Estimating Causal Effects

Yonghan Jung
Purdue University
yonghanjung.me

2022.07.12

University of Seoul

Introduction

Outline

Outline

- **Part 1**: Double/Debiased Machine Learning (DML)

  Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning." The Econometrics Journal 21.1 (2018).

Outline

- **Part 1**: Double/Debiased Machine Learning (DML)

  Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning." The Econometrics Journal 21.1 (2018).

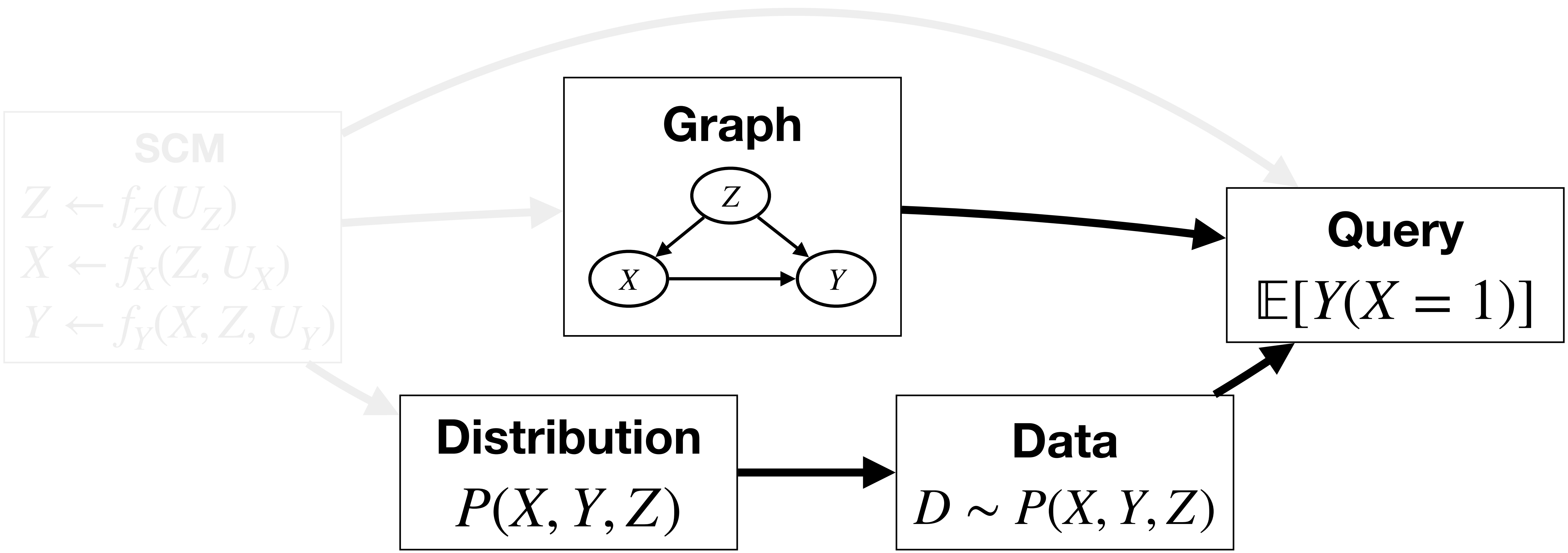
- **Part 2**: DML for any identifiable functional

  Jung, Tian, Bareinboim (2021a). Estimating Identifiable Causal Effects through Double Machine Learning. In Proceedings of the 35th AAAI Conference on AI, 2021.

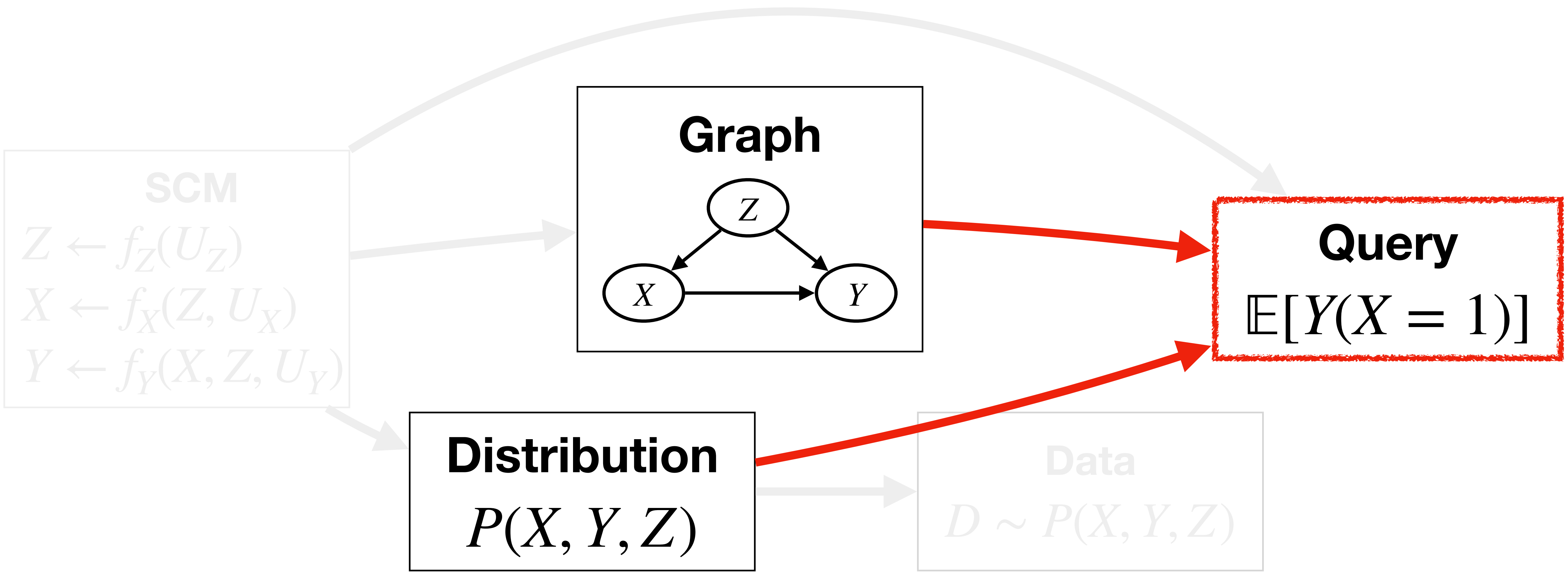
Part I.
Double/Debiased Machine Learning

Big Picture In Causal Inference

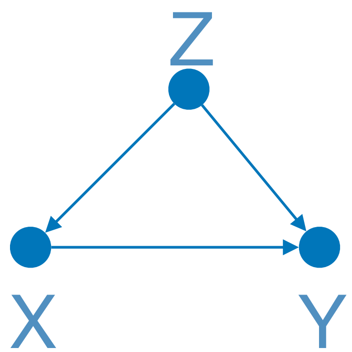
Big Picture for Causal Inference: Inaccessibility to SCMs



Causal Effect Identification: Big Picture (1)



Causal Effect Identification



Causal graph (G)

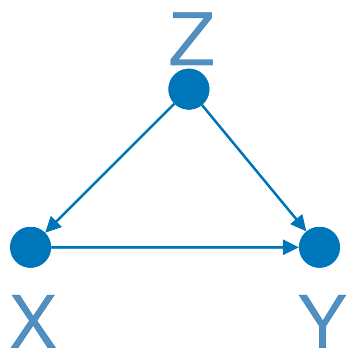
$$P(Z, X, Y)$$

Distribution on G (P)

$$Q_0 := \mathbb{E}[Y \mid do(x)]$$

Causal Query (Q_0)

Causal Effect Identification



Causal graph (G)

$P(Z, X, Y)$

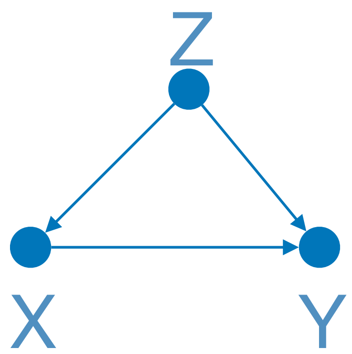
Distribution on G (P)

$Q_0 := \mathbb{E}[Y \mid do(x)]$

Causal Query (Q_0)

Given {G, P, Q_0},

Causal Effect Identification



Causal graph (G)

$$P(Z, X, Y)$$

Distribution on G (P)

$$Q_0 := \mathbb{E}[Y \mid do(x)]$$

Causal Query (Q_0)

Given {G, P, Q_0},

the causal effect identification (ID) task

$$\mathbf{ID}(G, P, Q_0)$$

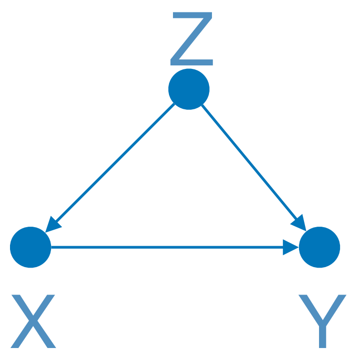
ID algorithm

Causal Effect Identification

Given {G, P, Q_0},

the causal effect identification (ID) task

finds the functional $C(P)$ s.t. $C(P) = Q_0$.



Causal graph (G)

$P(Z, X, Y)$

Distribution on G (P)

$Q_0 := \mathbb{E}[Y \mid do(x)]$

Causal Query (Q_0)

ID(G, P, Q_0)

ID algorithm

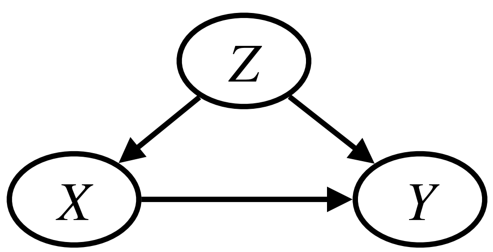
$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$

Causal functional

$C(P)$ **s.t.** $C(P) = Q_0$

Causal Effect Identification: Definition

Graph (G)



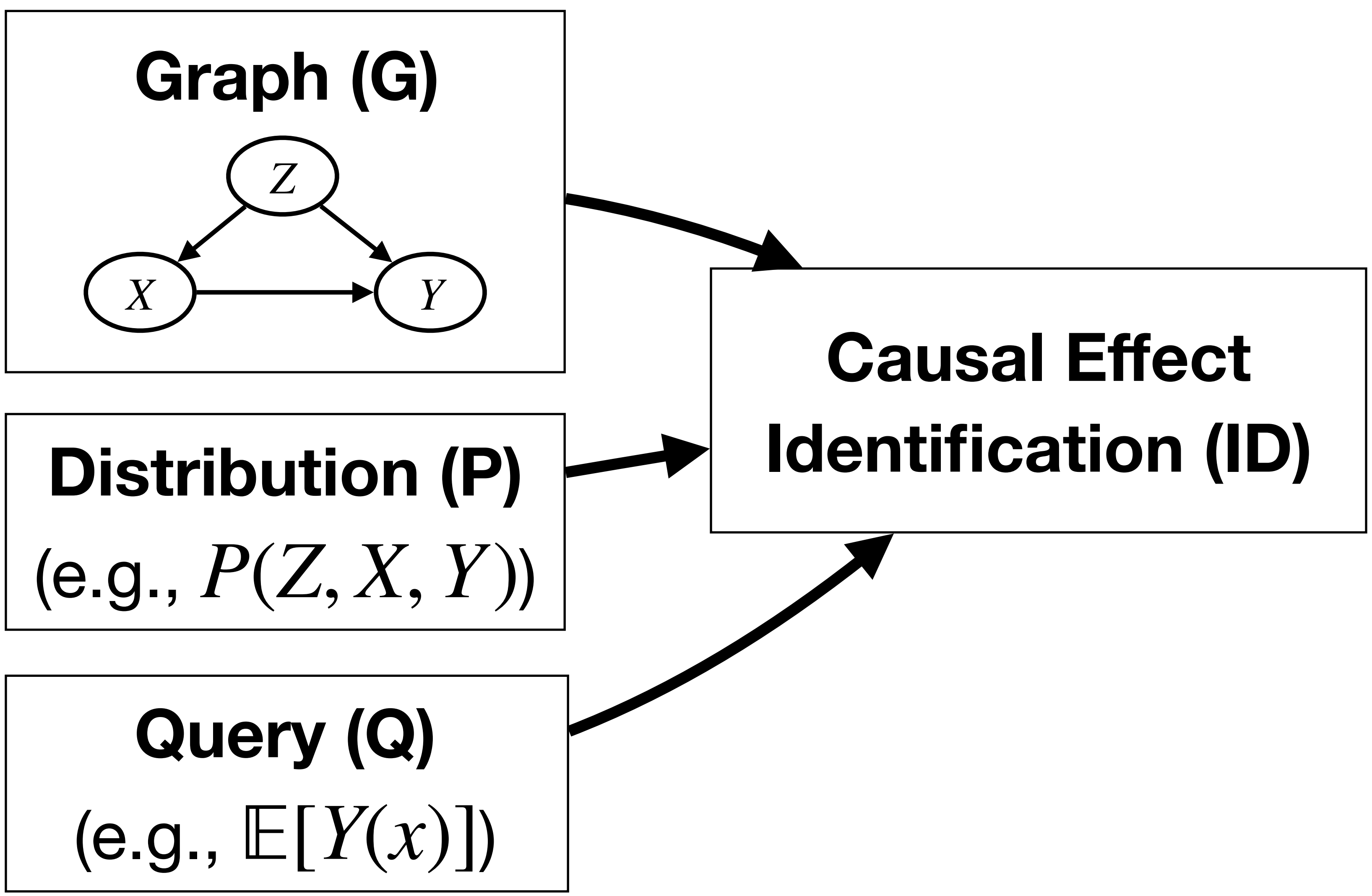
Distribution (P)

(e.g., $P(Z, X, Y)$)

Query (Q)

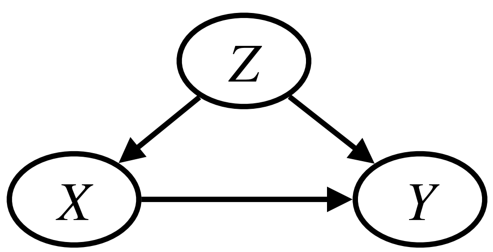
(e.g., $\mathbb{E}[Y(x)]$)

Causal Effect Identification: Definition



Causal Effect Identification: Definition

Graph (G)



Distribution (P)

(e.g., $P(Z, X, Y)$)

Query (Q)

(e.g., $\mathbb{E}[Y(x)]$)

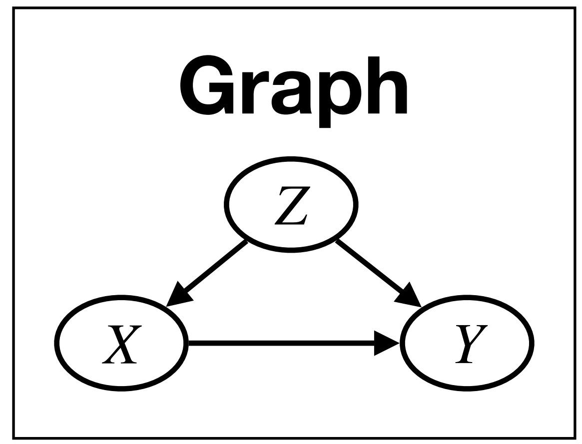
Causal Effect Identification (ID)

Solution

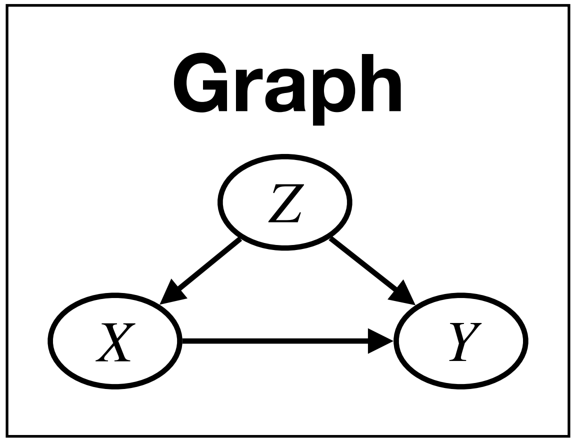
YES: A function $C(P)$ s.t. $C(P) = Q$.

No: If no such functions exist.

Causal Effect Estimation: Back-door Adjustment

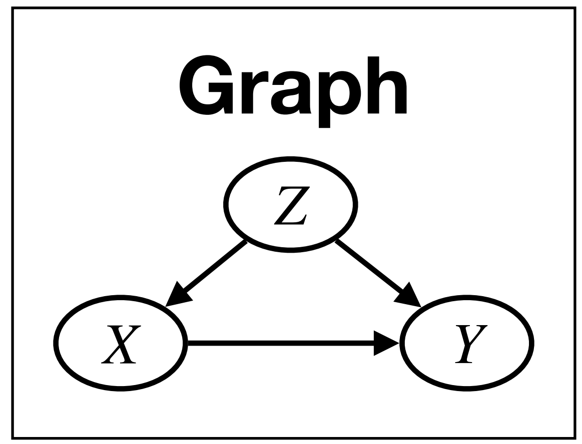


Causal Effect Estimation: Back-door Adjustment



Back-door Adjustment

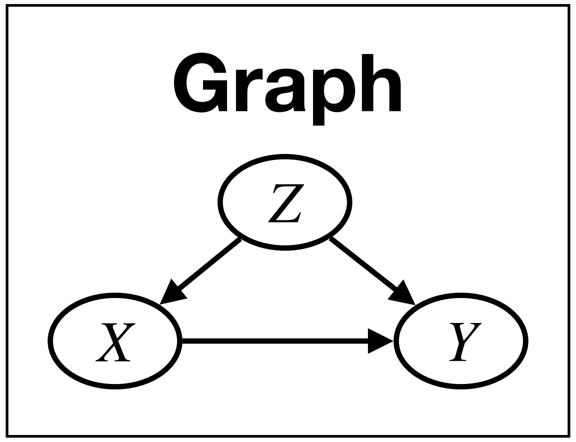
Causal Effect Estimation: Back-door Adjustment



Back-door Adjustment

↳ If there exists Z s.t. (1) Z is non-descendent of $\{X, Y\}$ and (2) $(Y \perp\!\!\!\perp X \,|\, Z)_{\underline{X}}$, then

Causal Effect Estimation: Back-door Adjustment

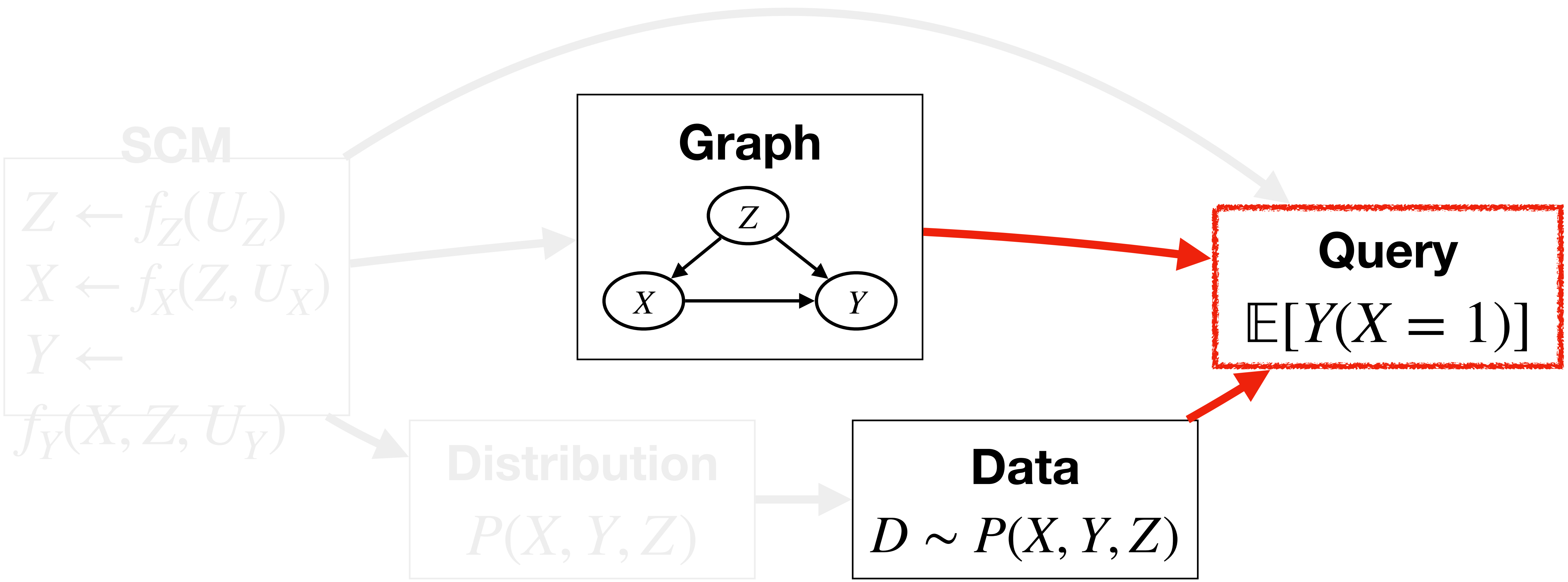


Back-door Adjustment

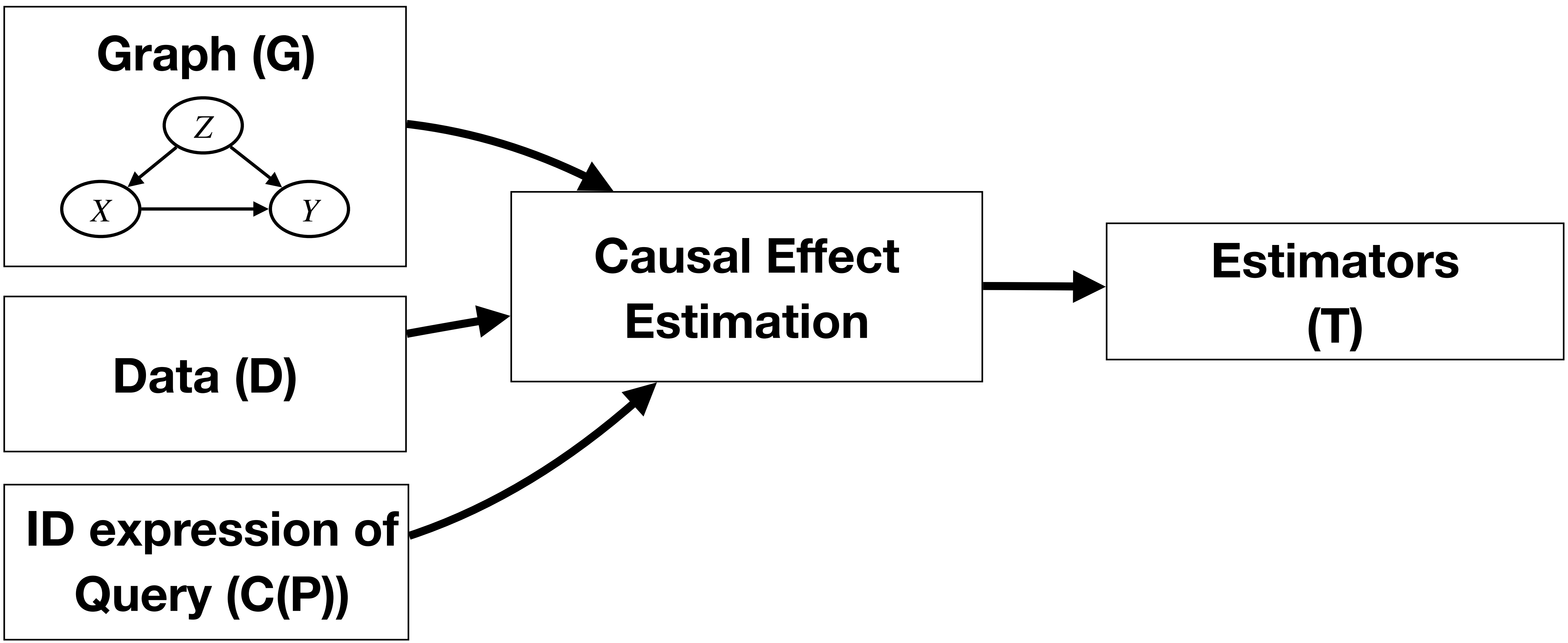
↳ If there exists Z s.t. (1) Z is non-descendent of $\{X, Y\}$ and (2) $(Y \perp\!\!\!\perp X \mid Z)_{\underline{X}}$, then

$$\mathbb{E}[Y(x)] = \sum_z \mathbb{E}[Y \mid x, z] P(z).$$

Causal Effect Estimation: Big Picture (1)

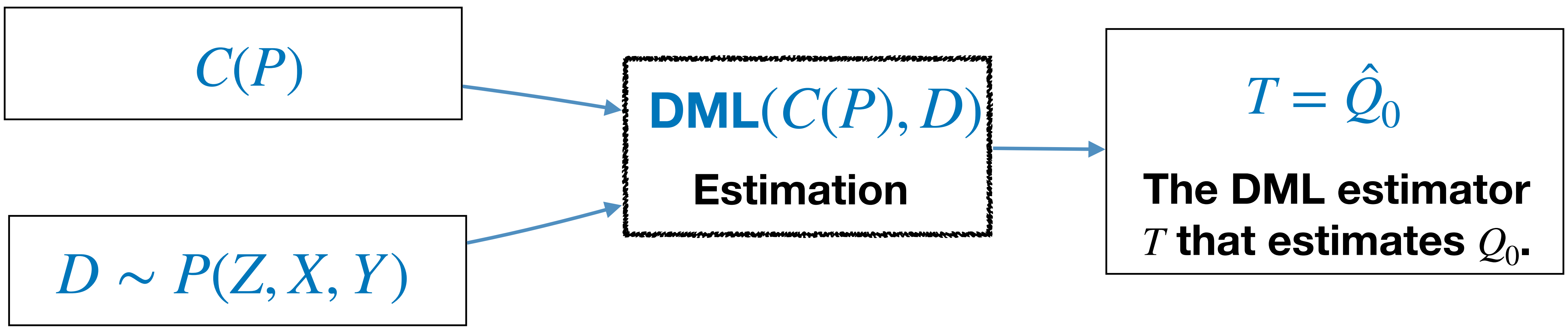


Causal Effect Estimation: Definition



Toward Double/Debiased Machine Learning

Double/Debiased Machine Learning (DML) [Chernozhukov et al., 2018] is a framework of constructing the estimator T.



Goal of the part I

Goal of the part I

We will understand the mechanism of the DML estimator with the BD adjustment example:

$$C(P) = \sum_z \mathbb{E}[Y | x, z] P(z) = \mathbb{E}[Y | do(x)]$$

Goal of the part I

We will understand the mechanism of the DML estimator with the BD adjustment example:

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z) = \mathbb{E}[Y \mid do(x)]$$

We assume the followings in the lecture.

$X \in \{0,1\}$ a binary treatment variable.

$P(v) > 0$ for any v.

Z can be multivariate (continuous/discrete)

Goal of the part I

We will understand the mechanism of the DML estimator with the BD adjustment example:

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z) = \mathbb{E}[Y \mid do(x)]$$

$= P(y \mid do(x))$ when $Y \leftarrow I_y(Y)$, an indicator function that is 1 when $Y = y$

We assume the followings in the lecture.

$X \in \{0,1\}$ a binary treatment variable.

$P(v) > 0$ for any v.

Z can be multivariate (continuous/discrete)

Estimand: Amenable Expression for Estimation

Challenges in estimating $C(P)$

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

Challenges in estimating $C(P)$

$$C(P) = \sum_z \mathbb{E}[Y \,|\, x, z] P(z)$$

Estimating $C(P)$ directly is challenging when Z is

- high-dimensional or
- a mixture of continuous/discrete variables.

Causal Estimand

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

Causal Estimand

$$C(P) = \sum_z \mathbb{E}[Y | x, z] P(z)$$

Instead of directly estimating $C(P)$, we find an *alternative representation* (***Estimand***) of $C(P)$, denoted $f(V; \eta)$:

Causal Estimand

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

Instead of directly estimating $C(P)$, we find an *alternative representation* (***Estimand***) of $C(P)$, denoted $f(V; \eta)$:

Estimand $f(V; \eta)$

An arbitrary function of V, η s.t. $\mathbb{E}[f(V; \eta_0)] = C(P)$ when $\eta = \eta_0$ for some η_0.

Causal Estimand

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

Instead of directly estimating $C(P)$, we find an *alternative representation* (***Estimand***) of $C(P)$, denoted $f(V; \eta)$:

Estimand $f(V; \eta)$

An arbitrary function of V, η s.t. $\mathbb{E}[f(V; \eta_0)] = C(P)$ when $\eta = \eta_0$ for some η_0.

$V := \{Z, X, Y\}$ all variables; and η is called "*nuisance*".

Causal Estimand

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

Instead of directly estimating $C(P)$, we find an *alternative representation* (***Estimand***) of $C(P)$, denoted $f(V; \eta)$:

Estimand $f(V; \eta)$

An arbitrary function of V, η s.t. $\mathbb{E}[f(V; \eta_0)] = C(P)$ when $\eta = \eta_0$ for some η_0.

$V := \{Z, X, Y\}$ all variables; and η is called "*nuisance*".

Next, we will introduce three estimands for the BD adjustment case.

Regression (REG)-based Estimand

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Regression (REG)-based Estimand

$$C(P) := \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$= \mathbb{E}_Z \left[\mathbb{E}[Y \mid x, Z] \right]$$

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Regression (REG)-based Estimand

$$C(P) := \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$= \mathbb{E}_Z \left[\mathbb{E}[Y \mid x, Z] \right]$$

Expectation over Z

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Regression (REG)-based Estimand

$$C(P) := \sum_z \mathbb{E}[Y \mid x, z] P(z)$$
$$= \mathbb{E}_Z \left[\mathbb{E}[Y \mid x, Z] \right]$$

Expectation over Z

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Let $\mu(X, Z)$ be any arbitrary function of $\{X, Z\}$, and $\mu_0(X, Z) := \mathbb{E}[Y \mid X, Z]$.

Regression (REG)-based Estimand

$$C(P) := \sum_{z} \mathbb{E}[Y \mid x, z] P(z)$$

$$= \mathbb{E}_Z \left[\mathbb{E}[Y \mid x, Z] \right]$$

Expectation over Z

$$C(P) = \sum_{z} \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Let $\mu(X, Z)$ be any arbitrary function of $\{X, Z\}$, and $\mu_0(X, Z) := \mathbb{E}[Y \mid X, Z]$.

The REG estimand is defined as,

$$f^{REG}(V; \mu) := \mu(x, Z)$$

Regression (REG)-based Estimand

$$C(P) := \sum_z \mathbb{E}[Y \mid x, z] P(z)$$
$$= \mathbb{E}_Z \left[\mathbb{E}[Y \mid x, Z] \right]$$

Expectation over Z

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$
$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Let $\mu(X, Z)$ be any arbitrary function of $\{X, Z\}$, and $\mu_0(X, Z) := \mathbb{E}[Y \mid X, Z]$.

The REG estimand is defined as,

$$f^{REG}(V; \mu) := \mu(x, Z)$$

Then, with the true nuisance μ_0,

$$\mathbb{E}[f^{REG}(V; \mu_0)] = \mathbb{E}[\mu_0(x, Z)] = \mathbb{E}[\mathbb{E}[Y \mid x, Z]] = C(P)$$

Inverse Probability Weighting (IPW)-based Estimand - 1

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Inverse Probability Weighting (IPW)-based Estimand - 1

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

$$C(P) = \sum_{z} \mathbb{E}[Y \mid x, z]P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Inverse Probability Weighting (IPW)-based Estimand - 1

$$C(P) = \sum_{z} \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

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

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

Indicator s.t. 1 when x' is equal to x

Inverse Probability Weighting (IPW)-based Estimand - 1

$$C(P) = \sum_{z} \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

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

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

Indicator s.t. 1 when x' is equal to x

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

Inverse Probability Weighting (IPW)-based Estimand - 1

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

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

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

Indicator s.t. 1 when x' is equal to x

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

P(z,x,y) = P(y|x,z)P(x|z)P(z)

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

Inverse Probability Weighting (IPW)-based Estimand - 1

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

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

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

Indicator s.t. 1 when x’ is equal to x

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

P(z,x,y) = P(y|x,z)P(x|z)P(z)

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

$$= \mathbb{E}_{X,Y,Z}\left[\frac{I_x(X)}{P(X \mid Z)} Y\right]$$

Inverse Probability Weighting (IPW)-based Estimand - 1

$$C(P) = \sum_{z} \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

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

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

Indicator s.t. 1 when x' is equal to x

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

P(z,x,y) = P(y|x,z)P(x|z)P(z)

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

$$= \mathbb{E}_{X,Y,Z} \left[\frac{I_x(X)}{P(X \mid Z)} Y \right] \quad \text{(Shortly, } = \mathbb{E} \left[\frac{I_x(X)}{P(X \mid Z)} Y \right] \text{)}$$

Inverse Probability Weighting (IPW)-based Estimand - 2

$$C(P) = \mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} Y\right]$$

from the prev. slide.

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

Inverse Probability Weighting (IPW)-based Estimand - 2

$$C(P) = \mathbb{E}\left[\frac{I_x(X)}{P(X|Z)}Y\right]$$

from the prev. slide.

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

Let $\pi(X|Z)$ be an arbitrary positive function and $\pi_0(X|Z) := P(X|Z)$.

Inverse Probability Weighting (IPW)-based Estimand - 2

$$C(P) = \mathbb{E}\left[\frac{I_x(X)}{P(X|Z)}Y\right]$$

from the prev. slide.

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

Let $\pi(X|Z)$ be an arbitrary positive function and $\pi_0(X|Z) := P(X|Z)$.

The IPW estimand is defined as

$$f^{IPW}(V;\pi) := \frac{I_x(X)}{\pi(X|Z)}Y$$

Inverse Probability Weighting (IPW)-based Estimand - 2

$$C(P) = \mathbb{E}\left[\frac{I_x(X)}{P(X|Z)}Y\right]$$

from the prev. slide.

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

Let $\pi(X|Z)$ be an arbitrary positive function and $\pi_0(X|Z) := P(X|Z)$.

The IPW estimand is defined as

$$f^{IPW}(V;\pi) := \frac{I_x(X)}{\pi(X|Z)}Y$$

Then, with the true nuisance π_0,

$$\mathbb{E}[f^{IPW}(V;\pi_0)] = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}Y\right] = \mathbb{E}\left[\frac{I_x(X)}{P(X|Z)}Y\right] = C(P)$$

Toward a
Doubly Robust Estimand

Toward a Doubly Robust Estimand

Now, we will introduce a "*doubly robust (DR)*" estimand, denoted

$$f^{DR}(V; \{\pi, \mu\})$$

where the name is from a property known as "*doubly robustness*":

Toward a Doubly Robust Estimand

Now, we will introduce a "*doubly robust (DR)*" estimand, denoted

$$f^{DR}(V; \{\pi, \mu\})$$

where the name is from a property known as "*doubly robustness*":

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] = C(P)$$

If either $\pi = \pi_0$ or $\mu = \mu_0$

Toward a Doubly Robust Estimand

Now, we will introduce a "*doubly robust (DR)*" estimand, denoted

$$f^{DR}(V; \{\pi, \mu\})$$

where the name is from a property known as "*doubly robustness*":

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] = C(P)$$

If either $\pi = \pi_0$ or $\mu = \mu_0$

Doubly robustness implies we have double chance for being correct in terms of the estimation of the nuisances parameters.

Derivation of Doubly Robust Estimand - 1

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

$$f^{REG}(V; \eta_0 := \mu) := \mu(x, Z)$$

$$f^{IPW}(V; \eta := \pi) := \frac{I_x(X)}{\pi(X \mid Z)} Y$$

Derivation of Doubly Robust Estimand - 1

$$C(P) = \mathbb{E}\left[f^{IPW}(V; \pi_0)\right]$$

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}[f(V; \eta_0)] = C(P)$$

$$f^{REG}(V; \eta_0 := \mu) := \mu(x, Z)$$

$$f^{IPW}(V; \eta := \pi) := \frac{I_x(X)}{\pi(X \mid Z)} Y$$

Derivation of Doubly Robust Estimand - 1

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right]$$

$$+\mathbb{E}[f^{REG}(V;\mu_0)]$$

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

$$f^{REG}(V;\eta_0 := \mu) := \mu(x,Z)$$

$$f^{IPW}(V;\eta := \pi) := \frac{I_x(X)}{\pi(X|Z)}Y$$

Derivation of Doubly Robust Estimand - 1

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right]$$

$$+\mathbb{E}[f^{REG}(V;\mu_0)]$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

$$f^{REG}(V;\eta_0 := \mu) := \mu(x,Z)$$

$$f^{IPW}(V;\eta := \pi) := \frac{I_x(X)}{\pi(X|Z)}Y$$

Derivation of Doubly Robust Estimand - 1

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] = \text{C(P)}$$

$$+\mathbb{E}[f^{REG}(V;\mu_0)]$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

$$f^{REG}(V;\eta_0 := \mu) := \mu(x,Z)$$

$$f^{IPW}(V;\eta := \pi) := \frac{I_x(X)}{\pi(X|Z)}Y$$

Derivation of Doubly Robust Estimand - 1

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right]$$ = C(P)

$$+\mathbb{E}[f^{REG}(V;\mu_0)]$$ = C(P)

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

$$f^{REG}(V;\eta_0 := \mu) := \mu(x,Z)$$

$$f^{IPW}(V;\eta := \pi) := \frac{I_x(X)}{\pi(X|Z)}Y$$

Derivation of Doubly Robust Estimand - 1

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right]$$ = C(P)

$$+\mathbb{E}[f^{REG}(V;\mu_0)]$$ = C(P)

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

= C(P), as shown next

$$C(P) = \sum_z \mathbb{E}[Y|x,z]P(z)$$

$$\mathbb{E}[f(V;\eta_0)] = C(P)$$

$$f^{REG}(V;\eta_0 := \mu) := \mu(x,Z)$$

$$f^{IPW}(V;\eta := \pi) := \frac{I_x(X)}{\pi(X|Z)}Y$$

Derivation of Doubly Robust Estimand - 2

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

Derivation of Doubly Robust Estimand - 2

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X \mid Z)}\mu_0(X, Z)\right] = \sum_{x', z} \frac{I_x(x')}{\pi_0(x' \mid z)} \underbrace{\mu_0(x', z)}_{=\mathbb{E}[Y|x,z]} \underbrace{P(x' \mid z)}_{=\pi_0(x'|z)} P(z)$$

Derivation of Doubly Robust Estimand - 2

$$C(P) = \sum_z \mathbb{E}[Y \,|\, x, z] P(z)$$

$$\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X \,|\, Z)} \mu_0(X, Z)\right] = \sum_{x',z} \frac{I_x(x')}{\pi_0(x' \,|\, z)} \underbrace{\mu_0(x', z)}_{=\mathbb{E}[Y|x,z]} \underbrace{P(x' \,|\, z)}_{=\pi_0(x'|z)} P(z)$$

$$= \sum_z \mathbb{E}[Y \,|\, x, z] P(z)$$

Derivation of Doubly Robust Estimand - 2

$$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X \mid Z)} \mu_0(X, Z)\right] = \sum_{x', z} \frac{I_x(x')}{\pi_0(x' \mid z)} \underbrace{\mu_0(x', z)}_{=\mathbb{E}[Y|x,z]} \underbrace{P(x' \mid z)}_{=\pi_0(x'|z)} P(z)$$

$$= \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

$$= C(P)$$

Derivation of Doubly Robust Estimand - 3

Derivation of Doubly Robust Estimand - 3

$$C(P) = \mathbb{E}\left[f^{IPW}(V; \pi_0)\right] + \mathbb{E}[f^{REG}(V; \mu_0)] - \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X \mid Z)} \mu_0(X, Z)\right]$$

Derivation of Doubly Robust Estimand - 3

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] + \mathbb{E}[f^{REG}(V;\mu_0)] - \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X\,|\,Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0) + f^{REG}(V;\mu_0) - \frac{I_x(X)}{\pi_0(X\,|\,Z)}\mu_0(X,Z)\right]$$

Derivation of Doubly Robust Estimand - 3

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] + \mathbb{E}[f^{REG}(V;\mu_0)] - \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X\,|\,Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0) + f^{REG}(V;\mu_0) - \frac{I_x(X)}{\pi_0(X\,|\,Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X\,|\,Z)}Y + \mu_0(x,Z) - \frac{I_x(X)}{\pi_0(X\,|\,Z)}\mu_0(X,Z)\right]$$

Derivation of Doubly Robust Estimand - 3

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] + \mathbb{E}[f^{REG}(V;\mu_0)] - \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0) + f^{REG}(V;\mu_0) - \frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}Y + \mu_0(x,Z) - \frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\left\{Y - \mu_0(X,Z)\right\} + \mu_0(x,Z)\right]$$

Derivation of Doubly Robust Estimand - 3

$$C(P) = \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] + \mathbb{E}[f^{REG}(V;\mu_0)] - \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0) + f^{REG}(V;\mu_0) - \frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}Y + \mu_0(x,Z) - \frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\left\{Y - \mu_0(X,Z)\right\} + \mu_0(x,Z)\right]$$

$$f^{DR}(V;\{\pi,\mu\}) := \frac{I_x(X)}{\pi(X|Z)}\{Y - \mu(X,Z)\} + \mu(x,Z)$$

Doubly Robustness - Proof 1

Doubly Robustness - Proof 1

$\mathbb{E}[f^{DR}(V; \{\pi_0, \mu\})] = C(P)$ for any μ

Doubly Robustness - Proof 1

$$\mathbb{E}[f^{DR}(V;\{\pi_0,\mu\})] = C(P) \text{ for any } \mu$$

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi_0,\mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] \quad \text{—— A}$$

$$+\mathbb{E}[f^{REG}(V;\mu)] \quad \text{—— B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right] \quad \text{—— C}$$

Doubly Robustness - Proof 1

$$\mathbb{E}[f^{DR}(V;\{\pi_0,\mu\})] = C(P) \text{ for any } \mu$$

$$A = \mathbb{E}[f^{IPW}(V;\pi_0)] = C(P)$$

$$\mathbb{E}[f^{DR}(V;\eta = \{\pi_0,\mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] \text{ ——— A}$$

$$+\mathbb{E}[f^{REG}(V;\mu)] \text{ ——— B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right] \text{ ——— C}$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})] = C(P) \text{ for any } \mu$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})]$$
$$= \mathbb{E}\left[f^{IPW}(V; \pi_0)\right] \text{ ——— A}$$
$$+\mathbb{E}[f^{REG}(V; \mu)] \text{ ——— B}$$
$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X \mid Z)}\mu(X, Z)\right] \text{ ——— C}$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi_0,\mu\})] = C(P) \text{ for any } \mu$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right]$$

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi_0,\mu\})]$$
$$= \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] \text{ — A}$$
$$+\mathbb{E}[f^{REG}(V;\mu)] \text{ — B}$$
$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right] \text{ — C}$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})] = C(P) \text{ for any } \mu$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X \mid Z)}\mu(X, Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi_0(x' \mid z)}\mu(x', z)\underbrace{P(x' \mid z)}_{=\pi_0(x'|z)}P(z)$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V; \pi_0)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V; \mu)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X \mid Z)}\mu(X, Z)\right] \text{ — C}$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})] = C(P) \text{ for any } \mu$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi_0(x'|z)}\mu(x',z)\underbrace{P(x'|z)}_{=\pi_0(x'|z)}P(z)$$

$$= \sum_z \mu(x,z)P(z) = \mathbb{E}[\mu(x,Z)]$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V; \pi_0)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V; \mu)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right] \text{ — C}$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})] = C(P) \text{ for any } \mu$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X, Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi_0(x'|z)}\mu(x', z)\underbrace{P(x'|z)}_{=\pi_0(x'|z)} P(z)$$

$$= \sum_{z} \mu(x, z)P(z) = \mathbb{E}[\mu(x, Z)]$$

$$= \mathbb{E}[f^{REG}(V; \mu)]$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V; \pi_0)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V; \mu)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X, Z)\right] \text{ — C}$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi_0,\mu\})] = C(P) \text{ for any } \mu$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi_0(x'|z)}\mu(x',z)\underbrace{P(x'|z)}_{=\pi_0(x'|z)}P(z)$$

$$= \sum_z \mu(x,z)P(z) = \mathbb{E}[\mu(x,Z)]$$

$$= \mathbb{E}[f^{REG}(V;\mu)]$$

$$= B$$

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi_0,\mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V;\mu)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right] \text{ — C}$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})] = C(P) \text{ for any } \mu$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi_0(x'|z)}\mu(x',z)\underbrace{P(x'|z)}_{=\pi_0(x'|z)}P(z)$$

$$= \sum_z \mu(x,z)P(z) = \mathbb{E}[\mu(x,Z)]$$

$$= \mathbb{E}[f^{REG}(V;\mu)]$$

$$= B$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi_0)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V;\mu)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right] \text{ — C}$$

$$= C(P) + B - B$$

Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})] = C(P) \text{ for any } \mu$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi_0(x'|z)}\mu(x',z)\underbrace{P(x'|z)}_{=\pi_0(x'|z)}P(z)$$

$$= \sum_z \mu(x,z)P(z) = \mathbb{E}[\mu(x,Z)]$$

$$= \mathbb{E}[f^{REG}(V;\mu)]$$

$$= B$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi_0, \mu\})]$$

$$= \mathbb{E}\left[f^{IPW}(V; \pi_0)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V;\mu)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi_0(X|Z)}\mu(X,Z)\right] \text{ — C}$$

$$= C(P) + B - B$$

$$= C(P)$$

Doubly Robustness - Proof 3

$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})] = C(P)$ for any positive π

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})]$$
$$= \mathbb{E}\left[f^{IPW}(V; \pi)\right]$$
$$+\mathbb{E}[f^{REG}(V; \mu_0)]$$
$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X\,|\,Z)}\mu_0(X, Z)\right]$$

Doubly Robustness - Proof 3

$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})] = C(P)$ for any positive π

$$B = \mathbb{E}[f^{REG}(V; \mu_0)] = C(P)$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})]$$

$$= \mathbb{E}\left[f^{IPW}(V; \pi)\right] \quad \text{——— A}$$

$$+\mathbb{E}[f^{REG}(V; \mu_0)] \quad \text{——— B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X, Z)\right] \quad \text{——— C}$$

Doubly Robustness - Proof 4

$\mathbb{E}[f^{DR}(V;\eta=\{\pi,\mu_0\})]=C(P)$ for any positive π

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi,\mu_0\})]$$
$$=\mathbb{E}\left[f^{IPW}(V;\pi)\right] \text{ ——— A}$$
$$+\mathbb{E}[f^{REG}(V;\mu_0)] \text{ ——— B}$$
$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ ——— C}$$

Doubly Robustness - Proof 4

$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})] = C(P)$ for any positive π

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right]$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})]$$

$$= \mathbb{E}\left[f^{IPW}(V; \pi)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V; \mu_0)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ — C}$$

Doubly Robustness - Proof 4

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi,\mu_0\})] = C(P) \text{ for any positive } \pi$$

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi(x'|z)} \underbrace{\mu_0(x',z)}_{=\sum_y yP(y|x',z)} P(x'|z)P(z)$$

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi,\mu_0\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi)\right] \text{ —— A}$$

$$+\mathbb{E}[f^{REG}(V;\mu_0)] \text{ —— B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ —— C}$$

Doubly Robustness - Proof 4

$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})] = C(P)$ for any positive π

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi(x'|z)} \underbrace{\mu_0(x',z)}_{=\sum_y yP(y|x',z)} P(x'|z)P(z)$$

$$= \sum_{z,x',y} \frac{I_x(x')}{\pi(x'|z)} yP(z,x',y)$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V;\mu_0)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ — C}$$

Doubly Robustness - Proof 4

$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})] = C(P)$ for any positive π

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi(x'|z)} \underbrace{\mu_0(x',z)}_{=\sum_y yP(y|x',z)} P(x'|z)P(z)$$

$$= \sum_{z,x',y} \frac{I_x(x')}{\pi(x'|z)} yP(z,x',y)$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}Y\right] = \mathbb{E}[f^{IPW}(V;\pi)]$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V;\mu_0)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ — C}$$

Doubly Robustness - Proof 4

$\mathbb{E}[f^{DR}(V;\eta = \{\pi, \mu_0\})] = C(P)$ for any positive π

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi(x'|z)} \underbrace{\mu_0(x',z)}_{=\sum_y yP(y|x',z)} P(x'|z)P(z)$$

$$= \sum_{z,x',y} \frac{I_x(x')}{\pi(x'|z)} yP(z,x',y)$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}Y\right] = \mathbb{E}[f^{IPW}(V;\pi)]$$

$$= A$$

$$\mathbb{E}[f^{DR}(V;\eta = \{\pi, \mu_0\})]$$

$$= \mathbb{E}\left[f^{IPW}(V;\pi)\right] \text{ — A}$$

$$+\mathbb{E}[f^{REG}(V;\mu_0)] \text{ — B}$$

$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ — C}$$

Doubly Robustness - Proof 4

$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})] = C(P)$ for any positive π

$$C = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right]$$

$$= \sum_{z,x'} \frac{I_x(x')}{\pi(x'|z)} \underbrace{\mu_0(x',z)}_{=\sum_y yP(y|x',z)} P(x'|z)P(z)$$

$$= \sum_{z,x',y} \frac{I_x(x')}{\pi(x'|z)} yP(z,x',y)$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}Y\right] = \mathbb{E}[f^{IPW}(V;\pi)]$$

$$= A$$

$$\mathbb{E}[f^{DR}(V; \eta = \{\pi, \mu_0\})]$$
$$= \mathbb{E}\left[f^{IPW}(V;\pi)\right] \text{ — A}$$
$$+\mathbb{E}[f^{REG}(V;\mu_0)] \text{ — B}$$
$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ — C}$$

$$= A + C(P) - A$$

Doubly Robustness - Proof 4

$\mathbb{E}[f^{DR}(V;\eta=\{\pi,\mu_0\})]=C(P)$ for any positive π

$$C=\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right]$$

$$=\sum_{z,x'}\frac{I_x(x')}{\pi(x'|z)}\underbrace{\mu_0(x',z)}_{=\sum_y yP(y|x',z)}P(x'|z)P(z)$$

$$=\sum_{z,x',y}\frac{I_x(x')}{\pi(x'|z)}yP(z,x',y)$$

$$=\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}Y\right]=\mathbb{E}[f^{IPW}(V;\pi)]$$

$$=A$$

$$\mathbb{E}[f^{DR}(V;\eta=\{\pi,\mu_0\})]$$
$$=\mathbb{E}\left[f^{IPW}(V;\pi)\right] \text{ — A}$$
$$+\mathbb{E}[f^{REG}(V;\mu_0)] \text{ — B}$$
$$-\mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\mu_0(X,Z)\right] \text{ — C}$$

$$=A+C(P)-A$$

$$=C(P)$$

Intermediate Summary - Estimands

Intermediate Summary - Estimands

So far…

Intermediate Summary - Estimands

So far…

1. We defined an estimand $f(V;\eta)$, a functional for estimating the causal effects.

Intermediate Summary - Estimands

So far…

1. We defined an estimand $f(V; \eta)$, a functional for estimating the causal effects.

2. For the BD adjustment, we illustrated three estimands (REG, IPW, DR), and showed that the DR estimand has the doubly-robustness property.

Intermediate Summary - Estimands

So far…

1. We defined an estimand $f(V; \eta)$, a functional for estimating the causal effects.

2. For the BD adjustment, we illustrated three estimands (REG, IPW, DR), and showed that the DR estimand has the doubly-robustness property.

Next, we will introduce a general principle for choosing an estimand.

Orthogonal Estimand

Idea of an Orthogonal Estimand

Idea of an Orthogonal Estimand

If $\mathbb{E}[f(V;\eta)]$ is invariant to the small perturbation of η,

Idea of an Orthogonal Estimand

If $\mathbb{E}[f(V;\eta)]$ is invariant to the small perturbation of η,

… despite the error of η,

Idea of an Orthogonal Estimand

If $\mathbb{E}[f(V;\eta)]$ is invariant to the small perturbation of η,

… despite the error of η,

… $\mathbb{E}[f(V;\eta)]$ will not suffer from the error.

Idea of an Orthogonal Estimand

If $\mathbb{E}[f(V;\eta)]$ is invariant to the small perturbation of η,

… despite the error of η,

… $\mathbb{E}[f(V;\eta)]$ will not suffer from the error.

We will formalize this idea by considering the *directional derivative* of $\mathbb{E}[f(V;\eta)]$.

Directional Derivative & Orthogonal Estimand

Directional Derivative & Orthogonal Estimand

Directional Derivative:

For a function $g(\eta)$, its derivative along the direction h at η_0 is given as

$$D_\eta g(\eta_0)\{h\} := \frac{\partial}{\partial t} g(\eta_0 + th)\Big|_{t=0}$$

Directional Derivative & Orthogonal Estimand

Directional Derivative:

For a function $g(\eta)$, its derivative along the direction h at η_0 is given as

$$D_\eta g(\eta_0)\{h\} := \frac{\partial}{\partial t} g(\eta_0 + th)\Big|_{t=0}$$

Orthogonal Estimand:

$f(V;\eta)$ is an *orthogonal estimand* if

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Directional Derivative & Orthogonal Estimand

Directional Derivative:

For a function $g(\eta)$, its derivative along the direction h at η_0 is given as

$$D_\eta g(\eta_0)\{h\} := \frac{\partial}{\partial t} g(\eta_0 + th)\bigg|_{t=0}$$

Orthogonal Estimand:

$f(V;\eta)$ is an *orthogonal estimand* if

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

The estimand is invariant along the error $(\eta - \eta_0)$ at the true nuisance η_0

Debiasedness of Orthogonal Estimands

Orthogonal Estimand:

$f(V;\eta)$ is an *orthogonal estimand* if

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness of Orthogonal Estimands

Orthogonal Estimand:

$f(V;\eta)$ is an *orthogonal estimand* if

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Debiasedness of Orthogonal Estimands

Orthogonal Estimand:

$f(V;\eta)$ is an *orthogonal estimand* if

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

If η converges to η_0 at some slow rate, say $N^{-1/4}$

Debiasedness of Orthogonal Estimands

Orthogonal Estimand:

$f(V;\eta)$ is an *orthogonal estimand* if

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

If η converges to η_0 at some slow rate, say $N^{-1/4}$

$\mathbb{E}[f(V;\eta)]$ converges to $C(P)$ at $(N^{-1/4})^2 = N^{-1/2}$ rate.

Debiasedness of Orthogonal Estimands - Proof

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Debiasedness of Orthogonal Estimands - Proof

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Then, by Taylor's expansion (up to the 2nd order),

Debiasedness of Orthogonal Estimands - Proof

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Then, by Taylor's expansion (up to the 2nd order),

$$\mathbb{E}[f(V;\eta)] - \mathbb{E}[f(V;\eta_0)] = D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} + \frac{1}{2} D_\eta^2 \mathbb{E}[f(V;\eta)]\{\eta - \eta_0\}^2$$

Debiasedness of Orthogonal Estimands - Proof

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Then, by Taylor's expansion (up to the 2nd order),

$$\mathbb{E}[f(V;\eta)] - \mathbb{E}[f(V;\eta_0)] = \cancel{D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta - \eta_0\}} + \frac{1}{2}D_\eta^2 \mathbb{E}[f(V;\eta)]\{\eta - \eta_0\}^2$$

0

Debiasedness of Orthogonal Estimands - Proof

$$D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta-\eta_0\} = 0$$

Debiasedness

For any η, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Then, by Taylor's expansion (up to the 2nd order),

$$\mathbb{E}[f(V;\eta)] - \mathbb{E}[f(V;\eta_0)] = \underbrace{D_\eta \mathbb{E}[f(V;\eta_0)]\{\eta-\eta_0\}}_{0} + \frac{1}{2} D_\eta^2 \mathbb{E}[f(V;\eta)]\{\eta-\eta_0\}^2$$

$O_P\left(\| \eta - \eta_0 \|_{L_2(P)}^2\right) := O_P\left(\mathbb{E}[(\eta-\eta_0)^2]\right)$, shortly, $O_P\left(\| \eta - \eta_0 \|^2\right)$.

Orthogonal Estimand - Two nuisances

Orthogonal Estimand - Two nuisances

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right) + O_P\left(\| \eta^b - \eta_0^b \|^2\right) + O_P\left(\| \eta^a - \eta_0^a \| \| \eta^b - \eta_0^b \|\right)$$

Orthogonal Estimand - Two nuisances

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta^a_0 \|^2\right) + O_P\left(\| \eta^b - \eta^b_0 \|^2\right) + O_P\left(\| \eta^a - \eta^a_0 \|\| \eta^b - \eta^b_0 \|\right)$$

If $\eta = \{\eta^a, \eta^b\}$ converges to $\eta_0 := \{\eta^a_0, \eta^b_0\}$ at some rate, say $N^{-1/4}$

Orthogonal Estimand - Two nuisances

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right) + O_P\left(\| \eta^b - \eta_0^b \|^2\right) + O_P\left(\| \eta^a - \eta_0^a \|\| \eta^b - \eta_0^b \|\right)$$

If $\eta = \{\eta^a, \eta^b\}$ converges to $\eta_0 := \{\eta_0^a, \eta_0^b\}$ at some rate, say $N^{-1/4}$

$\mathbb{E}[f(V;\eta)]$ converges to $C(P)$ at $(N^{-1/4})^2 = N^{-1/2}$ rate.

Orthogonal Estimand - Two nuisances (Proof)

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right) + O_P\left(\| \eta^b - \eta_0^b \|^2\right) + O_P\left(\| \eta^a - \eta_0^a \| \| \eta^b - \eta_0^b \|\right)$$

Orthogonal Estimand - Two nuisances (Proof)

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right) + O_P\left(\| \eta^b - \eta_0^b \|^2\right) + O_P\left(\| \eta^a - \eta_0^a \| \| \eta^b - \eta_0^b \|\right)$$

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - \mathbb{E}[f(V;\{\eta_0^a . \eta_0^b\})]$$
$$= D_{\eta^a}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^a - \eta_0^a\}$$
$$+D_{\eta^b}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^b - \eta_0^b\}$$
$$+\frac{1}{2}D^2_{\eta^a}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta - \eta_0\}^2$$
$$+\frac{1}{2}D^2_{\eta^b}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta^b - \eta_0^b\}^2$$
$$+D_{\eta_a}D_{\eta_b}\mathbb{E}[f(V;\{\eta^a,\eta^b\}]\{\eta^a - \eta_0^a, \eta^b - \eta_0^b\}$$

Orthogonal Estimand - Two nuisances (Proof)

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right) + O_P\left(\| \eta^b - \eta_0^b \|^2\right) + O_P\left(\| \eta^a - \eta_0^a \| \| \eta^b - \eta_0^b \|\right)$$

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - \mathbb{E}[f(V;\{\eta_0^a . \eta_0^b\})]$$
$$= D_{\eta^a}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^a - \eta_0^a\}$$
$$+D_{\eta^b}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^b - \eta_0^b\}$$
$$+\frac{1}{2}D^2_{\eta^a}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta - \eta_0\}^2$$
$$+\frac{1}{2}D^2_{\eta^b}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta^b - \eta_0^b\}^2$$
$$+D_{\eta_a}D_{\eta_b}\mathbb{E}[f(V;\{\eta^a,\eta^b\}]\{\eta^a - \eta_0^a, \eta^b - \eta_0^b\}$$

Orthogonal Estimand - Two nuisances (Proof)

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\|\eta^a-\eta_0^a\|^2\right) + O_P\left(\|\eta^b-\eta_0^b\|^2\right) + O_P\left(\|\eta^a-\eta_0^a\|\|\eta^b-\eta_0^b\|\right)$$

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - \mathbb{E}[f(V;\{\eta_0^a.\eta_0^b\})]$$
$$= D_{\eta^a}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^a-\eta_0^a\}$$
$$+D_{\eta^b}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^b-\eta_0^b\}$$
$$+\frac{1}{2}D^2_{\eta^a}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta-\eta_0\}^2$$
$$+\frac{1}{2}D^2_{\eta^b}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta^b-\eta_0^b\}^2$$
$$+D_{\eta_a}D_{\eta_b}\mathbb{E}[f(V;\{\eta^a,\eta^b\}]\{\eta^a-\eta_0^a,\eta^b-\eta_0^b\}$$

0

First-order derivative

Orthogonal Estimand - Two nuisances (Proof)

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right) + O_P\left(\| \eta^b - \eta_0^b \|^2\right) + O_P\left(\| \eta^a - \eta_0^a \|\| \eta^b - \eta_0^b \|\right)$$

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - \mathbb{E}[f(V;\{\eta_0^a . \eta_0^b\})]$$
$$= D_{\eta^a}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^a - \eta_0^a\}$$
$$+ D_{\eta^b}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^b - \eta_0^b\}$$

First-order derivative

0

$$+ \frac{1}{2}D^2_{\eta^a}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta - \eta_0\}^2$$

$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right)$$

$$+ \frac{1}{2}D^2_{\eta^b}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta^b - \eta_0^b\}^2$$
$$+ D_{\eta_a}D_{\eta_b}\mathbb{E}[f(V;\{\eta^a,\eta^b\}]\{\eta^a - \eta_0^a, \eta^b - \eta_0^b\}$$

Orthogonal Estimand - Two nuisances (Proof)

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right) + O_P\left(\| \eta^b - \eta_0^b \|^2\right) + O_P\left(\| \eta^a - \eta_0^a \| \| \eta^b - \eta_0^b \|\right)$$

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - \mathbb{E}[f(V;\{\eta_0^a . \eta_0^b\})]$$
$$= D_{\eta^a}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^a - \eta_0^a\}$$
$$+ D_{\eta^b}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^b - \eta_0^b\}$$
$$+ \frac{1}{2}D^2_{\eta^a}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta - \eta_0\}^2$$
$$+ \frac{1}{2}D^2_{\eta^b}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta^b - \eta_0^b\}^2$$
$$+ D_{\eta_a}D_{\eta_b}\mathbb{E}[f(V;\{\eta^a,\eta^b\}]\{\eta^a - \eta_0^a, \eta^b - \eta_0^b\}$$

0

First-order derivative

$$= O_P\left(\| \eta^a - \eta_0^a \|^2\right)$$

$$= O_P\left(\| \eta^b - \eta_0^b \|^2\right)$$

Orthogonal Estimand - Two nuisances (Proof)

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - C(P)$$
$$= O_P\left(\|\eta^a-\eta_0^a\|^2\right) + O_P\left(\|\eta^b-\eta_0^b\|^2\right) + O_P\left(\|\eta^a-\eta_0^a\|\|\eta^b-\eta_0^b\|\right)$$

$$\mathbb{E}[f(V;\{\eta^a,\eta^b\})] - \mathbb{E}[f(V;\{\eta_0^a.\eta_0^b\})]$$
$$= D_{\eta^a}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^a-\eta_0^a\}$$
$$+D_{\eta^b}\mathbb{E}[f(V;\{\eta_0^a,\eta_0^b\})]\{\eta^b-\eta_0^b\}$$

First-order derivative

0

$$+\frac{1}{2}D^2_{\eta^a}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta-\eta_0\}^2 \quad = O_P\left(\|\eta^a-\eta_0^a\|^2\right)$$
$$+\frac{1}{2}D^2_{\eta^b}\mathbb{E}[f(V;\{\eta^a,\eta^b\})]\{\eta^b-\eta_0^b\}^2 \quad = O_P\left(\|\eta^b-\eta_0^b\|^2\right)$$
$$+D_{\eta_a}D_{\eta_b}\mathbb{E}[f(V;\{\eta^a,\eta^b\}]\{\eta^a-\eta_0^a,\eta^b-\eta_0^b\} \quad = O_P\left(\|\eta^a-\eta_0^a\|\|\eta^b-\eta_0^b\|\right)$$

Is the REG orthogonal?

Is the REG orthogonal?

$f^{REG}(V; \mu)$ estimand is non-orthogonal.

Is the REG orthogonal?

$f^{REG}(V;\mu)$ estimand is non-orthogonal.

$$
\begin{aligned}
D_\mu \mathbb{E}[f^{REG}(V;\mu_0)]\{\mu-\mu_0\} &:= \frac{\partial}{\partial t}\mathbb{E}\left[f^{REG}(V;\mu+t(\mu-\mu_0)\right]|_{t=0} \\
&= \mathbb{E}\left[\frac{\partial}{\partial t} f^{REG}(V;\mu+t(\mu-\mu_0)\,|_{t=0}\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial t}\{\mu+t(\mu-\mu_0)\}\,|_{t=0}\right] \\
&= \mathbb{E}[\mu(x,Z)-\mu_0(x,Z)]
\end{aligned}
$$

Is the REG orthogonal?

$f^{REG}(V;\mu)$ estimand is non-orthogonal.

$$
\begin{aligned}
D_\mu \mathbb{E}[f^{REG}(V;\mu_0)]\{\mu-\mu_0\} &:= \frac{\partial}{\partial t}\mathbb{E}\left[f^{REG}(V;\mu+t(\mu-\mu_0)\right] |_{t=0} \\
&= \mathbb{E}\left[\frac{\partial}{\partial t} f^{REG}(V;\mu+t(\mu-\mu_0)\,|_{t=0}\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial t}\{\mu+t(\mu-\mu_0)\}\,|_{t=0}\right] \\
&= \mathbb{E}[\mu(x,Z)-\mu_0(x,Z)] \\
&\neq 0
\end{aligned}
$$

Is the IPW orthogonal?

Is the IPW orthogonal?

$f^{IPW}(V;\pi)$ estimand is non-orthogonal.

Is the IPW orthogonal?

$f^{IPW}(V;\pi)$ estimand is non-orthogonal.

$$D_\pi \mathbb{E}[f^{IPW}(V;\pi_0)]\{\pi-\pi_0\} := \frac{\partial}{\partial t}\mathbb{E}\left[f^{IPW}(V;\pi+t(\pi-\pi_0)\right]|_{t=0}$$

$$= \mathbb{E}\left[\frac{\partial}{\partial t}f^{IPW}(V;\pi+t(\pi-\pi_0)\,|_{t=0}\right]$$

$$= \mathbb{E}\left[(\pi-\pi_0)\frac{\partial}{\partial \pi}f^{IPW}(V;\pi)\,|_{\pi=\pi_0}\right]$$

$$= -\,\mathbb{E}\left[\{\pi-\pi_0\}\left\{\frac{I_x(X)}{\pi_0^2(X\,|\,Z)}Y\right\}\right]$$

Is the IPW orthogonal?

$f^{IPW}(V;\pi)$ estimand is non-orthogonal.

$$D_\pi \mathbb{E}[f^{IPW}(V;\pi_0)]\{\pi - \pi_0\} := \frac{\partial}{\partial t}\mathbb{E}\left[f^{IPW}(V;\pi + t(\pi - \pi_0)\right]|_{t=0}$$

$$= \mathbb{E}\left[\frac{\partial}{\partial t} f^{IPW}(V;\pi + t(\pi - \pi_0)\,|_{t=0}\right]$$

$$= \mathbb{E}\left[(\pi - \pi_0)\frac{\partial}{\partial \pi} f^{IPW}(V;\pi)\,|_{\pi=\pi_0}\right]$$

$$= -\,\mathbb{E}\left[\{\pi - \pi_0\}\left\{\frac{I_x(X)}{\pi_0^2(X\,|\,Z)}Y\right\}\right]$$

$$\neq 0$$

Is the DR estimand orthogonal? - 1

Is the DR estimand orthogonal? - 1

$f^{DR}(V; \{\pi, \mu_0\})$ is an orthogonal estimand

Is the DR estimand orthogonal? - 1

$f^{DR}(V; \{\pi, \mu_0\})$ is an orthogonal estimand

$$D_\mu \mathbb{E}[f^{DR}(V; \{\pi, \mu_0\})]\{\mu - \mu_0\} := \frac{\partial}{\partial t} \mathbb{E}\left[f^{DR}(V; \{\pi_0, \mu + t(\mu - \mu_0)\}\right] |_{t=0}$$

$$= \mathbb{E}\left[\frac{\partial}{\partial t} f^{DR}(V; \{\pi_0, \mu + t(\mu - \mu_0)\}) |_{t=0}\right]$$

$$= \mathbb{E}\left[(\mu - \mu_0) \frac{\partial}{\partial \mu} f^{DR}(V; \{\pi, \mu\}) |_{\mu=\mu_0}\right]$$

$$= \mathbb{E}\left[\{\mu - \mu_0\}\left\{-\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X, Z) + \mu_0(x, Z)\right\}\right]$$

$$= \mathbb{E}\left[\{\mu - \mu_0\}\left\{-\mu_0(x, Z) + \mu_0(x, Z)\right\}\right]$$

Is the DR estimand orthogonal? - 1

$f^{DR}(V;\{\pi,\mu_0\})$ is an orthogonal estimand

$$
\begin{aligned}
D_\mu \mathbb{E}[f^{DR}(V;\{\pi,\mu_0\})]\{\mu-\mu_0\} := & \frac{\partial}{\partial t}\mathbb{E}\left[f^{DR}(V;\{\pi_0,\mu+t(\mu-\mu_0)\}\right]|_{t=0} \\
&= \mathbb{E}\left[\frac{\partial}{\partial t} f^{DR}(V;\{\pi_0,\mu+t(\mu-\mu_0)\})|_{t=0}\right] \\
&= \mathbb{E}\left[(\mu-\mu_0)\frac{\partial}{\partial \mu} f^{DR}(V;\{\pi,\mu\})|_{\mu=\mu_0}\right] \\
&= \mathbb{E}\left[\{\mu-\mu_0\}\left\{-\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X,Z)+\mu_0(x,Z)\right\}\right] \\
&= \mathbb{E}\left[\{\mu-\mu_0\}\{-\mu_0(x,Z)+\mu_0(x,Z)\}\right]
\end{aligned}
$$

0

Is the DR estimand orthogonal? - 1

$f^{DR}(V; \{\pi, \mu_0\})$ is an orthogonal estimand

$$
\begin{aligned}
D_\mu \mathbb{E}[f^{DR}(V; \{\pi, \mu_0\})]\{\mu - \mu_0\} &:= \frac{\partial}{\partial t} \mathbb{E}\left[f^{DR}(V; \{\pi_0, \mu + t(\mu - \mu_0)\}\right] |_{t=0} \\
&= \mathbb{E}\left[\frac{\partial}{\partial t} f^{DR}(V; \{\pi_0, \mu + t(\mu - \mu_0)\}) |_{t=0}\right] \\
&= \mathbb{E}\left[(\mu - \mu_0)\frac{\partial}{\partial \mu} f^{DR}(V; \{\pi, \mu\}) |_{\mu=\mu_0}\right] \\
&= \mathbb{E}\left[\{\mu - \mu_0\}\left\{-\frac{I_x(X)}{\pi_0(X|Z)}\mu_0(X, Z) + \mu_0(x, Z)\right\}\right] \\
&= \mathbb{E}\left[\{\mu - \mu_0\}\{-\mu_0(x, Z) + \mu_0(x, Z)\}\right] \\
&= 0
\end{aligned}
$$

0

Is the DR estimand orthogonal? - 2

Is the DR estimand orthogonal? - 2

$f^{DR}(V; \{\pi, \mu_0\})$ is an orthogonal estimand

Is the DR estimand orthogonal? - 2

$f^{DR}(V; \{\pi, \mu_0\})$ is an orthogonal estimand

$$
\begin{aligned}
D_\pi \mathbb{E}[f^{DR}(V; \{\pi, \mu_0\})]\{\pi - \pi_0\} := & \frac{\partial}{\partial t}\mathbb{E}\left[f^{DR}(V; \{\pi + t(\pi - \pi_0), \mu_0\}\right] |_{t=0} \\
&= \mathbb{E}\left[\frac{\partial}{\partial t} f^{DR}(V; \{\pi + t(\pi - \pi_0), \mu_0\}) |_{t=0}\right] \\
&= \mathbb{E}\left[(\pi - \pi_0)\frac{\partial}{\partial \pi} f^{DR}(V; \{\pi, \mu_0\}) |_{\pi=\pi_0}\right] \\
&= \mathbb{E}\left[\{\pi - \pi_0\}\left\{-\frac{I_x(X)}{\pi_0^2(X|Z)}\{Y - \mu_0(X, Z)\}\right\}\right] \\
&= \mathbb{E}\left[\{\pi - \pi_0\}\left\{-\frac{I_x(X)}{\pi_0^2(X|Z)}\{\mu_0(X, Z) - \mu_0(X, Z)\}\right\}\right]
\end{aligned}
$$

Is the DR estimand orthogonal? - 2

$f^{DR}(V; \{\pi, \mu_0\})$ is an orthogonal estimand

$$D_\pi \mathbb{E}[f^{DR}(V; \{\pi, \mu_0\})]\{\pi - \pi_0\} := \frac{\partial}{\partial t}\mathbb{E}\left[f^{DR}(V; \{\pi + t(\pi - \pi_0), \mu_0\}\right] |_{t=0}$$

$$= \mathbb{E}\left[\frac{\partial}{\partial t} f^{DR}(V; \{\pi + t(\pi - \pi_0), \mu_0\}) |_{t=0}\right]$$

$$= \mathbb{E}\left[(\pi - \pi_0)\frac{\partial}{\partial \pi} f^{DR}(V; \{\pi, \mu_0\}) |_{\pi=\pi_0}\right]$$

$$= \mathbb{E}\left[\{\pi - \pi_0\}\left\{-\frac{I_x(X)}{\pi_0^2(X|Z)}\{Y - \mu_0(X, Z)\}\right\}\right]$$

$$= \mathbb{E}\left[\{\pi - \pi_0\}\left\{-\frac{I_x(X)}{\pi_0^2(X|Z)}\{\mu_0(X, Z) - \mu_0(X, Z)\}\right\}\right]$$

0

Is the DR estimand orthogonal? - 2

$f^{DR}(V;\{\pi,\mu_0\})$ is an orthogonal estimand

$$
\begin{aligned}
D_\pi \mathbb{E}[f^{DR}(V;\{\pi,\mu_0\})]\{\pi-\pi_0\} &:= \frac{\partial}{\partial t}\mathbb{E}\left[f^{DR}(V;\{\pi+t(\pi-\pi_0),\mu_0\}\right]|_{t=0} \\
&= \mathbb{E}\left[\frac{\partial}{\partial t} f^{DR}(V;\{\pi+t(\pi-\pi_0),\mu_0\})\,|_{t=0}\right] \\
&= \mathbb{E}\left[(\pi-\pi_0)\frac{\partial}{\partial \pi} f^{DR}(V;\{\pi,\mu_0\})\,|_{\pi=\pi_0}\right] \\
&= \mathbb{E}\left[\{\pi-\pi_0\}\left\{-\frac{I_x(X)}{\pi_0^2(X|Z)}\{Y-\mu_0(X,Z)\}\right\}\right] \\
&= \mathbb{E}\left[\{\pi-\pi_0\}\left\{-\frac{I_x(X)}{\pi_0^2(X|Z)}\{\mu_0(X,Z)-\mu_0(X,Z)\}\right\}\right] \quad 0 \\
&= 0
\end{aligned}
$$

Debiasedness and Doubly Robustness

Debiasedness:

If $f(V;\eta)$ is *orthogonal*, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Converging at $N^{-1/2}$ if π, μ converges at $N^{-1/4}$ ("*debiasedness*")

Debiasedness and Doubly Robustness

Debiasedness:

If $f(V;\eta)$ is *orthogonal*, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Converging at $N^{-1/2}$ if π, μ converges at $N^{-1/4}$ ("*debiasedness*")

For the BD adjustment, $f^{DR}(V;\{\pi,\mu\})$ is doubly robust & orthogonal estimand

Debiasedness and Doubly Robustness

Debiasedness:

If $f(V;\eta)$ is *orthogonal*, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\parallel \eta - \eta_0 \parallel_2^2\right)$

Converging at $N^{-1/2}$ if π, μ converges at $N^{-1/4}$ ("*debiasedness*")

For the BD adjustment, $f^{DR}(V;\{\pi,\mu\})$ is doubly robust & orthogonal estimand

Debiasedness and Doubly-Robust Estimand

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\parallel \pi - \pi_0 \parallel \parallel \mu - \mu_0 \parallel\right)$$

Debiasedness and Doubly Robustness

Debiasedness:

If $f(V;\eta)$ is *orthogonal*, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Converging at $N^{-1/2}$ if π, μ converges at $N^{-1/4}$ ("*debiasedness*")

For the BD adjustment, $f^{DR}(V; \{\pi, \mu\})$ is doubly robust & orthogonal estimand

Debiasedness and Doubly-Robust Estimand

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

Converging at $N^{-1/2}$ if π, μ converges at $N^{-1/4}$ ("*debiasedness*")

Debiasedness and Doubly Robustness

Debiasedness:

If $f(V;\eta)$ is *orthogonal*, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Converging at $N^{-1/2}$ if π, μ converges at $N^{-1/4}$ ("*debiasedness*")

For the BD adjustment, $f^{DR}(V;\{\pi,\mu\})$ is doubly robust & orthogonal estimand

Debiasedness and Doubly-Robust Estimand

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

Converging at $N^{-1/2}$ if π, μ converges at $N^{-1/4}$ ("*debiasedness*")

No errors if either $\pi = \pi_0$ or $\mu = \mu_0$ ("doubly-robustness")

Debiasedness and Doubly Robustness - Proof 1

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

Debiasedness and Doubly Robustness - Proof 1

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] - C(P) = \mathbb{E}[f^{DR}(V; \{\pi, \mu\}) - f^{DR}(V; \{\pi_0, \mu_0\})]$$

Debiasedness and Doubly Robustness - Proof 1

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\|\pi-\pi_0\|\|\mu-\mu_0\|\right)$$

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = \mathbb{E}[f^{DR}(V;\{\pi,\mu\}) - f^{DR}(V;\{\pi_0,\mu_0\})]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{Y-\mu(X,Z)\} + \mu(x,Z) - \frac{I_x(X)}{\pi_0(X|Z)}\{Y-\mu_0(X,Z)\} - \mu_0(x,Z)\right]$$

Debiasedness and Doubly Robustness - Proof 1

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\|\, \pi - \pi_0 \,\|\|\, \mu - \mu_0 \,\|\right)$$

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = \mathbb{E}[f^{DR}(V;\{\pi,\mu\}) - f^{DR}(V;\{\pi_0,\mu_0\})]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{Y - \mu(X,Z)\} + \mu(x,Z) - \frac{I_x(X)}{\pi_0(X|Z)}\{Y - \mu_0(X,Z)\} - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z) - \mu(X,Z)\} + \mu(x,Z) - \frac{I_x(X)}{\pi_0(X|Z)}\{\mu_0(X,Z) - \mu_0(X,Z)\} - \mu_0(x,Z)\right]$$

Debiasedness and Doubly Robustness - Proof 1

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\parallel \pi - \pi_0 \parallel \parallel \mu - \mu_0 \parallel\right)$$

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = \mathbb{E}[f^{DR}(V;\{\pi,\mu\}) - f^{DR}(V;\{\pi_0,\mu_0\})]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{Y - \mu(X,Z)\} + \mu(x,Z) - \frac{I_x(X)}{\pi_0(X|Z)}\{Y - \mu_0(X,Z)\} - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z) - \mu(X,Z)\} + \mu(x,Z) - \frac{I_x(X)}{\pi_0(X|Z)}\{\mu_0(X,Z) - \mu_0(X,Z)\} - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z) - \mu(X,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

Debiasedness and Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

Debiasedness and Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z) - \mu(X,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

Debiasedness and Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

$$\mathbb{E}[f^{DR}(V; \{\pi, \mu\})] - C(P) = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z) - \mu(X,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\frac{\pi_0(x|Z)}{\pi(x|Z)}\{\mu_0(x,Z) - \mu(x,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

Debiasedness and Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\| \pi - \pi_0 \| \| \mu - \mu_0 \|\right)$$

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z) - \mu(X,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\frac{\pi_0(x|Z)}{\pi(x|Z)}\{\mu_0(x,Z) - \mu(x,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\left\{\frac{\pi_0(x|Z)}{\pi(x|Z)} - 1\right\}\{\mu_0(x,Z) - \mu(x,Z)\}\right]$$

Debiasedness and Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\|\ \pi - \pi_0\ \|\|\ \mu - \mu_0\ \|\right)$$

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z) - \mu(X,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\frac{\pi_0(x|Z)}{\pi(x|Z)}\{\mu_0(x,Z) - \mu(x,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\left\{\frac{\pi_0(x|Z)}{\pi(x|Z)} - 1\right\}\{\mu_0(x,Z) - \mu(x,Z)\}\right]$$

$$= \mathbb{E}\left[\frac{1}{\pi(x|Z)}\left\{\pi_0(x|Z) - \pi(x|Z)\right\}\left\{\mu_0(x,Z) - \mu(x,Z)\right\}\right]$$

Debiasedness and Doubly Robustness - Proof 2

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = O_P\left(\|\pi-\pi_0\|\|\mu-\mu_0\|\right)$$

$$\mathbb{E}[f^{DR}(V;\{\pi,\mu\})] - C(P) = \mathbb{E}\left[\frac{I_x(X)}{\pi(X|Z)}\{\mu_0(X,Z)-\mu(X,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\frac{\pi_0(x|Z)}{\pi(x|Z)}\{\mu_0(x,Z)-\mu(x,Z)\} + \mu(x,Z) - \mu_0(x,Z)\right]$$

$$= \mathbb{E}\left[\left\{\frac{\pi_0(x|Z)}{\pi(x|Z)} - 1\right\}\{\mu_0(x,Z)-\mu(x,Z)\}\right]$$

$$= \mathbb{E}\left[\frac{1}{\pi(x|Z)}\left\{\pi_0(x|Z)-\pi(x|Z)\right\}\left\{\mu_0(x,Z)-\mu(x,Z)\right\}\right]$$

$$= O_P\left(\|\pi_0-\pi\|\|\mu-\mu_0\|\right)$$

Intermediate Summary - Orthogonal Estimand

Intermediate Summary - Orthogonal Estimand

So far…

Intermediate Summary - Orthogonal Estimand

So far…

1. We defined an *orthogonal estimand (OE).*

Intermediate Summary - Orthogonal Estimand

So far…

1. We defined an *orthogonal estimand (OE).*

2. (Debiasedness) We showed that an expectation of the OE converges at $N^{-1/2}$ rate if the nuisance converges at $N^{-1/4}$ rate.

Intermediate Summary - Orthogonal Estimand

So far…

1. We defined an *orthogonal estimand (OE).*

2. (Debiasedness) We showed that an expectation of the OE converges at $N^{-1/2}$ rate if the nuisance converges at $N^{-1/4}$ rate.

3. For the BD adjustment, $f^{DR}(V; \{\mu, \pi\})$ estimand has doubly robustness & debiasedness.

Intermediate Summary - Orthogonal Estimand

So far…

1. We defined an *orthogonal estimand (OE).*

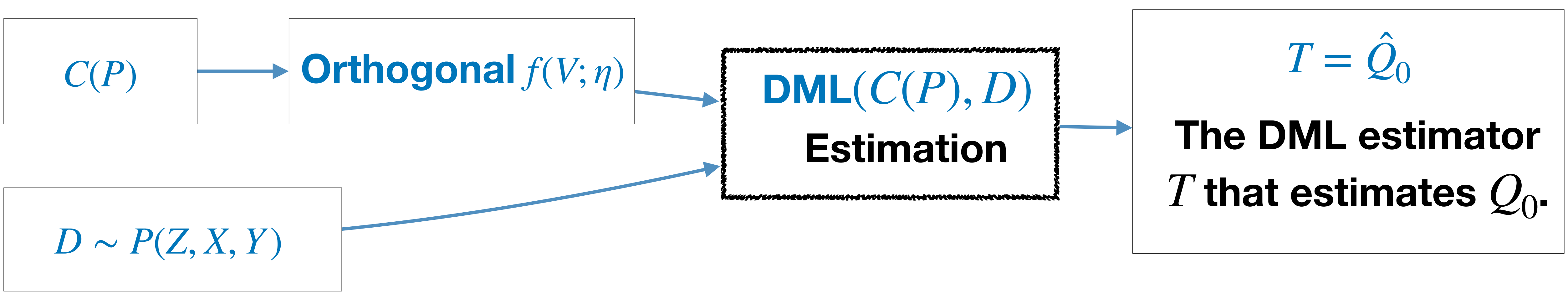
2. (Debiasedness) We showed that an expectation of the OE converges at $N^{-1/2}$ rate if the nuisance converges at $N^{-1/4}$ rate.

3. For the BD adjustment, $f^{DR}(V; \{\mu, \pi\})$ estimand has doubly robustness & debiasedness.

Next, we will study finite sample properties.

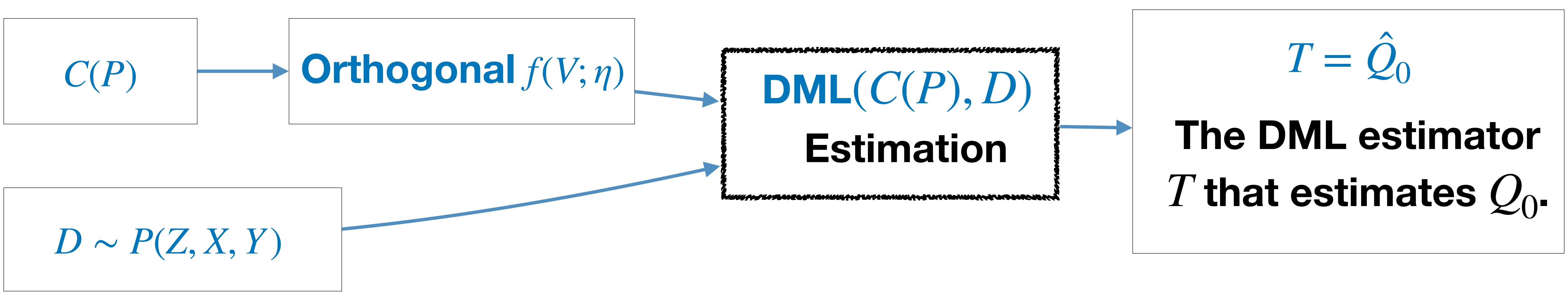
Estimating with finite samples

Estimating with finite samples



Estimating with finite samples

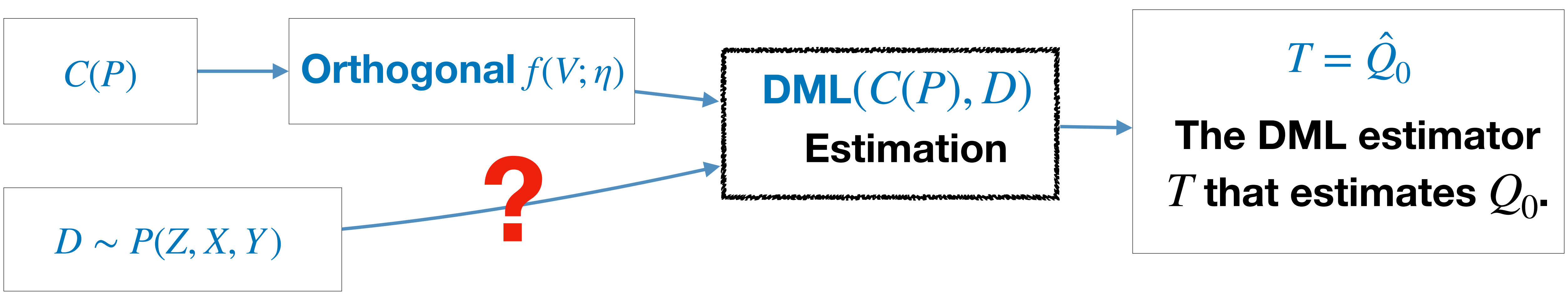
So far, we study the power of the orthogonal estimand.



Estimating with finite samples

So far, we study the power of the orthogonal estimand.

Now, we connect the estimand to the estimation task using finite samples.



Estimating with Finite Samples

Debiasedness:

If $f(V; \eta)$ is orthogonal, $\mathbb{E}[f(V; \eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

Estimating with Finite Samples

Debiasedness:

If $f(V;\eta)$ is orthogonal, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

An estimator of $C(P)$ using an estimand $f(V;\eta)$ is

$$\mathbb{E}_D\left[f(V;\hat{\eta})\right] := \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}),$$

where $\hat{\eta}$ denotes an estimated nuisance.

Estimating with Finite Samples

Debiasedness:

If $f(V;\eta)$ is orthogonal, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

An estimator of $C(P)$ using an estimand $f(V;\eta)$ is

$$\mathbb{E}_D\left[f(V;\hat{\eta})\right] := \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}),$$

where $\hat{\eta}$ denotes an estimated nuisance.

$\mathbb{E}_D[f(V;\hat{\eta})] - C(P)$

Estimating with Finite Samples

Debiasedness:

If $f(V;\eta)$ is orthogonal, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

An estimator of $C(P)$ using an estimand $f(V;\eta)$ is

$$\mathbb{E}_D\left[f(V;\hat{\eta})\right] := \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}),$$

where $\hat{\eta}$ denotes an estimated nuisance.

$$\begin{aligned}&\mathbb{E}_D[f(V;\hat{\eta})] - C(P)\\ &= \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)\end{aligned}$$

Estimating with Finite Samples

Debiasedness:

If $f(V;\eta)$ is orthogonal, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

An estimator of $C(P)$ using an estimand $f(V;\eta)$ is

$$\mathbb{E}_D\left[f(V;\hat{\eta})\right] := \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}),$$

where $\hat{\eta}$ denotes an estimated nuisance.

$$\mathbb{E}_D[f(V;\hat{\eta})] - C(P)$$
$$= \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)$$

Estimating with Finite Samples

Debiasedness:

If $f(V;\eta)$ is orthogonal, $\mathbb{E}[f(V;\eta)] - C(P) = O_P\left(\| \eta - \eta_0 \|_2^2\right)$

An estimator of $C(P)$ using an estimand $f(V;\eta)$ is

$$\mathbb{E}_D\left[f(V;\hat{\eta})\right] := \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}),$$

where $\hat{\eta}$ denotes an estimated nuisance.

$$\mathbb{E}_D[f(V;\hat{\eta})] - C(P)$$
$$= \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)$$

We will focus on analyzing the remaining term: $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$.

Law of Large Numbers (LLN)

Law of Large Numbers (LLN)

Remaining term:

$$\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] = \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}) - \mathbb{E}_P[f(V;\hat{\eta})].$$

Law of Large Numbers (LLN)

Remaining term:

$$\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] = \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}) - \mathbb{E}_P[f(V;\hat{\eta})].$$

Law of Large Numbers (LLN)

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)]$ converges to $\mathbb{E}_P[f(V;\eta_*)]$.

Law of Large Numbers (LLN)

Remaining term:

$$\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] = \frac{1}{N}\sum_{i=1}^{N} f(V_i;\hat{\eta}) - \mathbb{E}_P[f(V;\hat{\eta})].$$

Law of Large Numbers (LLN)

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)]$ converges to $\mathbb{E}_P[f(V;\eta_*)]$.

Concentration inequalities (e.g., Hoeffding's inequality)

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)]$ converges to $\mathbb{E}_P[f(V;\eta_*)]$ at $N^{-1/2}$ rate.

Challenges in LLN

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

Challenges in LLN

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

Consider $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$, where $\hat{\eta} := \hat{\eta}(D)$ is an estimate using samples D.

Challenges in LLN

For any fixed η_*, $\mathbb{E}_D[f(V; \eta_*)] - \mathbb{E}_P[f(V; \eta_*)] \to 0$ at $N^{-1/2}$ rate.

Consider $\mathbb{E}_D[f(V; \hat{\eta})] - \mathbb{E}_P[f(V; \hat{\eta})]$, where $\hat{\eta} := \hat{\eta}(D)$ is an estimate using samples D.

… Equivalently, consider $\frac{1}{N}\sum_{i=1}^{N} f(V_i, \hat{\eta}(N)) - \mathbb{E}[f(V; \hat{\eta}(N)]$

Challenges in LLN

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

Consider $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$, where $\hat{\eta} := \hat{\eta}(D)$ is an estimate using samples D.

… Equivalently, consider $\frac{1}{N}\sum_{i=1}^{N} f(V_i, \hat{\eta}(N)) - \mathbb{E}[f(V;\hat{\eta}(N)]$

The LLN is not applicable since $\hat{\eta}$ is varying w.r.t. D (and N).

Challenges in LLN

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

Consider $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$, where $\hat{\eta} := \hat{\eta}(D)$ is an estimate using samples D.

… Equivalently, consider $\frac{1}{N}\sum_{i=1}^{N} f(V_i, \hat{\eta}(N)) - \mathbb{E}[f(V;\hat{\eta}(N)]$

The LLN is not applicable since $\hat{\eta}$ is varying w.r.t. D (and N).

… Without any special treatises, $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$ is *not necessarily converging to 0.*

Uniform Convergence

Uniform Convergence

To guarantee $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0$, we should have a following worst-case convergence guarantee:

Uniform convergence:

$$\sup_{\eta \in H} \left(\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \right) \to 0$$

Uniform Convergence

To guarantee $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0$, we should have a following worst-case convergence guarantee:

Uniform convergence:

$$\sup_{\eta \in H} \left(\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \right) \to 0$$

Then, $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0$ obviously holds.

Donsker Class

Donsker Class

For some function class $H(\ni \eta)$, the uniform convergence holds.

Donsker Class

For some function class $H(\ni \eta)$, the uniform convergence holds.

Donsker class:

A class H s.t. $\sup_{\eta \in H} \left(\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \right) \to 0$ at $N^{-1/2}$ rate

For the concrete definition, see [van der Vaart, 2000]

Donsker Class

For some function class $H(\ni \eta)$, the uniform convergence holds.

Donsker class:

A class H s.t. $\sup_{\eta \in H} \left(\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \right) \to 0$ at $N^{-1/2}$ rate

For the concrete definition, see [van der Vaart, 2000]

Example: A function class with bounded VC-dimension (called VC-class).

Donsker Class

For some function class $H(\ni \eta)$, the uniform convergence holds.

Donsker class:

A class H s.t. $\sup_{\eta \in H} \left(\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \right) \to 0$ at $N^{-1/2}$ rate

For the concrete definition, see [van der Vaart, 2000]

Example: A function class with bounded VC-dimension (called VC-class).

: differentiable functions

Error analysis under Donsker

Error analysis under Donsker

If a nuisance function class H is "Donsker", $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0$ at $N^{-1/2}$ rate.

Error analysis under Donsker

If a nuisance function class H is "Donsker", $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \rightarrow 0$ at $N^{-1/2}$ rate.

$$\mathbb{E}_D[f(V;\hat{\eta})] - C(P) = \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)$$

Error analysis under Donsker

If a nuisance function class H is "Donsker", $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0$ at $N^{-1/2}$ rate.

$$\mathbb{E}_D[f(V;\hat{\eta})] - C(P) = \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)$$

$\to 0$ at $N^{-1/2}$ rate

Error analysis under Donsker

If a nuisance function class H is “Donsker”, $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0$ at $N^{-1/2}$ rate.

$$\mathbb{E}_D[f(V;\hat{\eta})] - C(P) = \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)$$

$\to$ 0 at $N^{-1/2}$ rate

$= O\left(\| \hat{\eta} - \eta_0 \|_2^2\right)$

Error analysis under Donsker

If a nuisance function class H is "Donsker", $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0$ at $N^{-1/2}$ rate.

$$\mathbb{E}_D[f(V;\hat{\eta})] - C(P) = \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)$$

$\to$ 0 at $N^{-1/2}$ rate

$= O\left(\| \hat{\eta} - \eta_0 \|_2^2\right)$

Then, the estimator $\mathbb{E}_D[f(V;\hat{\eta})]$ *converges to* $C(P)$ *fast even if* $\hat{\eta}$ *converges slow*

Limitation of Donsker

Limitation of Donsker

However, the Donsker class assumption is restrictive!

Limitation of Donsker

However, the Donsker class assumption is restrictive!

No guarantee that modern ML methods (e.g., deep and complicated neural networks) fall into the Donsker class.

Limitation of Donsker

However, the Donsker class assumption is restrictive!

No guarantee that modern ML methods (e.g., deep and complicated neural networks) fall into the Donsker class.

Releasing Donsker Assumption

Releasing Donsker Assumption

Recall the Law of Large Numbers and the concentration inequality:

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

Releasing Donsker Assumption

Recall the Law of Large Numbers and the concentration inequality:

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

This can be rewritten as a following principle ([Robins et al., 1997, Kennedy et al., 2019], etc.)

For any η independent to samples D,
$\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \to 0$ at $N^{-1/2}$ rate.

Releasing Donsker Assumption

Recall the Law of Large Numbers and the concentration inequality:

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

This can be rewritten as a following principle ([Robins et al., 1997, Kennedy et al., 2019], etc.)

For any η independent to samples D,
$\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \to 0$ at $N^{-1/2}$ rate.

Suppose $\hat{\eta}$ is estimated from a separate dataset D' that is independent to D. Then,

$$\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0 \text{ at } N^{-1/2} \text{ rate.}$$

Releasing Donsker Assumption

Recall the Law of Large Numbers and the concentration inequality:

For any fixed η_*, $\mathbb{E}_D[f(V;\eta_*)] - \mathbb{E}_P[f(V;\eta_*)] \to 0$ at $N^{-1/2}$ rate.

This can be rewritten as a following principle ([Robins et al., 1997, Kennedy et al., 2019], etc.)

For any η independent to samples D,
$\mathbb{E}_D[f(V;\eta)] - \mathbb{E}_P[f(V;\eta)] \to 0$ at $N^{-1/2}$ rate.

Suppose $\hat{\eta}$ is estimated from a separate dataset D' that is independent to D. Then,

$$\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] \to 0 \text{ at } N^{-1/2} \text{ rate.}$$

This doesn't require the Donsker class assumption!

Sample Splitting

Sample Splitting

Sample splitting procedure

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

For $k \in \{0,1\}$,

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

For $k \in \{0,1\}$,

Let $\hat{\eta}_k$ denote the estimated nuisance using D_k.

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

For $k \in \{0,1\}$,

Let $\hat{\eta}_k$ denote the estimated nuisance using D_k.

Let $T_k := \mathbb{E}_{D_{1-k}}\left[f(V;\hat{\eta}_k)\right]$

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

For $k \in \{0,1\}$,

Let $\hat{\eta}_k$ denote the estimated nuisance using D_k.

Let $T_k := \mathbb{E}_{D_{1-k}}\left[f(V; \hat{\eta}_k)\right]$

Let $T := (T_0 + T_1)/2$.

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

For $k \in \{0,1\}$,

Let $\hat{\eta}_k$ denote the estimated nuisance using D_k.

Let $T_k := \mathbb{E}_{D_{1-k}}\left[f(V;\hat{\eta}_k)\right]$

Let $T := (T_0 + T_1)/2$.

Then, $T - \mathbb{E}\left[f(V;\hat{\eta})\right] \rightarrow 0$ at $N^{-1/2}$ rate

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

For $k \in \{0,1\}$,

Let $\hat{\eta}_k$ denote the estimated nuisance using D_k.

Let $T_k := \mathbb{E}_{D_{1-k}}\left[f(V;\hat{\eta}_k)\right]$

Let $T := (T_0 + T_1)/2$.

Then, $T - \mathbb{E}\left[f(V;\hat{\eta})\right] \to 0$ at $N^{-1/2}$ rate

Donsker class assumption is dropped.

Sample Splitting

Sample splitting procedure

Split D into two random halves D_0, D_1.

For $k \in \{0,1\}$,

Let $\hat{\eta}_k$ denote the estimated nuisance using D_k.

Let $T_k := \mathbb{E}_{D_{1-k}}\left[f(V; \hat{\eta}_k)\right]$

Let $T := (T_0 + T_1)/2$.

Then, $T - \mathbb{E}\left[f(V; \hat{\eta})\right] \to 0$ at $N^{-1/2}$ rate

Donsker class assumption is dropped.

Any ML models can be employed!

Intermediate Summary - Sample Splitting

Intermediate Summary - Sample Splitting

For the empirical term $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$ to converges to zero at a fast rate, we need

Intermediate Summary - Sample Splitting

For the empirical term $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$ to converges to zero at a fast rate, we need

1. (Donsker Class) the nuisance function class should be Donsker, or

Intermediate Summary - Sample Splitting

For the empirical term $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$ to converges to zero at a fast rate, we need

1. (Donsker Class) the nuisance function class should be Donsker, or

2. (Sample splitting) Samples used for training $\hat{\eta}$ and evaluating the empirical average $\mathbb{E}_D$ should be independent.

Intermediate Summary - Sample Splitting

For the empirical term $\mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})]$ to converges to zero at a fast rate, we need

1. (Donsker Class) the nuisance function class should be Donsker, or
2. (Sample splitting) Samples used for training $\hat{\eta}$ and evaluating the empirical average $\mathbb{E}_D$ should be independent.

Now we are ready to formally define the DML estimator.

So, what’s the DML?

Definition of DML

Definition of DML

Double/Debiased Machine Learning (DML)

Definition of DML

Double/Debiased Machine Learning (DML)

Given a target quantity $C(P)$ and the data D, a DML estimator T is an estimator derived from

$C(P)$

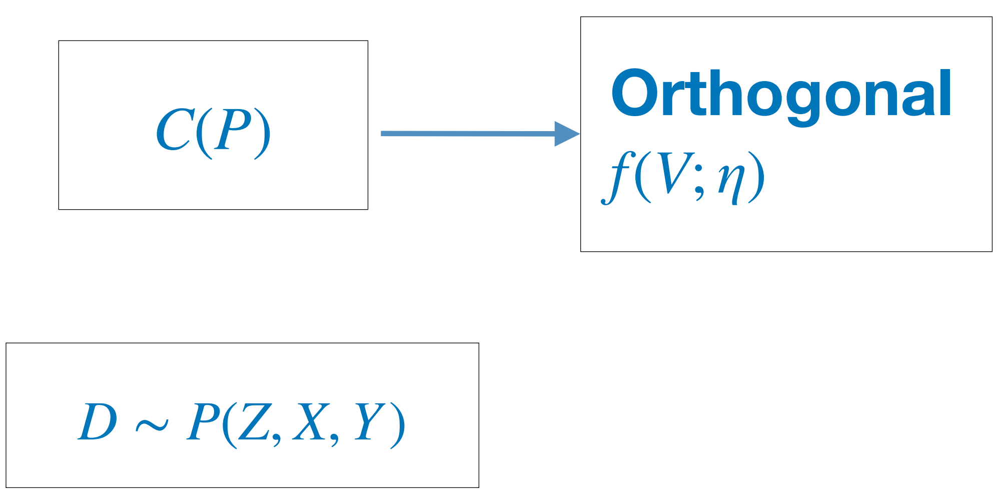
$D \sim P(Z, X, Y)$

Definition of DML

Double/Debiased Machine Learning (DML)

Given a target quantity $C(P)$ and the data D, a DML estimator T is an estimator derived from

1. an orthogonal estimand $f(V;\eta)$, and

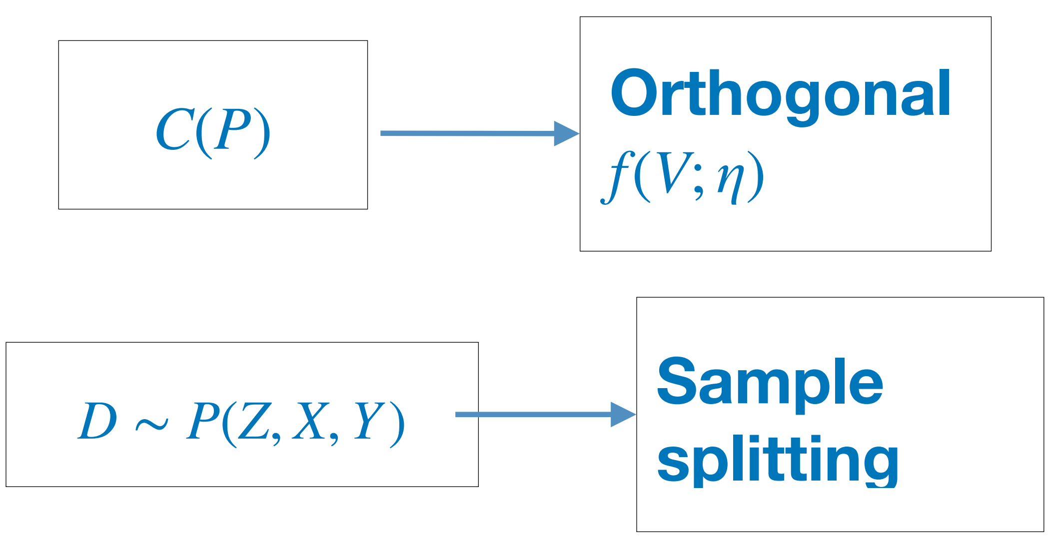


Definition of DML

Double/Debiased Machine Learning (DML)

Given a target quantity $C(P)$ and the data D, a DML estimator T is an estimator derived from

1. an orthogonal estimand $f(V;\eta)$, and
2. the sample-splitting procedure.

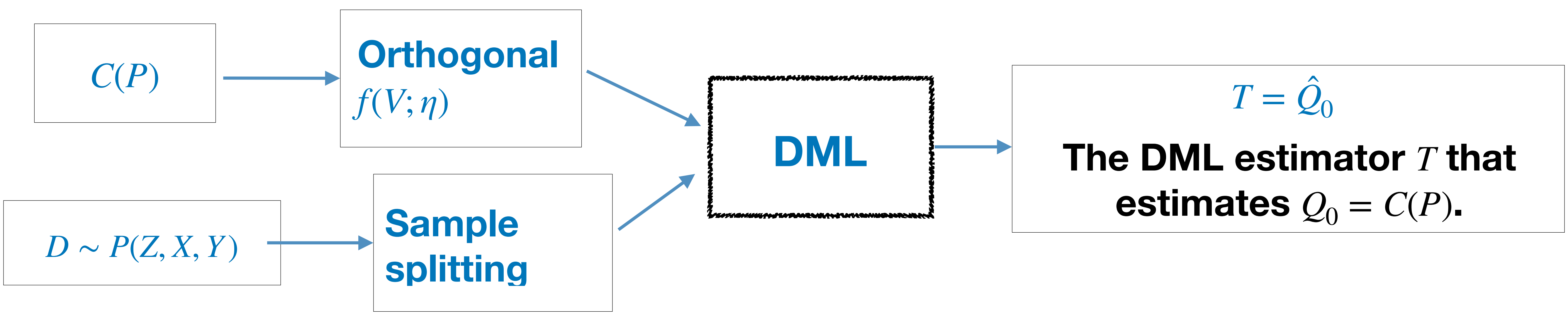


Definition of DML

Double/Debiased Machine Learning (DML)

Given a target quantity $C(P)$ and the data D, a DML estimator T is an estimator derived from

1. an orthogonal estimand $f(V;\eta)$, and
2. the sample-splitting procedure.



Summary

Summary

The DML estimator is an estimator based on

Summary

The DML estimator is an estimator based on

1. (Orthogonal estimand) — An estimand s.t. its 1st derivative is 0.

Summary

The DML estimator is an estimator based on

1. (Orthogonal estimand) — An estimand s.t. its 1st derivative is 0.
2. (Sample splitting) Samples used for training $\hat{\eta}$ and evaluating the empirical average $\mathbb{E}_D$ should be independent.

Summary

The DML estimator is an estimator based on

1. (Orthogonal estimand) — An estimand s.t. its 1st derivative is 0.
2. (Sample splitting) Samples used for training $\hat{\eta}$ and evaluating the empirical average $\mathbb{E}_D$ should be independent.

The DML estimator achieves debiasedness property! It converges at $N^{-1/2}$ rate if $\hat{\eta}$ converges to η_0 at $N^{-1/4}$ rate.

Summary

The DML estimator is an estimator based on

1. (Orthogonal estimand) — An estimand s.t. its 1st derivative is 0.
2. (Sample splitting) Samples used for training $\hat{\eta}$ and evaluating the empirical average $\mathbb{E}_D$ should be independent.

The DML estimator achieves debiasedness property! It converges at $N^{-1/2}$ rate if $\hat{\eta}$ converges to η_0 at $N^{-1/4}$ rate.

Now we are ready to formally define the DML estimator.

Any Questions?

Estimating Causal Effects

Yonghan Jung
Purdue University
yonghanjung.me

2022.07.12

University of Seoul

Introduction

Outline

Outline

- **Part 1**: Double/Debiased Machine Learning (DML)

  Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning." The Econometrics Journal 21.1 (2018).

Outline

- **Part 1**: Double/Debiased Machine Learning (DML)

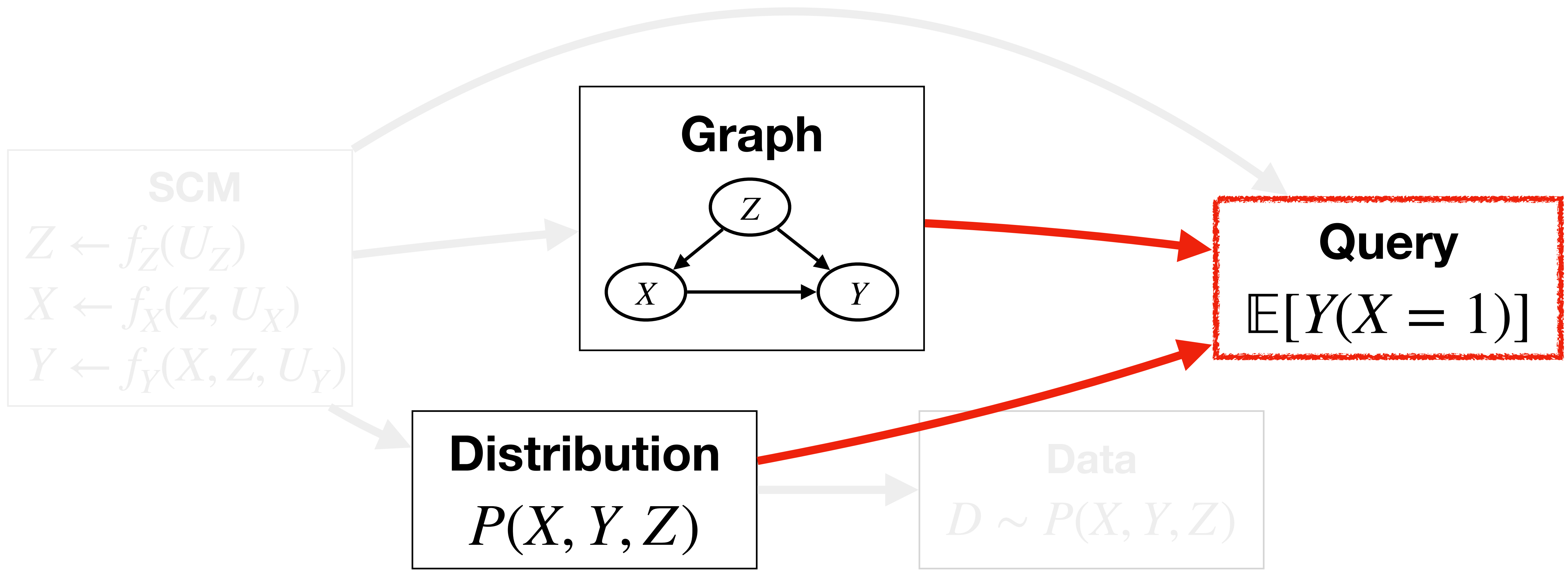
  Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning." The Econometrics Journal 21.1 (2018).

- **Part 2**: DML for any identifiable functional

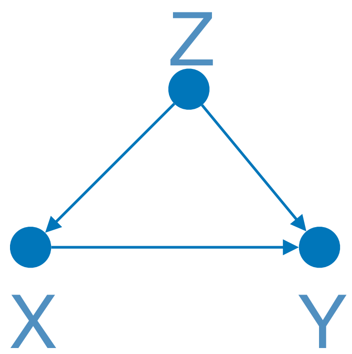
  Jung, Tian, Bareinboim (2021a). Estimating Identifiable Causal Effects through Double Machine Learning. In Proceedings of the 35th AAAI Conference on AI, 2021.

Part II.
Double ML-ID

1. Causal Effect Identification: Big Picture (1)



Causal Effect Identification



Causal graph (G)

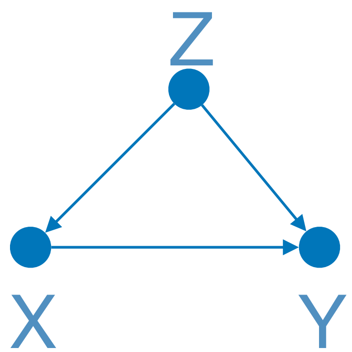
$$P(Z, X, Y)$$

Distribution on G (P)

$$Q_0 := \mathbb{E}[Y \mid do(x)]$$

Causal Query (Q_0)

Causal Effect Identification



Causal graph (G)

$$P(Z, X, Y)$$

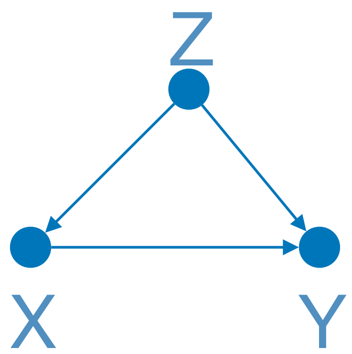
Distribution on G (P)

$$Q_0 := \mathbb{E}[Y \mid do(x)]$$

Causal Query (Q_0)

Given {G, P, Q_0},

Causal Effect Identification



Causal graph (G)

$P(Z, X, Y)$

Distribution on G (P)

$Q_0 := \mathbb{E}[Y \mid do(x)]$

Causal Query (Q_0)

Given $\{G, P, Q_0\}$,

the causal effect identification (ID) task

$\mathbf{ID}(G, P, Q_0)$

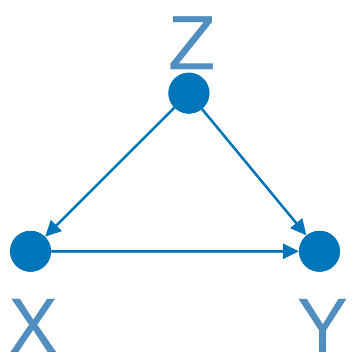
ID algorithm

Causal Effect Identification

Given {G, P, Q_0},

the causal effect identification (ID) task

finds the functional $C(P)$ s.t. $C(P) = Q_0$.



Causal graph (G)

$P(Z, X, Y)$

Distribution on G (P)

$Q_0 := \mathbb{E}[Y \mid do(x)]$

Causal Query (Q_0)

ID(G, P, Q_0)

ID algorithm

$C(P) = \sum_z \mathbb{E}[Y \mid x, z] P(z)$

Causal functional

$C(P)$ **s.t.** $C(P) = Q_0$

Causal Effect Estimation: Back-door Adjustment

$$\mathbb{E}[Y(x)] = \sum_z \mathbb{E}[Y | x, z]P(z).$$

Causal Effect Estimation: Back-door Adjustment

$$\mathbb{E}[Y(x)] = \sum_z \mathbb{E}[Y \mid x, z]P(z).$$

Double/Debiased Machine Learning Estimator:

Causal Effect Estimation: Back-door Adjustment

$$\mathbb{E}[Y(x)] = \sum_z \mathbb{E}[Y \mid x, z] P(z).$$

Double/Debiased Machine Learning Estimator:

↳ The estimator T for $\mathbb{E}[Y(x)]$ composed of estimates $\eta_1 := \hat{\mathbb{E}}[Y \mid x, z]$ and $\eta_2 := \hat{P}(x \mid z)$; i.e., $T = T(\eta_1, \eta_2)$.

Causal Effect Estimation: Back-door Adjustment

$$\mathbb{E}[Y(x)] = \sum_z \mathbb{E}[Y \mid x, z] P(z).$$

Double/Debiased Machine Learning Estimator:

↳ The estimator T for $\mathbb{E}[Y(x)]$ composed of estimates $\eta_1 := \hat{\mathbb{E}}[Y \mid x, z]$ and $\eta_2 := \hat{P}(x \mid z)$; i.e., $T = T(\eta_1, \eta_2)$.

↳ $T(\eta_1, \eta_2)$ converges fast even when either η_1 or η_2 is correct; or η_1, η_2 converge slowly.

Causal Effect Estimation: Limitation

Causal Effect Estimation: Limitation

When the identification expression is beyond the back-door adjustment, virtually no estimators are available.

Causal Effect Estimation: Limitation

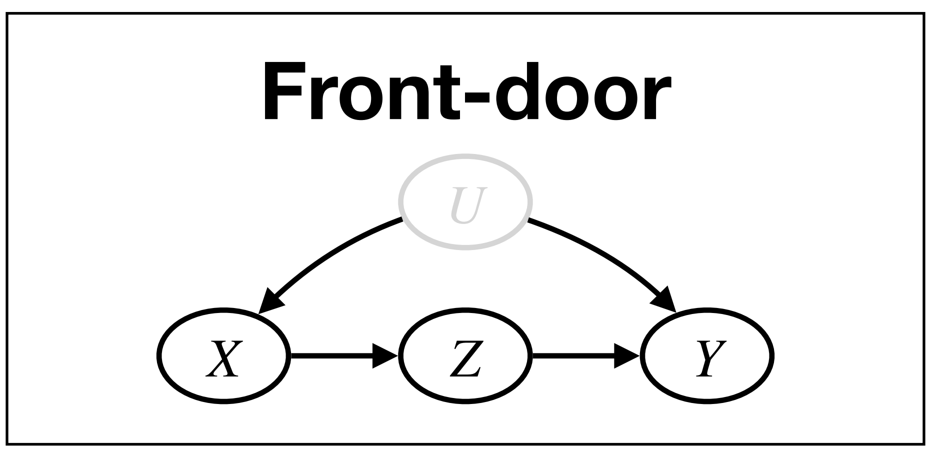
When the identification expression is beyond the back-door adjustment, virtually no estimators are available.

↳ For example, when the graph satisfies the “front-door criteria” , the expression is not a BD adjustment, so the existing estimators are not applicable.

Causal Effect Estimation: Limitation

When the identification expression is beyond the back-door adjustment, virtually no estimators are available.

↳ For example, when the graph satisfies the “front-door criteria” , the expression is not a BD adjustment, so the existing estimators are not applicable.



$$\mathbb{E}[Y(x)] = \sum_{z} P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', z] P(x')$$

Causal Effect Estimation: Our Tasks

Causal Effect Estimation: Our Tasks

My research interest centers around developing estimators for any identifiable causal effects.

Causal Effect Estimation: Our Tasks

My research interest centers around developing estimators for any identifiable causal effects.

↳ For example, in [AAAI-21], we extended the DML estimator for generally identifiable causal effects.

Causal Effect Estimation: Our Tasks

My research interest centers around developing estimators for any identifiable causal effects.

↳ For example, in [AAAI-21], we extended the DML estimator for generally identifiable causal effects.

↳ The estimator has the DML robustness property.

Goal of the talk

Goal of the talk

We will study the construction of the DML estimator for general causal functional

$$C(P) = P_{\mathbf{x}}(\mathbf{y})$$

Goal of the talk

We will study the construction of the DML estimator for general causal functional

$$C(P) = P_{\mathbf{x}}(\mathbf{y})$$

We assume the followings in the lecture.

Goal of the talk

We will study the construction of the DML estimator for general causal functional

$$C(P) = P_{\mathbf{x}}(\mathbf{y})$$

We assume the followings in the lecture.

$\mathbf{V}$ is a set of discrete variables.

Goal of the talk

We will study the construction of the DML estimator for general causal functional

$$C(P) = P_{\mathbf{x}}(\mathbf{y})$$

We assume the followings in the lecture.

$\mathbf{V}$ is a set of discrete variables.

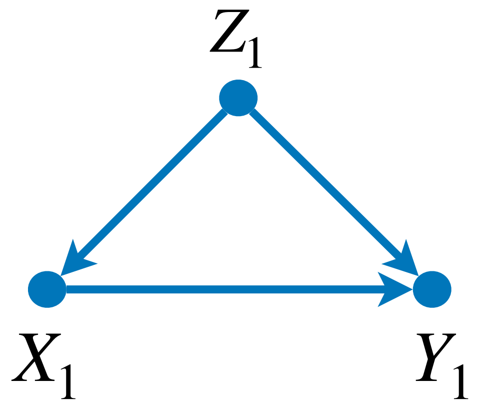
$P(\mathbf{v}) > 0$ for any v.

Multi-outcome sequential BD (mSBD)

Jung, Yonghan, Jin Tian, and Elias Bareinboim. "Estimating causal effects using weighting-based estimators." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 06. 2020.

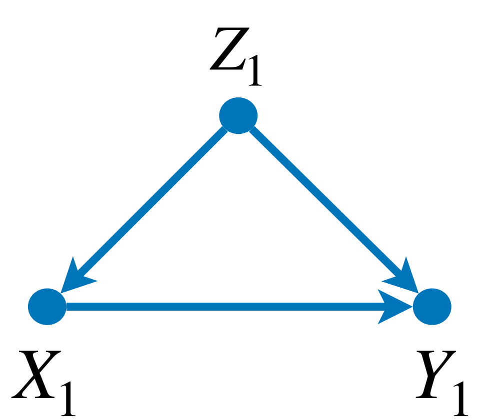
Multi-outcome sequential BD (mSBD)

Multi-outcome sequential BD (mSBD)



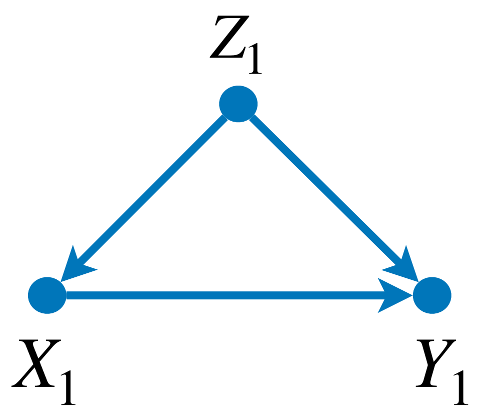
$i = 1$

Multi-outcome sequential BD (mSBD)

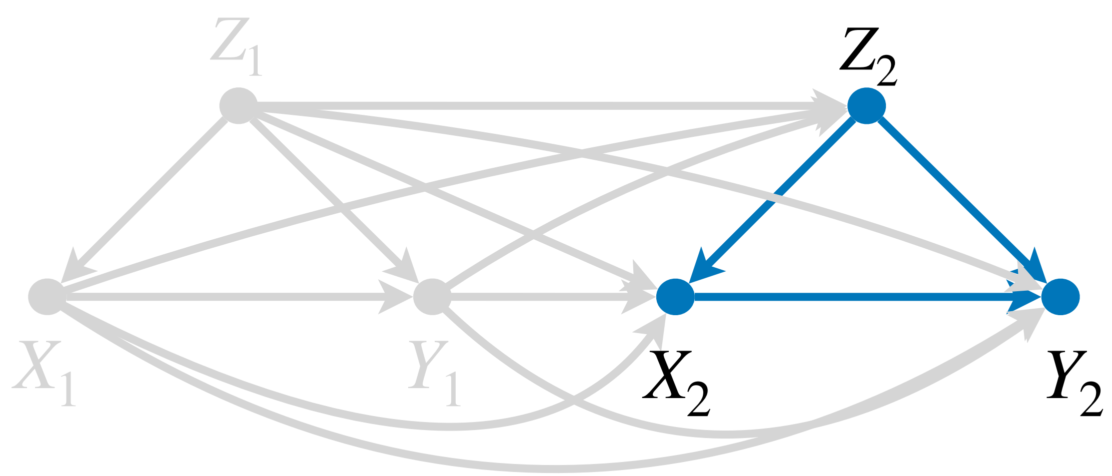


$i = 1$

Multi-outcome sequential BD (mSBD)



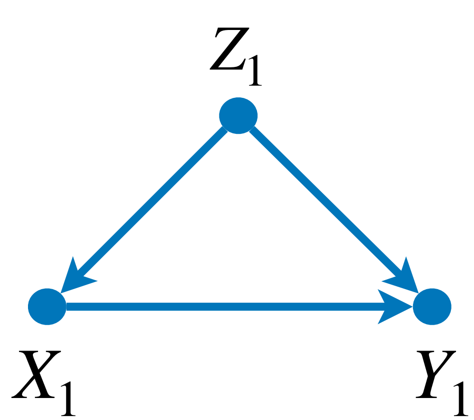
$i = 1$



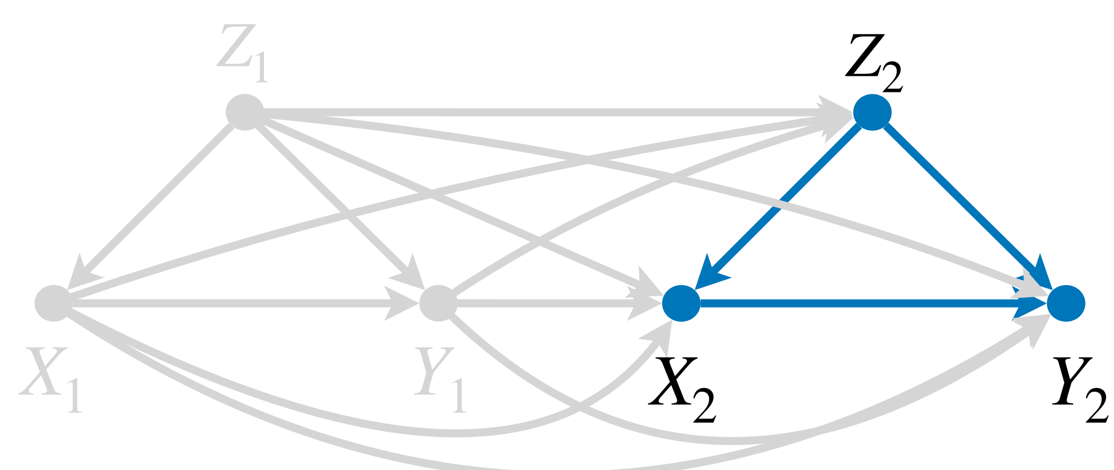
$i = 2$

Multi-outcome sequential BD (mSBD)

Multi-outcome sequential BD (mSBD): An extension of BD, where, at i th round, Z_i satisfies the BD criterion relative to $\{X_i, Y_i\}$ conditioned on previous variables $\{(X_j, Y_j, Z_j)\}_{j=1}^{i-1}$.



$i = 1$

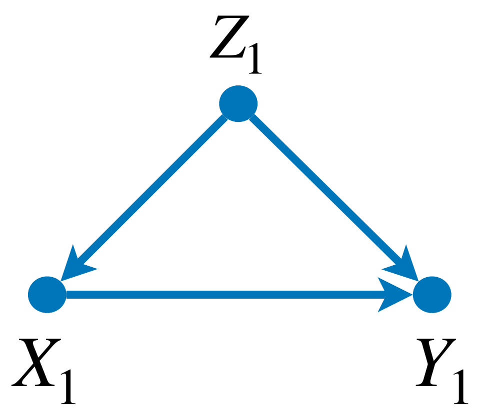


$i = 2$

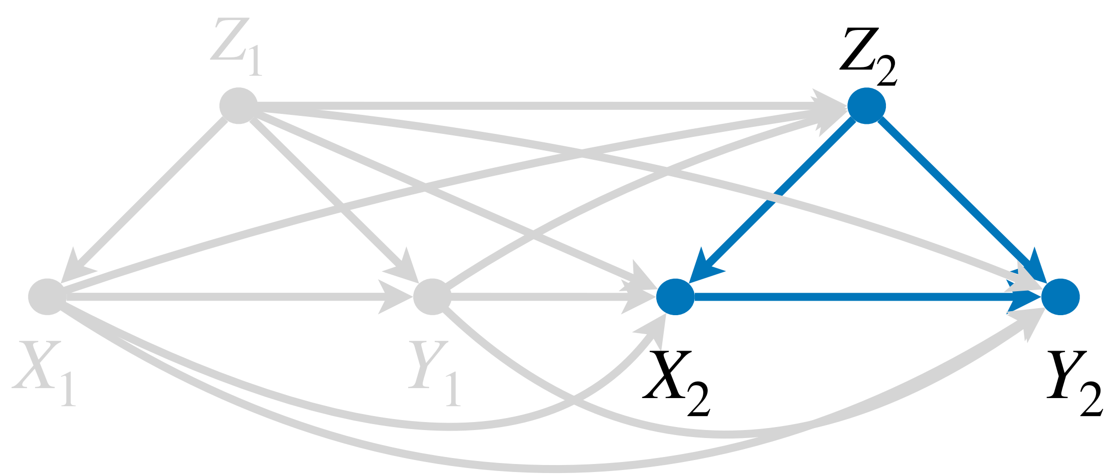
Multi-outcome sequential BD (mSBD)

Multi-outcome sequential BD (mSBD): An extension of BD, where, at i th round, Z_i satisfies the BD criterion relative to $\{X_i, Y_i\}$ conditioned on previous variables $\{(X_j, Y_j, Z_j)\}_{j=1}^{i-1}$.

- This can cover the case where there are *no unmeasured confounder* between $\mathbf{X} = \{X_1, \cdots, X_n\}$ and $\mathbf{Y} = \{Y_1, \cdots, Y_n\}$.



$i = 1$



$i = 2$

Multi-outcome sequential BD (mSBD)

Multi-outcome sequential BD (mSBD)

mSBD adjustment: If $\mathbf{Z} = \{Z_1, \cdots, Z_n\}$ satisfies the mSBD criterion relative to $(\mathbf{X}, \mathbf{Y})$,

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{Y_i \in \mathbf{Y}} P(y_i \mid \mathbf{x}^{(i)}, \mathbf{z}^{(i)}, \mathbf{y}^{(i-1)}) \prod_{Z_i \in \mathbf{Z}} P(z_i \mid \mathbf{x}^{(i-1)}, \mathbf{z}^{(i-1)}, \mathbf{y}^{(i-1)})$$

Multi-outcome sequential BD (mSBD)

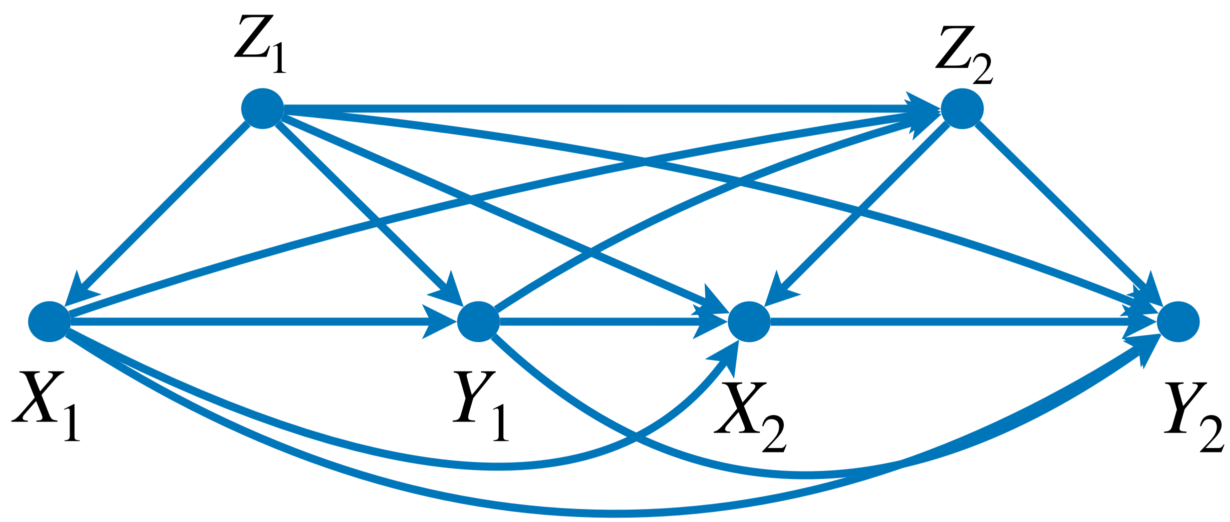
mSBD adjustment: If $\mathbf{Z} = \{Z_1, \cdots, Z_n\}$ satisfies the mSBD criterion relative to $(\mathbf{X}, \mathbf{Y})$,

$\mathbf{X}^{(i)} := \{X_1, \cdots, X_i\}$

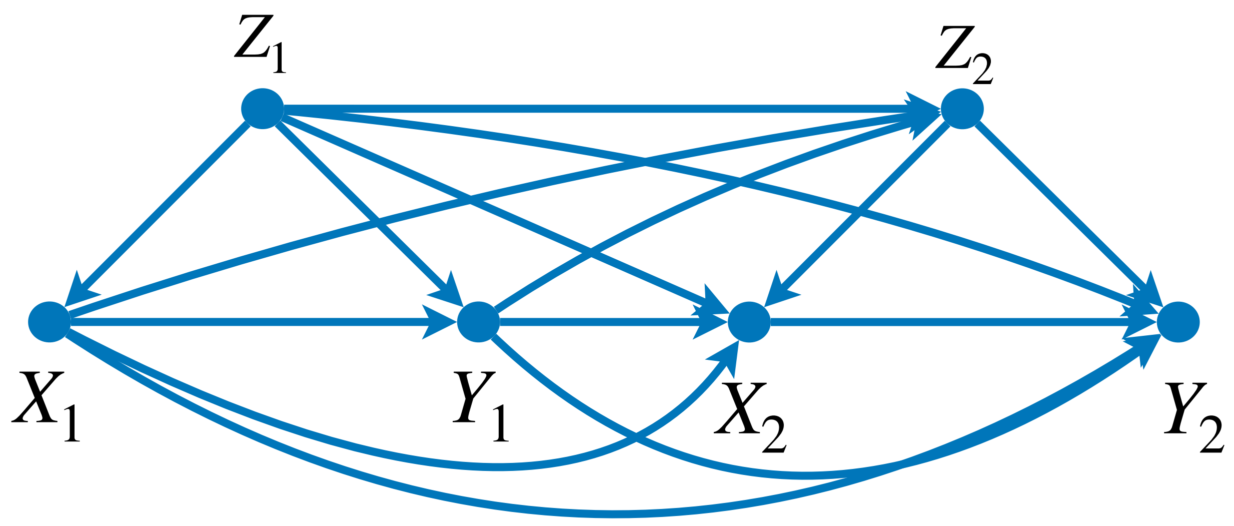
$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{Y_i \in \mathbf{Y}} P(y_i \mid \mathbf{x}^{(i)}, \mathbf{z}^{(i)}, \mathbf{y}^{(i-1)}) \prod_{Z_i \in \mathbf{Z}} P(z_i \mid \mathbf{x}^{(i-1)}, \mathbf{z}^{(i-1)}, \mathbf{y}^{(i-1)})$$

mSBD - example

mSBD - example

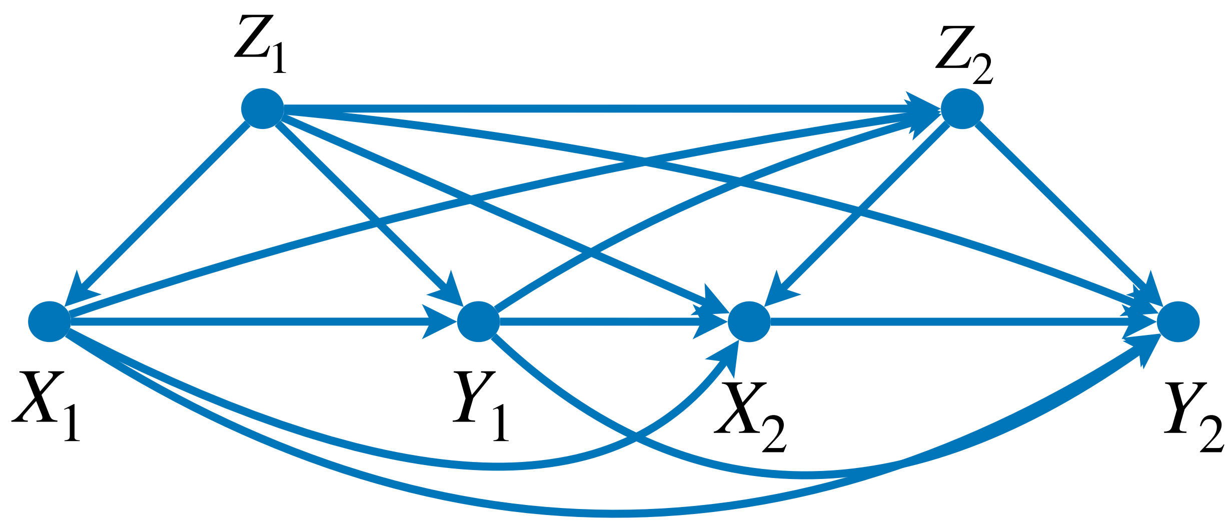


mSBD - example



$\{Z_1, Z_2\}$ satisfies mSBD criterion relative to $(\{X_1, X_2\}, \{Y_1, Y_2\})$.

mSBD - example



$\{Z_1, Z_2\}$ satisfies mSBD criterion relative to $(\{X_1, X_2\}, \{Y_1, Y_2\})$.

$$P_{x_1,x_2}(y_1, y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \,|\, x_1, z_1)P(z_2 \,|\, z_1, x_1, y_1)P(y_2 \,|\, z_1, x_1, y_1, z_2, x_2)$$

Regression (REG)-based Estimand for mSBD

Regression (REG)-based Estimand for mSBD

$f^{REG}(\mathbf{V}; \mathbf{H}_0) = H_0^1(x_1)$, defined as follow:

Regression (REG)-based Estimand for mSBD

$f^{REG}(\mathbf{V}; \mathbf{H}_0) = H_0^1(x_1)$, defined as follow:

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

Regression (REG)-based Estimand for mSBD

$f^{REG}(\mathbf{V}; \mathbf{H}_0) = H_0^1(x_1)$, defined as follow:

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

For all $k = n, n-1, \cdots, 1$

Regression (REG)-based Estimand for mSBD

$f^{REG}(\mathbf{V}; \mathbf{H}_0) = H_0^1(x_1)$, defined as follow:

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

For all $k = n, n-1, \cdots, 1$

$$H_0^k(X_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, X_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

Regression (REG)-based Estimand for mSBD

$f^{REG}(\mathbf{V}; \mathbf{H}_0) = H_0^1(x_1)$, defined as follow:

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

For all $k = n, n-1, \cdots, 1$

$$H_0^k(X_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, X_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

$$H_0^k(x_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, x_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

REG Estimand for mSBD - Proof

REG Estimand for mSBD - Proof

$$\mathbb{E}[f^{REG}(\mathbf{V}; \mathbf{H}_0)] = \mathbb{E}[H_0^1(x_1)] = C(P)$$

REG Estimand for mSBD - Proof

$$\mathbb{E}[f^{REG}(\mathbf{V};\mathbf{H}_0)] = \mathbb{E}[H_0^1(x_1)] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

REG Estimand for mSBD - Proof

$$\mathbb{E}[f^{REG}(\mathbf{V};\mathbf{H}_0)] = \mathbb{E}[H_0^1(x_1)] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$H_0^3(x_3) := I_{y_1,y_2}(Y_1,Y_2)$$

REG Estimand for mSBD - Proof

$$\mathbb{E}[f^{REG}(\mathbf{V};\mathbf{H}_0)] = \mathbb{E}[H_0^1(x_1)] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$H_0^3(x_3) := I_{y_1,y_2}(Y_1,Y_2)$$

$$H_0^2(x_2) = \mathbb{E}[H_0^3 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)}] = I_{y_1}(Y_1)P(y_2 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)})$$

REG Estimand for mSBD - Proof

$$\mathbb{E}[f^{REG}(\mathbf{V};\mathbf{H}_0)] = \mathbb{E}[H_0^1(x_1)] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$H_0^3(x_3) := I_{y_1,y_2}(Y_1,Y_2)$$

$$H_0^2(x_2) = \mathbb{E}[H_0^3 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)}] = I_{y_1}(Y_1)P(y_2 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)})$$

$$H_0^1(x_1) = \mathbb{E}[H_0^2 \mid x_1, Z_1] = \sum_{z_2} P(y_2 \mid \mathbf{x}^{(2)}, y_1, z_2, Z_1)P(z_2 \mid x_1, y_1, Z_1)P(y_1 \mid x_1, Z_1)$$

REG Estimand for mSBD - Proof

$$\mathbb{E}[f^{REG}(\mathbf{V};\mathbf{H}_0)] = \mathbb{E}[H_0^1(x_1)] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$H_0^3(x_3) := I_{y_1,y_2}(Y_1,Y_2)$$

$$H_0^2(x_2) = \mathbb{E}[H_0^3 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)}] = I_{y_1}(Y_1)P(y_2 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)})$$

$$H_0^1(x_1) = \mathbb{E}[H_0^2 \mid x_1, Z_1] = \sum_{z_2} P(y_2 \mid \mathbf{x}^{(2)}, y_1, z_2, Z_1)P(z_2 \mid x_1, y_1, Z_1)P(y_1 \mid x_1, Z_1)$$

$$\mathbb{E}[H_0^1(x_1)] = \mathbb{E}[f^{REG}(\mathbf{V};\eta_0)] = \sum_{z_1,z_2} P(y_2 \mid \mathbf{x}^{(2)}, y_1, z_2, z_1)P(z_2 \mid x_1, y_1, z_1)P(y_1 \mid x_1, z_1)P(z_1)$$

REG Estimand for mSBD - Proof

$$\mathbb{E}[f^{REG}(\mathbf{V};\mathbf{H}_0)] = \mathbb{E}[H_0^1(x_1)] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$H_0^3(x_3) := I_{y_1,y_2}(Y_1,Y_2)$$

$$H_0^2(x_2) = \mathbb{E}[H_0^3 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)}] = I_{y_1}(Y_1)P(y_2 \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)})$$

$$H_0^1(x_1) = \mathbb{E}[H_0^2 \mid x_1, Z_1] = \sum_{z_2} P(y_2 \mid \mathbf{x}^{(2)}, y_1, z_2, Z_1)P(z_2 \mid x_1, y_1, Z_1)P(y_1 \mid x_1, Z_1)$$

$$\mathbb{E}[H_0^1(x_1)] = \mathbb{E}[f^{REG}(\mathbf{V};\eta_0)] = \sum_{z_1,z_2} P(y_2 \mid \mathbf{x}^{(2)}, y_1, z_2, z_1)P(z_2 \mid x_1, y_1, z_1)P(y_1 \mid x_1, z_1)P(z_1)$$

$$= P_{x_1,x_2}(y_1,y_2)$$

IPW Estimand for mSBD

IPW Estimand for mSBD

$f^{IPW}(\mathbf{V};\Pi) := W_0^n I_{\mathbf{y}}(\mathbf{Y})$ for

IPW Estimand for mSBD

$$f^{IPW}(\mathbf{V}; \Pi) := W_0^n I_{\mathbf{y}}(\mathbf{Y}) \text{ for}$$

$$W_0^n = \prod_{p=1}^{n} \frac{I_{x_p}(X_p)}{\pi_0^p(\mathbf{X}^{(p-1)}, \mathbf{Z}^{(p)}, \mathbf{Y}^{(p-1)})}, \text{ where}$$

IPW Estimand for mSBD

$$f^{IPW}(\mathbf{V};\Pi) := W_0^n I_{\mathbf{y}}(\mathbf{Y})$$ for

$$W_0^n = \prod_{p=1}^{n} \frac{I_{x_p}(X_p)}{\pi_0^p(\mathbf{X}^{(p-1)}, \mathbf{Z}^{(p)}, \mathbf{Y}^{(p-1)})},$$ where

$$\pi_0^p(\mathbf{X}^{(p-1)}, \mathbf{Z}^{(p)}, \mathbf{Y}^{(p-1)}) := P(X_p \,|\, \mathbf{X}^{(p-1)}, \mathbf{Z}^{(p)}, \mathbf{Y}^{(p-1)})$$

IPW Estimand for mSBD - Proof

$$\mathbb{E}[f^{IPW}(\mathbf{V};\Pi)] = \mathbb{E}[W_0^n I_{\mathbf{y}}(\mathbf{Y})] = C(P)$$

IPW Estimand for mSBD - Proof

$$\mathbb{E}[f^{IPW}(\mathbf{V};\Pi)] = \mathbb{E}[W_0^n I_{\mathbf{y}}(\mathbf{Y})] = C(P)$$

$$P_{x_1,x_2}(y_1, y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1, z_1)P(z_2 \mid z_1, x_1, y_1)P(y_2 \mid z_1, x_1, y_1, z_2, x_2)$$

IPW Estimand for mSBD - Proof

$$\mathbb{E}[f^{IPW}(\mathbf{V};\Pi)] = \mathbb{E}[W_0^n I_{\mathbf{y}}(\mathbf{Y})] = C(P)$$

$$P_{x_1,x_2}(y_1, y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1, z_1)P(z_2 \mid z_1, x_1, y_1)P(y_2 \mid z_1, x_1, y_1, z_2, x_2)$$

$$\mathbb{E}[W_0^2 I_{y_1,y_2}(Y_1, Y_2)] = \mathbb{E}\left[\frac{I_{x_1,x_2}(X_1, X_2)}{P(X_1 \mid Z_1)P(X_2 \mid \mathbf{Z}, X_1, Y_1)} I_{y_1,y_2}(Y_1, Y_2)\right]$$

IPW Estimand for mSBD - Proof

$$\mathbb{E}[f^{IPW}(\mathbf{V};\Pi)] = \mathbb{E}[W_0^n I_{\mathbf{y}}(\mathbf{Y})] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$\mathbb{E}[W_0^2 I_{y_1,y_2}(Y_1,Y_2)] = \mathbb{E}\left[\frac{I_{x_1,x_2}(X_1,X_2)}{P(X_1 \mid Z_1)P(X_2 \mid \mathbf{Z},X_1,Y_1)} I_{y_1,y_2}(Y_1,Y_2)\right]$$

$$= \sum_{\mathbf{y}',\mathbf{x}',\mathbf{z}} \frac{P(y_2,x_2,z_2,y_1,x_1,z_1)}{P(x_1' \mid z_1)P(x_2' \mid \mathbf{z},x_1',y_1')} I_{y_1,y_2}(y_1',y_2') I_{x_1,x_2}(x_1',x_2')$$

IPW Estimand for mSBD - Proof

$$\mathbb{E}[f^{IPW}(\mathbf{V};\Pi)] = \mathbb{E}[W_0^n I_{\mathbf{y}}(\mathbf{Y})] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$\mathbb{E}[W_0^2 I_{y_1,y_2}(Y_1,Y_2)] = \mathbb{E}\left[\frac{I_{x_1,x_2}(X_1,X_2)}{P(X_1 \mid Z_1)P(X_2 \mid \mathbf{Z},X_1,Y_1)} I_{y_1,y_2}(Y_1,Y_2)\right]$$

$$= \sum_{\mathbf{y}',\mathbf{x}',\mathbf{z}} \frac{P(y_2,x_2,z_2,y_1,x_1,z_1)}{P(x_1' \mid z_1)P(x_2' \mid \mathbf{z},x_1',y_1')} I_{y_1,y_2}(y_1',y_2') I_{x_1,x_2}(x_1',x_2')$$

$$= \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

IPW Estimand for mSBD - Proof

$$\mathbb{E}[f^{IPW}(\mathbf{V};\Pi)] = \mathbb{E}[W_0^n I_{\mathbf{y}}(\mathbf{Y})] = C(P)$$

$$P_{x_1,x_2}(y_1,y_2) = \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$\mathbb{E}[W_0^2 I_{y_1,y_2}(Y_1,Y_2)] = \mathbb{E}\left[\frac{I_{x_1,x_2}(X_1,X_2)}{P(X_1 \mid Z_1)P(X_2 \mid \mathbf{Z},X_1,Y_1)} I_{y_1,y_2}(Y_1,Y_2)\right]$$

$$= \sum_{\mathbf{y}',\mathbf{x}',\mathbf{z}} \frac{P(y_2,x_2,z_2,y_1,x_1,z_1)}{P(x_1' \mid z_1)P(x_2' \mid \mathbf{z},x_1',y_1')} I_{y_1,y_2}(y_1',y_2') I_{x_1,x_2}(x_1',x_2')$$

$$= \sum_{z_1,z_2} P(z_1)P(y_1 \mid x_1,z_1)P(z_2 \mid z_1,x_1,y_1)P(y_2 \mid z_1,x_1,y_1,z_2,x_2)$$

$$= P_{x_1,x_2}(y_1,y_2)$$

DR Estimand for mSBD

DR Estimand for mSBD

$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \Pi_0\}) = H_0^1(x_1) + \sum_{k=1}^{n} W_0^k \{H_0^{k+1}(x_{k+1}) - H_0^k(X_k)\},$$

DR Estimand for mSBD

$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \Pi_0\}) = H_0^1(x_1) + \sum_{k=1}^{n} W_0^k \{H_0^{k+1}(x_{k+1}) - H_0^k(X_k)\},$$

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

DR Estimand for mSBD

$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \Pi_0\}) = H_0^1(x_1) + \sum_{k=1}^{n} W_0^k \{H_0^{k+1}(x_{k+1}) - H_0^k(X_k)\},$$

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

For all $k = n, n-1, \cdots, 1$

DR Estimand for mSBD

$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \Pi_0\}) = H_0^1(x_1) + \sum_{k=1}^{n} W_0^k \{H_0^{k+1}(x_{k+1}) - H_0^k(X_k)\},$$

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

For all $k = n, n-1, \cdots, 1$

$$H_0^k(X_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, X_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

DR Estimand for mSBD

$$f^{DR}(\mathbf{V};\{\mathbf{H}_0,\Pi_0\}) = H_0^1(x_1) + \sum_{k=1}^{n} W_0^k\{H_0^{k+1}(x_{k+1}) - H_0^k(X_k)\},$$

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

For all $k = n, n-1, \cdots, 1$

$$H_0^k(X_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, X_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

$$H_0^k(x_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, x_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

DR Estimand for mSBD

$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \Pi_0\}) = H_0^1(x_1) + \sum_{k=1}^{n} W_0^k \{H_0^{k+1}(x_{k+1}) - H_0^k(X_k)\},$$

$H_0^{n+1}(x_{n+1}) := I_{\mathbf{y}}(\mathbf{Y})$ (for notational convenience)

For all $k = n, n-1, \cdots, 1$

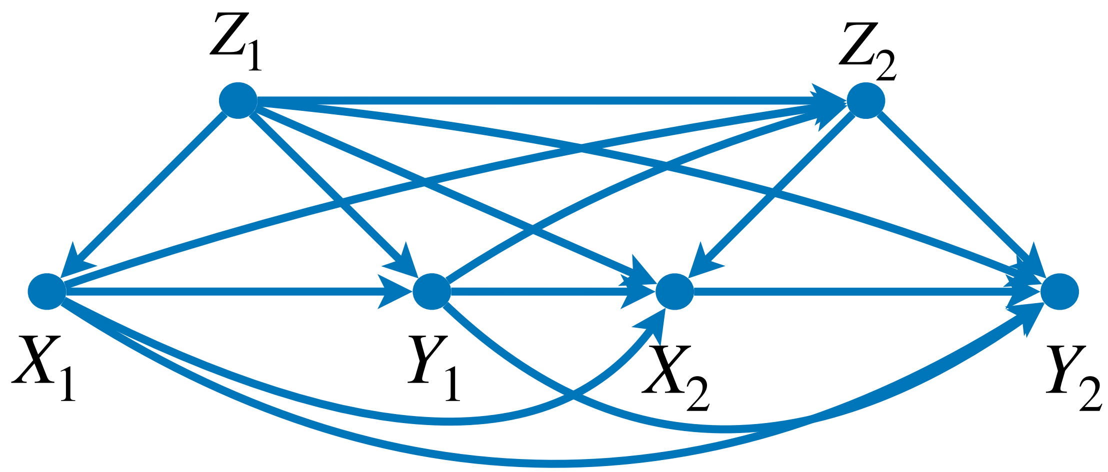
$$H_0^k(X_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, X_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

$$H_0^k(x_k) = \mathbb{E}[H_0^{k+1}(x_{k+1}) \,|\, x_k, \mathbf{X}^{(k-1)}, \mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k)}]$$

$$W_0^k = \prod_{p=1}^{k} \frac{I_{x_p}(X_p)}{\pi_0^p(\mathbf{X}^{(p-1)}, \mathbf{Z}^{(p)}, \mathbf{Y}^{(p-1)})},$$ where

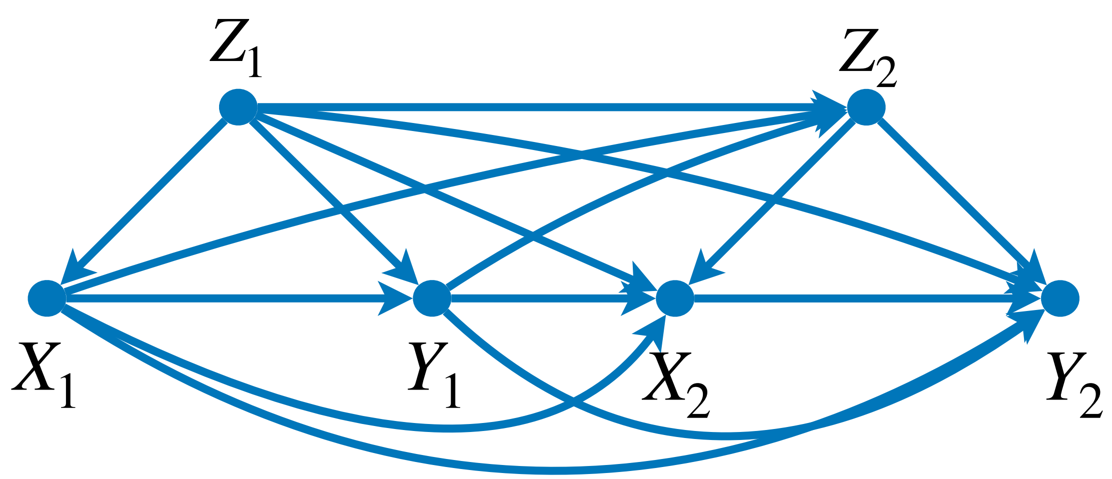
$$\pi_0^p(\mathbf{X}^{(p-1)}, \mathbf{Z}^{(p)}, \mathbf{Y}^{(p-1)}) := P(X_p \,|\, \mathbf{X}^{(p-1)}, \mathbf{Z}^{(p)}, \mathbf{Y}^{(p-1)})$$

DR Estimand for mSBD — Example 1



$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\}) = H_0^1(x_1) + W_0^1(H_0^2(x_2) - H_0^1(X_1)) + W_0^2(I_{\mathbf{y}}(\mathbf{Y}) - H_0^2(X_2))$$

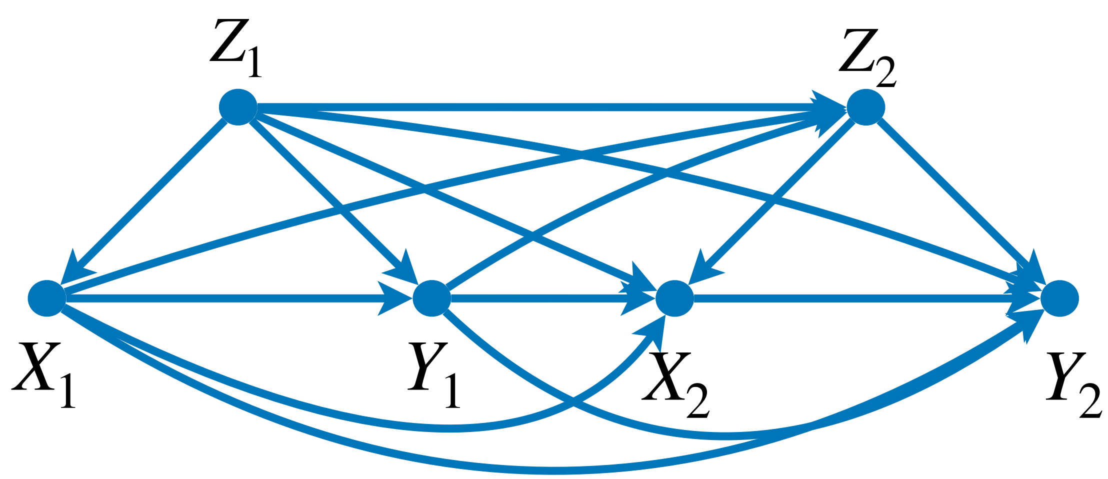
DR Estimand for mSBD — Example 1



$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\}) = H_0^1(x_1) + W_0^1(H_0^2(x_2) - H_0^1(X_1)) + W_0^2(I_{\mathbf{y}}(\mathbf{Y}) - H_0^2(X_2))$$

$$H_0^2(X_2) = \mathbb{E}[I_{\mathbf{y}}(\mathbf{Y}) \mid \mathbf{X}^{(2)}, Y_1, \mathbf{Z}^{(2)}]$$

DR Estimand for mSBD — Example 1

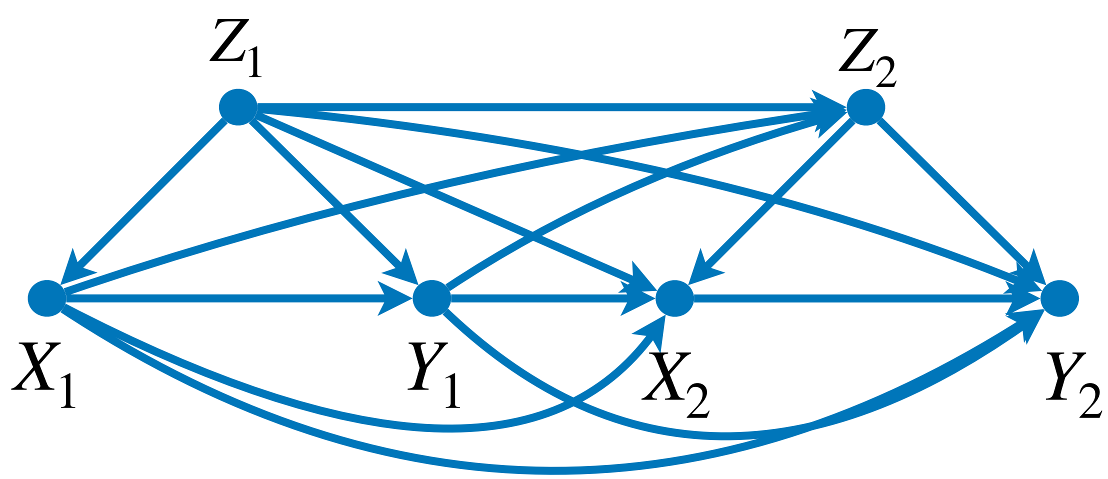


$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\}) = H_0^1(x_1) + W_0^1(H_0^2(x_2) - H_0^1(X_1)) + W_0^2(I_{\mathbf{y}}(\mathbf{Y}) - H_0^2(X_2))$$

$$H_0^2(X_2) = \mathbb{E}[I_{\mathbf{y}}(\mathbf{Y}) \,|\, \mathbf{X}^{(2)}, Y_1, \mathbf{Z}^{(2)}]$$

$$H_0^2(x_2) = \mathbb{E}[I_{\mathbf{y}}(\mathbf{Y}) \,|\, x_2, X_1, Y_1, \mathbf{Z}^{(2)}]$$

DR Estimand for mSBD — Example 1



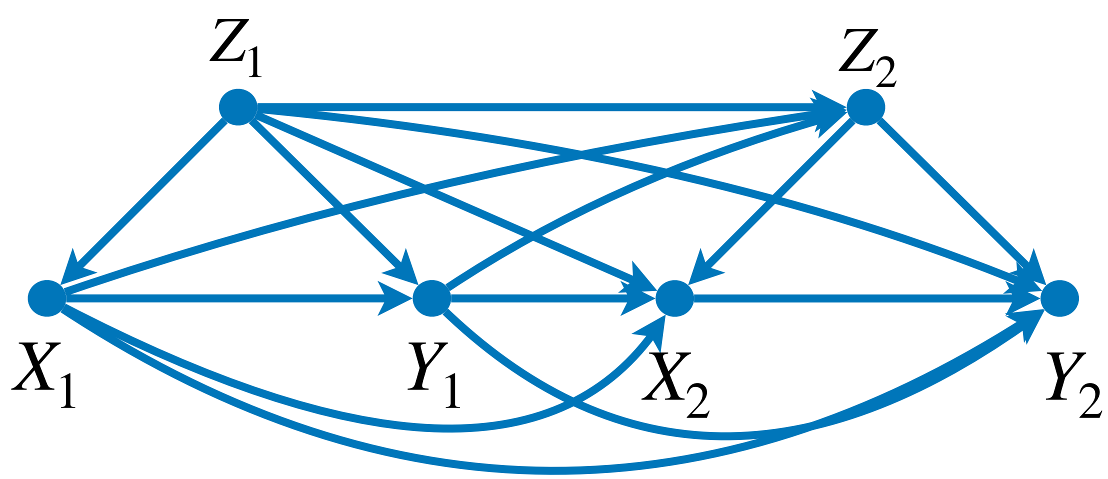
$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\}) = H_0^1(x_1) + W_0^1(H_0^2(x_2) - H_0^1(X_1)) + W_0^2(I_{\mathbf{y}}(\mathbf{Y}) - H_0^2(X_2))$$

$$H_0^2(X_2) = \mathbb{E}[I_{\mathbf{y}}(\mathbf{Y}) \mid \mathbf{X}^{(2)}, Y_1, \mathbf{Z}^{(2)}]$$

$$H_0^1(X_1) = \mathbb{E}[H_0^2(x_2) \mid X_1, Z_1]$$

$$H_0^2(x_2) = \mathbb{E}[I_{\mathbf{y}}(\mathbf{Y}) \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)}]$$

DR Estimand for mSBD — Example 1



$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\}) = H_0^1(x_1) + W_0^1(H_0^2(x_2) - H_0^1(X_1)) + W_0^2(I_{\mathbf{y}}(\mathbf{Y}) - H_0^2(X_2))$$

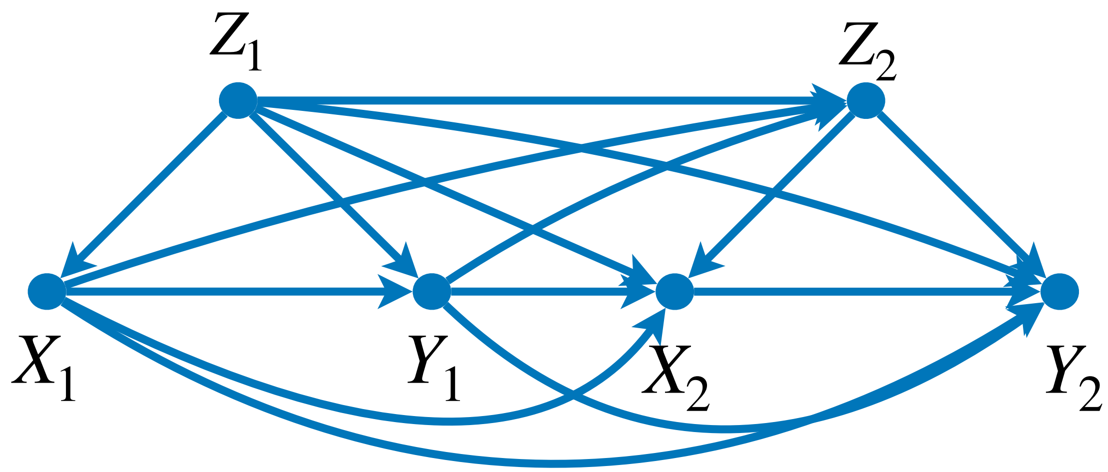
$$H_0^2(X_2) = \mathbb{E}[I_{\mathbf{y}}(\mathbf{Y}) \mid \mathbf{X}^{(2)}, Y_1, \mathbf{Z}^{(2)}]$$

$$H_0^1(X_1) = \mathbb{E}[H_0^2(x_2) \mid X_1, Z_1]$$

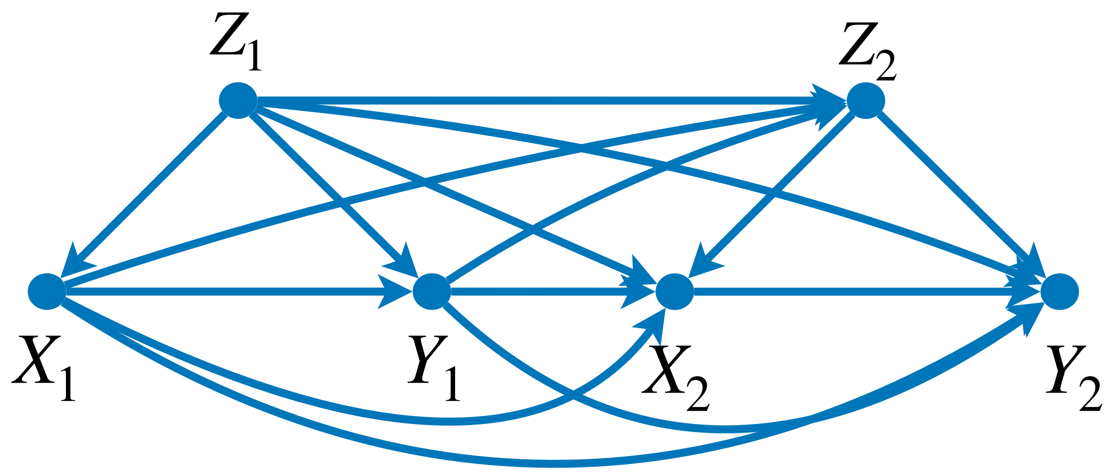
$$H_0^2(x_2) = \mathbb{E}[I_{\mathbf{y}}(\mathbf{Y}) \mid x_2, X_1, Y_1, \mathbf{Z}^{(2)}]$$

$$H_0^1(x_1) = \mathbb{E}[H_0^2(x_2) \mid x_1, Z_1]$$

DR Estimand for mSBD — Example 2

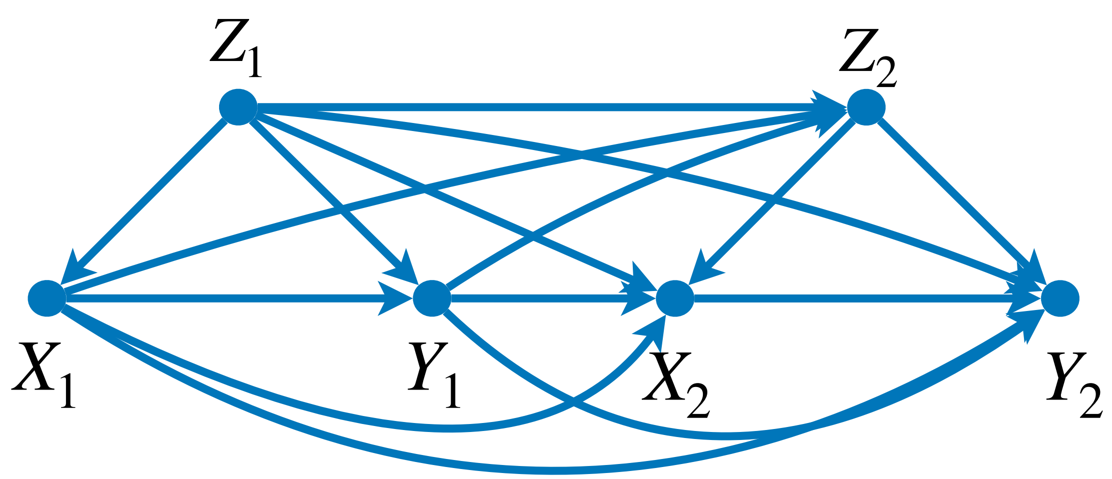


DR Estimand for mSBD — Example 2



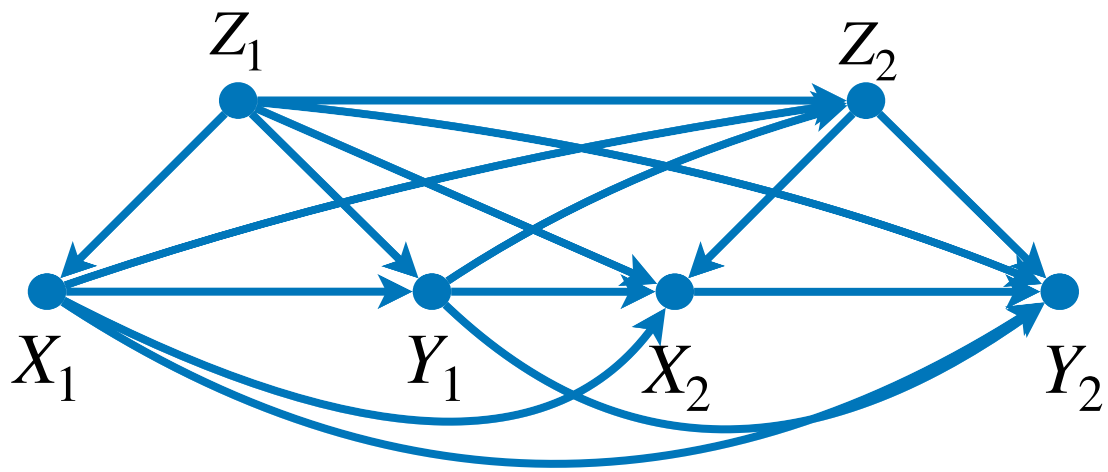
$$W_0^2 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)} \frac{I_{x_2}(X_2)}{\pi_0^2(X_1, \mathbf{Z}^{(2)}, Y_1)}$$

DR Estimand for mSBD — Example 2



$$W_0^2 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)} \frac{I_{x_2}(X_2)}{\pi_0^2(X_1, \mathbf{Z}^{(2)}, Y_1)} \qquad W_0^1 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)}$$

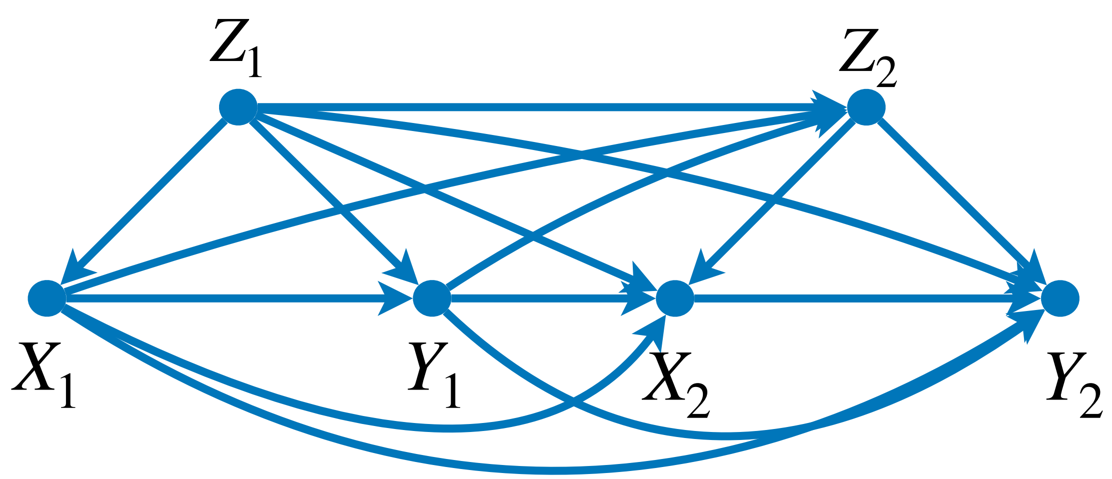
DR Estimand for mSBD — Example 2



$$W_0^2 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)} \frac{I_{x_2}(X_2)}{\pi_0^2(X_1, \mathbf{Z}^{(2)}, Y_1)} \qquad W_0^1 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)}$$

$$\pi_0^1(Z_1) = P(X_1 \mid Z_1)$$

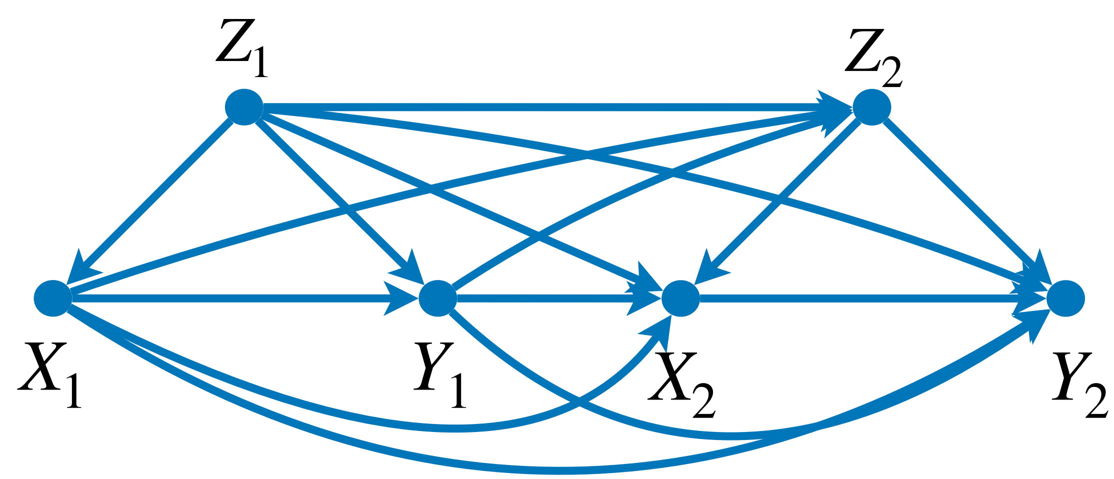
DR Estimand for mSBD — Example 2



$$W_0^2 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)} \frac{I_{x_2}(X_2)}{\pi_0^2(X_1, \mathbf{Z}^{(2)}, Y_1)} \qquad W_0^1 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)}$$

$$\pi_0^1(Z_1) = P(X_1 \mid Z_1) \qquad \pi_0^2(X_1, \mathbf{Z}^{(2)}, Y_1) = P(X_2 \mid X_1, \mathbf{Z}^{(2)}, Y_1)$$

DR Estimand for mSBD — Example 2



$$f^{DR}(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\}) = H_0^1(x_1) + W_0^1(H_0^2(x_2) - H_0^1(X_1)) + W_0^2(I_{\mathbf{y}}(\mathbf{Y}) - H_0^2(X_2))$$

$$W_0^2 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)} \frac{I_{x_2}(X_2)}{\pi_0^2(X_1, \mathbf{Z}^{(2)}, Y_1)} \qquad W_0^1 = \frac{I_{x_1}(X_1)}{\pi_0^1(Z_1)}$$

$$\pi_0^1(Z_1) = P(X_1 \mid Z_1) \qquad \pi_0^2(X_1, \mathbf{Z}^{(2)}, Y_1) = P(X_2 \mid X_1, \mathbf{Z}^{(2)}, Y_1)$$

Orthogonality of the DR estimand - Proof Sketch 1

$f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$ is an orthogonal estimand.

Orthogonality of the DR estimand - Proof Sketch 1

$f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$ is an orthogonal estimand.

For all $i = 1,2,\cdots, n$

Orthogonality of the DR estimand - Proof Sketch 1

$f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$ is an orthogonal estimand.

For all $i = 1,2,\cdots,n$

$$\frac{\partial}{\partial W_i}\mathbb{E}[f(\mathbf{V};\eta)\,|_{\eta=\eta_0} = \mathbb{E}\left[\frac{\partial}{\partial W_i}\{H^1 + \sum_{j\neq i}^{n} W_j(H^{j+1} - H^j) + W_i\{H^{i+1} - H^i\}\}\,|_{\eta=\eta_0}\right]$$

Orthogonality of the DR estimand - Proof Sketch 1

$$f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$$ is

an orthogonal estimand.

For all $i = 1,2,\cdots, n$

$$\frac{\partial}{\partial W_i}\mathbb{E}[f(\mathbf{V};\eta)\,|_{\eta=\eta_0} = \mathbb{E}\left[\frac{\partial}{\partial W_i}\{H^1 + \sum_{j\neq i}^{n} W_j(H^{j+1} - H^j) + W_i\{H^{i+1} - H^i\}\}\,|_{\eta=\eta_0}\right]$$

$$= \mathbb{E}[H_0^{i+1} - H_0^i]$$

Orthogonality of the DR estimand - Proof Sketch 1

$f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$ is an orthogonal estimand.

For all $i = 1,2,\cdots, n$

$$\frac{\partial}{\partial W_i}\mathbb{E}[f(\mathbf{V};\eta)\,|_{\eta=\eta_0} = \mathbb{E}\left[\frac{\partial}{\partial W_i}\{H^1 + \sum_{j\neq i}^{n} W_j(H^{j+1} - H^j) + W_i\{H^{i+1} - H^i\}\}\,|_{\eta=\eta_0}\right]$$

$$= \mathbb{E}[H_0^{i+1} - H_0^i]$$

$$= 0$$

Orthogonality of the DR estimand - Proof Sketch 2

$$f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i \{H^{i+1}(x_{i+1}) - H^i(X_i)\}$$

is an orthogonal estimand.

Orthogonality of the DR estimand - Proof Sketch 2

$$f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$$

is an orthogonal estimand.

For all $i = 1, 2, \cdots, n$

Orthogonality of the DR estimand - Proof Sketch 2

$$f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$$

is an orthogonal estimand.

For all $i = 1,2,\cdots,n$

$$\frac{\partial}{\partial H^i}\mathbb{E}[f(\mathbf{V};\eta)\,|_{\eta=\eta_0} = \mathbb{E}\left[\frac{\partial}{\partial H^i}\{W^i(H^{i+1} - H^i) + W^{i-1}(H^i - H^{i-1}) + \cdots\}\,|_{\eta=\eta_0}\right]$$

Orthogonality of the DR estimand - Proof Sketch 2

$$f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i \{H^{i+1}(x_{i+1}) - H^i(X_i)\}$$

is an orthogonal estimand.

For all $i = 1,2,\cdots,n$

$$\frac{\partial}{\partial H^i}\mathbb{E}[f(\mathbf{V};\eta)\,|_{\eta=\eta_0} = \mathbb{E}\left[\frac{\partial}{\partial H^i}\{W^i(H^{i+1} - H^i) + W^{i-1}(H^i - H^{i-1}) + \cdots\}\,|_{\eta=\eta_0}\right]$$

$$= \mathbb{E}[-W_0^{i+1} + W_0^i]$$

Orthogonality of the DR estimand - Proof Sketch 2

$$f(\mathbf{V};\eta=\{\mathbf{H},\Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i\{H^{i+1}(x_{i+1}) - H^i(X_i)\}$$

is an orthogonal estimand.

For all $i = 1,2,\cdots,n$

$$\frac{\partial}{\partial H^i}\mathbb{E}[f(\mathbf{V};\eta)\,|_{\eta=\eta_0} = \mathbb{E}\left[\frac{\partial}{\partial H^i}\{W^i(H^{i+1} - H^i) + W^{i-1}(H^i - H^{i-1}) + \cdots\}\,|_{\eta=\eta_0}\right]$$

$$= \mathbb{E}[-W_0^{i+1} + W_0^i]$$

$$= 0$$

Debiasedness &
Doubly Robustness

Debiasedness & Doubly Robustness

$$\mathbb{E}[f^{DR}(\mathbf{V};\{\mathbf{H},\Pi\})] - C(P) = \sum_{i=1}^{n} O_P\left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \|\right)$$

A proof is omitted due to its complexness. Check [Rotnitzky, Robins, Babino, 2017] for the detailed proof.

Debiasedness & Doubly Robustness

$$\mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\})] - C(P) = \sum_{i=1}^{n} O_P \left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \| \right)$$

A proof is omitted due to its complexness. Check [Rotnitzky, Robins, Babino, 2017] for the detailed proof.

1. **Debiasedness**: $\mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\})]$ converges to the mSBD adjustment $C(P)$ at $N^{-1/2}$ rate if H^i, π^i converge to H_0^i, π_0^i at $N^{-1/4}$ rate.

Debiasedness & Doubly Robustness

$$\mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\})] - C(P) = \sum_{i=1}^{n} O_P\left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \|\right)$$

A proof is omitted due to its complexness. Check [Rotnitzky, Robins, Babino, 2017] for the detailed proof.

1. **Debiasedness**: $\mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\})]$ converges to the mSBD adjustment $C(P)$ at $N^{-1/2}$ rate if H^i, π^i converge to H_0^i, π_0^i at $N^{-1/4}$ rate.

2. **Doubly robustness**: $\mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\})] = C(P)$ if either $H^i = H_0^i$ or $\pi^i = \pi_0^i$ for all $i \in \{1, 2, \cdots, n\}$.

Finite samples

Finite samples

$$\mathbb{E}[f^{DR}(\mathbf{V}; \hat{\eta} = \{\hat{\mathbf{H}}, \hat{\Pi}\})] - C(P)$$

Finite samples

$$\mathbb{E}[f^{DR}(\mathbf{V}; \hat{\eta} = \{\hat{\mathbf{H}}, \hat{\Pi}\})] - C(P)$$

$$= \mathbb{E}_D[f(V;\hat{\eta})] - \mathbb{E}_P[f(V;\hat{\eta})] + \mathbb{E}_P[f(V;\hat{\eta})] - C(P)$$

Finite samples

$$\mathbb{E}[f^{DR}(\mathbf{V}; \hat{\eta} = \{\hat{\mathbf{H}}, \hat{\Pi}\})] - C(P)$$

$$= \mathbb{E}_D[f(V; \hat{\eta})] - \mathbb{E}_P[f(V; \hat{\eta})] + \mathbb{E}_P[f(V; \hat{\eta})] - C(P)$$

1

1 $$\mathbb{E}[f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\})] - C(P) = \sum_{i=1}^{n} O_P\left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \|\right)$$

Finite samples

$$\mathbb{E}[f^{DR}(\mathbf{V}; \hat{\eta} = \{\hat{\mathbf{H}}, \hat{\Pi}\})] - C(P)$$

$$= \mathbb{E}_D[f(V; \hat{\eta})] - \mathbb{E}_P[f(V; \hat{\eta})] + \mathbb{E}_P[f(V; \hat{\eta})] - C(P)$$

2 (first term) 1 (second term)

1 $$\mathbb{E}[f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\})] - C(P) = \sum_{i=1}^{n} O_P\left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \|\right)$$

2 $$\mathbb{E}_D[f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\})] - \mathbb{E}[f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\})] = O_P(N^{-1/2})$$

If samples for training η and evaluating (D) are independent (achieved by the ***sample-splitting***)

Finite samples

$$\mathbb{E}[f^{DR}(\mathbf{V}; \hat{\eta} = \{\hat{\mathbf{H}}, \hat{\Pi}\})] - C(P)$$

$$= \mathbb{E}_D[f(V; \hat{\eta})] - \mathbb{E}_P[f(V; \hat{\eta})] + \mathbb{E}_P[f(V; \hat{\eta})] - C(P)$$

2 1

$$= O_P(N^{-1/2}) + \sum_{i=1}^{n} O_P\left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \|\right)$$

1 $$\mathbb{E}[f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\})] - C(P) = \sum_{i=1}^{n} O_P\left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \|\right)$$

2 $$\mathbb{E}_D[f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\})] - \mathbb{E}[f(\mathbf{V}; \eta = \{\mathbf{H}, \Pi\})] = O_P(N^{-1/2})$$

If samples for training η and evaluating (D) are independent (achieved by the ***sample-splitting***)

Intermediate Summary - mSBD

Intermediate Summary - mSBD

So far,

Intermediate Summary - mSBD

So far,

1. (mSBD adjustment) We defined the mSBD adjustment, which is an extension of the BD adjustment.

Intermediate Summary - mSBD

So far,

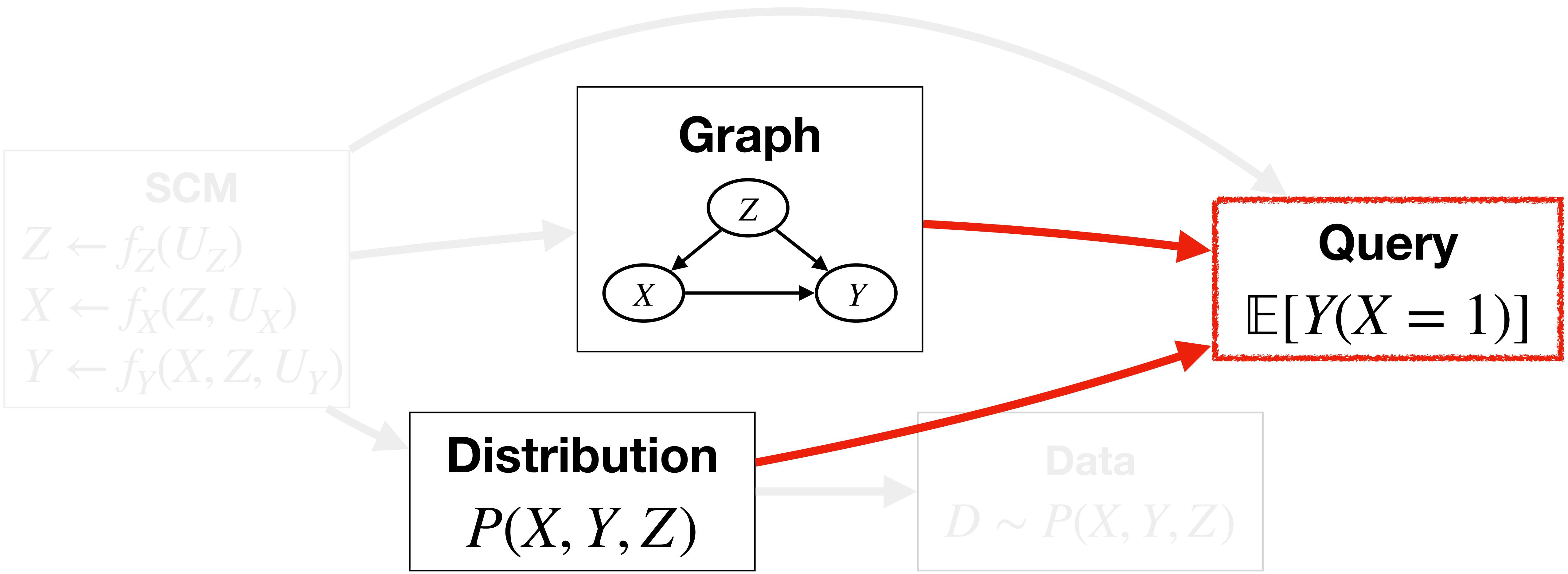
1. (mSBD adjustment) We defined the mSBD adjustment, which is an extension of the BD adjustment.

2. (Orthogonal Estimand) We defined the orthogonal estimand of the mSBD adjustment.

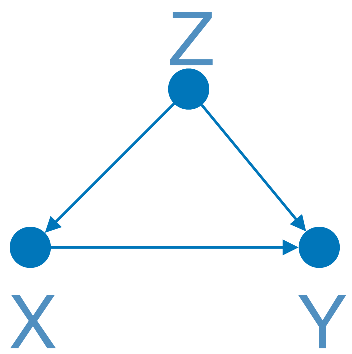
General case — Causal functional represented as a function of mSBDs

Algorithmic approach to Identifiability

Recap: Causal Effect Identification: Big Picture (1)



Recap: Causal Effect Identification



Causal graph (G)

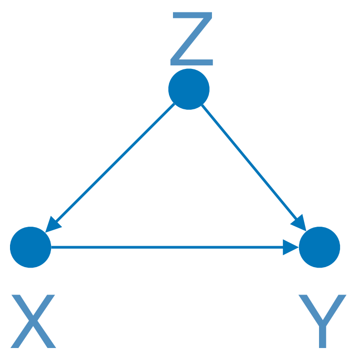
$P(Z, X, Y)$

Distribution on G (P)

$$Q_0 := \mathbb{E}[Y \mid do(x)]$$

Causal Query (Q_0)

Recap: Causal Effect Identification



Causal graph (G)

$$P(Z, X, Y)$$

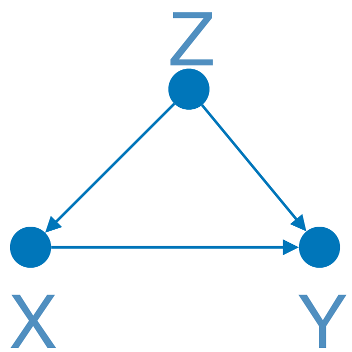
Distribution on G (P)

$$Q_0 := \mathbb{E}[Y \,|\, do(x)]$$

Causal Query (Q_0)

Given {G, P, Q_0},

Recap: Causal Effect Identification



Causal graph (G)

$P(Z, X, Y)$

Distribution on G (P)

$Q_0 := \mathbb{E}[Y \mid do(x)]$

Causal Query (Q_0)

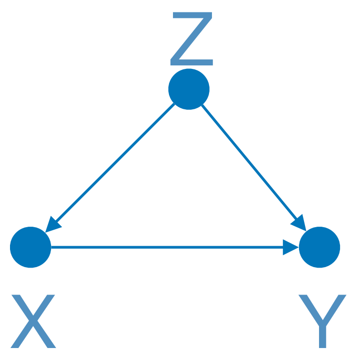
Given {G, P, Q_0},

the causal effect identification (ID) task

$\textbf{ID}(G, P, Q_0)$

ID algorithm

Recap: Causal Effect Identification



Causal graph (G)

$P(Z, X, Y)$

Distribution on G (P)

$Q_0 := \mathbb{E}[Y \mid do(x)]$

Causal Query (Q_0)

Given {G, P, Q_0},

the causal effect identification (ID) task

finds the functional $C(P)$ s.t. $C(P) = Q_0$.

ID(G, P, Q_0)

ID algorithm

$C(P) = \sum_z \mathbb{E}[Y \mid x, z]P(z)$

Causal functional

$C(P)$ **s.t.** $C(P) = Q_0$

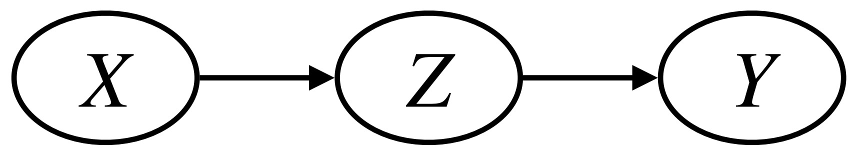
Conditional Independence and d-Separation

Conditional Independence and d-Separation

If $\{G, P\}$ are induced by the SCM, then, the conditional independence in P is represented as a d-separation in G.

Conditional Independence and d-Separation

If $\{G, P\}$ are induced by the SCM, then, the conditional independence in P is represented as a d-separation in G.



D-separation: The path from X to Y is blocked by Z;

$$(X \perp\!\!\!\perp Y \mid Z)_G$$

Conditional independence: Then, in the corresponding distribution P induced by the SCM,

$$(X \perp\!\!\!\perp Y \mid Z)_P\text{; i.e., } P(Y \mid X, Z) = P(Y \mid Z).$$

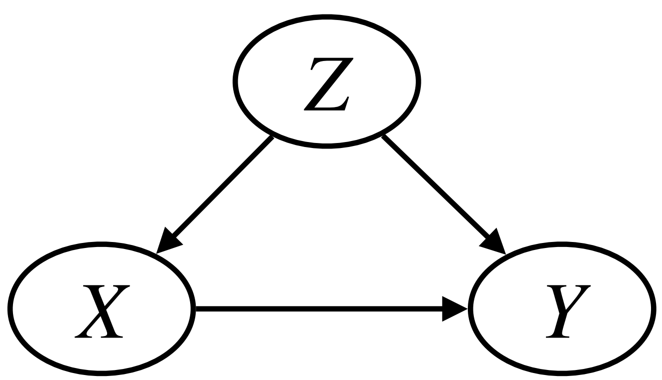
Intervention and Causal Graph $G_{\overline{X}}$

Intervention and Causal Graph $G_{\overline{X}}$

$G_{\overline{X}}$: Intervention $do(X=1)$ (replacing $X \leftarrow f_X$ to $X \leftarrow x$) induces the graph where the incoming edges to X is cut.

Intervention and Causal Graph $G_{\overline{X}}$

$G_{\overline{X}}$: Intervention $do(X = 1)$ (replacing $X \leftarrow f_X$ to $X \leftarrow x$) induces the graph where the incoming edges to X is cut.

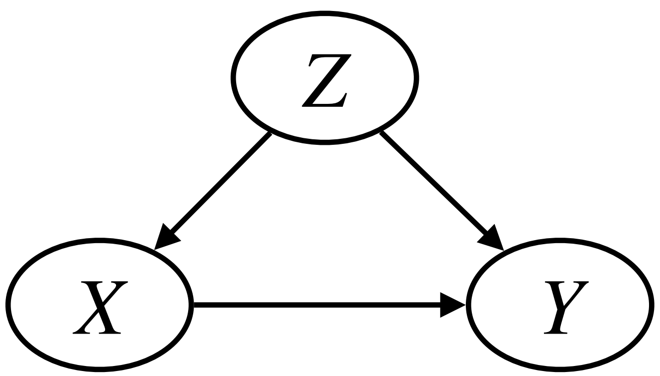


SCM

$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y \leftarrow f_Y(Z, X, U_Y)$$

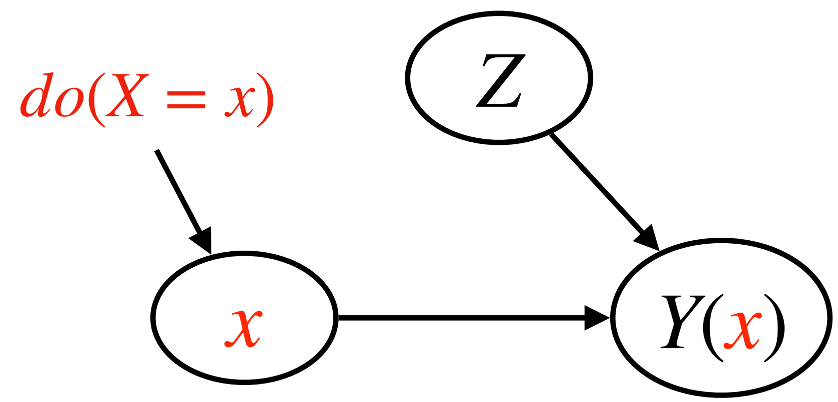
Intervention and Causal Graph $G_{\overline{X}}$

$G_{\overline{X}}$: Intervention $do(X=1)$ (replacing $X \leftarrow f_X$ to $X \leftarrow x$) induces the graph where the incoming edges to X is cut.



SCM

$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y \leftarrow f_Y(Z, X, U_Y)$$



SCM ($do(X=1)$)

$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow x$$
$$Y(x) \leftarrow f_Y(Z, X=1, U_Y)$$

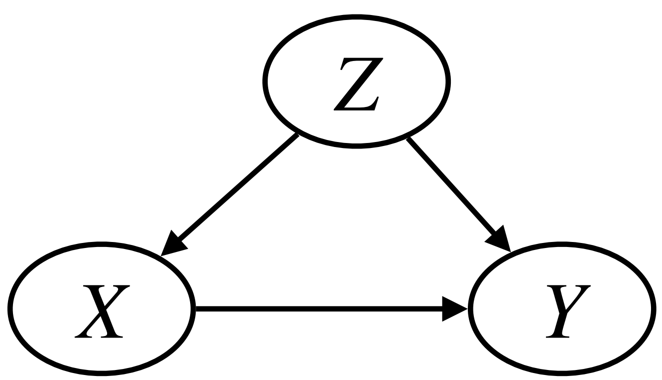
Counterfactual and Causal Graph $G_{\underline{X}}$

Counterfactual and Causal Graph $G_{\underline{X}}$

$G_{\underline{X}}$: The counterfactual variable $Y(x)$ can be expressed in the graph by cutting outgoing edges of the node X.

Counterfactual and Causal Graph $G_{\underline{X}}$

$G_{\underline{X}}$: The counterfactual variable $Y(x)$ can be expressed in the graph by cutting outgoing edges of the node X.



SCM

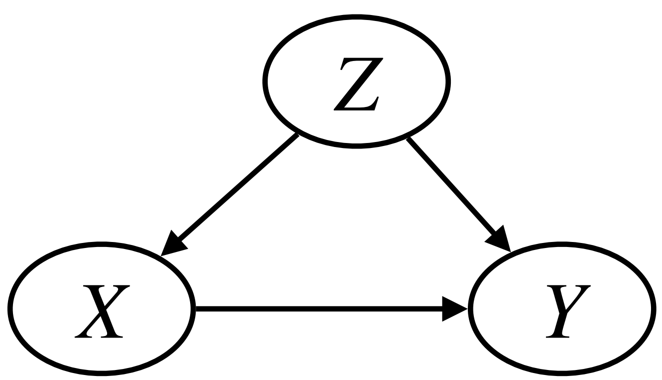
$Z \leftarrow f_Z(U_Z)$

$X \leftarrow f_X(Z, U_X)$

$Y \leftarrow f_Y(Z, X, U_Y)$

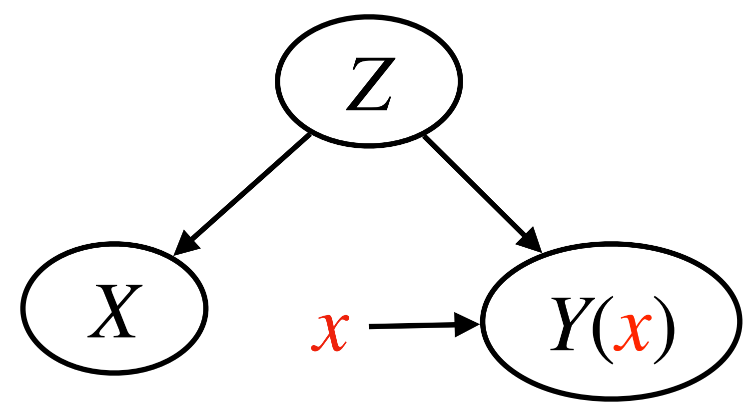
Counterfactual and Causal Graph $G_{\underline{X}}$

$G_{\underline{X}}$: The counterfactual variable $Y(x)$ can be expressed in the graph by cutting outgoing edges of the node X.



SCM

$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y \leftarrow f_Y(Z, X, U_Y)$$



SCM'

$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(U_X, Z)$$
$$Y(x) \leftarrow f_Y(Z, X = 1, U_Y)$$

do-Calculus - R1: Excluding Observation

do-Calculus - R1: Excluding Observation

Rule 1: If $(Y \perp\!\!\!\perp Z \,|\, X, W)_{G_{\overline{X}}}$, then

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

$G_{\overline{X}}$ is a graph for $P(\,\cdot\,|\, do(x))$.

do-Calculus - R1: Excluding Observation

Rule 1: If $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$, then

$$P(y \mid do(x), z, W) = P(y \mid do(x), w)$$

$G_{\overline{X}}$ is a graph for $P(\,\cdot \mid do(x))$.

$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$ means the conditional independence of Y, Z given W on $P(y \mid do(x))$.

do-Calculus - R2: Exchanging Intervention/Observation

do-Calculus - R2: Exchanging Intervention/Observation

Rule 2: If $(Y \perp\!\!\!\perp Z \,|\, X, W)_{G_{\overline{X}\underline{Z}}}$, then

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

$G_{\overline{X}\underline{Z}}$ is a graph for $P(Y(z), Z, W(z) \,|\, do(x))$.

do-Calculus - R2: Exchanging Intervention/Observation

Rule 2: If $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$, then

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

$G_{\overline{X}\underline{Z}}$ is a graph for $P(Y(z), Z, W(z) \mid do(x))$.

$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$ means the conditional independence of counterfactuals $Y(z), Z$ given W on $P(\,\cdot \mid do(x))$.

do-Calculus - R3: Excluding Intervention

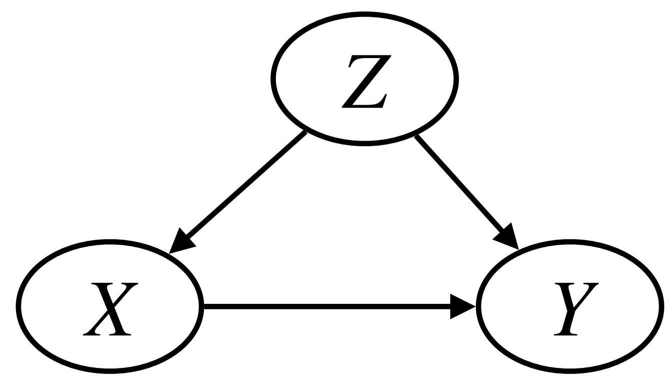
do-Calculus - R3: Excluding Intervention

Rule 3: If $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\overline{Z(W)}}}$, where $Z(W) := Z \backslash An(W)_{G_{\overline{X}}}$

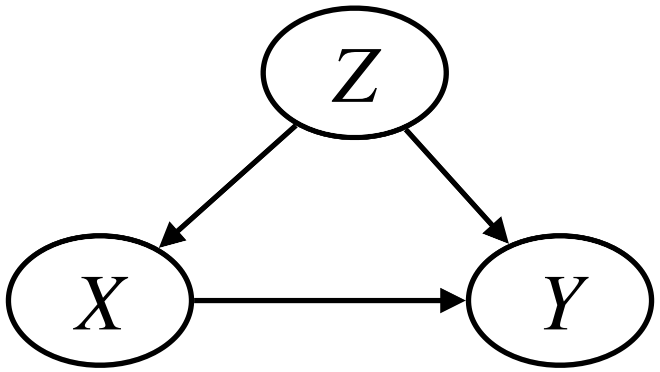
$$P(y \mid do(x), do(z), w) = P(y \mid do(x), w)$$

If (Y, Z) are conditionally independent after intervening $\{X, Z\}$, then Z is a redundant intervention, so it can be removed.

do-Calculus: Example to Back-door



do-Calculus:
Example to Back-door

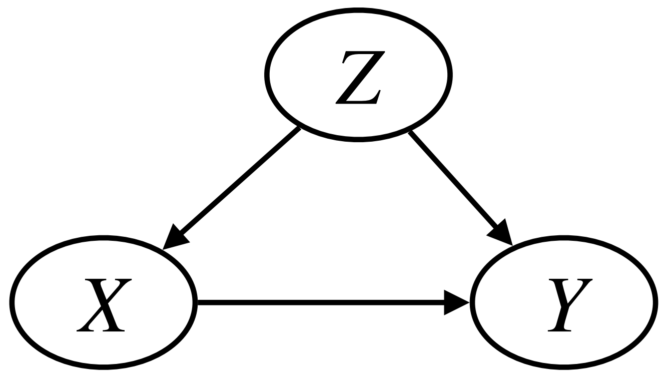


$$\mathbb{E}[Y \mid do(x)] = \sum_z \mathbb{E}[Y \mid do(x), z] P(z \mid do(x))$$ Marginalization over $Z = z$

$$= \sum_z \mathbb{E}[Y \mid do(x), z] P(z)$$

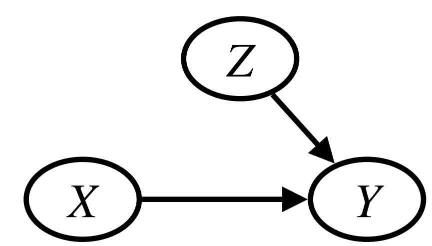
$$= \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

do-Calculus: Example to Back-door



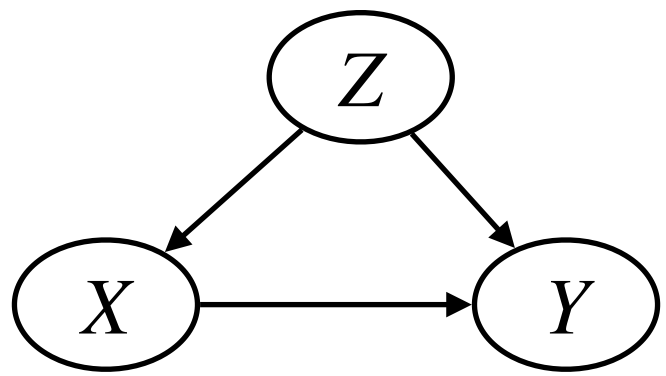
$$\mathbb{E}[Y \mid do(x)] = \sum_z \mathbb{E}[Y \mid do(x), z] P(z \mid do(x))$$ Marginalization over $Z = z$

$$= \sum_z \mathbb{E}[Y \mid do(x), z] P(z)$$ R3: $(Z \perp\!\!\!\perp X)_{G_{\overline{X}}}$



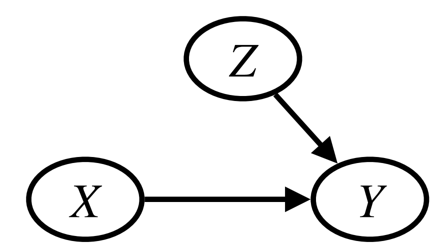
$$= \sum_z \mathbb{E}[Y \mid x, z] P(z)$$

do-Calculus: Example to Back-door

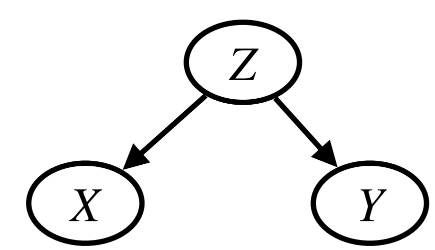


$$\mathbb{E}[Y \mid do(x)] = \sum_z \mathbb{E}[Y \mid do(x), z] P(z \mid do(x))$$ Marginalization over $Z = z$

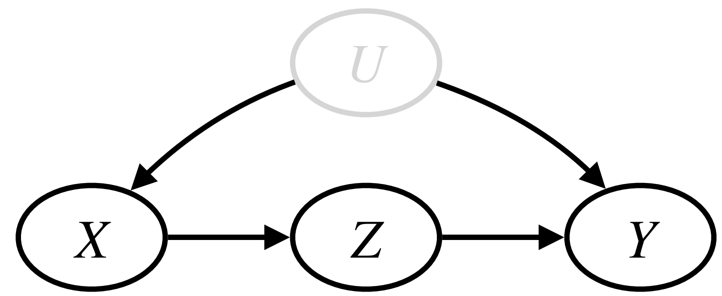
$$= \sum_z \mathbb{E}[Y \mid do(x), z] P(z)$$ R3: $(Z \perp\!\!\!\perp X)_{G_{\overline{X}}}$



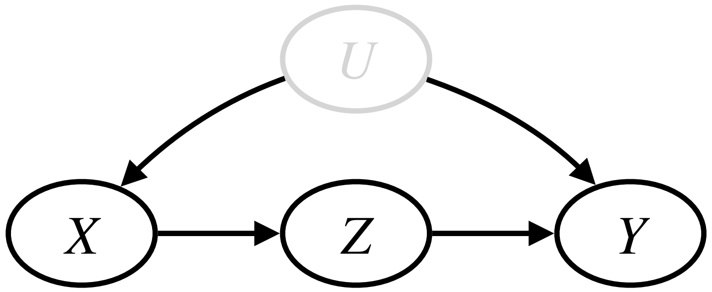
$$= \sum_z \mathbb{E}[Y \mid x, z] P(z)$$ R2: $(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}}$



do-Calculus: Example to Front-door



do-Calculus: Example to Front-door



$$\mathbb{E}[Y \mid do(x)] = \sum_z P(z \mid do(x))\mathbb{E}[Y \mid do(x), z]$$

Marginalization over $Z = z$

$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(x), z]$$

$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(x, z)]$$

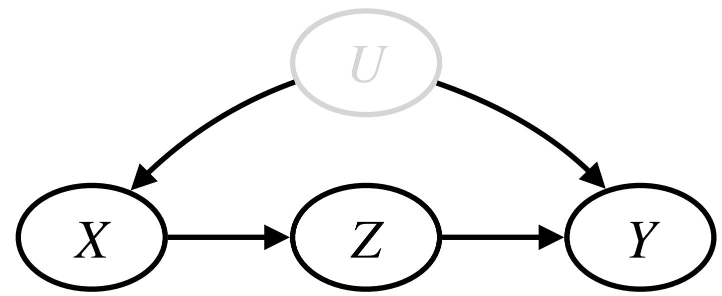
$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(z)]$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)]P(x' \mid do(z))$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)]P(x')$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', z]P(x')$$

do-Calculus: Example to Front-door

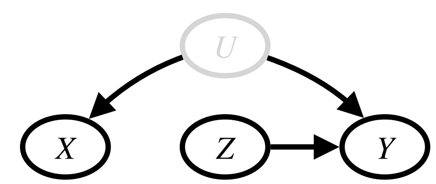


$$\mathbb{E}[Y \mid do(x)] = \sum_z P(z \mid do(x))\mathbb{E}[Y \mid do(x), z]$$

Marginalization over $Z = z$

$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(x), z]$$

R2: $(Z \perp\!\!\!\perp X)_{G_{\underline{X}}}$



$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(x, z)]$$

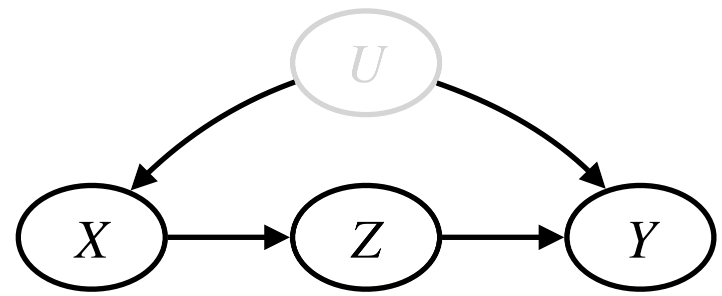
$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(z)]$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)]P(x' \mid do(z))$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)]P(x')$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', z]P(x')$$

do-Calculus: Example to Front-door

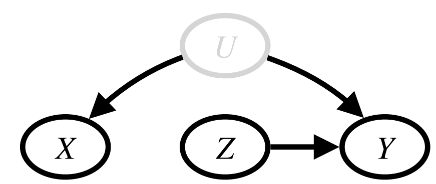


$$\mathbb{E}[Y \mid do(x)] = \sum_z P(z \mid do(x)) \mathbb{E}[Y \mid do(x), z]$$

Marginalization over $Z = z$

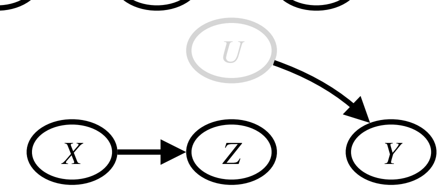
$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(x), z]$$

R2: $(Z \perp\!\!\!\perp X)_{G_{\underline{X}}}$



$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(x, z)]$$

R3: $(Y \perp\!\!\!\perp Z \mid X)_{G_{\overline{X}\underline{Z}}}$



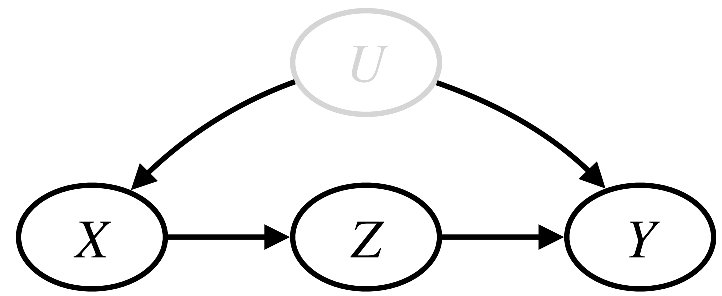
$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(z)]$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)] P(x' \mid do(z))$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)] P(x')$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', z] P(x')$$

do-Calculus: Example to Front-door

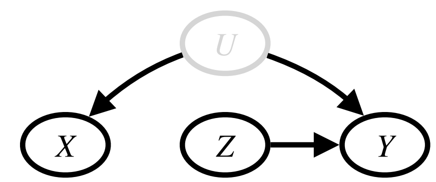


$$\mathbb{E}[Y \mid do(x)] = \sum_z P(z \mid do(x))\mathbb{E}[Y \mid do(x), z]$$

Marginalization over $Z = z$

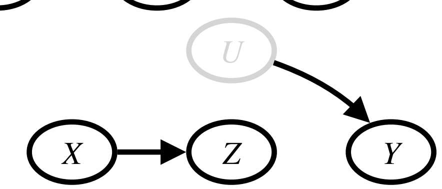
$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(x), z]$$

R2: $(Z \perp\!\!\!\perp X)_{G_{\underline{X}}}$



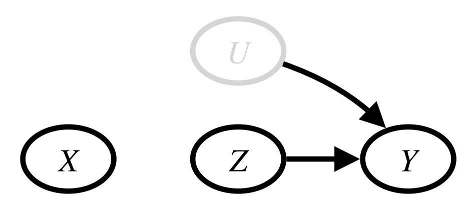
$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(x, z)]$$

R3: $(Y \perp\!\!\!\perp Z \mid X)_{G_{\overline{X}\underline{Z}}}$



$$= \sum_z P(z \mid x)\mathbb{E}[Y \mid do(z)]$$

R3: $(Y \perp\!\!\!\perp X \mid Z)_{G_{\overline{X}\overline{Z}}}$



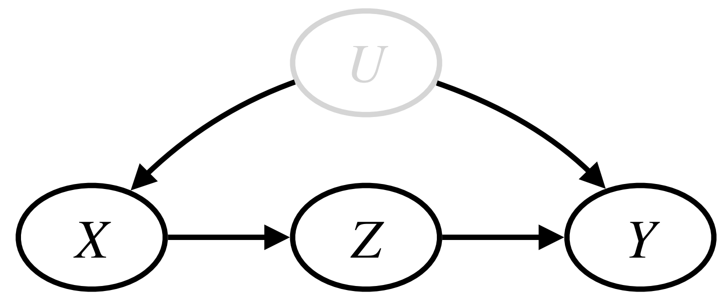
$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)]P(x' \mid do(z))$$

Marginalization over $X = x$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)]P(x')$$

$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', z]P(x')$$

do-Calculus: Example to Front-door

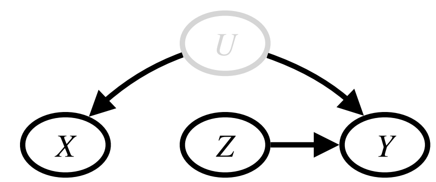


$$\mathbb{E}[Y \mid do(x)] = \sum_z P(z \mid do(x)) \mathbb{E}[Y \mid do(x), z]$$

Marginalization over $Z = z$

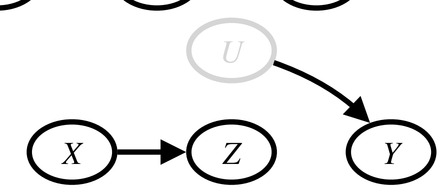
$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(x), z]$$

R2: $(Z \perp\!\!\!\perp X)_{G_{\underline{X}}}$



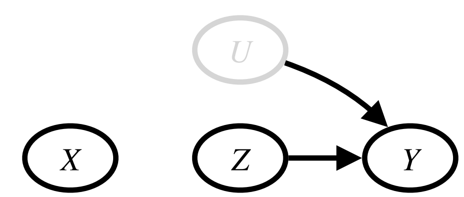
$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(x, z)]$$

R3: $(Y \perp\!\!\!\perp Z \mid X)_{G_{\overline{X}\underline{Z}}}$



$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(z)]$$

R3: $(Y \perp\!\!\!\perp X \mid Z)_{G_{\overline{X}\overline{Z}}}$

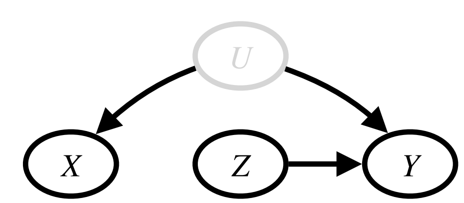


$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)] P(x' \mid do(z))$$

Marginalization over $X = x$

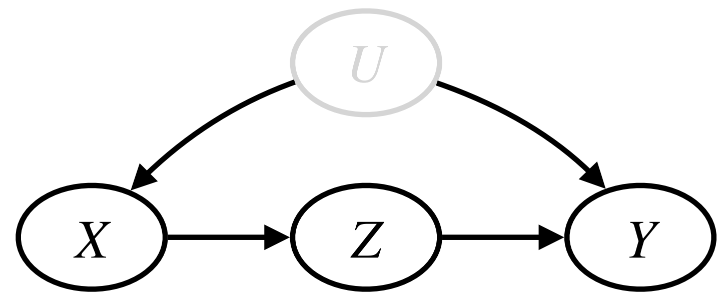
$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)] P(x')$$

R3: $(X \perp\!\!\!\perp Z)_{G_{\overline{Z}}}$



$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', z] P(x')$$

do-Calculus: Example to Front-door

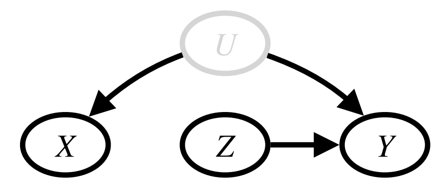


$$\mathbb{E}[Y \mid do(x)] = \sum_z P(z \mid do(x)) \mathbb{E}[Y \mid do(x), z]$$

Marginalization over $Z = z$

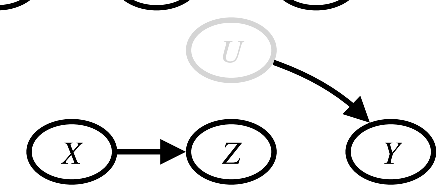
$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(x), z]$$

R2: $(Z \perp\!\!\!\perp X)_{G_{\underline{X}}}$



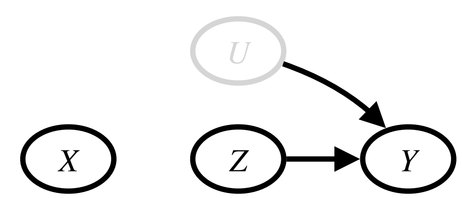
$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(x, z)]$$

R3: $(Y \perp\!\!\!\perp Z \mid X)_{G_{\overline{X}\underline{Z}}}$



$$= \sum_z P(z \mid x) \mathbb{E}[Y \mid do(z)]$$

R3: $(Y \perp\!\!\!\perp X \mid Z)_{G_{\overline{X}\overline{Z}}}$

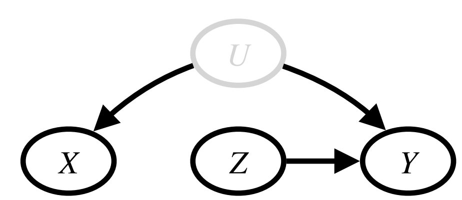


$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)] P(x' \mid do(z))$$

Marginalization over $X = x$

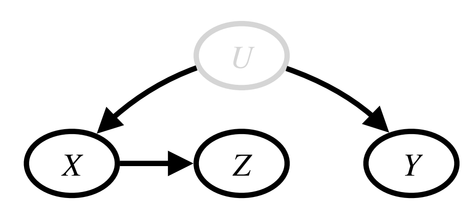
$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', do(z)] P(x')$$

R3: $(X \perp\!\!\!\perp Z)_{G_{\overline{Z}}}$



$$= \sum_z P(z \mid x) \sum_{x'} \mathbb{E}[Y \mid x', z] P(x')$$

R2: $(Y \perp\!\!\!\perp Z \mid X)_{G_{\underline{Z}}}$



do-Calculus:
Soundness and Completeness

do-Calculus: Soundness and Completeness

The causal effect is identifiable (i.e., $\mathbb{E}[Y | do(x)]$ can be written as a function $C(P)$) if-and-only-if there is a series of do-calculus application.

do-Calculus: Algorithmic Procedure

do-Calculus: Algorithmic Procedure

There is an algorithm (e.g., CausalFusion) that finds such rules for identifying causality.

Soundness and Completeness of DML-ID

Soundness and Completeness of DML-ID

Soundness and Completeness of DML-ID

Soundness and Completeness of DML-ID

Soundness and Completeness of DML-ID

A causal effect $P_{\mathbf{x}}(\mathbf{y})$ is identifiable if-and-only-if It can be written as a function of mSBD adjustments; i.e.,

Soundness and Completeness of DML-ID

Soundness and Completeness of DML-ID

A causal effect $P_{\mathbf{x}}(\mathbf{y})$ is identifiable if-and-only-if It can be written as a function of mSBD adjustments; i.e.,

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\})$$

Soundness and Completeness of DML-ID

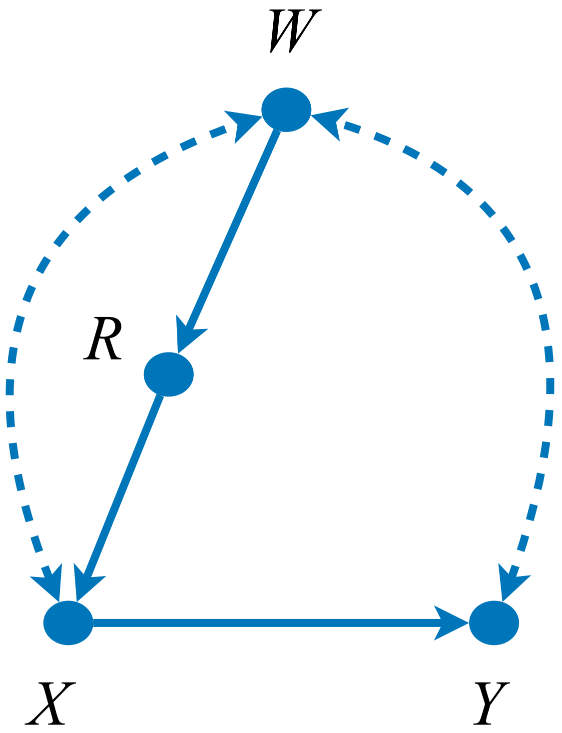
Soundness and Completeness of DML-ID

A causal effect $P_{\mathbf{x}}(\mathbf{y})$ is identifiable if-and-only-if It can be written as a function of mSBD adjustments; i.e.,

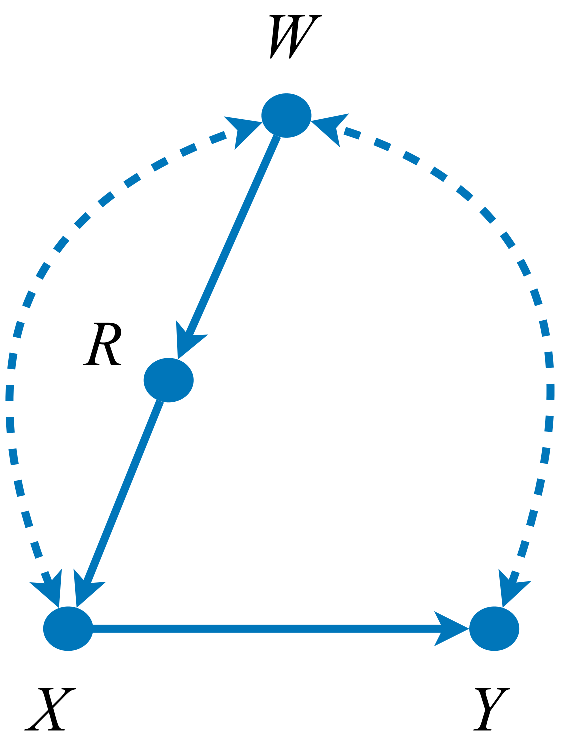
$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\})$$

for a function M^a.

Example of DML-ID: Napkin



Example of DML-ID: Napkin



$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)} = \frac{M^b}{M^a}$$

Construction of DML estimator

What we learned so far

What we learned so far

mSBD adjustment:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{Y_i \in \mathbf{Y}} P(y_i \mid \mathbf{x}^{(i)}, \mathbf{z}^{(i)}, \mathbf{y}^{(i-1)}) \prod_{Z_i \in \mathbf{Z}} P(z_i \mid \mathbf{x}^{(i-1)}, \mathbf{z}^{(i-1)}, \mathbf{y}^{(i-1)})$$

What we learned so far

mSBD adjustment:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{Y_i \in \mathbf{Y}} P(y_i \mid \mathbf{x}^{(i)}, \mathbf{z}^{(i)}, \mathbf{y}^{(i-1)}) \prod_{Z_i \in \mathbf{Z}} P(z_i \mid \mathbf{x}^{(i-1)}, \mathbf{z}^{(i-1)}, \mathbf{y}^{(i-1)})$$

Orthogonal (and Doubly-Robust) estimand for mSBD:

$$f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i \{H^{i+1}(x_{i+1}) - H^i(X_i)\},$$

What we learned so far

mSBD adjustment:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{Y_i \in \mathbf{Y}} P(y_i \mid \mathbf{x}^{(i)}, \mathbf{z}^{(i)}, \mathbf{y}^{(i-1)}) \prod_{Z_i \in \mathbf{Z}} P(z_i \mid \mathbf{x}^{(i-1)}, \mathbf{z}^{(i-1)}, \mathbf{y}^{(i-1)})$$

Orthogonal (and Doubly-Robust) estimand for mSBD:

$$f^{DR}(\mathbf{V}; \{\mathbf{H}, \Pi\}) = H^1(x_1) + \sum_{i=1}^{n} W^i \{H^{i+1}(x_{i+1}) - H^i(X_i)\},$$

Result of DML-ID: Any identifiable causal effect can be represented as an arithmetic function of mSBDs:

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\}).$$

Deriving an Orthogonal Estimand - 1

Result of DML-ID: Any identifiable causal effect can be represented as an arithmetic function of mSBDs:

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\})$$

Deriving an Orthogonal Estimand - 1

Result of DML-ID: Any identifiable causal effect can be represented as an arithmetic function of mSBDs:

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\})$$

$A(\cdot)$ is an arithmetic operator

M_0^a is an individual mSBD adjustment that composes the ID expression of $P_{\mathbf{x}}(\mathbf{y})$

Deriving an Orthogonal Estimand - 1

Result of DML-ID: Any identifiable causal effect can be represented as an arithmetic function of mSBDs:

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\})$$

$A(\,\cdot\,)$ is an arithmetic operator

M_0^a is an individual mSBD adjustment that composes the ID expression of $P_{\mathbf{x}}(\mathbf{y})$

For an mSBD operator M, let $\mu_{M_0} := \mathbb{E}[f^{DR}(V; \{\mathbf{H}_0, \Pi_0\})]$ denote an expectation of the DR (mSBD) estimand with true nuisances $\mathbf{H}_0, \Pi_0$.

Deriving an Orthogonal Estimand - 2

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\}_{a=1}^m)$$

Deriving an Orthogonal Estimand - 2

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\}_{a=1}^m)$$

Recall that, for a mSBD adjustment M, $\mu_{M_0} := \mathbb{E}[f(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\})] = M$

Deriving an Orthogonal Estimand - 2

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\}_{a=1}^m)$$

Recall that, for a mSBD adjustment M, $\mu_{M_0} := \mathbb{E}[f(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\})] = M$

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{\mu_{M_0^a}\}_{a=1}^m)$$

Deriving an Orthogonal Estimand - 2

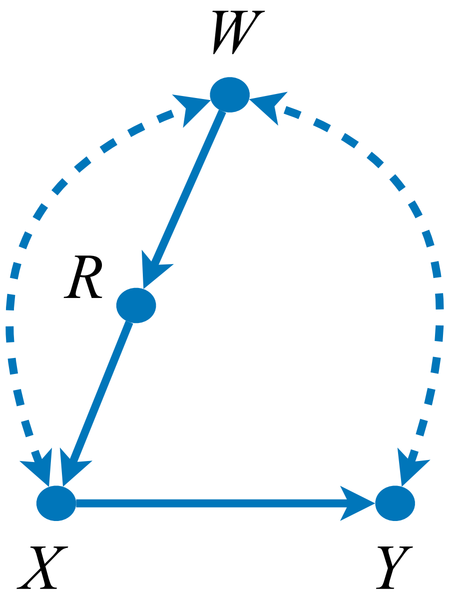
$$P_{\mathbf{x}}(\mathbf{y}) = A(\{M^a\}_{a=1}^m)$$

Recall that, for a mSBD adjustment M, $\mu_{M_0} := \mathbb{E}[f(\mathbf{V}; \{\mathbf{H}_0, \mathbf{W}_0\})] = M$

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{\mu_{M_0^a}\}_{a=1}^m)$$

$P_{\mathbf{x}}(\mathbf{y})$ can be represented as an expectation of the DR estimands of the mSBD adjustments.

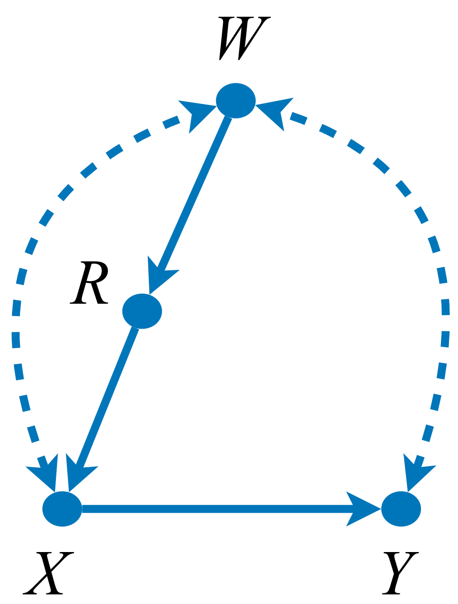
Example for deriving an Orthogonal Estimand - 1



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

Example for deriving an Orthogonal Estimand - 1

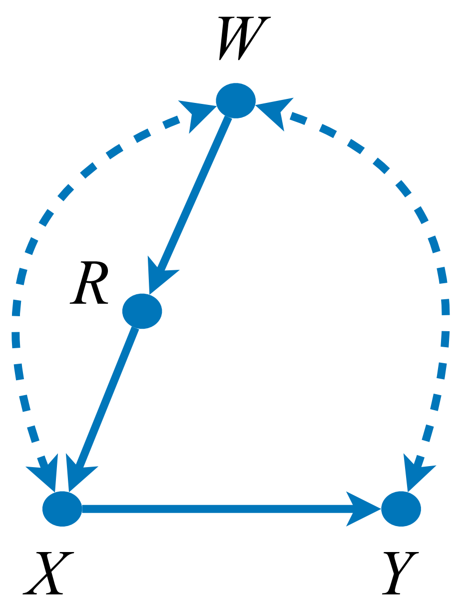


$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^a = \mu_{M_0^a} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^a, \Pi_0\})]$, where

Example for deriving an Orthogonal Estimand - 1



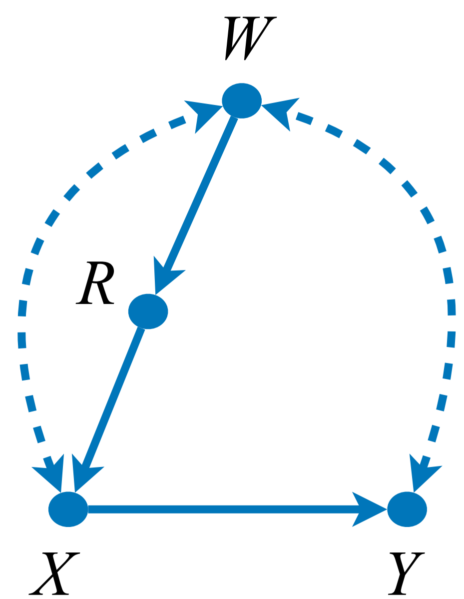
$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y\,|\,r;w]}{M[x\,|\,r;w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^a = \mu_{M_0^a} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^a, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\}) = \frac{I_r(R)}{\pi(R\,|\,W)}\left(I_{x,y}(X,Y) - H^a(R)\right) + H^a(r)$$

Example for deriving an Orthogonal Estimand - 1



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r;w]}{M[x \mid r;w]}$$

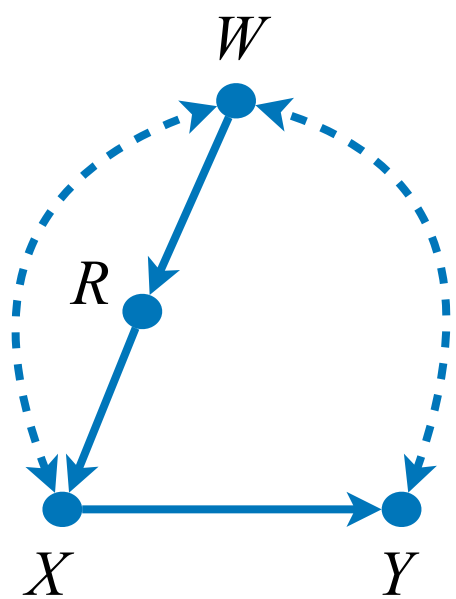
$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^a = \mu_{M_0^a} := \mathbb{E}[f^{DR}(\mathbf{V};\{\mathbf{H}_0^a, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V};\{\mathbf{H}^a, \Pi\}) = \frac{I_r(R)}{\pi(R \mid W)}\Big(I_{x,y}(X,Y) - H^a(R)\Big) + H^a(r)$$

$$H_0^a(R) := \mathbb{E}[I_{x,y}(X,Y) \mid R, W],$$

Example for deriving an Orthogonal Estimand - 1



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y\,|\,r;w]}{M[x\,|\,r;w]}$$

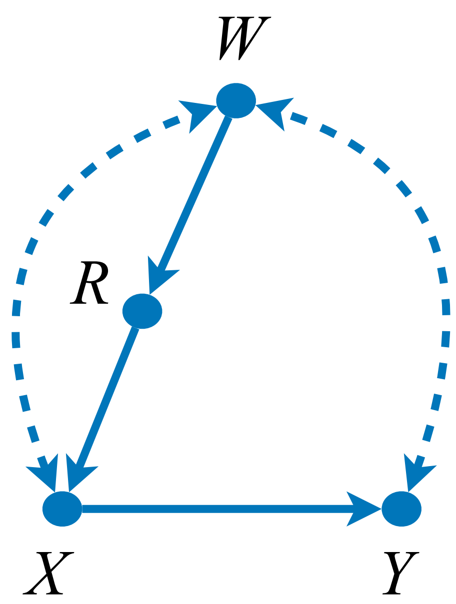
$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^a = \mu_{M_0^a} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^a, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\}) = \frac{I_r(R)}{\pi(R\,|\,W)}\Big(I_{x,y}(X,Y) - H^a(R)\Big) + H^a(r)$$

$$H_0^a(R) := \mathbb{E}[I_{x,y}(X,Y)\,|\,R,W], \quad H_0^a(r) := \mathbb{E}[I_{x,y}(X,Y)\,|\,r,W],$$

Example for deriving an Orthogonal Estimand - 1



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

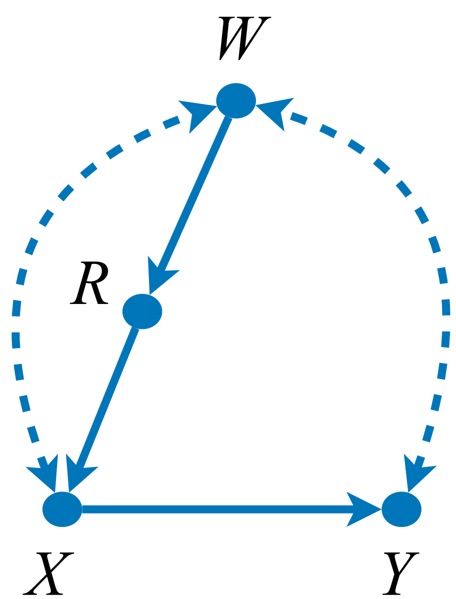
$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^a = \mu_{M_0^a} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^a, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\}) = \frac{I_r(R)}{\pi(R \mid W)} \Big(I_{x,y}(X,Y) - H^a(R) \Big) + H^a(r)$$

$$H_0^a(R) := \mathbb{E}[I_{x,y}(X,Y) \mid R, W], \quad H_0^a(r) := \mathbb{E}[I_{x,y}(X,Y) \mid r, W], \quad \pi_0(R \mid W) := P(R \mid W)$$

Example for deriving an Orthogonal Estimand - 1



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r;w]}{M[x \mid r;w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

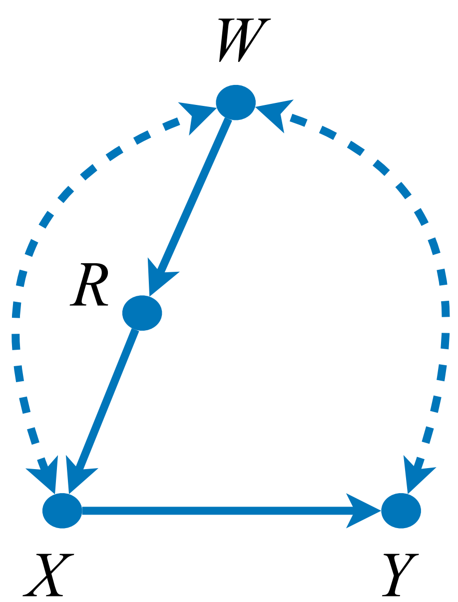
$M^a = \mu_{M_0^a} := \mathbb{E}[f^{DR}(\mathbf{V};\{\mathbf{H}_0^a, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V};\{\mathbf{H}^a,\Pi\}) = \frac{I_r(R)}{\pi(R \mid W)}\left(I_{x,y}(X,Y) - H^a(R)\right) + H^a(r)$$

$$H_0^a(R) := \mathbb{E}[I_{x,y}(X,Y) \mid R, W], \quad H_0^a(r) := \mathbb{E}[I_{x,y}(X,Y) \mid r, W], \quad \pi_0(R \mid W) := P(R \mid W)$$

$$\mu_{M_0^a} := \mathbb{E}[f^a(\mathbf{V};\{\mathbf{H}_0^a, \Pi_0\})] = M^a$$

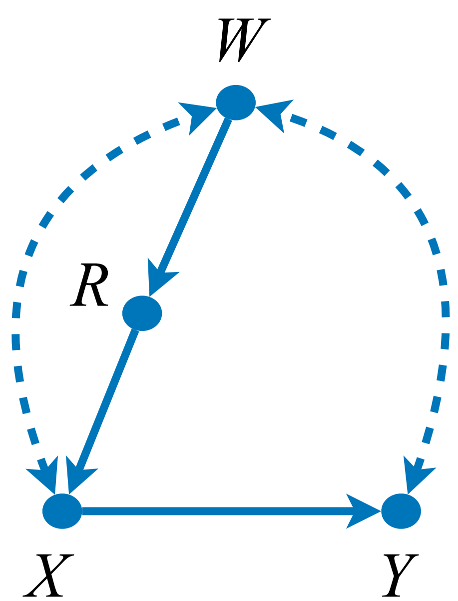
Example for deriving an Orthogonal Estimand - 2



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y\,|\,r;w]}{M[x\,|\,r;w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

Example for deriving an Orthogonal Estimand - 2

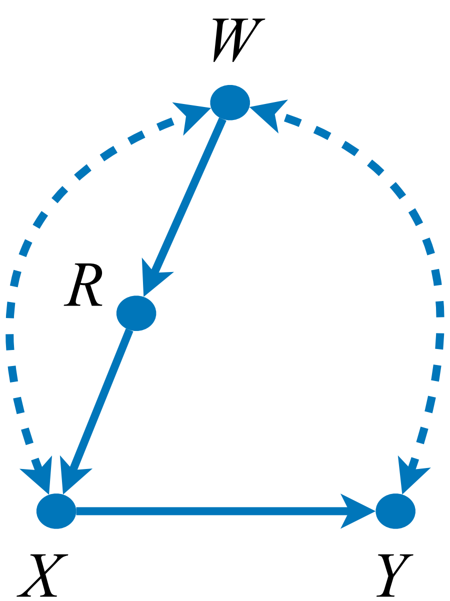


$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^b = \mu_{M_0^b} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^b, \Pi_0\})]$, where

Example for deriving an Orthogonal Estimand - 2



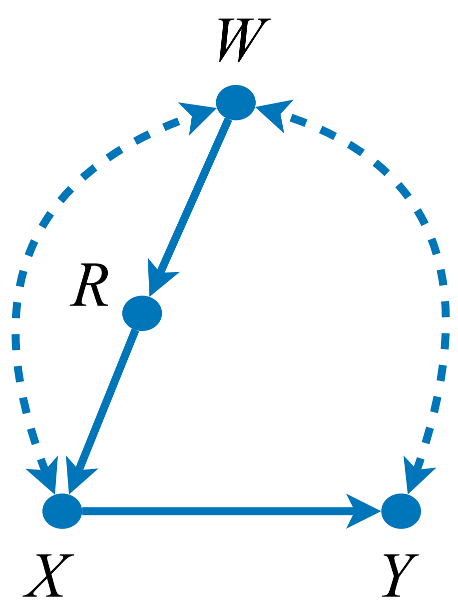
$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^b = \mu_{M_0^b} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^b, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\}) = \frac{I_r(R)}{\pi(R \mid W)} \left(I_x(X) - H^b(R)\right) + H^b(r)$$

Example for deriving an Orthogonal Estimand - 2



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

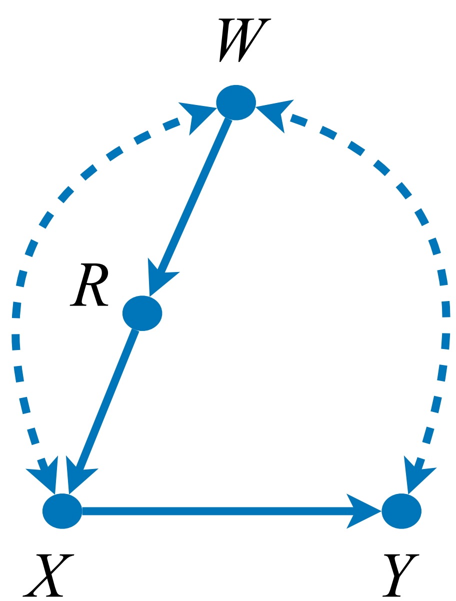
$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^b = \mu_{M_0^b} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^b, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\}) = \frac{I_r(R)}{\pi(R \mid W)} \left(I_x(X) - H^b(R)\right) + H^b(r)$$

$H_0^b(R) := \mathbb{E}[I_x(X) \mid R, W]$,

Example for deriving an Orthogonal Estimand - 2



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x, y \,|\, r; w]}{M[x \,|\, r; w]}$$

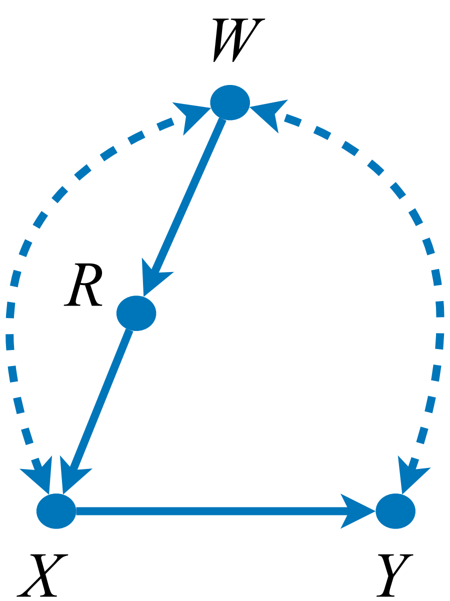
$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^b = \mu_{M_0^b} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^b, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\}) = \frac{I_r(R)}{\pi(R \,|\, W)} \left(I_x(X) - H^b(R)\right) + H^b(r)$$

$$H_0^b(R) := \mathbb{E}[I_x(X) \,|\, R, W], \qquad H_0^b(r) := \mathbb{E}[I_x(X) \,|\, r, W],$$

Example for deriving an Orthogonal Estimand - 2



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y\,|\,r;w]}{M[x\,|\,r;w]}$$

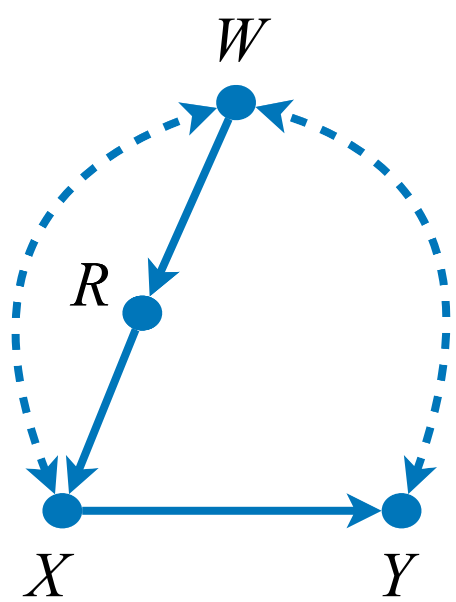
$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$M^b = \mu_{M_0^b} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^b, \Pi_0\})]$, where

$$f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\}) = \frac{I_r(R)}{\pi(R\,|\,W)} \left(I_x(X) - H^b(R)\right) + H^b(r)$$

$$H_0^b(R) := \mathbb{E}[I_x(X)\,|\,R, W], \qquad H_0^b(r) := \mathbb{E}[I_x(X)\,|\,r, W], \qquad \pi_0(R\,|\,W) := P(R\,|\,W)$$

Example for deriving an Orthogonal Estimand - 2



$$Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

$$= A(M^1, M^2) = \frac{M^a}{M^b}$$

$$M^b = \mu_{M_0^b} := \mathbb{E}[f^{DR}(\mathbf{V}; \{\mathbf{H}_0^b, \Pi_0\})]$$, where

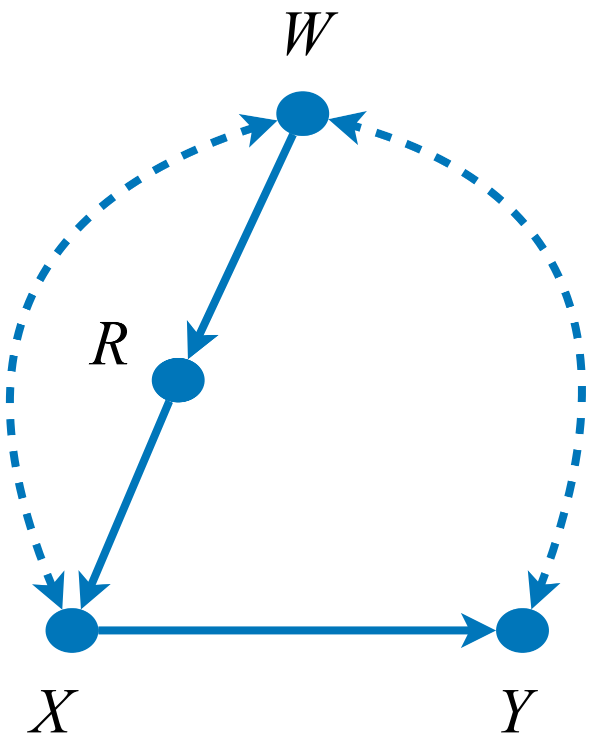
$$f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\}) = \frac{I_r(R)}{\pi(R \mid W)} \left(I_x(X) - H^b(R)\right) + H^b(r)$$

$$H_0^b(R) := \mathbb{E}[I_x(X) \mid R, W], \qquad H_0^b(r) := \mathbb{E}[I_x(X) \mid r, W], \qquad \pi_0(R \mid W) := P(R \mid W)$$

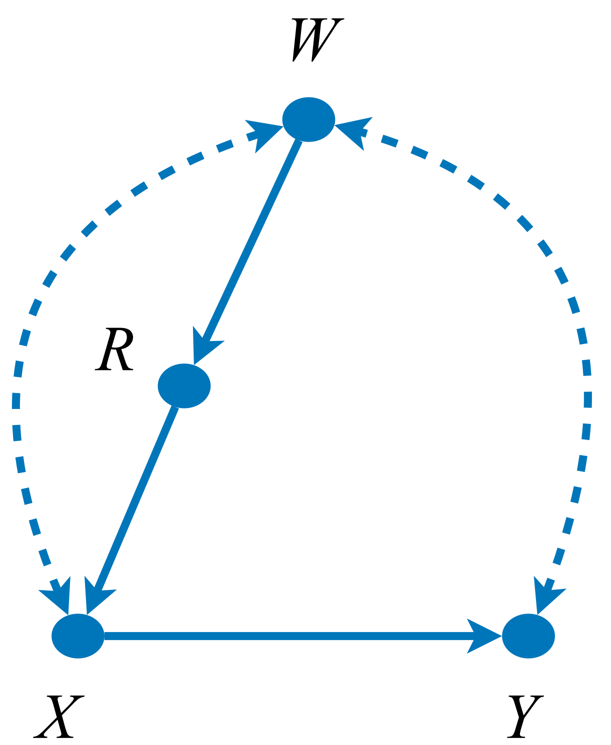
$$\mu_{M_0^b} := \mathbb{E}[f^b(\mathbf{V}; \{\eta_0\})] = M^b$$

Example for deriving an Orthogonal Estimand - 3

Example for deriving an Orthogonal Estimand - 3

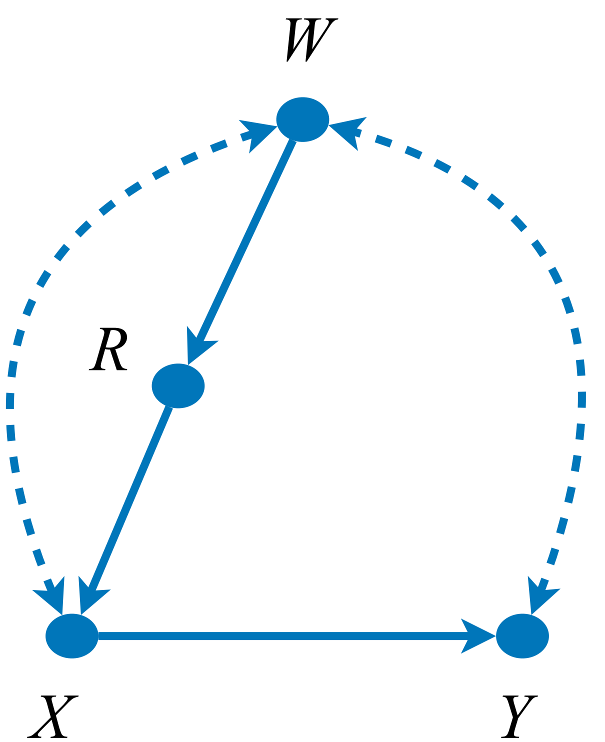


Example for deriving an Orthogonal Estimand - 3



$$P_x(y) = Q[Y] = \frac{Q[X, Y]}{Q[Y]} = \frac{M[x, y \mid r; w]}{M[x \mid r; w]}$$

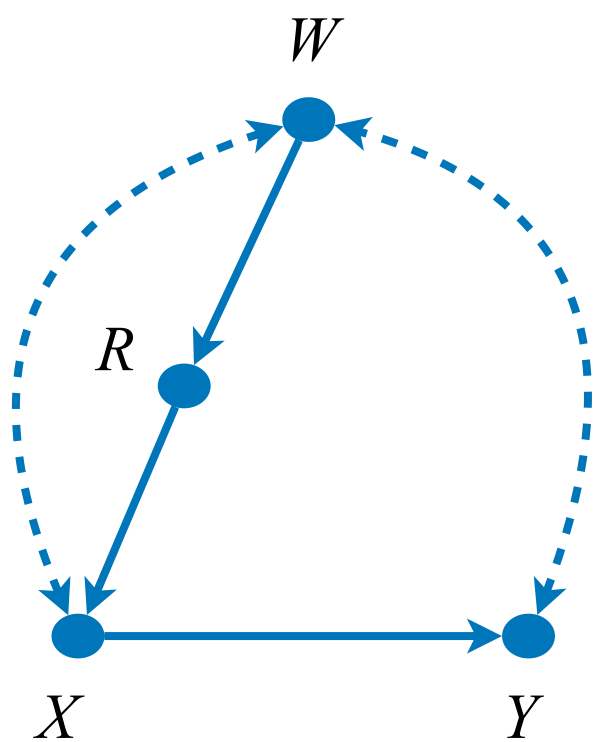
Example for deriving an Orthogonal Estimand - 3



$$P_x(y) = Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r; w]}{M[x \mid r; w]}$$

$$P_x(y) = A(M^a, M^b) = \frac{M^a}{M^b}$$

Example for deriving an Orthogonal Estimand - 3



$$P_x(y) = Q[Y] = \frac{Q[X,Y]}{Q[Y]} = \frac{M[x,y \mid r;w]}{M[x \mid r;w]}$$

$$P_x(y) = A(M^a, M^b) = \frac{M^a}{M^b}$$

$$= A(\mu_{M^a}, \mu_{M^b}) = \frac{\mu_{M_0^a}}{\mu_{M_0^b}}$$

Constructing DML estimators

Constructing DML estimators

Given

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{\mu_{M_0^a}\}_{a=1}^{m}),$$

the DML estimator is given as follow:

Constructing DML estimators

Given

$$P_{\mathbf{x}}(\mathbf{y}) = A(\{\mu_{M_0^a}\}_{a=1}^{m}),$$

the DML estimator is given as follow:

$$T = A(\{\hat{\mu}_{M^a}\}_{a=1}^{m}),$$

for $\hat{\mu}_{M^a} := \mathbb{E}_D\left[f^{DR}(\mathbf{V}; \{\hat{\mathbf{H}}, \hat{\mathbf{W}}\})\right]$ s.t. $\hat{\eta}$ are trained using samples independent to D

Error Analysis - 1

$$T = A(\{\hat{\mu}_{M^a}\}_{a=1}^{m}) \text{ for } \hat{\mu}_{M^a} := \mathbb{E}_D\left[f^{DR}(\mathbf{V}; \{\hat{\mathbf{H}}, \hat{\mathbf{W}}\})\right]$$

Error Analysis - 1

$$T = A(\{\hat{\mu}_{M^a}\}_{a=1}^{m}) \text{ for } \hat{\mu}_{M^a} := \mathbb{E}_D \left[f^{DR}(\mathbf{V}; \{\hat{\mathbf{H}}, \hat{\mathbf{W}}\}) \right]$$

$$T - C(P) = O_P(N^{-1/2}) + O_P\left(\sum_{a=1}^{m} \epsilon(\hat{\mu}_{M^a}) \right)$$

$$\text{for } \epsilon\left(\hat{\mu}_{M^a}\right) := \sum_{i=1}^{n} \| H^i - H_0^i \| \| \pi^i - \pi_0^i \|,$$

Error Analysis - 1

$$T = A(\{\hat{\mu}_{M^a}\}_{a=1}^{m}) \text{ for } \hat{\mu}_{M^a} := \mathbb{E}_D\left[f^{DR}(\mathbf{V}; \{\hat{\mathbf{H}}, \hat{\mathbf{W}}\})\right]$$

$$T - C(P) = O_P(N^{-1/2}) + O_P\left(\sum_{a=1}^{m} \epsilon(\hat{\mu}_{M^a})\right)$$

$$\text{for } \epsilon\left(\hat{\mu}_{M^a}\right) := \sum_{i=1}^{n} \| H^i - H_0^i \| \| \pi^i - \pi_0^i \|,$$

(where $\{H^i, \pi^i\}_{i=1}^{n}$ are nuisances for $\hat{\mu}_{M^a}$)

Error Analysis - 2

$$T - C(P) = O_P(N^{-1/2}) + O_P\left(\sum_{a=1}^{m} \epsilon(\hat{\mu}_{M^a})\right)$$

for $\epsilon\left(\hat{\mu}_{M^a}\right) := \sum_{i=1}^{n} \| H^i - H_0^i \| \| \pi^i - \pi_0^i \|,$

(where $\{H^i, \pi^i\}_{i=1}^{n}$ are nuisances for $\hat{\mu}_{M^a}$)

Error Analysis - 2

$$T - C(P) = O_P(N^{-1/2}) + O_P\left(\sum_{a=1}^{m} \epsilon(\hat{\mu}_{M^a})\right)$$

for $\epsilon\left(\hat{\mu}_{M^a}\right) := \sum_{i=1}^{n} \| H^i - H_0^i \| \| \pi^i - \pi_0^i \|,$

(where $\{H^i, \pi^i\}_{i=1}^n$ are nuisances for $\hat{\mu}_{M^a}$)

1. **Debiasedness**: T converges to $C(P)$ at $N^{-1/2}$ rate if all nuisances for μ_{M^a} (i.e., $\{H^i, \pi^i\}_{i=1}^n$) converge at $N^{-1/4}$ rate.

Error Analysis - 2

$$T - C(P) = O_P(N^{-1/2}) + O_P\left(\sum_{a=1}^{m} \epsilon(\hat{\mu}_{M^a})\right)$$

$$\text{for } \epsilon\left(\hat{\mu}_{M^a}\right) := \sum_{i=1}^{n} \| H^i - H_0^i \| \| \pi^i - \pi_0^i \|,$$

(where $\{H^i, \pi^i\}_{i=1}^n$ are nuisances for $\hat{\mu}_{M^a}$)

1. **Debiasedness**: T converges to $C(P)$ at $N^{-1/2}$ rate if all nuisances for μ_{M^a} (i.e., $\{H^i, \pi^i\}_{i=1}^n$) converge at $N^{-1/4}$ rate.

2. **Doubly robustness**: $T = P_x(y)$ if, for all nuisances for μ_{M^a} (i.e., $\{H^i, \pi^i\}_{i=1}^n$), $H^i = H_0^i$ or $\pi^i = \pi_0^i$.

Error Analysis - Proof 1

$$T - P_{\mathbf{x}}(\mathbf{y}) = O_P\left(\sum_{a=1}^{m} \{\hat{\mu}_{M^a} - \mu_{M_0^a}\}\right)$$

Error Analysis - Proof 1

$$T - P_{\mathbf{x}}(\mathbf{y}) = O_P\left(\sum_{a=1}^{m}\{\hat{\mu}_{M^a} - \mu_{M_0^a}\}\right)$$

It can be witnessed by

Error Analysis - Proof 1

$$T - P_{\mathbf{x}}(\mathbf{y}) = O_P\left(\sum_{a=1}^{m} \{\hat{\mu}_{M^a} - \mu_{M_0^a}\}\right)$$

It can be witnessed by

$$T - P_{\mathbf{x}}(\mathbf{y}) = A(\{\hat{\mu}_{M^a}\}) - A(\{\mu_{M_0^a}\})$$

Error Analysis - Proof 1

$$T - P_{\mathbf{x}}(\mathbf{y}) = O_P\left(\sum_{a=1}^{m}\{\hat{\mu}_{M^a} - \mu_{M_0^a}\}\right)$$

It can be witnessed by

$$T - P_{\mathbf{x}}(\mathbf{y}) = A(\{\hat{\mu}_{M^a}\}) - A(\{\mu_{M_0^a}\})$$

$$= O_P\left(\sum_{a=1}^{m}(\hat{\mu}_{M^a} - \mu_0^a)\right)$$

Error Analysis - Proof 2

$$T - P_{\mathbf{x}}(\mathbf{y}) = O_P\left(\sum_{a=1}^{m} \{\hat{\mu}_{M^a} - \mu_{M_0^a}\} \right)$$

Note that

Error Analysis - Proof 2

$$T - P_{\mathbf{x}}(\mathbf{y}) = O_P\left(\sum_{a=1}^{m}\{\hat{\mu}_{M^a} - \mu_{M_0^a}\}\right)$$

Note that

$$\hat{\mu}_{M^a} - \mu_{M_0^a} = \mathbb{E}_D[f^{DR}(V; \{\mathbf{H}, \Pi\})] - C(P)$$

Error Analysis - Proof 2

$$T - P_{\mathbf{x}}(\mathbf{y}) = O_P\left(\sum_{a=1}^{m} \{\hat{\mu}_{M^a} - \mu_{M_0^a}\} \right)$$

Note that

$$\hat{\mu}_{M^a} - \mu_{M_0^a} = \mathbb{E}_D[f^{DR}(V; \{\mathbf{H}, \Pi\})] - C(P)$$
$$= O_P(N^{-1/2}) + \sum_{i=1}^{n} O_P\left(\| H^i - H_0^i \| \| \pi^i - \pi_0^i \|\right)$$

DML estimator for the Napkin Graph

$$P_x(y) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)} = \frac{\mu_{M_0^a}}{\mu_{M_0^b}}$$

DML estimator for the Napkin Graph

$$P_x(y) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)} = \frac{\mu_{M_0^a}}{\mu_{M_0^b}}$$

$$\hat{\mu}_{M^a} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_{x,y}(X, Y) - H^a(R)\} + H^a(r)\right]$$

DML estimator for the Napkin Graph

$$P_x(y) = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)} = \frac{\mu_{M_0^a}}{\mu_{M_0^b}}$$

$$\hat{\mu}_{M^a} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_{x,y}(X, Y) - H^a(R)\} + H^a(r)\right]$$

$$\hat{\mu}_{M^b} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_x(X) - H^b(R)\} + H^b(r)\right]$$

DML estimator for the Napkin Graph

$$P_x(y) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)} = \frac{\mu_{M_0^a}}{\mu_{M_0^b}}$$

$$\hat{\mu}_{M^a} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_{x,y}(X, Y) - H^a(R)\} + H^a(r)\right]$$

$$\hat{\mu}_{M^b} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_x(X) - H^b(R)\} + H^b(r)\right]$$

where samples for training nuisances and evaluating f^a, f^b are independent (sample-splitting).

DML estimator for the Napkin Graph

$$P_x(y) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)} = \frac{\mu_{M_0^a}}{\mu_{M_0^b}}$$

$$\hat{\mu}_{M^a} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_{x,y}(X, Y) - H^a(R)\} + H^a(r)\right]$$

$$\hat{\mu}_{M^b} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_x(X) - H^b(R)\} + H^b(r)\right]$$

where samples for training nuisances and evaluating f^a, f^b are independent (sample-splitting).

Then, the DML estimator for the example is

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

DML estimator for the Napkin Graph

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

$$\hat{\mu}_{M^a} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_{x,y}(X, Y) - H^a(R)\} + H^a(r)\right]$$

$$\hat{\mu}_{M^b} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_x(X) - H^b(R)\} + H^b(r)\right]$$

DML estimator for the Napkin Graph

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

$$\hat{\mu}_{M^a} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_{x,y}(X, Y) - H^a(R)\} + H^a(r)\right]$$

$$\hat{\mu}_{M^b} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_x(X) - H^b(R)\} + H^b(r)\right]$$

1. **Debiasedness**: T converges to $C(P)$ at $N^{-1/2}$ rate if all nuisances in $\hat{\mu}_{M^a}, \hat{\mu}_{M^b}$ are converging at $N^{-1/4}$ rate.

DML estimator for the Napkin Graph

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

$$\hat{\mu}_{M^a} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^a, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_{x,y}(X,Y) - H^a(R)\} + H^a(r)\right]$$

$$\hat{\mu}_{M^b} := \mathbb{E}_D[f^{DR}(\mathbf{V}; \{\mathbf{H}^b, \Pi\})] = \mathbb{E}_D\left[\frac{I_r(R)}{\pi(R \mid W)}\{I_x(X) - H^b(R)\} + H^b(r)\right]$$

1. **Debiasedness**: T converges to $C(P)$ at $N^{-1/2}$ rate if all nuisances in $\hat{\mu}_{M^a}, \hat{\mu}_{M^b}$ are converging at $N^{-1/4}$ rate.

2. **Doubly robustness**: $T = P_x(y)$ if $H^a = H^a_0$ or $\pi^i = \pi^i_0$; and $H^b = H^b_0$ or $\pi^i = \pi^i_0$

Error Analysis - Napkin Example 1

$$T - P_x(y) = O\left(\sum_{i \in \{a,b\}} \hat{\mu}_{M^i} - \mu_{M_0^i}\right)$$

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

Error Analysis - Napkin Example 1

$$T - P_x(y) = O\left(\sum_{i \in \{a,b\}} \hat{\mu}_{M^i} - \mu_{M_0^i} \right)$$

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

$$\begin{aligned}
T - P_x(y) &= \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}} - \frac{\mu_{M_0^a}}{\mu_{M_0^b}} \\
&= \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}} - \frac{\hat{\mu}_{M^a}}{\mu_{M_0^b}} + \frac{\hat{\mu}_{M^a}}{\mu_{M_0^b}} - \frac{\mu_{M_0^a}}{\mu_{M_0^b}} \\
&= \hat{\mu}_{M^a}\left(\frac{1}{\hat{\mu}_{M^b}} - \frac{1}{\mu_{M_0^b}} \right) + \frac{1}{\mu_{M_0^b}}(\hat{\mu}_{M^a} - \mu_{M_0^a}) \\
&= O(\sum_{i \in \{a,b\}} \hat{\mu}_{M^i} - \mu_{M_0^i})
\end{aligned}$$

Error Analysis - Napkin Example 2

$$T - P_x(y) = O\left(\sum_{i \in \{a,b\}} \hat{\mu}_{M^i} - \mu_{M_0^i} \right)$$

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

Error Analysis - Napkin Example 2

$$T - P_x(y) = O\left(\sum_{i \in \{a,b\}} \hat{\mu}_{M^i} - \mu_{M_0^i}\right)$$

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

Recall

$$\hat{\mu}_{M^a} - \mu_{M_0^a} = O_P(N^{-1/2}) + O_P\left(\| H^a - H_0^a \| \| \pi - \pi_0 \|\right)$$

$$\hat{\mu}_{M^b} - \mu_{M_0^b} = O_P(N^{-1/2}) + O_P\left(\| H^b - H_0^b \| \| \pi - \pi_0 \|\right),$$

where $\{H^i, \pi^i\}$ are nuisances for $M^i \in \{M^a, M^b\}$

Error Analysis - Napkin Example 2

$$T - P_x(y) = O\left(\sum_{i \in \{a,b\}} \hat{\mu}_{M^i} - \mu_{M_0^i} \right)$$

$$T := \frac{\hat{\mu}_{M^a}}{\hat{\mu}_{M^b}}$$

Recall

$$\hat{\mu}_{M^a} - \mu_{M_0^a} = O_P(N^{-1/2}) + O_P\left(\| H^a - H_0^a \| \| \pi - \pi_0 \|\right)$$

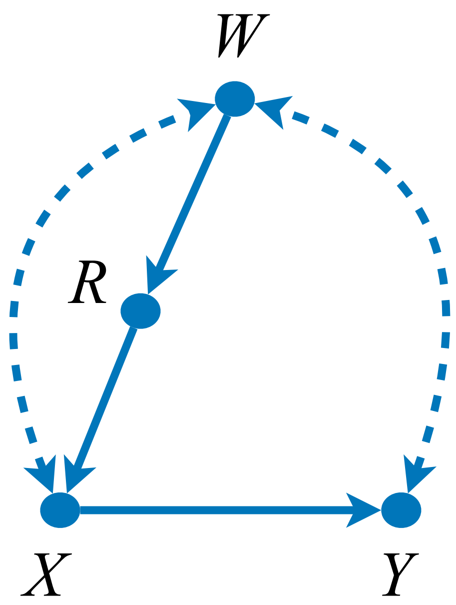
$$\hat{\mu}_{M^b} - \mu_{M_0^b} = O_P(N^{-1/2}) + O_P\left(\| H^b - H_0^b \| \| \pi - \pi_0 \|\right),$$

where $\{H^i, \pi^i\}$ are nuisances for $M^i \in \{M^a, M^b\}$

1. **Debiasedness**: T converges to $C(P)$ at $N^{-1/2}$ rate if H^a, H^b, π converge to H_0^a, H_0^b, π_0^i at $N^{-1/4}$ rate.
2. **Doubly robustness**: $T = P_x(y)$ if $H^a = H_0^a$ or $\pi^i = \pi_0^i$; and $H^b = H_0^b$ or $\pi^i = \pi_0^i$

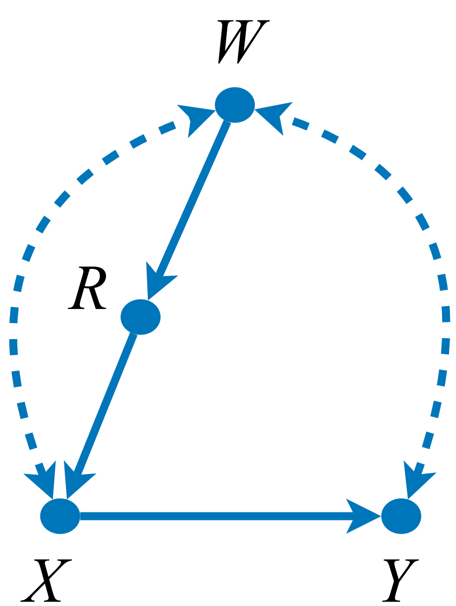
Empirical Evidence

Empirical result — Expected results



$$P_x(y) = C(P) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)}$$

Empirical result — Expected results



$$P_x(y) = C(P) = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

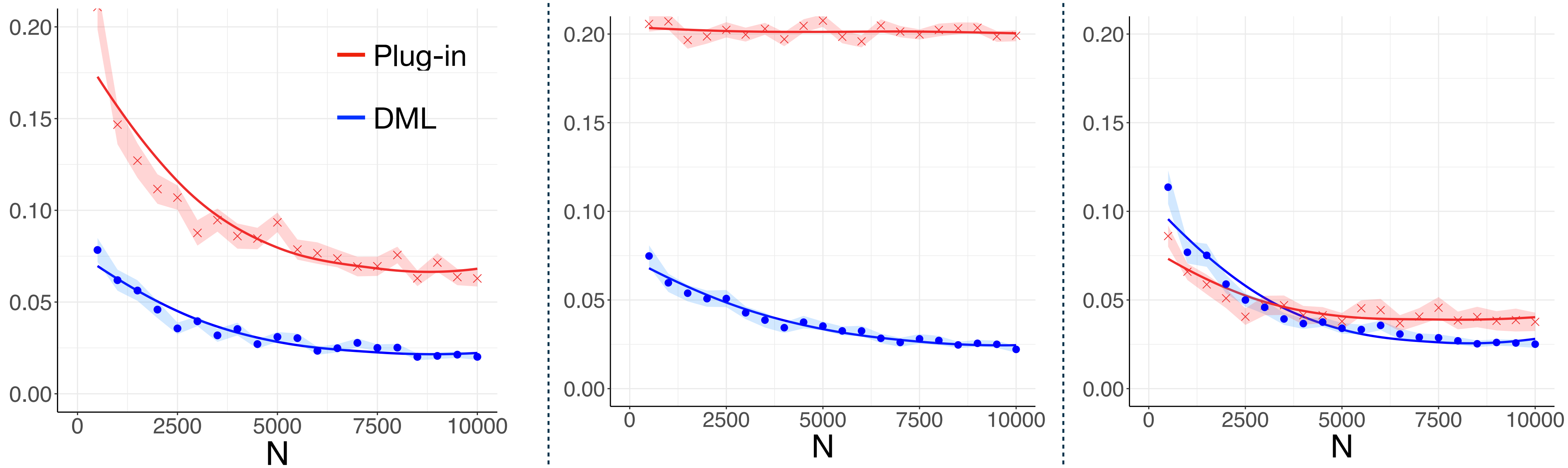
$$C(\hat{P}) = \frac{\sum_w \hat{P}(x, y \mid r, w) \hat{P}(w)}{\sum_w \hat{P}(x \mid r, w) \hat{P}(w)}$$

$C(\hat{P})$: the plug-in estimator, only viable estimator working for identifiable causal functional.

Empirical results

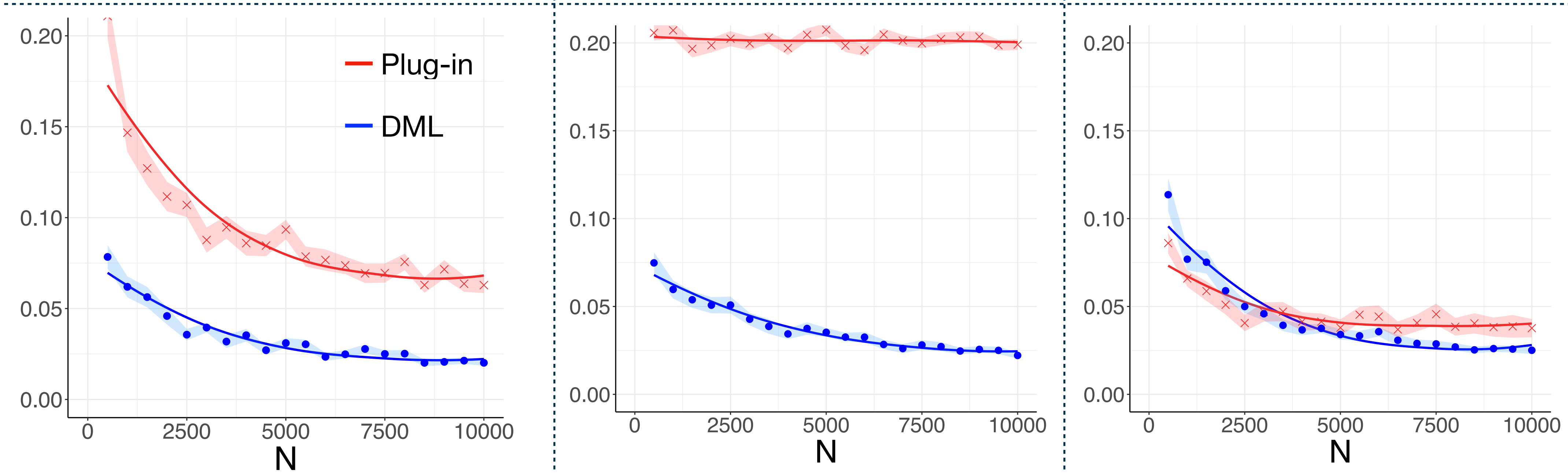
Empirical results

Both converge at rate $N^{-1/4}$	$P(x, y \mid r, w)$ misspecified.	$P(r \mid w)$ misspecified.



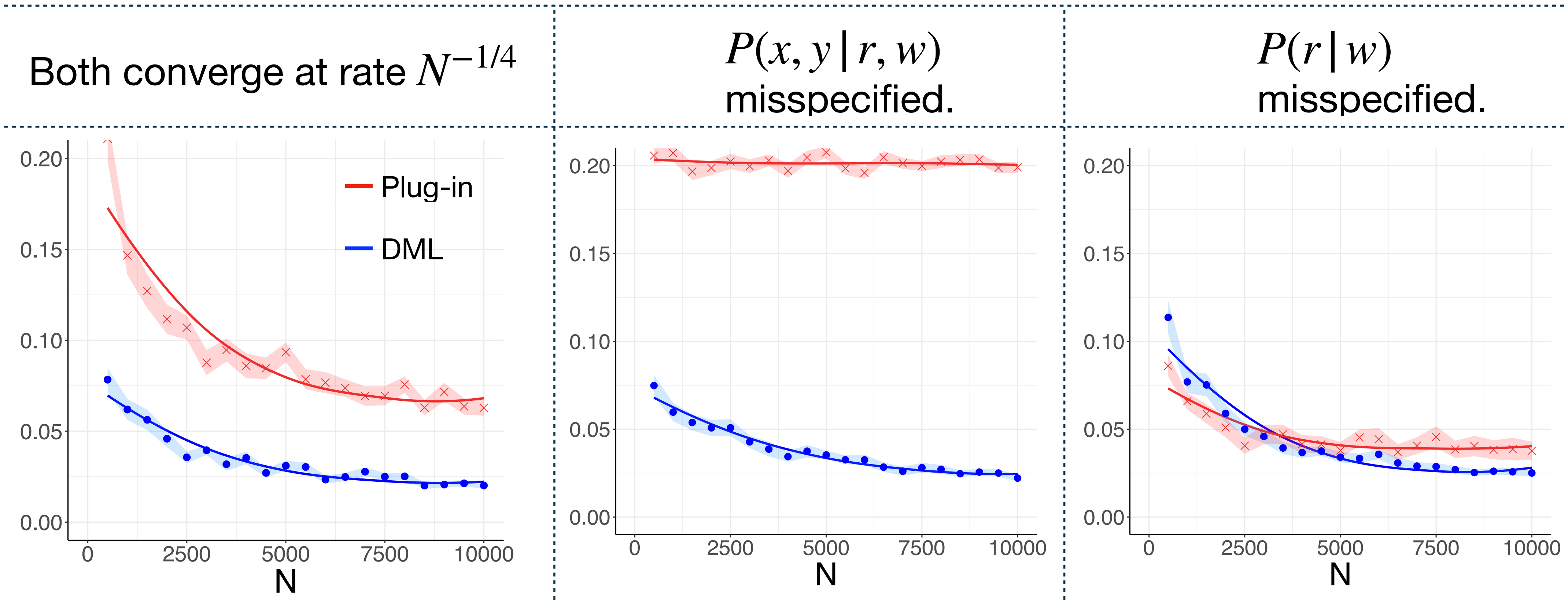
Empirical results

Both converge at rate $N^{-1/4}$	$P(x, y \mid r, w)$ misspecified.	$P(r \mid w)$ misspecified.



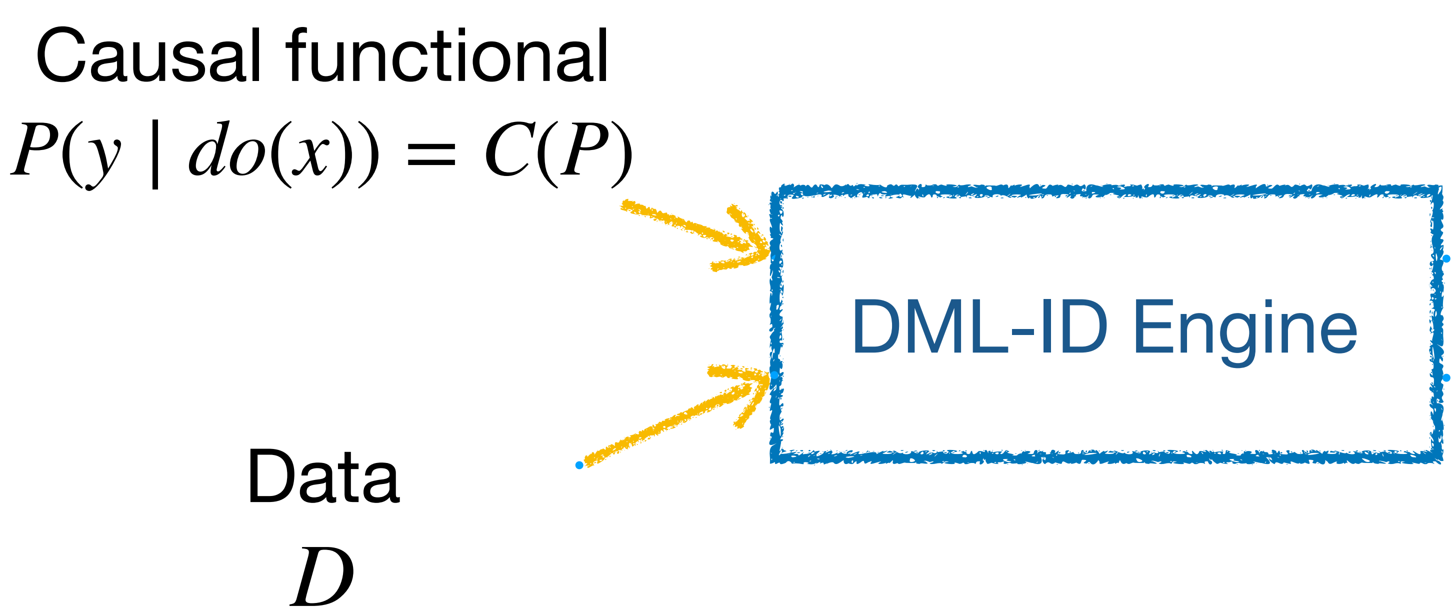
- **(Debiasedness; Left)** DML converges (i.e., the error decreases) faster even when nuisances converge slower rate ($N^{-1/4}$).

Empirical results



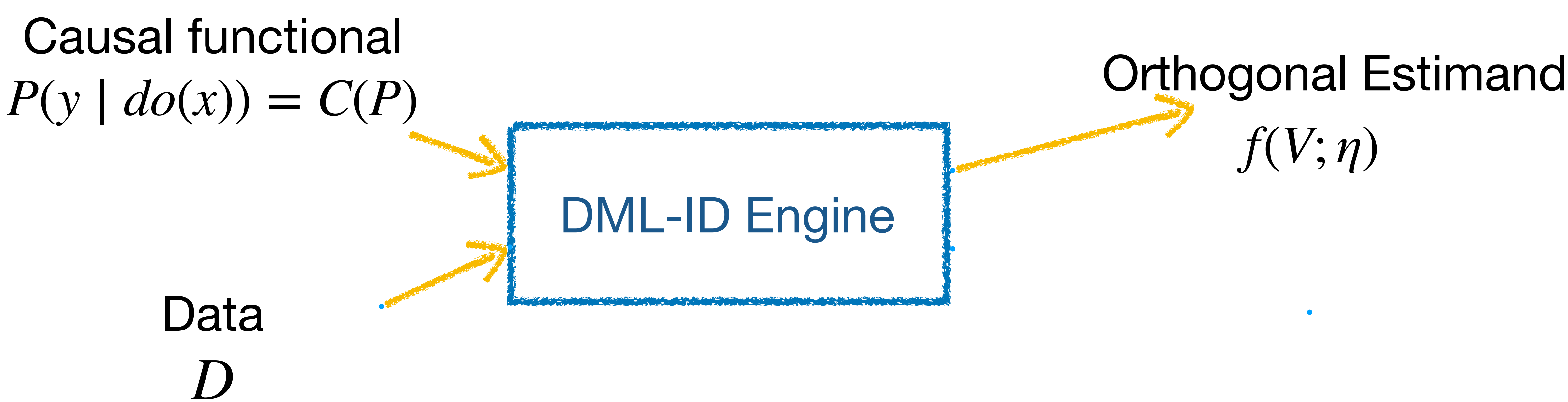
- (Debiasedness; Left) DML converges (i.e., the error decreases) faster even when nuisances converge slower rate ($N^{-1/4}$).
- (Doubly Robustness; (Center, Right)) DML converges even when models for either $P(x, y \mid r, w)$ (center) or $P(r \mid w)$ (right) is misspecified.

Conclusions



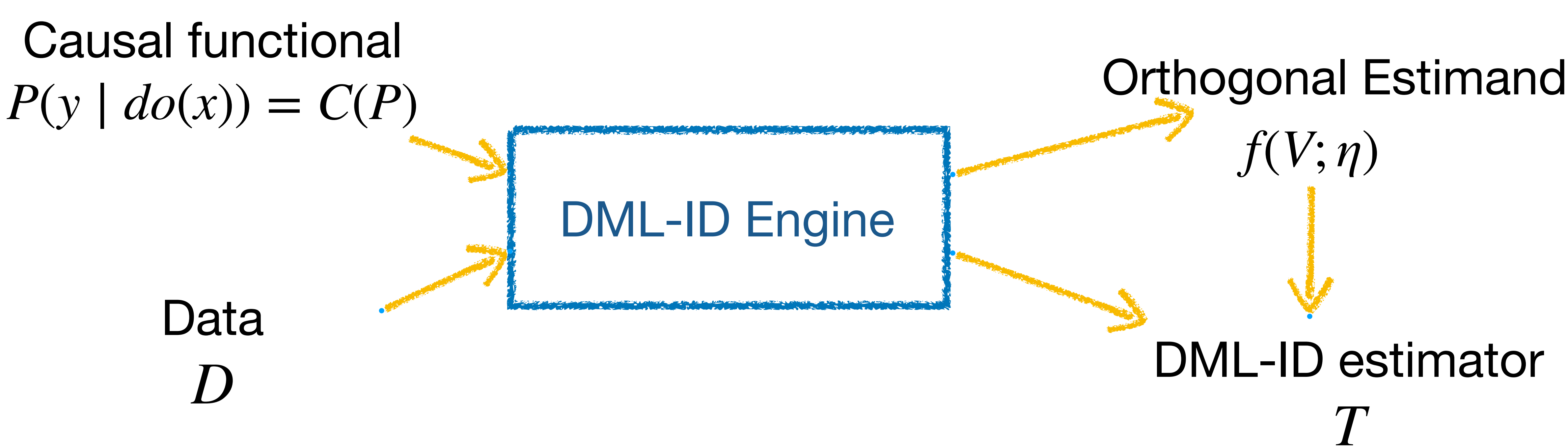
Conclusions

- We develop a systematic procedure for constructing a DML estimator for any identifiable causal effects.



Conclusions

- We develop a systematic procedure for constructing a DML estimator for any identifiable causal effects.
- A DML estimator enjoys *debiasedness* and *doubly robustness* against model misspecification and slow convergence rate.



Any Questions?