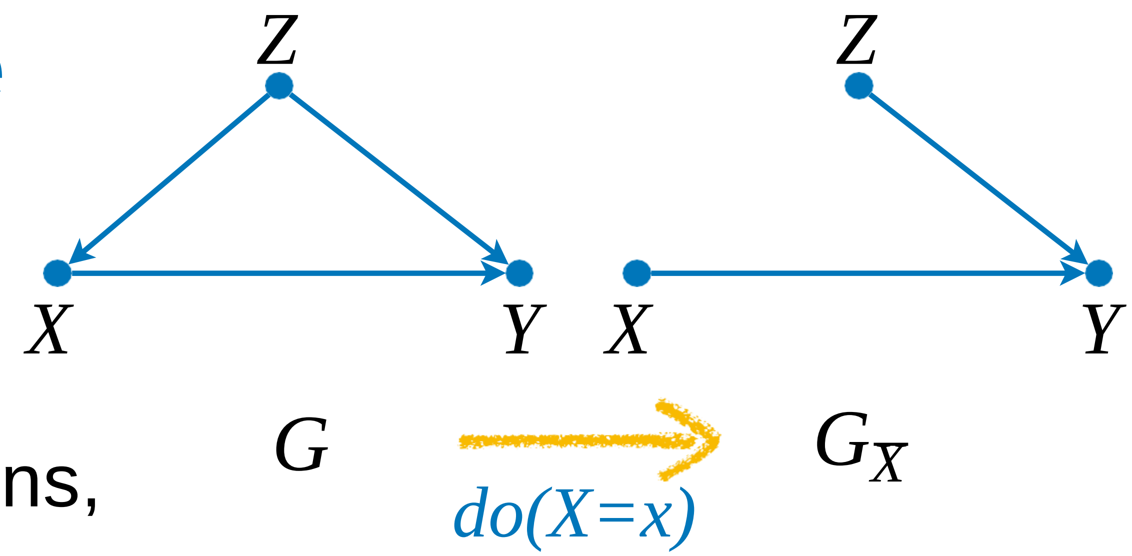
The team was able to adjust for several behaviors and characteristics that might influence people’s sleep patterns or dementia risk, said an author of the study, Séverine Sabia, an epidemiologist at Inserm, the French public-health research center. Those included smoking, alcohol consumption, how physically active people were, body mass index, fruit and vegetable consumption, education level, marital status and conditions like hypertension, diabetes and cardiovascular disease.

In addition, most participants were white and better educated and healthier than the overall British population. And in relying on electronic medical records for dementia diagnoses, researchers might have missed some cases. They also could not identify exact types of dementia.

Insight 2: Action/Observation Exchange

- Action/Observation Exchange

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



First, let's write the interventional distributions,

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

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

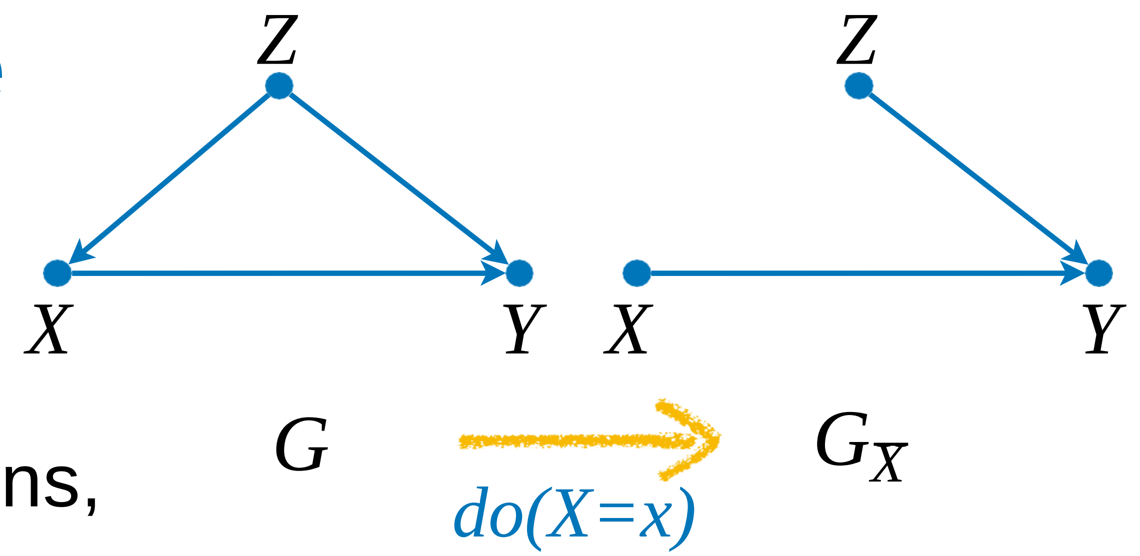
$$\boxed{P(y|do(x),z)} = \frac{P(z,y|do(x))}{P(z|do(x))} = \frac{P(z)P(y|x,z)}{P(z)} \boxed{= P(y|x,z)}$$

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

Insight 2: Action/Observation Exchange

- Action/Observation Exchange

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



First, let's write the interventional distributions,

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

Looks familiar?

BD perhaps?

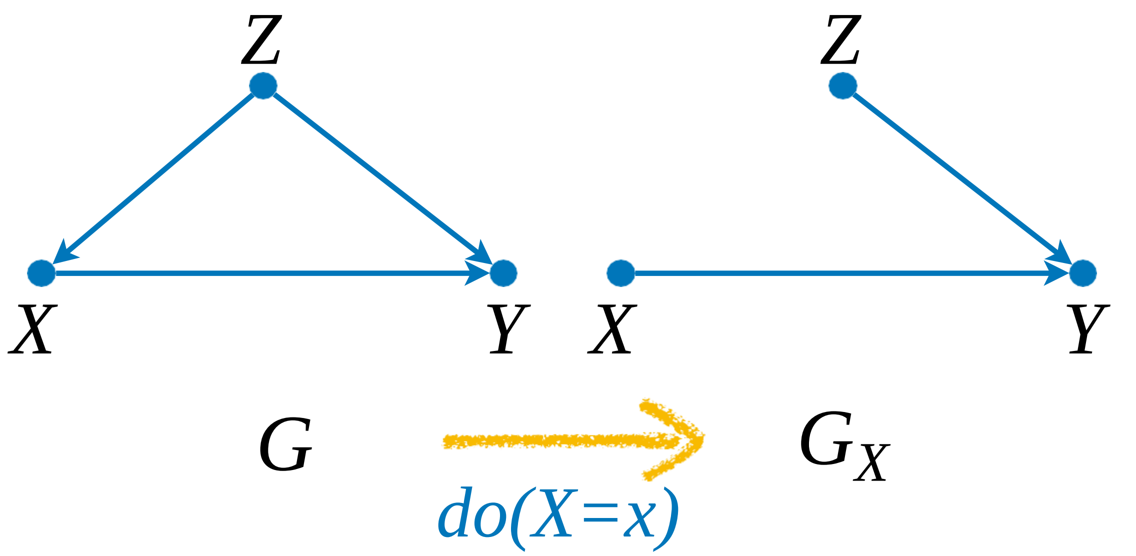
$$P(z|do(x)) = \sum_{y} P(z)P(y|x,z) = P(z)$$

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

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

Insight 2: Action/Observation Exchange

- Action/Observation Exchange



Great, but what about the equality

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

$$(Y \not\perp\!\!\!\perp X)_{G_{\underline{X}}}$$

Let's compare left and right-hand sides:

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

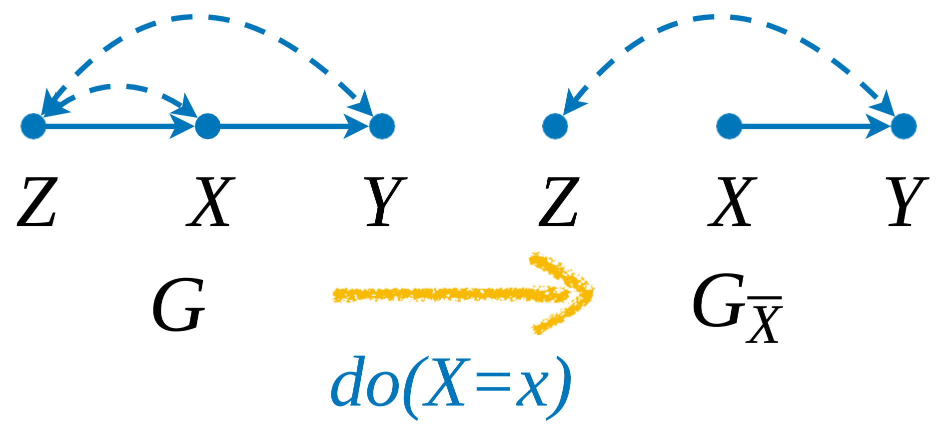
$$= \sum_z P(y \mid x, z) \boxed{P(z)}$$

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

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.



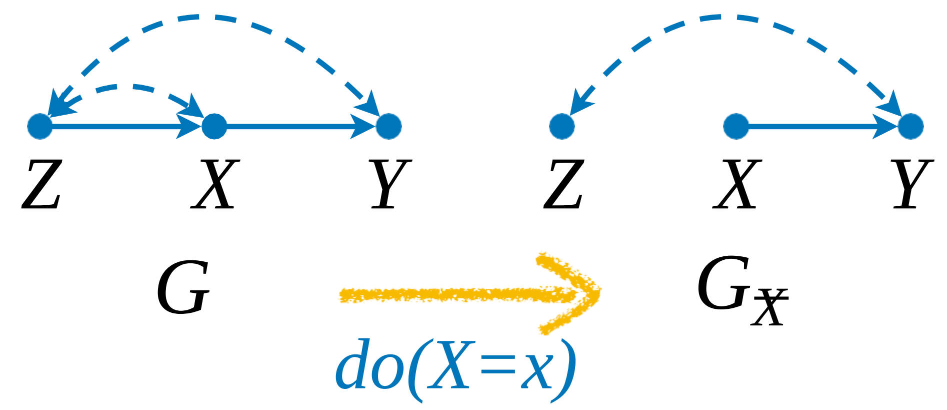
$$(Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \quad \Longrightarrow \quad P(z|do(x))=P(z)$$

Let's verify this equality!

Insight 3: Adding/Removing Actions

- Adding/Removing Actions

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



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

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

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

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

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 $G_{\overline{X}}$.

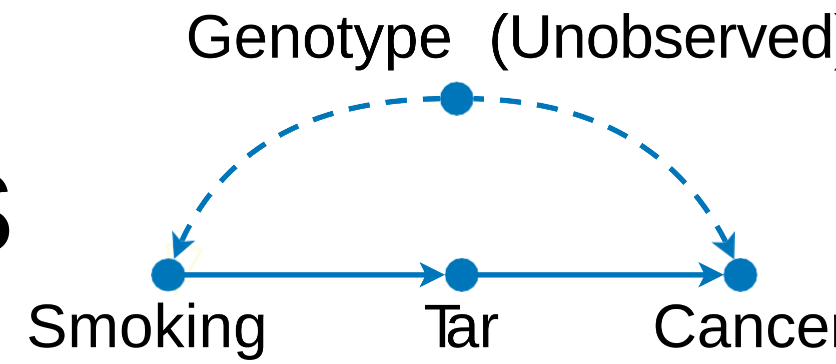
Properties of Do-Calculus

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.

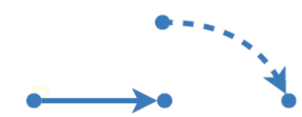
Syntactic goal: Re-express original Q without do()!

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 $(T \perp\!\!\!\perp C \mid S)_{G_{\underline{T}}}$

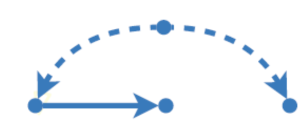


$$= \sum_t P(c \mid do(t)) P(t \mid do(s))$$ Rule 3 $(S \perp\!\!\!\perp C \mid T)_{G_{\overline{C,T}}}$

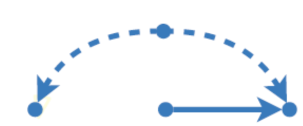
$$= \sum_t P(c \mid do(t)) P(t \mid s)$$ Rule 2 $(S \perp\!\!\!\perp T)_{G_{\underline{S}}}$

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



Non-identifiability Machinery

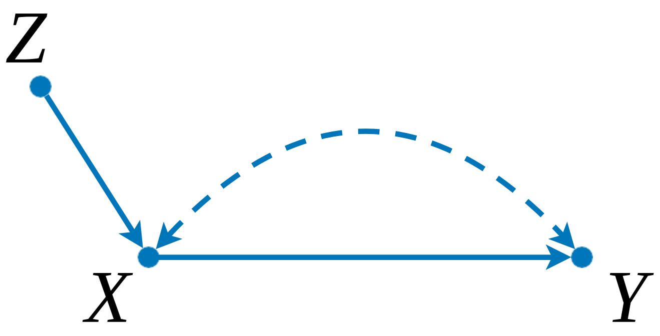
Lemma (Graph-subgraph ID (Tian and Pearl, 2002))

- 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 Machinery

- Proof idea. Suppose M_1, M_2 induce the same $P(\mathbf{v})$ but differ in $P(y|do(x))$. Construct two new models M_1', M_2' with any $P(z)$ and let
  $P_i'(x|z,u_{xy})=P_i(x|u_{xy}),\ i=1,2$.

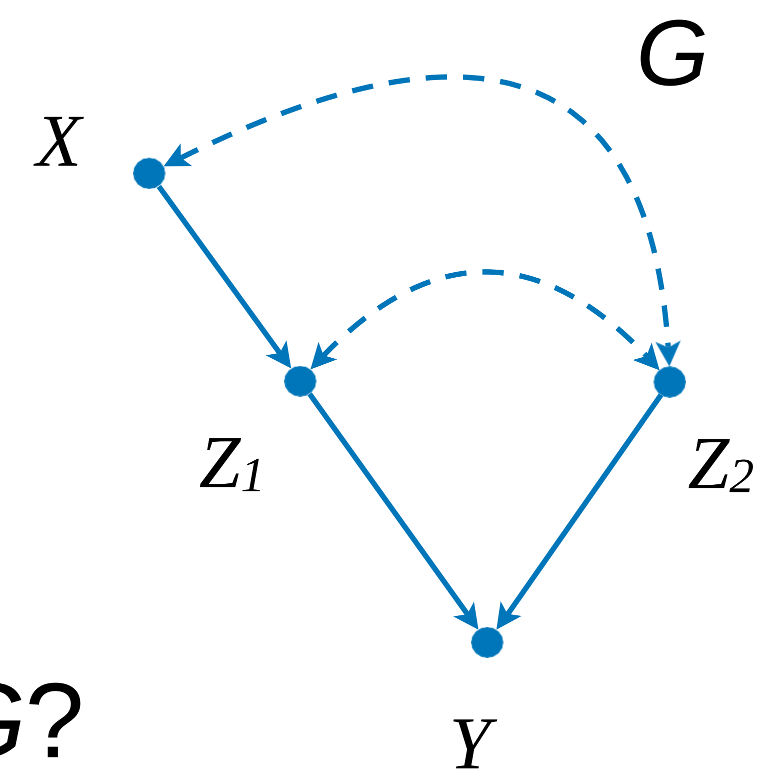
  This construction entails
  $P_1'(y|do(x))\neq P_2'(y|do(x))$.



Question: Do all non-ID models look like the bow graph?

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, z2)) identifiable from *G*?

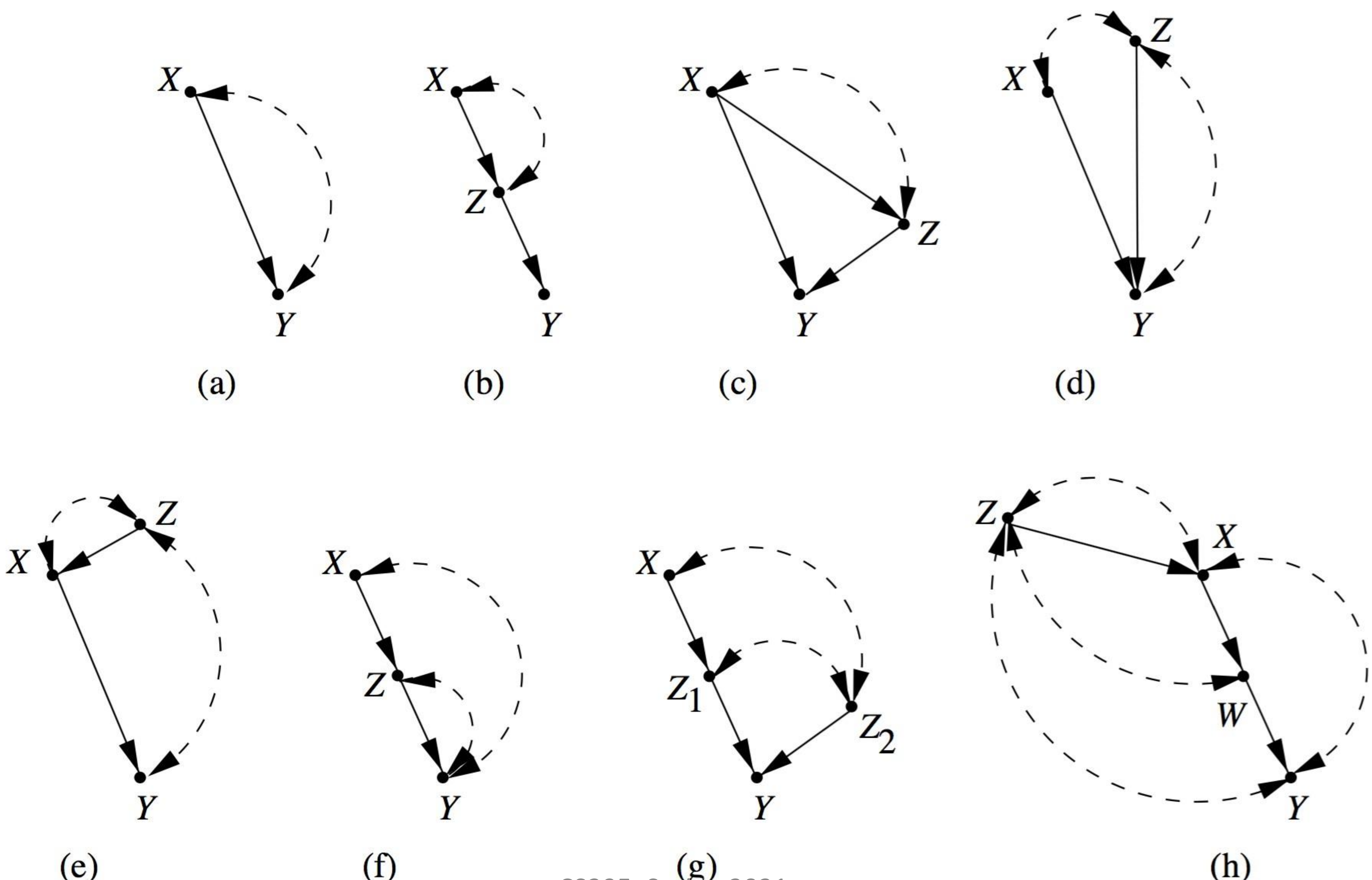


Non-Identifiability Criterion

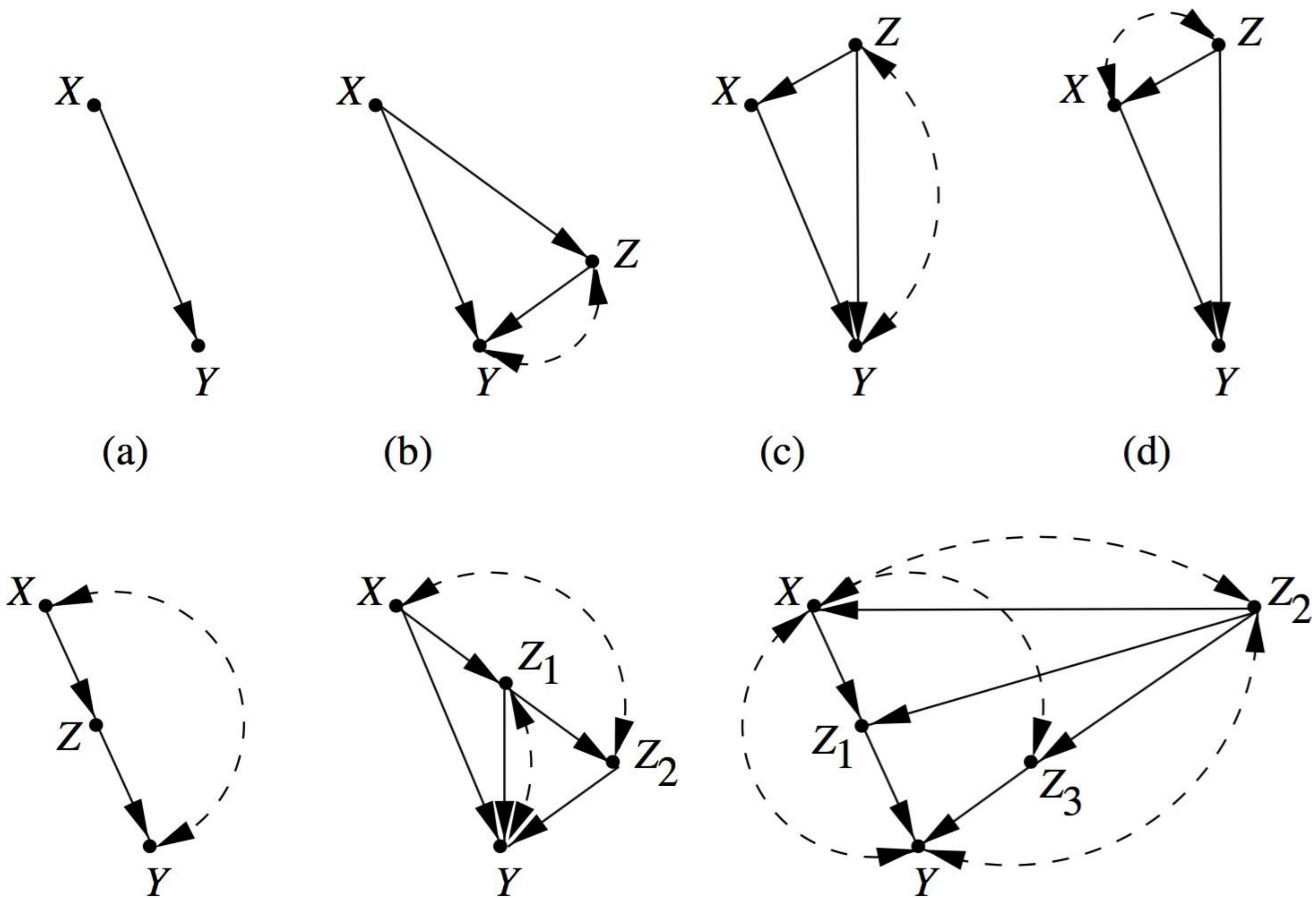
Theorem (Graphical criterion for non-identifiability of joint interventional distributions (Tian, 2002)).

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 Non-Identifiable Graphs



Some Identifiable Graphs



(a) (b) (c) (d)

(e) (f) (g)