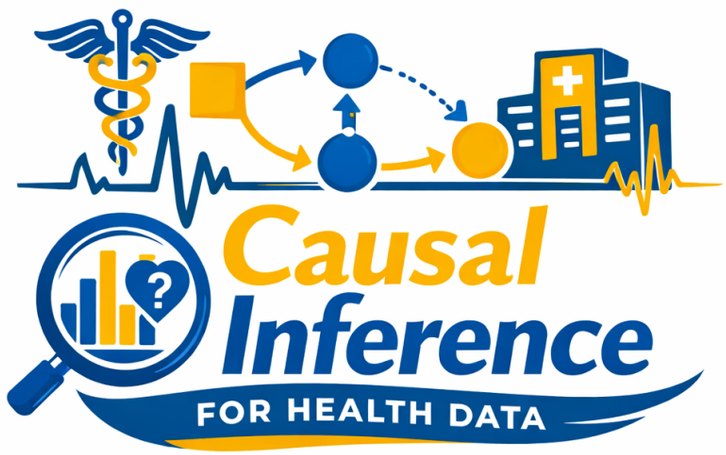
Causal Inference for Health Data

(STATS C160/C260 — Winter 2026)

Lectures 6 & 7: Estimation of Causal Effects Based on the Back-door Criterion

Drago Plečko

UCLA College | Physical Sciences Statistics & Data Science



Recap: Back-door Adjustment

Theorem (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:

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

Today: how do we use back-door in practice?

Recall: Simpson's Paradox Example

	HbA1c low (Y)	HbA1c high ($\neg Y$)		Sucess Rate
drug (X)	20	20	40	50%
no-drug ($\neg X$)	16	24	40	40%
	36	44		

$$P(Y \mid F, X) < P(Y \mid F, \neg X)$$

$$P(Y \mid \neg F, X) < P(Y \mid \neg F, \neg X)$$

but

$$P(Y \mid X) > P(Y \mid \neg X)\ !$$

($\neg F$)	HbA1c low (Y)	HbA1c high ($\neg Y$)		Sucess Rate
drug (X)	18	12	30	60%
no-drug ($\neg X$)	7	3	10	70%
	25	15	40	

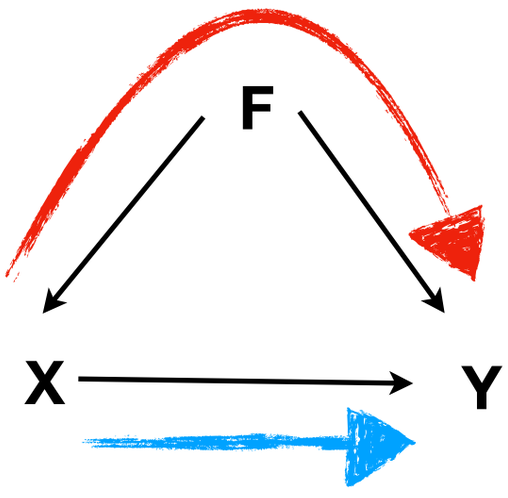
(F)	HbA1c low (Y)	HbA1c high ($\neg Y$)		Sucess Rate
drug (X)	2	8	10	20%
no-drug ($\neg X$)	9	21	30	30%
	11	29	40	

Resolving Simpson's Paradox with Back-Door Adjustment

($\neg F$)	HbA1c low (Y)	HbA1c high ($\neg Y$)		Sucess Rate
drug (X)	18	12	30	60%
no-drug ($\neg X$)	7	3	10	70%
	25	15	40	

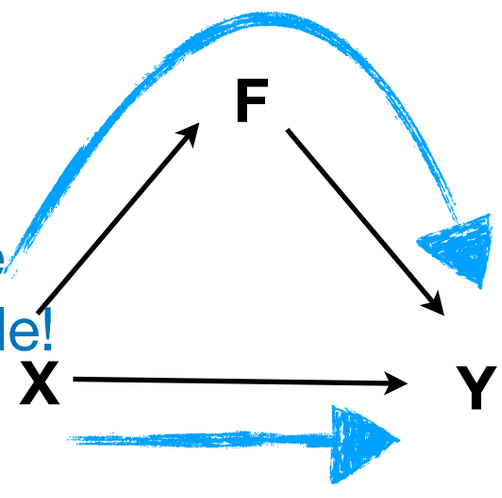
(F)	HbA1c low (Y)	HbA1c high ($\neg Y$)		Sucess Rate
drug (X)	2	8	10	20%
no-drug ($\neg X$)	9	21	30	30%
	11	29	40	

Sex-Stratified



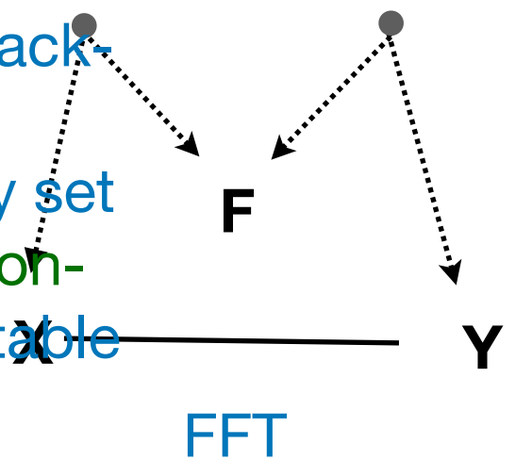
Weight Loss-Stratified

F is back-door; consult the stratified table!



Another Story?

F is not back-door; but empty set is! use non-stratified table

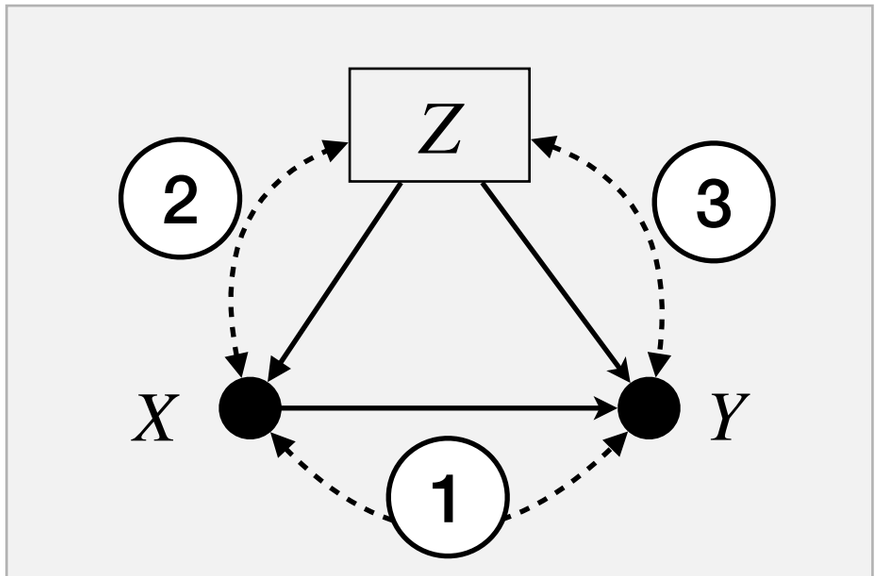


FFT

Most Common Back-door Setting: CG(3) Graphical Model

- What is the simplest setting in which back-door helps?
  Approach: put all Z variables into a single block!

Construct graph over three blocks



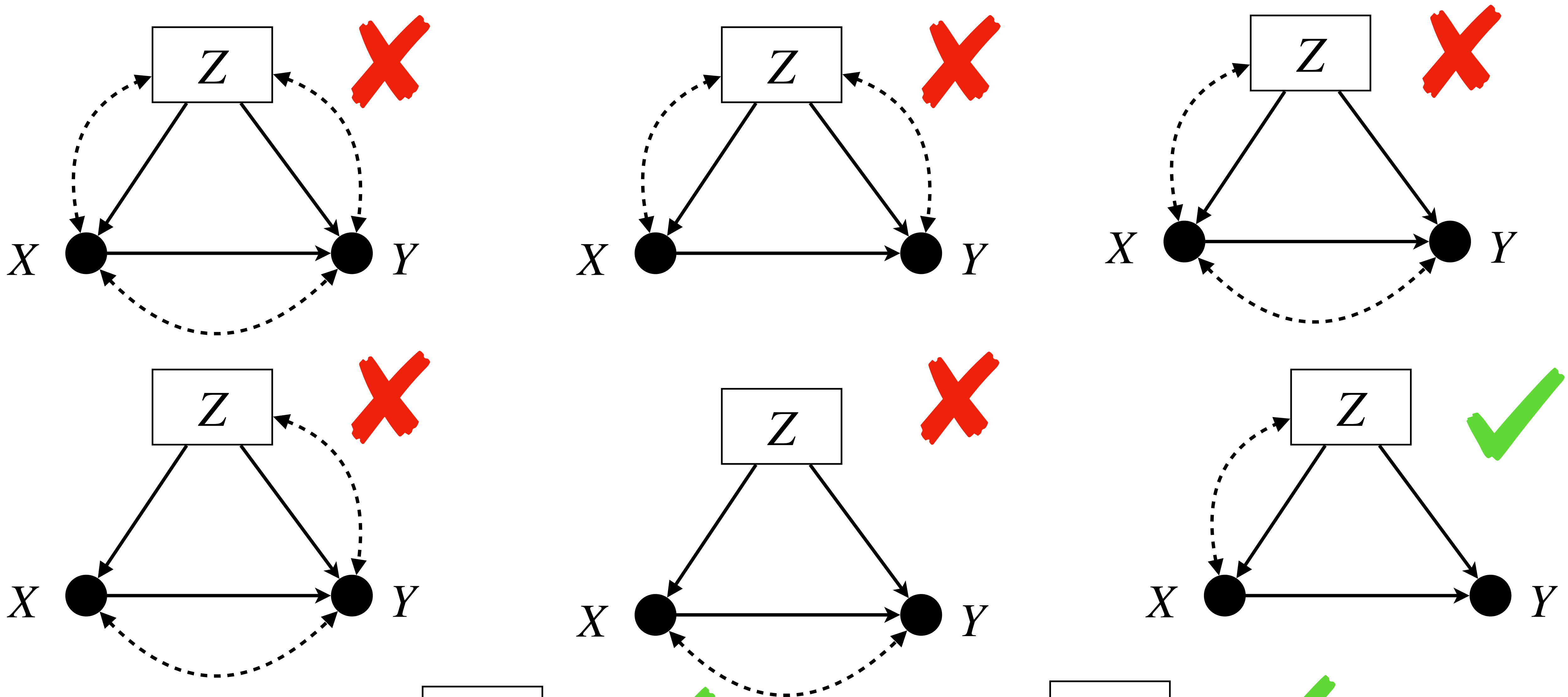
start with no assumptions — CG(3) model

Elicit Causal Knowledge Stepwise

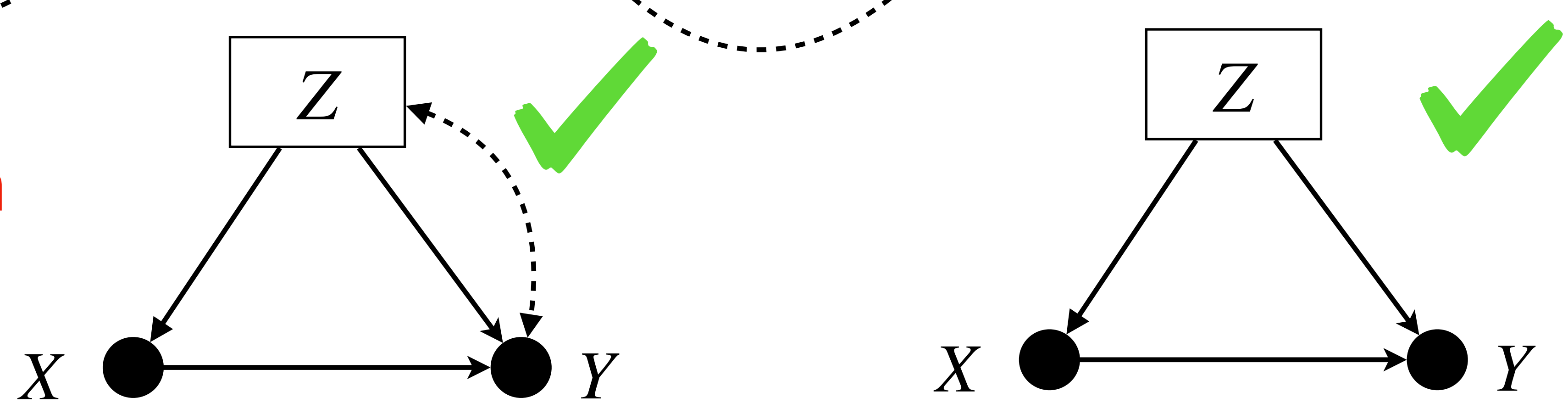
1. is there confounding X - Y?
2. is there confounding X - Z?
3. is there confounding Y - Z?

Causal Assumptions encoded in edge removals!

Assess Back-door: 8 possible cases



If Back-Door holds for Z, can use it for adjustment!



Evaluating BD adjustment

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

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

- In practice, how should such expressions be estimated?

- From the above, there are two important ingredients:
  - the term $P(\mathbf{y} \,|\, \mathbf{x}, \mathbf{z})$, which may be okay for $\mathbf{Y} \in \mathbb{R}$,
  - term $P(\mathbf{z})$, which may be hard if $\mathbf{Z}$ is higher-dim.,
  - evaluate their convolution, summing / integrating over a possibly high-dimensional $\boldsymbol{Z}$ (could be hard).

Estimation Techniques: IPW, Regression, AIPW, DML

- To address these challenges, there are several important techniques for estimation based on BD-expression

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

1. Inverse Propensity Weighing (IPW)
2. Regression-based Methods
3. Augmented Inverse Propensity Weighing (AIPW)
4. Double Machine Learning (DML)

Inverse Propensity Weighting (IPW)

- Let's rewrite the back-door expression,

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

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

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

note: $\frac{1}{P(\mathbf{x}, \mathbf{z})} P(\mathbf{z}) = \frac{1}{P(\mathbf{x} \mid \mathbf{z})}$

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

Entries of the joint distribution

Fit a function $\pi(z)$ that approximates this probability

Inverse Propensity Weighting (IPW)

- Assume we have N samples, then

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

replace with empirical distribution

replace with estimate π

$$\hat{E}(\mathbf{y} \mid do(\mathbf{X} = \mathbf{x})) = \sum_{\mathbf{z},\mathbf{y}} \frac{\frac{1}{N}\sum_{i=1}^{N} \mathbf{y} 1(\mathbf{Y}_i = \mathbf{y}, \mathbf{X}_i = \mathbf{x}, \mathbf{Z}_i = \mathbf{z})}{\pi(\mathbf{z})}$$

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

$$= \frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{Y}_i 1(\mathbf{X}_i = \mathbf{x})}{\pi(\mathbf{Z}_i)}$$

→ our IPW estimation expression.

IPW Causal Effect Estimate

- In the previous slide, we got a single potential outcome estimate $\hat{E}(\mathbf{y} \mid do(\mathbf{X} = 1))$, and for the causal effect we simply take a difference:

$$\hat{E}(\mathbf{y} \mid do(\mathbf{X} = 1)) - \hat{E}(\mathbf{y} \mid do(\mathbf{X} = 0)) \overset{(*)}{=} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\mathbf{Y}_i 1(\mathbf{X}_i = 1)}{\pi(\mathbf{Z}_i)} - \frac{\mathbf{Y}_i 1(\mathbf{X}_i = 0)}{1 - \pi(\mathbf{Z}_i)} \right)$$

where $\pi(\mathbf{z})$ is an estimator of $P(\mathbf{X} = 1 \mid \mathbf{Z} = \mathbf{z})$

S1 write code to obtain the π estimator

S2 implement expression $(*)$ to obtain effect estimate

Note on Positivity Assumption

- In our estimator, we are re-weighing samples with

$$\frac{1}{\pi(\mathbf{z})} \textbf{ or } \frac{1}{1-\pi(\mathbf{z})}$$

- Of course, divisions by 0 are not possible, and introduce issues (estimator not defined, exploding variance),

- Therefore, for effect estimation we require *positivity*:

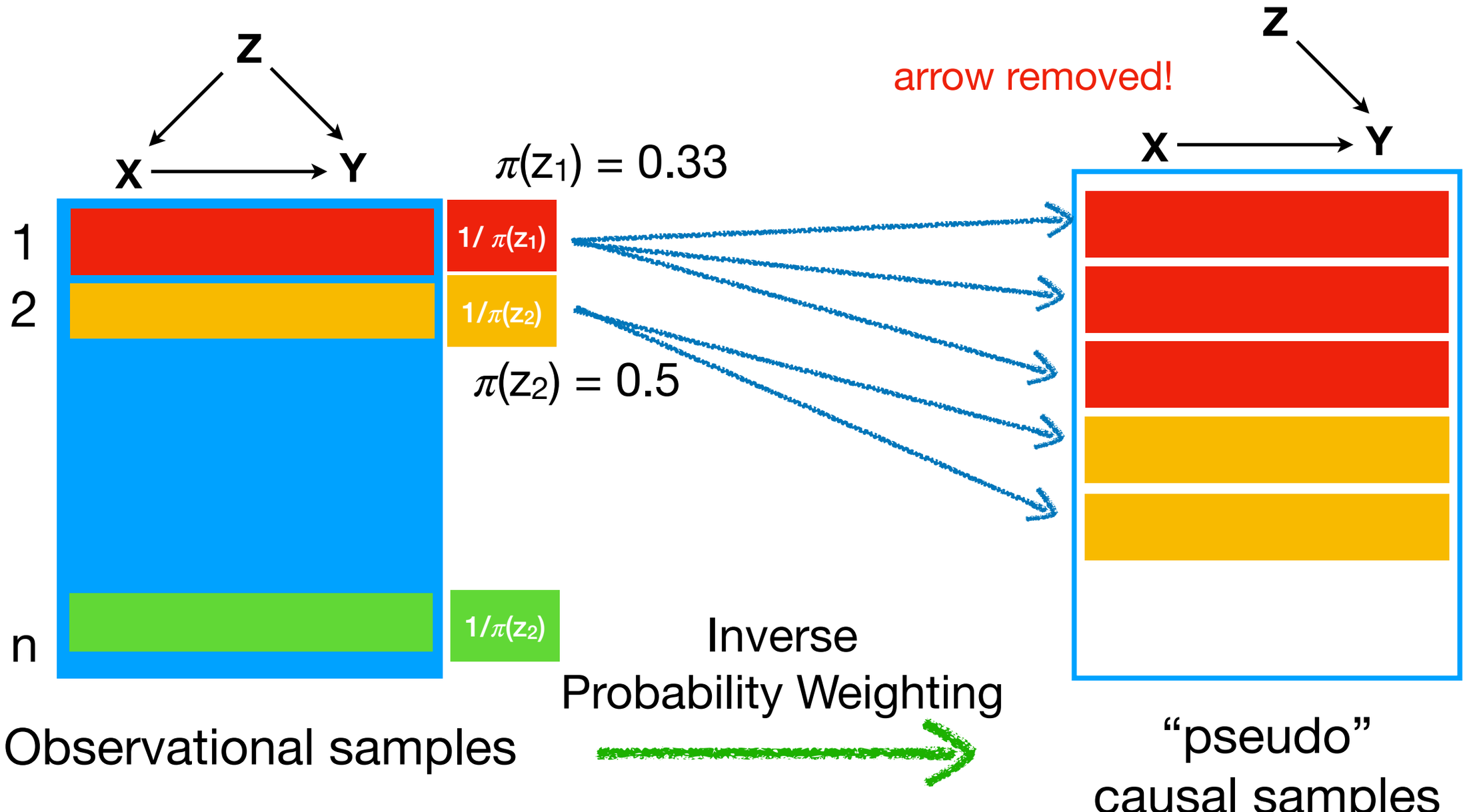
Assumption (Positivity). Positivity holds when $P(X = x \mid \mathbf{Z} = \mathbf{z})$ is bounded away from 0, that is $\exists \delta > 0$ such that $\forall\, x, \mathbf{z}$:

$$\delta < P(X = x \mid \mathbf{Z} = \mathbf{z}) < 1 - \delta\,.$$

IPW — Intuition

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

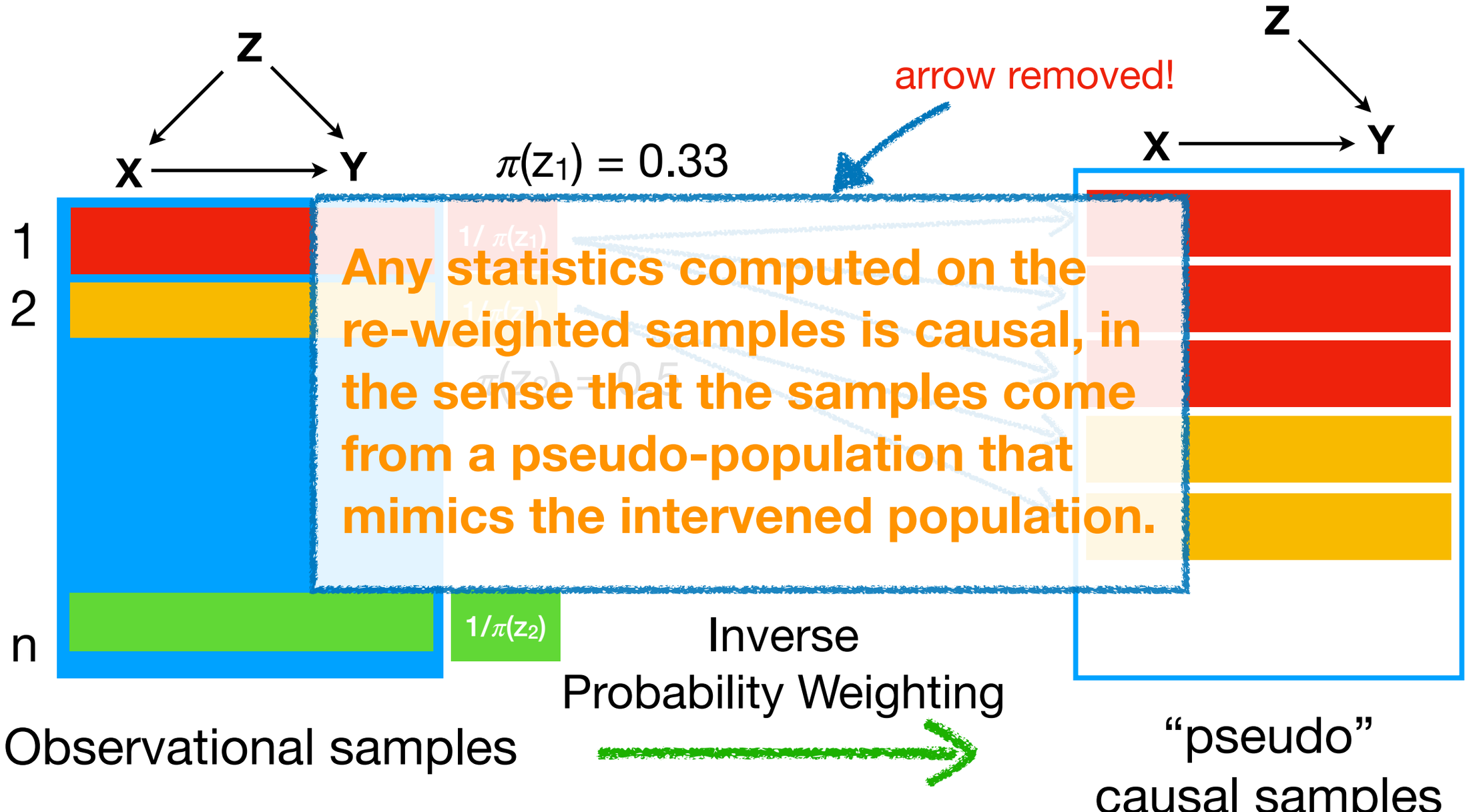
$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{Y}_i 1(\mathbf{X}_i = \mathbf{x})}{\pi(\mathbf{z})}$$



IPW — Intuition

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

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{Y}_i 1(\mathbf{X}_i = \mathbf{x})}{\pi(\mathbf{z})}$$



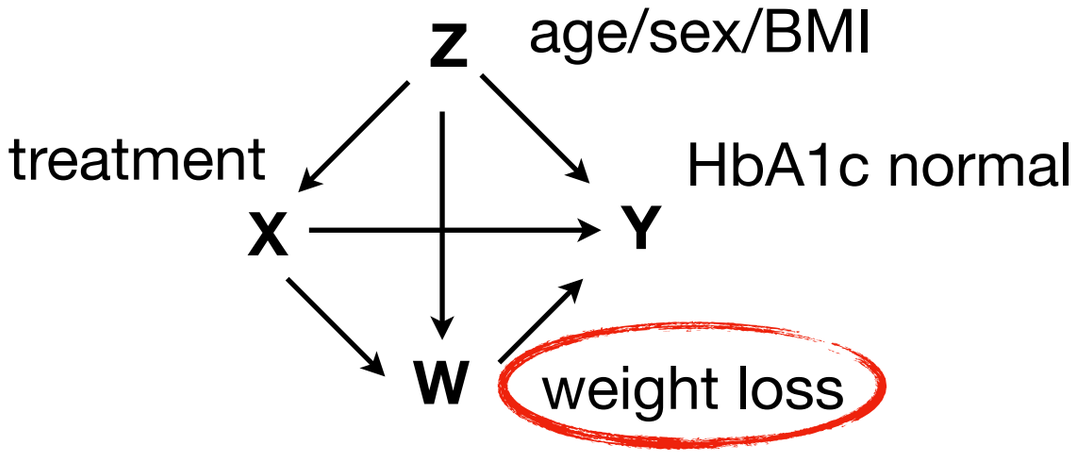
Remember Back-Door Lessons: #1 — Avoid Including *Mediators*

Available Data:

ID	semaglutide	age	*sex*	*BMI*	*weight loss*	*HbA1c normal*
200003	0	33	*1*	*33*	*1*	*1*
…	…	…	…	…	…	…
299999	1	47	*0*	*36*	*0*	*1*

Question: When computing $\pi_x(z)$, can I regress $X \sim \ldots$ - Y?

Back-door Lesson: No — you are at risk of including mediators.



Do not include weight loss!

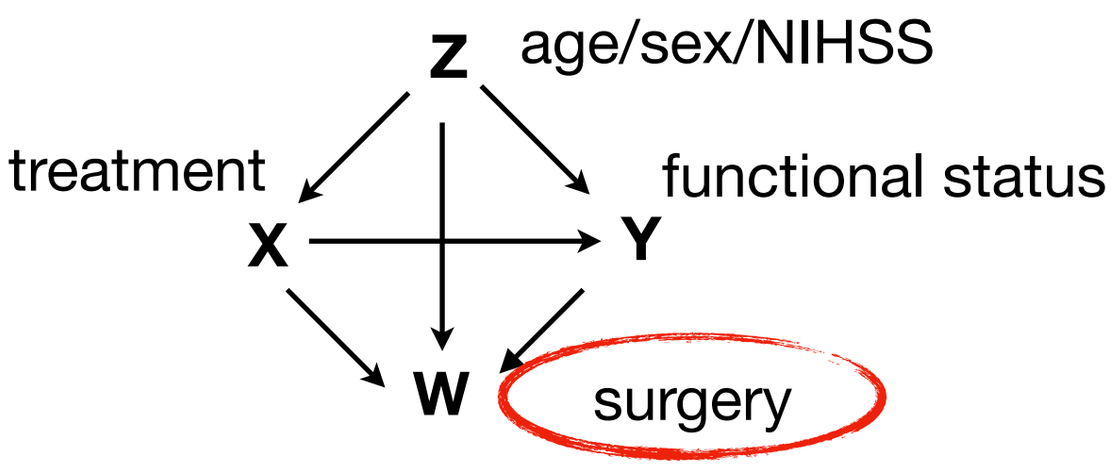
Remember Back-Door Lessons: #2 — Avoid Including *Colliders*

Available Data:

ID	thrombectomy	age	*sex*	*NIHSS*	*mRS (func. status)*	*surgery*
200003	0	58	*1*	*18*	*1*	*1*
…	…	…	…	…	…	…
299999	1	66	*0*	*34*	*0*	*1*

Question: When computing $\pi_x(z)$, can I regress $X \sim \ldots$ - Y?

Back-door Lesson: No — you are at risk of including colliders.



Do not include surgery information!

Note on Estimator Consistency

- Often, we are interested whether our estimator converges to the ground truth,

- For this, we need $\pi(\mathbf{z}) \to P(\mathbf{X} = 1 \mid \mathbf{Z} = \mathbf{z})$ in some appropriate way (e.g., uniform consistency),

- We mention the issue of model *mis-specification:*

SCM M

$$Z \leftarrow U_Z$$

$$X \leftarrow \text{Bern}\left(\frac{\exp(\beta^T Z + Z_1^2)}{1 + \exp(\beta^T Z + Z_1^2)}\right)$$

- but we fit

$$X \overset{\text{logistic}}{\sim} Z$$

fragile

$\Longrightarrow$ $\pi(\mathbf{z})$ not consistent for $P(\mathbf{X} \mid \mathbf{Z} = \mathbf{z})$!

$\Longrightarrow$ estimator not consistent

Regression-Based Approach

- Going to back our back-door expression,

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

why not replace $P(\mathbf{z})$ with empirical distribution?

$$= \sum_{\mathbf{z}} E(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) \frac{1}{N} \sum_{i=1}^{N} 1(\mathbf{Z}_i = \mathbf{z})$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{\mathbf{z}} E(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) 1(\mathbf{Z}_i = \mathbf{z})$$

$$= \frac{1}{N} \sum_{i=1}^{N} E(\mathbf{y} \mid \mathbf{x}, \mathbf{Z}_i)$$

Fit a function $\mu_x(z)$ that approximates this conditional expectation

Regression Causal Effect Estimate

- Once again, for the causal effect we simply take a difference:

$$\hat{E}(\mathbf{y} \,|\, do(\mathbf{X} = 1)) - \hat{E}(\mathbf{y} \,|\, do(\mathbf{X} = 0)) \overset{(**)}{=} \frac{1}{N} \sum_{i=1}^{N} \left(\mu_{x_1}(\mathbf{Z}_i) - \mu_{x_0}(\mathbf{Z}_i) \right)$$

where $\mu_x(\mathbf{z})$ is an estimator of $E(\mathbf{Y} \mid \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})$

S1 write code to obtain the μ_x estimator

S2 implement expression $(**)$ to obtain effect estimate

Note: Back-door lessons apply equally!

Note: Mis-specification of μ a concern

FFT: do we need positivity?

Augmented Inverse Propensity Weighting (AIPW): Best of Both Worlds

- is there a way to leverage both μ_x, π_x?
- consider the following random variable:

$$\phi(\mathbf{V}) = \frac{1(X = x)(Y - \tilde{\mu}_x)}{\tilde{\pi}_x(\mathbf{Z})} + \tilde{\mu}_x$$

think of $\tilde{\pi}_x, \tilde{\mu}_x$ as estimators of true π_x, μ_x

$$E[\phi] = E[E[\phi \mid \mathbf{Z}]] = E[E[\frac{1(X = x)(Y - \tilde{\mu}_x(\mathbf{Z}))}{\tilde{\pi}_x(\mathbf{Z})} + \tilde{\mu}_x(\mathbf{Z}) \mid \mathbf{Z}]]$$

$$= E[E[\frac{\pi_x(\mathbf{Z})(\mu_x(\mathbf{Z}) - \tilde{\mu}_x(\mathbf{Z}))}{\tilde{\pi}_x(\mathbf{Z})} + \mu_x(\mathbf{Z}) \mid \mathbf{Z}]]$$

$$= E[\frac{\pi_x(\mu_x - \tilde{\mu}_x)}{\tilde{\pi}_x} + \tilde{\mu}_x]$$

Augmented Inverse Propensity Weighting (AIPW): Best of Both Worlds

$$E[\phi] = E[\frac{\pi_x(\mu_x - \tilde{\mu}_x)}{\tilde{\pi}_x} + \tilde{\mu}_x]$$

For the estimator, we can use

$$\frac{1}{N}\sum_{i=1}^{N}\phi(\mathbf{V}_i)$$ **(+ LLN)**

Case 1: $\tilde{\pi}_x = \pi_x$ (correct propensity)

$$E[\phi] = E[\frac{\pi_x(\mu_x - \tilde{\mu}_x)}{\pi_x} - \tilde{\mu}_x]$$

$$= E[\mu_x - \tilde{\mu}_x + \tilde{\mu}_x]$$

$$= E[\mu_x] = \psi$$ **… even if $\tilde{\mu}_x \neq \mu_x$**

Case 2: $\tilde{\mu}_x = \mu_x$ (correct regression)

$$E[\phi] = E[\frac{\pi_x(\mu_x - \mu_x)}{\tilde{\pi}_x} + \mu_x]$$

$$= E[\mu_x] = \psi$$ **… even if $\tilde{\pi}_x \neq \pi_x$**

Need to get only one estimator correct: this is known as *double robustness*.

AIPW Causal Effect Estimate

- AIPW estimator is given by:

$$\text{ATE} \overset{(\top)}{=} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i(Y_i-\mu_{x_1}(\mathbf{Z}_i))}{\pi(\mathbf{Z}_i)}+\mu_{x_1}(\mathbf{Z}_i)-\frac{(1-X_i)(Y_i-\mu_{x_0}(\mathbf{Z}_i))}{1-\pi(\mathbf{Z}_i)}-\mu_{x_0}(\mathbf{Z}_i)\right)$$

where $\mu_x(\mathbf{z})$, $\pi(\mathbf{z})$ are the estimators as before,of $E(\mathbf{Y} \mid \mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z})$, $P(\mathbf{X}=1 \mid \mathbf{Z}=\mathbf{z})$, respectively.

S1 write code to obtain the μ_x estimator

S2 write code to obtain the π estimator

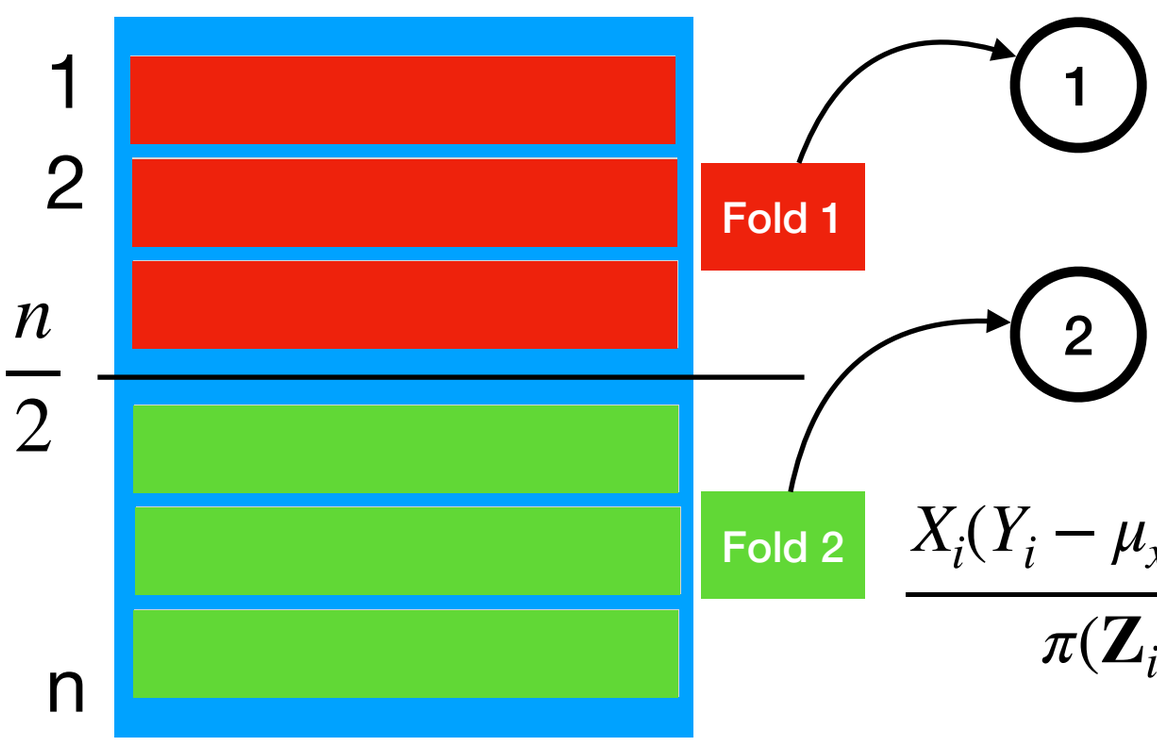
S3 implement expression (T) to obtain effect estimate

Note: Back-door lessons apply equally!

Note: Double robustness gained

Double Machine Learning

- What does Double Machine Learning (DML) do?
- It uses almost the same estimator as AIPW, but it performs *cross-fitting* or *sample-splitting:*



using $(X_i, Y_i, \mathbf{Z}_i)$ on Fold 1, fit estimators $\mu_x(\mathbf{Z})$ and $\pi(\mathbf{Z})$

using $(X_i, Y_i, \mathbf{Z}_i)$ on Fold 2, and μ_x, π from Fold 1, compute $\phi(\mathbf{V}_i)$:

$$\frac{X_i(Y_i - \mu_{x_1}(\mathbf{Z}_i))}{\pi(\mathbf{Z}_i)} + \mu_{x_1}(\mathbf{Z}_i) - \frac{(1 - X_i)(Y_i - \mu_{x_0}(\mathbf{Z}_i))}{1 - \pi(\mathbf{Z}_i)} - \mu_{x_0}(\mathbf{Z}_i)$$

Finally, use estimator $\frac{1}{N}\sum_{i=1}^{N} \phi(\mathbf{V}_i)$

DML Causal Effect Estimate

- DML estimator is given by:

$$\textbf{ATE} \overset{(\dagger)}{=} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i(Y_i-\mu_{x_1}(\mathbf{Z}_i))}{\pi(\mathbf{Z})}+\mu_{x_1}(\mathbf{Z}_i)-\frac{(1-X_i)(Y_i-\mu_{x_0}(\mathbf{Z}_i))}{1-\pi(\mathbf{Z})}-\mu_{x_0}(\mathbf{Z}_i)\right)$$

where $\mu_x(\mathbf{z})$, $\pi(\mathbf{z})$ are the *out-of-fold estimators* of $E(\mathbf{Y} \mid \mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z})$, $P(\mathbf{X}=1 \mid \mathbf{Z}=\mathbf{z})$, respectively.

S0 split data into folds 1, …, K

S1 by removing fold k, obtain estimator $\mu_x^{(-k)}$, $\pi^{(-k)}$

S2 obtain the values of $\phi(\mathbf{V}_i)$ on fold k using $\mu_x^{(-k)}$, $\pi^{(-k)}$

S3 average the values according to expression $(\dagger)$

Note: Double robustness retained

Note: Usually better sample efficiency