Estimating Identifiable Causal Effects through Double Machine Learning

Yonghan Jung
PURDUE UNIVERSITY

Jin Tian
IOWA STATE UNIVERSITY

Elias Bareinboim
COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Learning causal effect: 2-step procedures

Learning causal effect: 2-step procedures

Step 1.
(Identification, ID)

Represent the causal query **Q**=P(y|do(x)) as a function of P and set of causal assumption **A** (i.e., **Q**=**F**(P, A)).

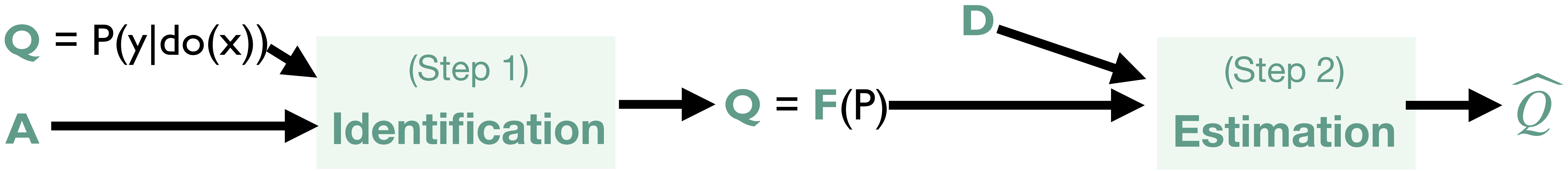
Learning causal effect: 2-step procedures

Step 1.
(Identification, ID)

Represent the causal query **Q**=P(y|do(x)) as a function of P and set of causal assumption **A** (i.e., **Q**=**F**(P, A)).

Step 2.
(Estimation)

Estimate the identified estimand **Q**=**F**(P, A) from finite samples **D**.



Current status of causal inference

Current status of causal inference

Strength

Completeness of Identification algorithm

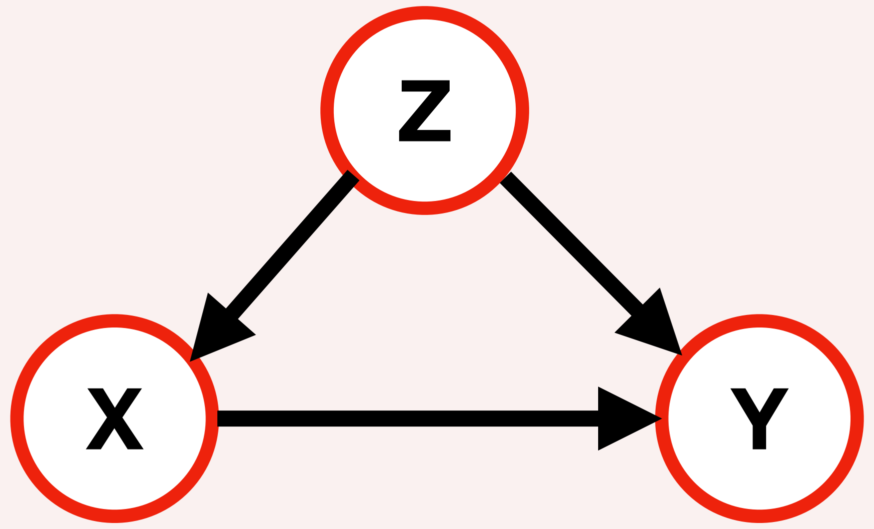
- There exist ***sound*** and ***complete*** identification algorithms for determining whether a causal query **Q** can be represented as a functional of P (i.e., Q=F(P, A)) from a given causal graph.

Current status of causal inference

Strength

Completeness of Identi

- There exist ***sound*** and **c** identification algorithms whether a causal query **Q** can be represented as a functional of P (i.e., Q=F(P, A)) from a given causal graph.



Backdoor graph

Weakness

No sample-efficient estimator

- Estimation has been mainly done on backdoor/ignorability assumption.

Q = P(y|do(x))

G (Causal graph) → **ID algorithm** → **Q** = **F**(P)

D → **Estimation** e.g., IPW,... → $\widehat{Q}$

Under backdoor assumption,

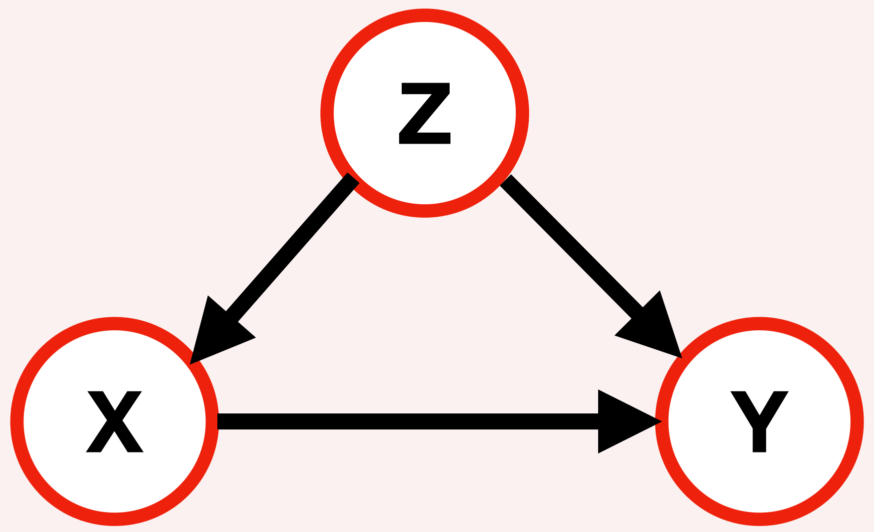
$$Q = \sum P(y \mid x, z) P(z)$$

Current status of causal inference

Strength

Completeness of Identi

- There exist ***sound*** and ***c*** identification algorithms whether a causal query Q can be represented as a functional of P (i.e., Q=F(P, A)) from a given causal graph.

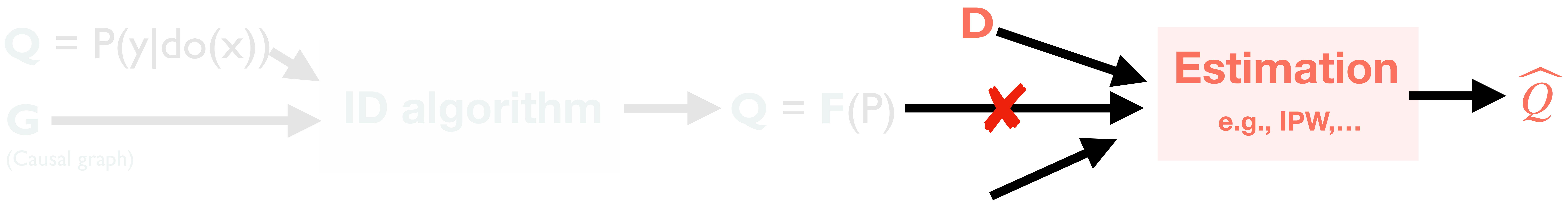


Backdoor graph

Weakness

No sample-efficient estimator

- Estimation has been mainly done on backdoor/ignorability assumption.
- For general identifiable estimands, it's not known (neither obvious) how to estimate the causal effect sample-efficiently.

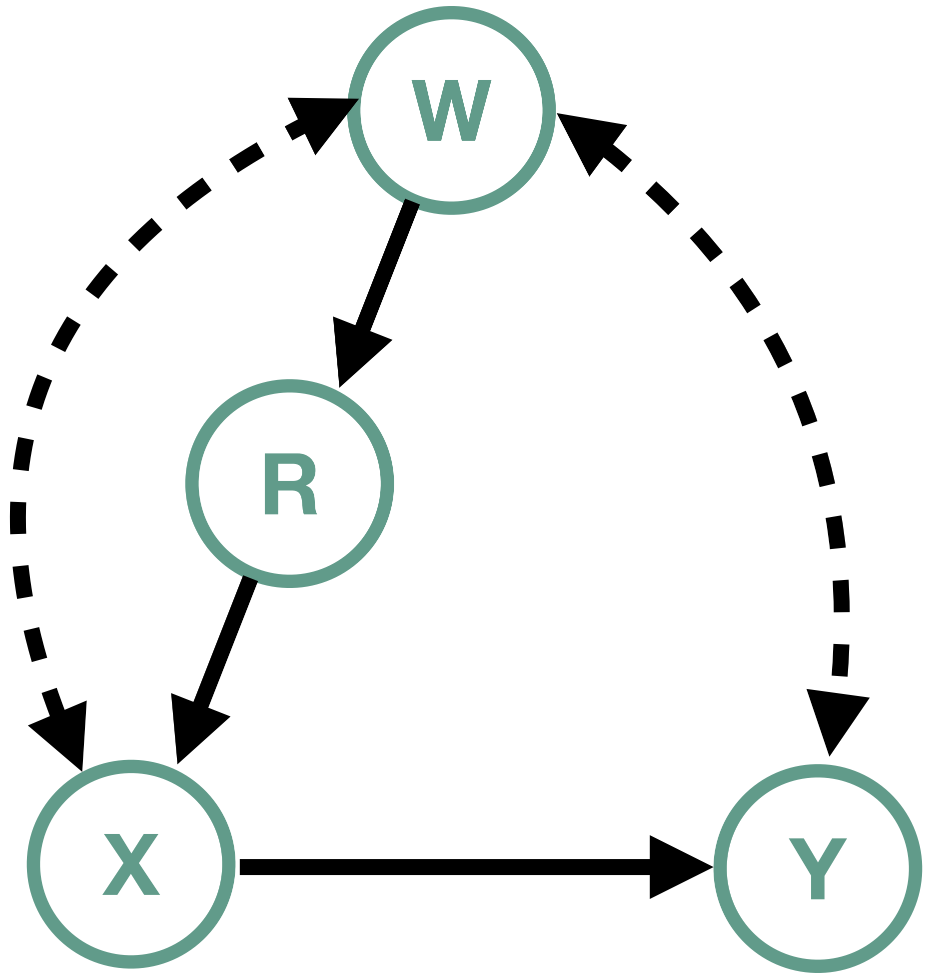


Under backdoor assumption,

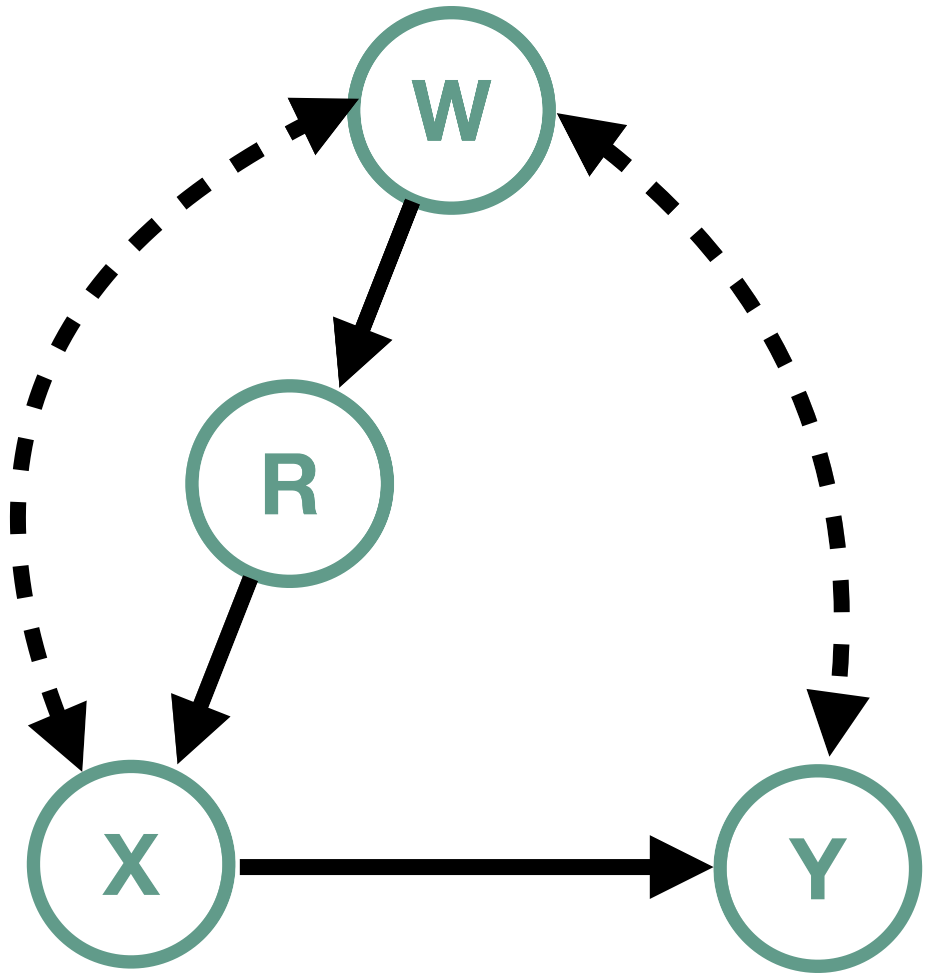
$$Q = \sum P(y \mid x, z) P(z)$$

Connecting Identification and Estimation

Connecting Identification and Estimation



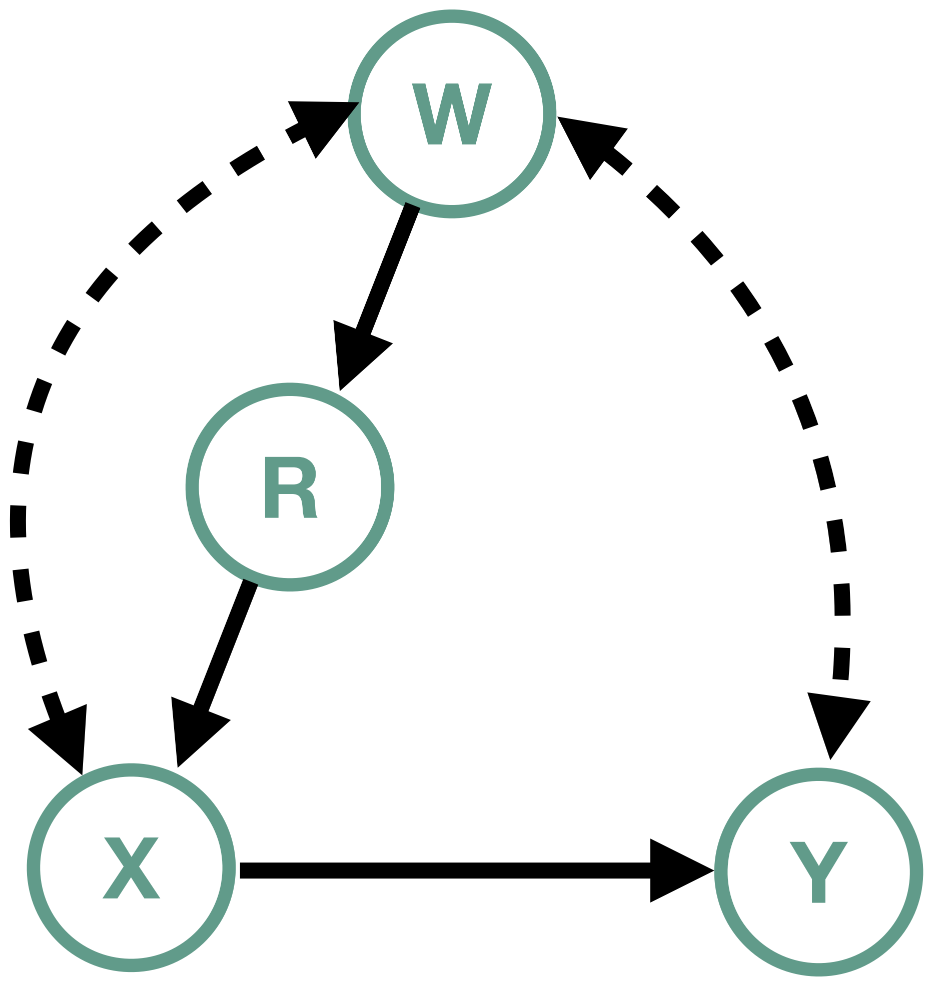
Connecting Identification and Estimation



$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

- Causal effect is identifiable and the functional is given as the below equation.

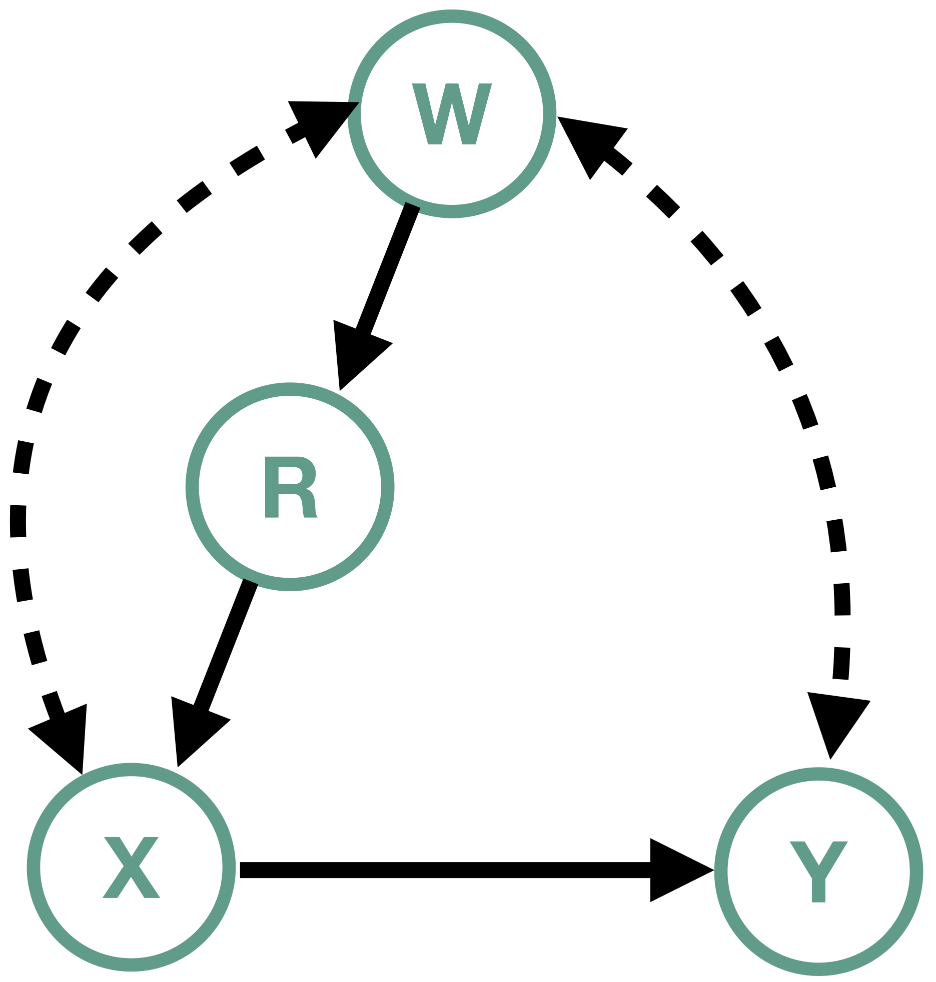
Connecting Identification and Estimation



$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)}$$

- Causal effect is identifiable and the functional is given as the below equation.
- Existing BD estimators are not applicable, b/c the functional is not in an adjustment form.

Connecting Identification and Estimation



$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

- Causal effect is identifiable and the functional is given as the below equation.
- Existing BD estimators are not applicable, b/c the functional is not in an adjustment form.
- **Observation** — Causal functional P is a arithmetic of two adjustments:

$P(x, y \mid r, w)$ adjusting over W=w

$P(x \mid r, w)$ adjusting over W=w

Connecting Identification and Estimation

Intuition

A causal functional is expressible as a function of adjustments. Then, a BD estimator might be applicable in estimating such functional.

$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

- **Observation** — Causal functional P is a arithmetic of two adjustments:

$P(x, y \mid r, w)$ adjusting over W=w

$P(x \mid r, w)$ adjusting over W=w

Connecting Identification and Estimation

Intuition

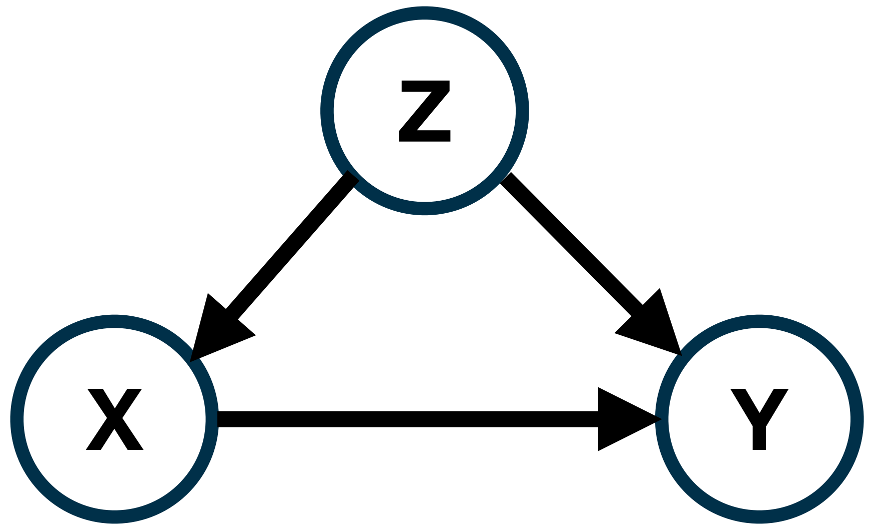
A causal functional is expressible as a function of adjustments. Then, a BD estimator might be applicable in estimating such functional.

- **Observation** — Causal functional P is a

$P(y\,|$

What BD estimators should we leverage?

Recap: Classic BD estimators

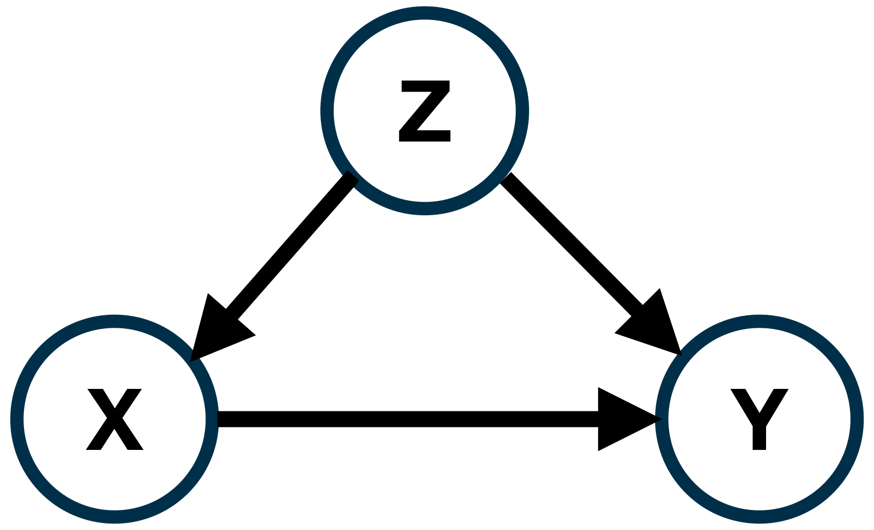


Backdoor graph

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

Estimand (Q)		
Estimator ($\widehat{Q}$)		
For correct estimation		
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

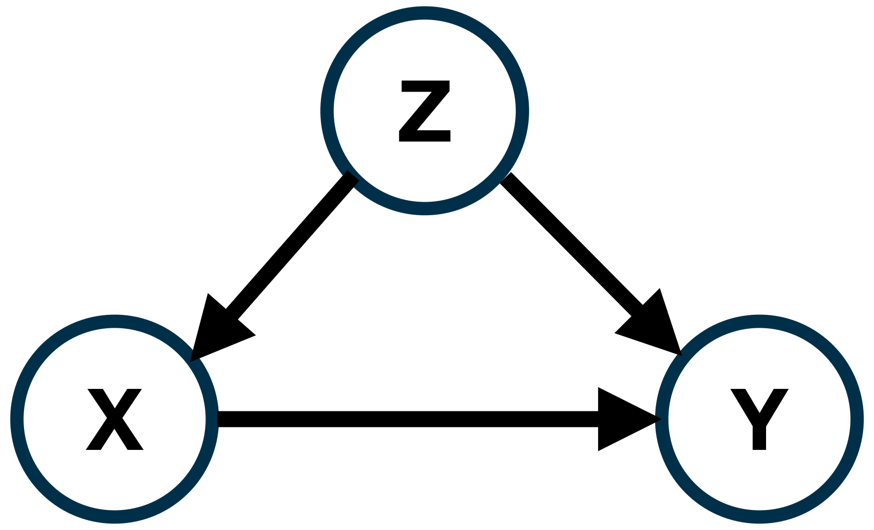


Backdoor graph

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

	Inverse Probability Weight	
Estimand (Q)		
Estimator ($\widehat{Q}$)		
For correct estimation		
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

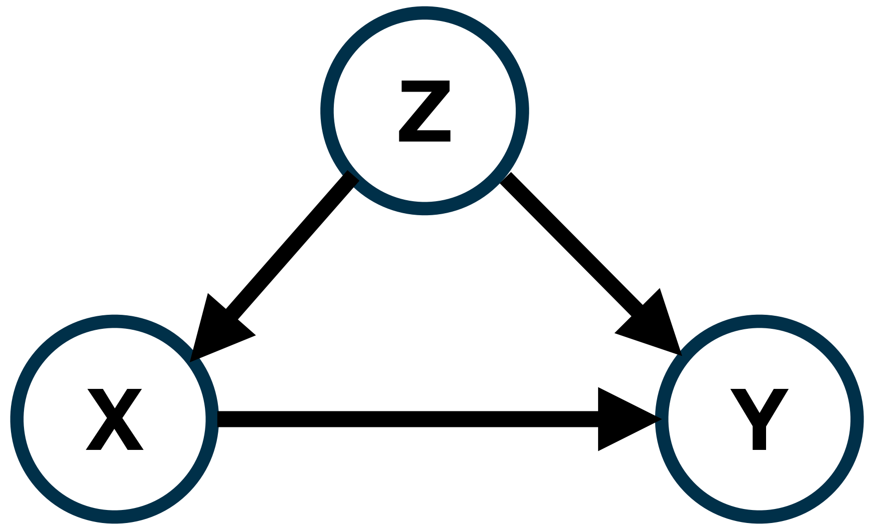


Backdoor graph

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

	Inverse Probability Weight	
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	
Estimator ($\widehat{Q}$)		
For correct estimation		
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

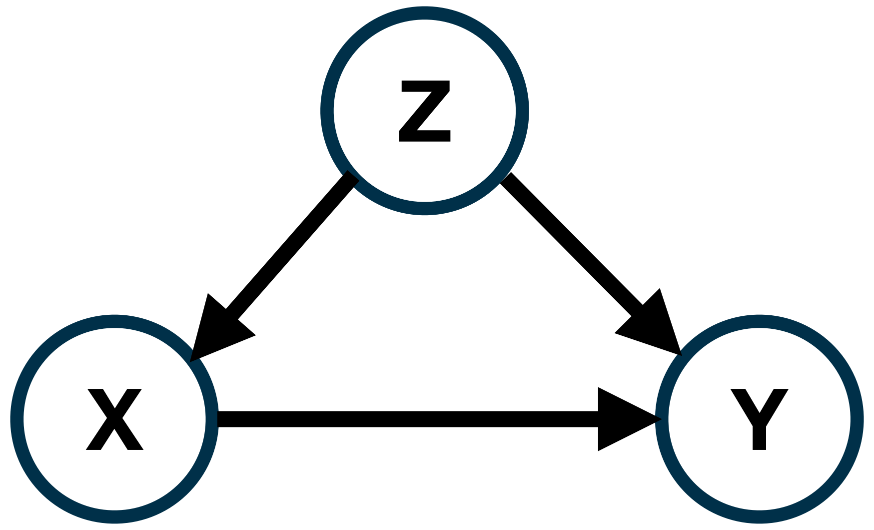


Backdoor graph

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

	Inverse Probability Weight	
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	
Estimator ($\widehat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	
For correct estimation		
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

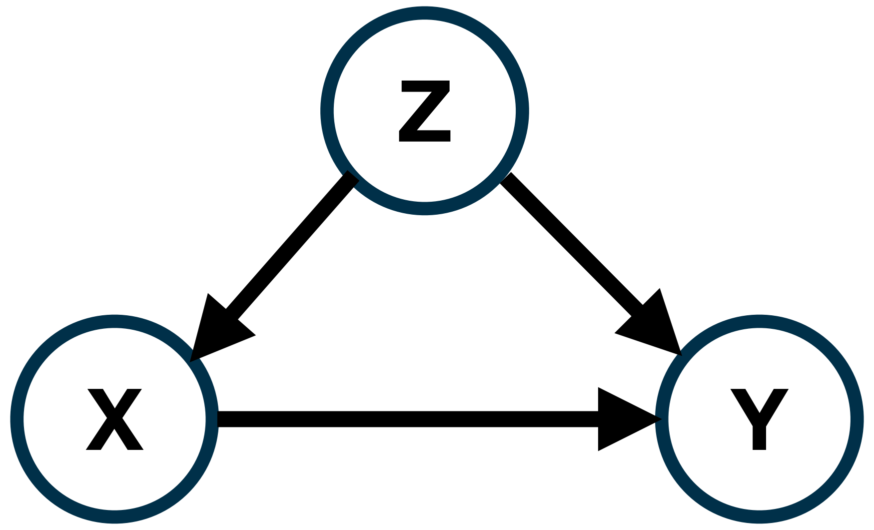


Backdoor graph

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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	
Estimator ($\widehat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N}\frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	
For correct estimation		
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

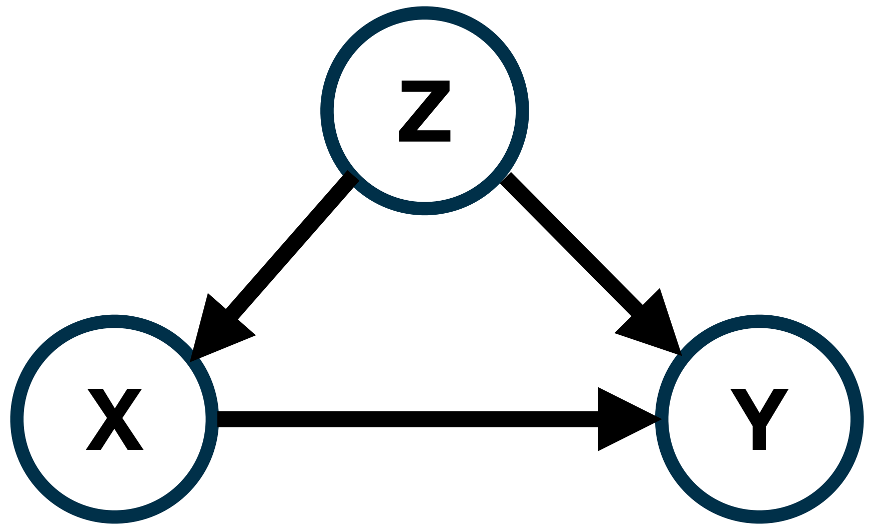


Backdoor graph

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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	$\mathbb{E}_{\mathbf{Z}}\left[P(y \mid x, Z)\right]$
Estimator ($\widehat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	
For correct estimation		
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

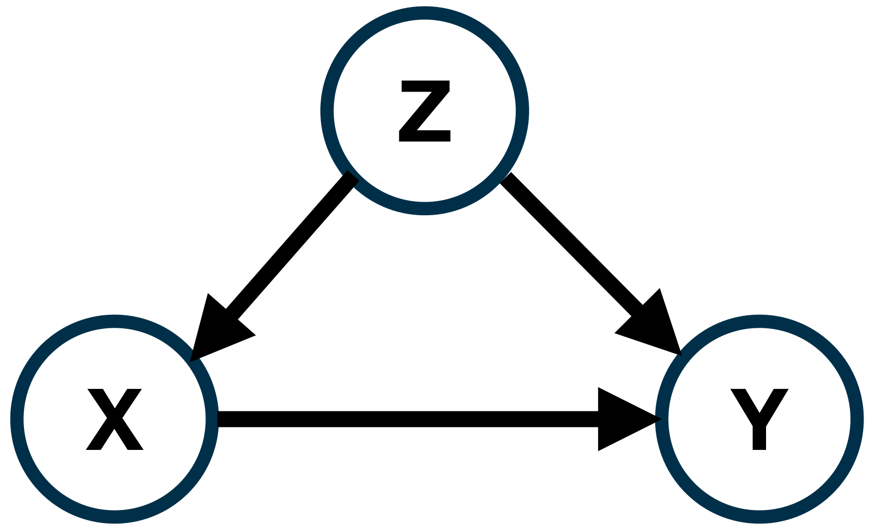


Backdoor graph

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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	$\mathbb{E}_{\mathbf{z}}\left[P(y \mid x, Z)\right]$
Estimator ($\widehat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	$\frac{1}{N}\sum_{i=1}^{N} \widehat{P}(y \mid x, Z_{(i)})$
For correct estimation		
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

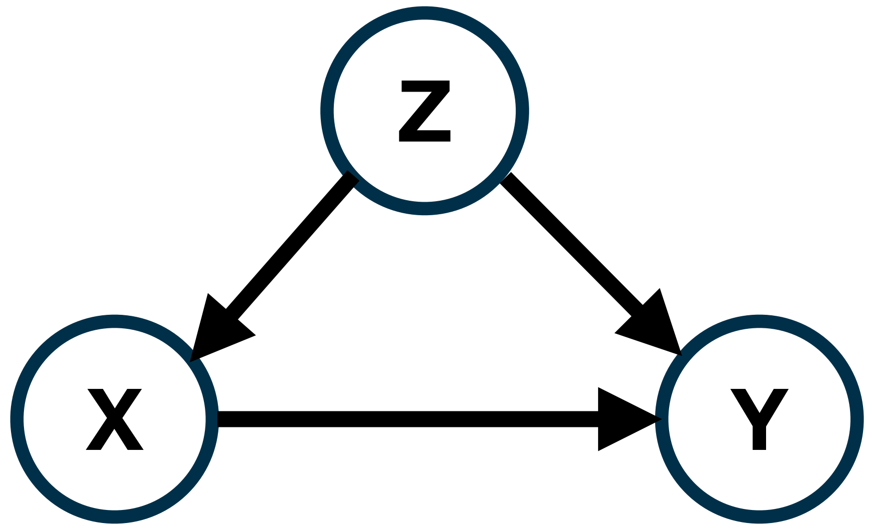


Backdoor graph

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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\dfrac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	$\mathbb{E}_{\mathbf{z}}\left[P(y \mid x, Z)\right]$
Estimator ($\widehat{Q}$)	$\dfrac{1}{N}\sum_{i=1}^{N} \dfrac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	$\dfrac{1}{N}\sum_{i=1}^{N} \widehat{P}(y \mid x, Z_{(i)})$
For correct estimation	Nuisance functional should be correctly estimated .	
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

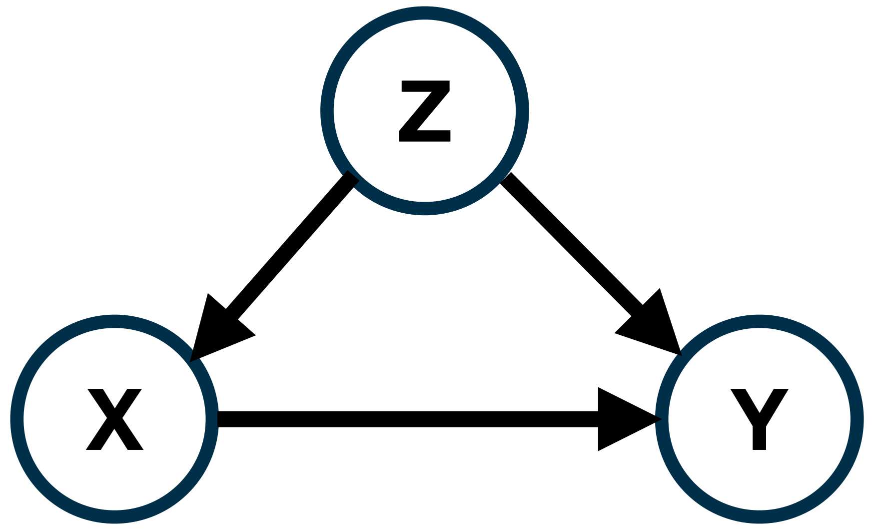


Backdoor graph

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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	$\mathbb{E}_{\mathbf{Z}}\left[P(y \mid x, Z)\right]$
Estimator ($\widehat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	$\frac{1}{N}\sum_{i=1}^{N} \widehat{P}(y \mid x, Z_{(i)})$
For correct estimation	Incorrectly estimated if estimates of nuisances are misspecified.	
For $\sqrt{N}$-consistency		

Recap: Classic BD estimators

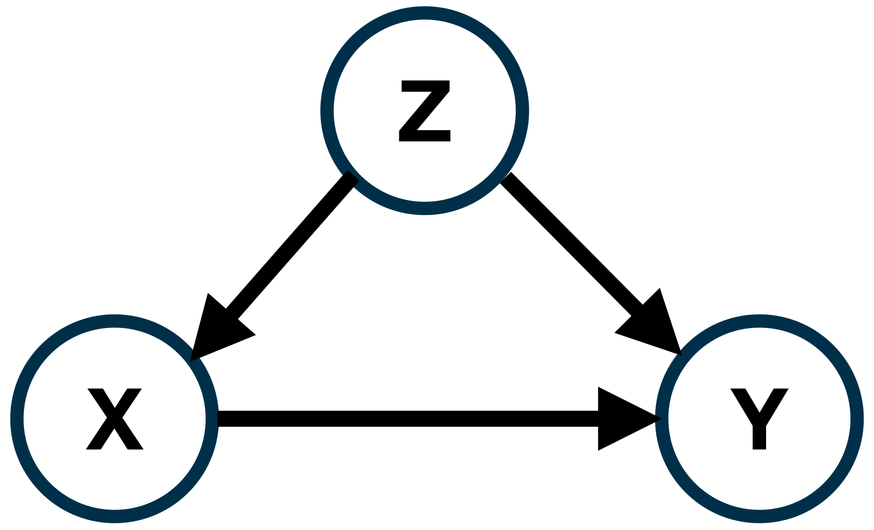


Backdoor graph

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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	$\mathbb{E}_{\mathbf{z}}\left[P(y \mid x, Z)\right]$
Estimator ($\widehat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	$\frac{1}{N}\sum_{i=1}^{N} \widehat{P}(y \mid x, Z_{(i)})$
For correct estimation	Incorrectly estimated if estimates of nuisances are misspecified.	
For $\sqrt{N}$-consistency	Estimates for nuisance converges at $o_P(N^{-1/2})$.	

Recap: Classic BD estimators

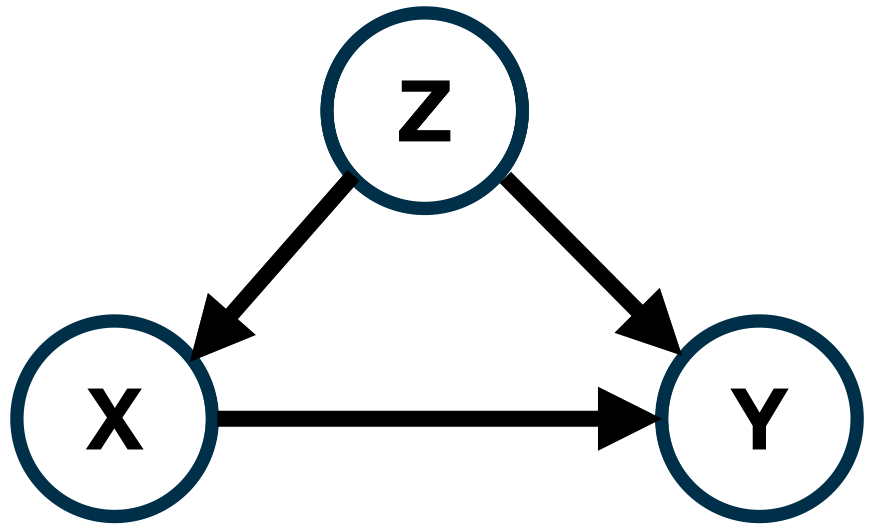


Backdoor graph

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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)} I_y(Y)\right]$	$\mathbb{E}_\mathbf{z}\left[P(y \mid x, Z)\right]$
Estimator ($\widehat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})} I_y(Y_{(i)})$	$\frac{1}{N}\sum_{i=1}^{N} \widehat{P}(y \mid x, Z_{(i)})$
For correct estimation	**Incorrectly estimated** if estimates of nuisances are **misspecified.**	
For $\sqrt{N}$-consistency	**Not $\sqrt{N}$-consistent** if estimates of nuisances converges **slower than** $o_P(N^{-1/2})$	

Recap: Classic BD estimators



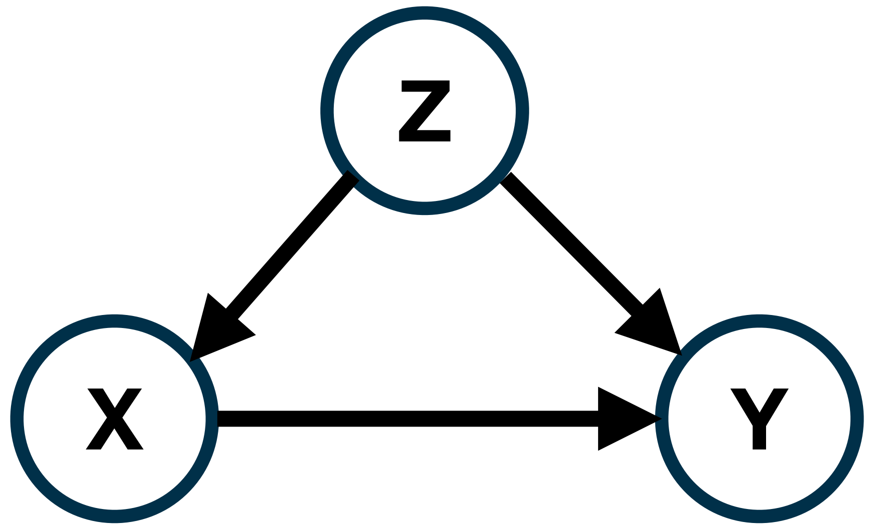
Backdoor graph

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

😈 **Risk of classic estimators**

Estimand (Q)	
Estimator ($\widehat{Q}$)	
For correct estimation	**Incorrectly estimated** if estimates of nuisances are **misspecified.**
For $\sqrt{N}$-consistency	**Not $\sqrt{N}$-consistent** if estimates of nuisances converges **slower than** $o_P(N^{-1/2})$

Recap: Classic BD estimators



Backdoor graph

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

Risk of classic estimators

- **Model misspecification** is common in practice when the data generating process is complicated $\Rightarrow$ Incorrect estimation.

Estimand (Q)

Estimator ($\widehat{Q}$)

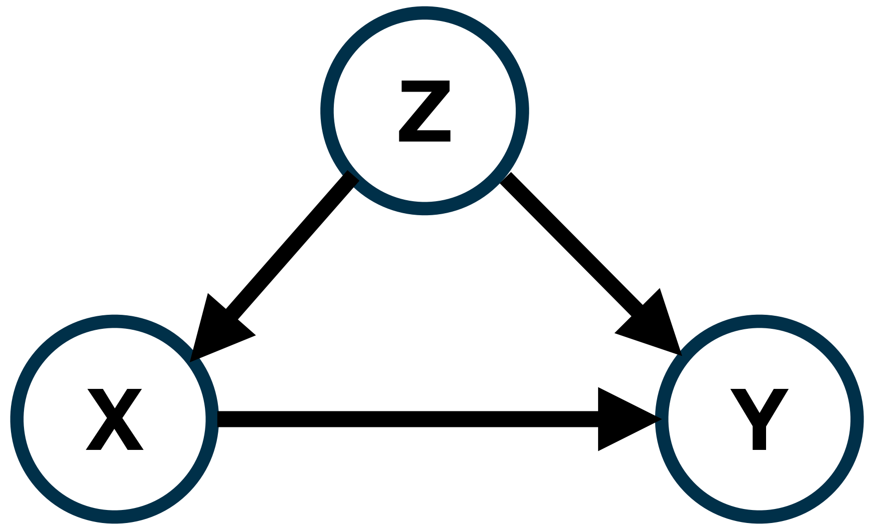
For correct estimation

Incorrectly estimated if estimates of nuisances are **misspecified.**

For $\sqrt{N}$-consistency

Not $\sqrt{N}$-consistent if estimates of nuisances converges **slower than** $o_P(N^{-1/2})$

Recap: Classic BD estimators



Backdoor graph

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

😈 Risk of classic estimators

- **Model misspecification** is common in practice when the data generating process is complicated $\Rightarrow$ Incorrect estimation.
- **Slow convergence** is common for flexible ML models $\Rightarrow$ Not $\sqrt{N}$-consistency.

Estimand (Q)	
Estimator ($\widehat{Q}$)	
For correct estimation	**Incorrectly estimated** if estimates of nuisances are **misspecified.**
For $\sqrt{N}$-consistency	**Not $\sqrt{N}$-consistent** if estimates of nuisances converges **slower than** $o_P(N^{-1/2})$

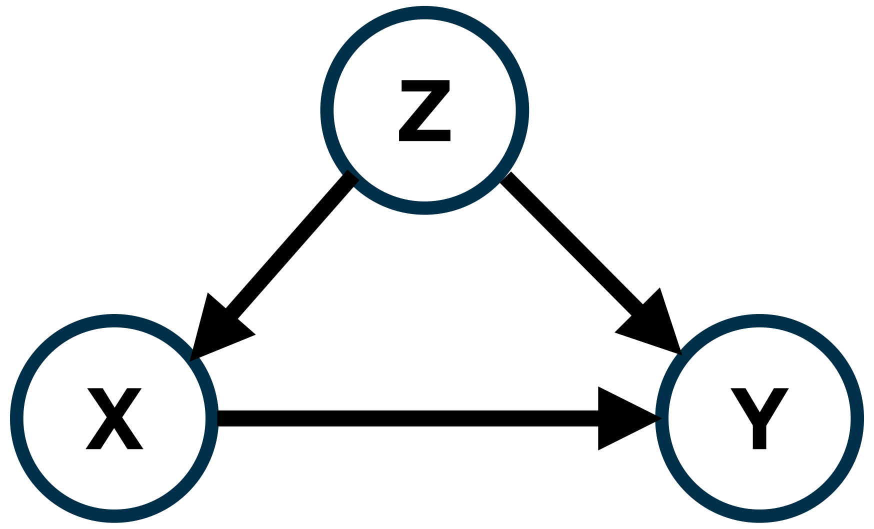
Recap: Classic BD estimators

tors

- **Model misspecification** is common in practice when the data generating process is complicated $\Longrightarrow$ Incorrect estimation.
- **Slow convergence** is common for flexible ML models $\Rightarrow$ Not $\sqrt{N}$-consistency.

Estimand (Q)	
Estimator ($\widehat{Q}$)	
For correct estimation	**Incorrectly estimated** if estimates of nuisances are **misspecified.**
For $\sqrt{N}$-consistency	**Not $\sqrt{N}$-consistent** if estimates of nuisances converges **slower than** $o_P(N^{-1/2})$

Recent advance: DML estimator

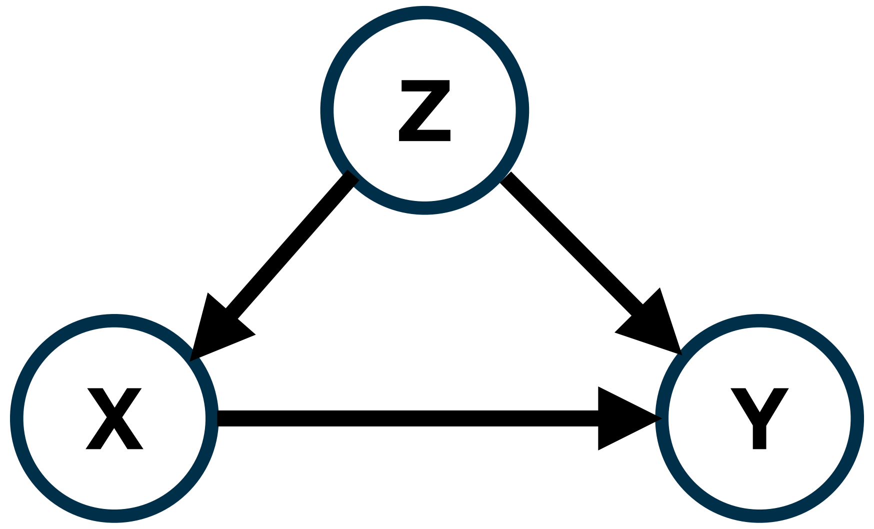


Backdoor graph

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

	Double Machine Learning Estimator
Estimand Representation of Q	
Estimators Estimator $\widehat{Q}$	
For correct estimation	
For $\sqrt{N}$-consistency	

Recent advance: DML estimator

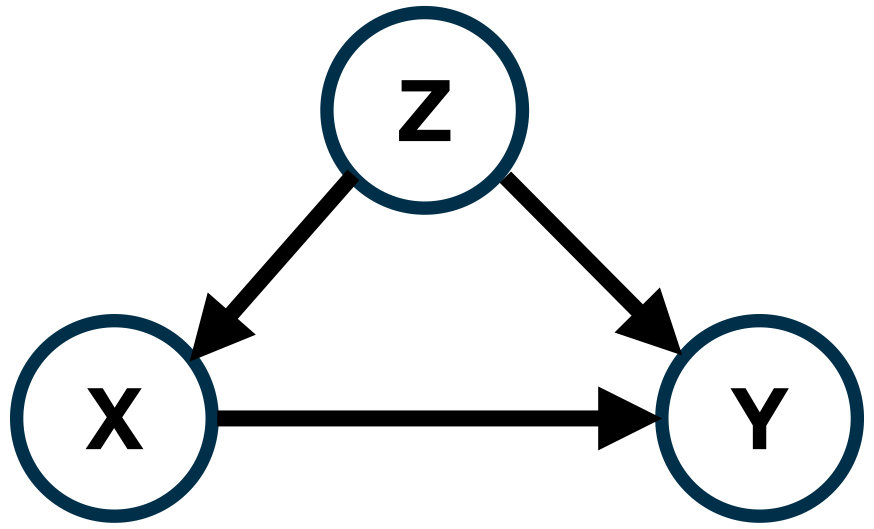


Backdoor graph

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

	Double Machine Learning Estimator
Estimand Representation of Q	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)}\left(I_y(Y) - P(y \mid x, Z)\right) + P(y \mid x, Z)\right]$
Estimators Estimator $\widehat{Q}$	
For correct estimation	
For $\sqrt{N}$-consistency	

Recent advance: DML estimator

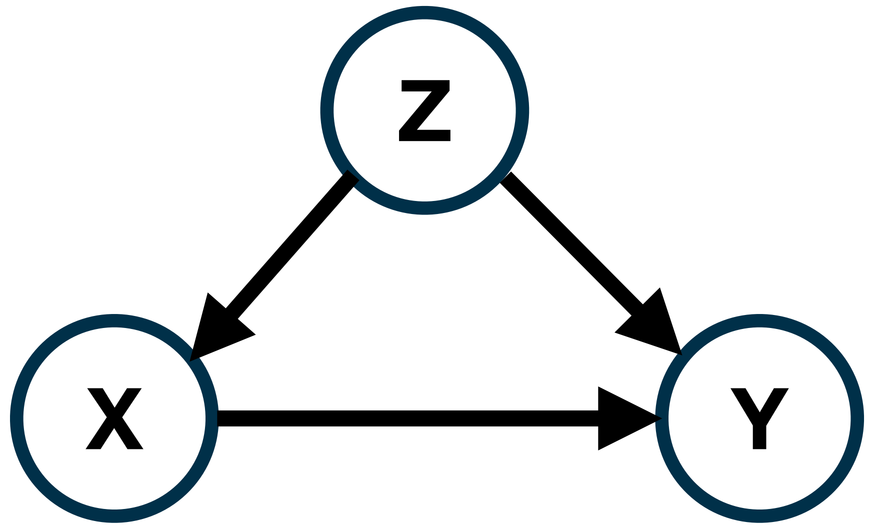


Backdoor graph

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

	Double Machine Learning Estimator
Estimand Representation of Q	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)}\left(I_y(Y) - P(y \mid x, Z)\right) + P(y \mid x, Z)\right]$
Estimators Estimator $\widehat{Q}$	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})}\left(I_y(Y) - \widehat{P}(y \mid x, Z)\right) + \widehat{P}(y \mid x, Z)$
For correct estimation	
For $\sqrt{N}$-consistency	

Recent advance: DML estimator



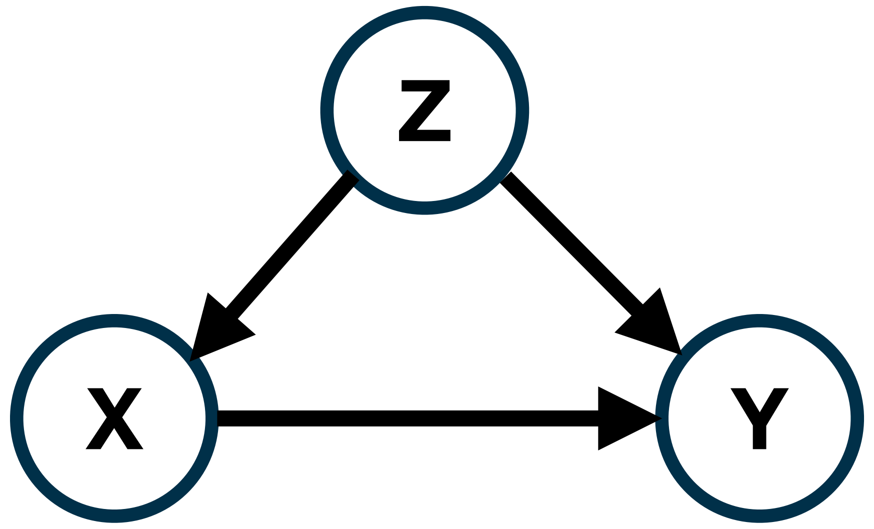
Backdoor graph

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

Double Machine Learning Estimator

Estimand Representation of Q	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)}\Big(I_y(Y) - P(y \mid x, Z)\Big) + P(y \mid x, Z)\right]$
Estimators Estimator $\widehat{Q}$	$\frac{1}{N}\sum_{i=1}^{N}\frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})}\Big(I_y(Y) - \widehat{P}(y \mid x, Z)\Big) + \widehat{P}(y \mid x, Z)$ where training and evaluation of $\widehat{P}$ are done with two distinct sets of samples (“***Cross fitting***”)
For correct estimation	
For $\sqrt{N}$-consistency	

Recent advance: DML estimator



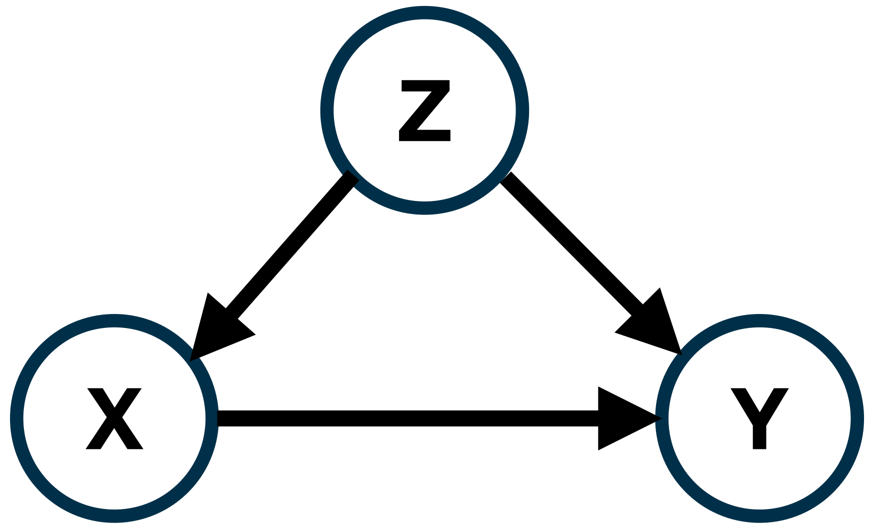
Backdoor graph

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

Double Machine Learning Estimator

Estimand Representation of Q	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)}\left(I_y(Y) - P(y \mid x, Z)\right) + P(y \mid x, Z)\right]$
Estimators Estimator $\widehat{Q}$	$\frac{1}{N}\sum_{i=1}^{N}\frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})}\left(I_y(Y) - \widehat{P}(y \mid x, Z)\right) + \widehat{P}(y \mid x, Z)$ where training and evaluation of $\widehat{P}$ are done with two distinct sets of samples ("***Cross fitting***")
For correct estimation	("Doubly robustness") **Either one of two nuisances** should be correctly **estimated**.
For $\sqrt{N}$-consistency	

Recent advance: DML estimator



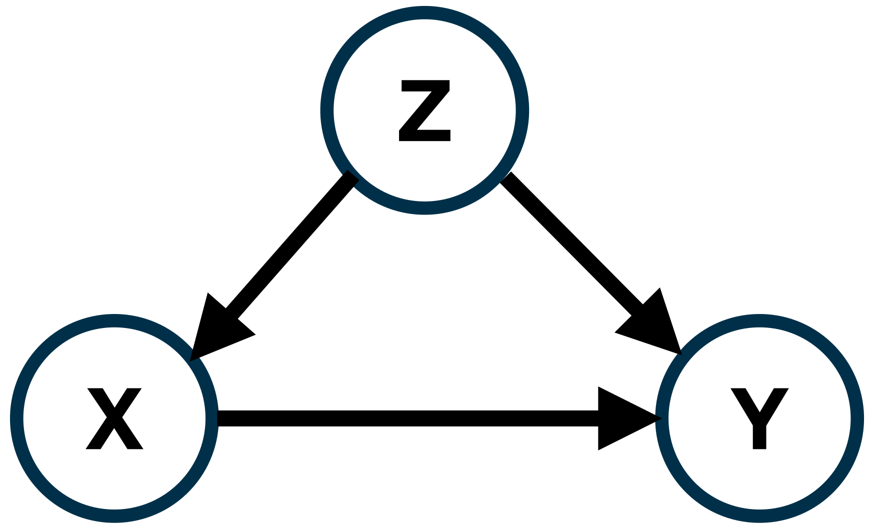
Backdoor graph

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

Double Machine Learning Estimator

Estimand Representation of Q	$\mathbb{E}\left[\frac{I_x(X)}{P(X \mid Z)}\left(I_y(Y) - P(y \mid x, Z)\right) + P(y \mid x, Z)\right]$	
Estimators Estimator $\widehat{Q}$	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\widehat{P}(X_{(i)} \mid Z_{(i)})}\left(I_y(Y) - \widehat{P}(y \mid x, Z)\right) + \widehat{P}(y \mid x, Z)$	where training and evaluation of $\widehat{P}$ are done with two distinct sets of samples ("***Cross fitting***")
For correct estimation	("Doubly robustness")	**Either one of two nuisances** should be correctly **estimated**.
For $\sqrt{N}$-consistency	("Debiasedness")	**Estimates** for **nuisance** converges at $o_P(N^{-1/4})$.

Recent advance: DML estimator

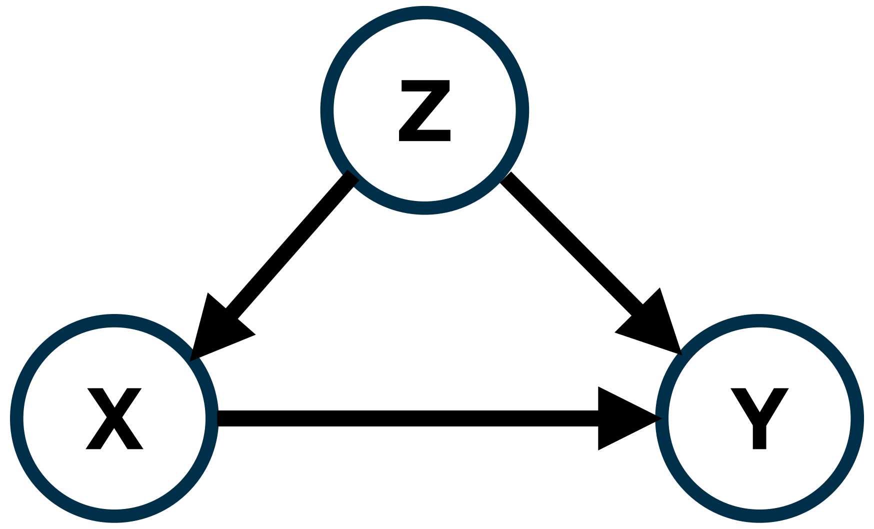


$$Q = \sum_z P(y \mid x, z) P(z)$$

DML estimator is robust!

Estimand Representation of Q		
Estimators Estimator $\widehat{Q}$		
For correct estimation	("Doubly robustness")	Either one of two nuisances should be correctly estimated.
For $\sqrt{N}$-consistency	("Debiasedness")	Estimates for nuisance converges at $o_P(N^{-1/4})$.

Recent advance: DML estimator



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

DML estimator is robust!

- Robust against **model misspecification — Either one of nuisance is misspecified, the estimate could be correct.**

Estimand

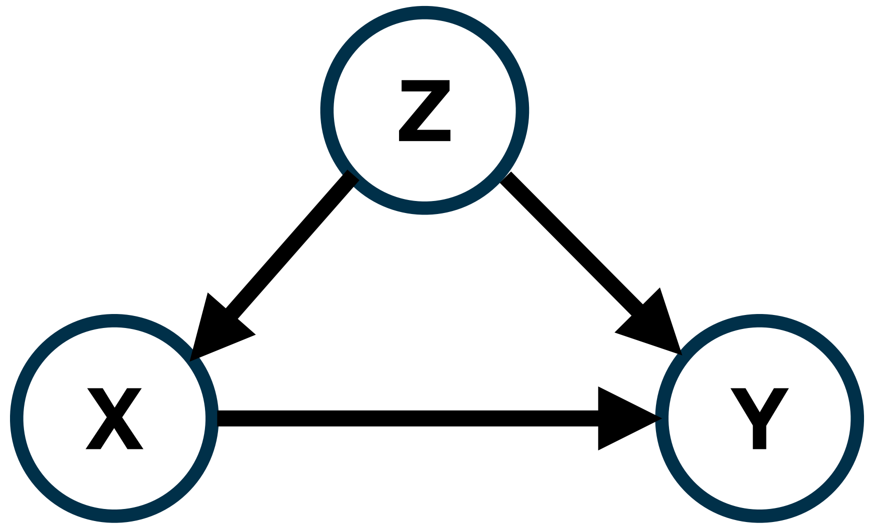
Representation of Q

Estimators

Estimator $\widehat{Q}$

For correct estimation	("Doubly robustness")	**Either one of two nuisances** should be correctly **estimated**.
For $\sqrt{N}$-consistency	("Debiasedness")	**Estimates** for **nuisance** converges at $o_P(N^{-1/4})$.

Recent advance: DML estimator



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

DML estimator is robust!

- Robust against **model misspecification — Either one of nuisance is misspecified, the estimate could be correct.**
- Robust against **slow convergence** — $\sqrt{N}$-consistency even when nuisance converges slowly.

Estimand

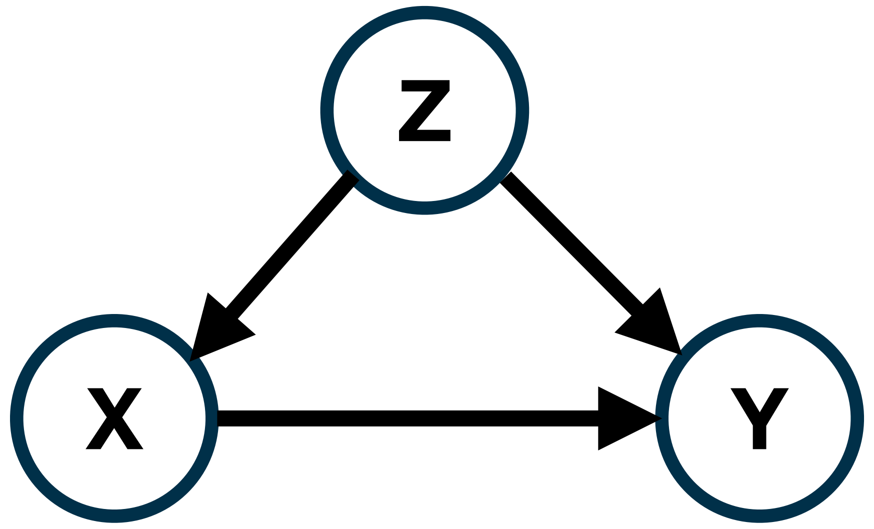
Representation of Q

Estimators

Estimator $\widehat{Q}$

For correct estimation	("Doubly robustness")	**Either one of two nuisances** should be correctly **estimated**.
For $\sqrt{N}$-consistency	("Debiasedness")	**Estimates** for **nuisance** converges at $o_P(N^{-1/4})$.

Recent advance: DML estimator



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

Estimand

Representation of Q

Estimators

Estimator $\widehat{Q}$

DML estimator is robust!

- Robust against **model misspecification** — **Either one of nuisance is misspecified, the estimate could be correct.**
- Robust against **slow convergence** — $\sqrt{N}$-consistency even when nuisance converges slowly.

Is DML estimators applicable for estimating any identifiable causal functional?

Preliminary — DML estimator (Chernozhukov, 2016)

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 is a true nuisance, and
2. (**Orthogonality**) $(\partial / \partial \eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right] \Big|_{\eta = \eta_0} = 0.$

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 is a true nuisance, and
2. (**Orthogonality**) $(\partial/\partial\eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right]|_{\eta=\eta_0} = 0$.

DML estimator

An estimator for ψ that is composed of $\widehat{\eta}$, such that

1. (**Neyman orthogonal score**) the estimand is based on Neyman orthogonal score; and
2. (**Cross-fitting**) training and evaluating $\widehat{\eta}$ is done with two distinct sets of samples.

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 is a true nuisance, and
2. (**Orthogonality**) $(\partial/\partial\eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right]|_{\eta=\eta_0} = 0.$

DML estimator

An estimator for ψ that is composed of $\widehat{\eta}$, such that

1. (**Neyman orthogonal score**) the estimand is based on Neyman orthogonal score; and
2. (**Cross-fitting**) training and evaluating $\widehat{\eta}$ is done with two distinct sets of samples.

Debiasedness

- *Without any smoothness & complexity constraints on the model for* $\widehat{\eta}$, DML estimator achieves debiasedness property:
  - (**Debiasedness**) $\sqrt{N}$-consistent whenever $\widehat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 is a true nuisance, and
2. (**Orthogonality**) $(\partial / \partial \eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right] \rvert_{\eta=\eta_0} = 0.$

DML estimator

An estimator for ψ that is composed of $\widehat{\eta}$, such that

1. (**Neyman orthogonal score**) the estimand is based on Neyman orthogonal score; and
2. (**Cross-fitting**) training and evaluating $\widehat{\eta}$ is done with two distinct sets of samples.

Debiasedness

- *Without any smoothness & complexity constraints on the model for* $\widehat{\eta}$, DML estimator achieves debiasedness property:
  - (**Debiasedness**) $\sqrt{N}$-consistent whenever $\widehat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 a true nuisance, and
2. (**Orthogonality**) $(\partial / \partial \eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right] \rvert_{\eta = \eta_0} = 0.$

DML estimator

An estimator for ψ that is composed of $\widehat{\eta}$, such that

1. (Neyman orthogonal score) the estimand

Debiasedness

- *Without any smoothness & complexity constraints on the model for $\widehat{\eta}$*, DML estimator achieves debiasedness property:
  - (**Debiasedness**) $\sqrt{N}$-consistent whenever $\widehat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

For achieving $\sqrt{N}$-consistency

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 a true nuisance, and
2. (**Orthogonality**) $(\partial / \partial \eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right] |_{\eta = \eta_0} = 0.$

DML estimator

An estimator for ψ that is composed of $\widehat{\eta}$, such that

1. (**Neyman orthogonal score**) the estimand

For achieving $\sqrt{N}$-consistency

- Classic BD estimators requires $\widehat{\eta}$ to be in the "***Donsker***" function class, which restricts smoothness / complexity of estimators.

Debiasedness

- *Without any smoothness & complexity constraints on the model for* $\widehat{\eta}$, DML estimator achieves debiasedness property:
  - (**Debiasedness**) $\sqrt{N}$-consistent whenever $\widehat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 a true nuisance, and
2. (**Orthogonality**) $(\partial / \partial \eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right]|_{\eta=\eta_0} = 0$.

DML estimator

An estimator for ψ that is composed of $\hat{\eta}$, such that

1. (Neyman orthogonal score) the estimand

For achieving $\sqrt{N}$-consistency

- Classic BD estimators requires $\hat{\eta}$ to be in the "***Donsker***" function class, which restricts smoothness / complexity of estimators.
- DML estimator doesn't have such constraints in achieving $\sqrt{N}$-consistency

Debiasedness

- *Without any smoothness & complexity constraints on the model for* $\hat{\eta}$, DML estimator achieves debiasedness property:
  - (**Debiasedness**) $\sqrt{N}$-consistent whenever $\hat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 a true nuisance, and
2. (**Orthogonality**) $(\partial / \partial \eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right]\big|_{\eta = \eta_0} = 0$.

DML estimator

An estimator for ψ that is composed of $\hat{\eta}$, such that

1. (Neyman orthogonal score) the estimand

Debiasedness

- *Without any smoothness & complexity constraints on the model for* $\hat{\eta}$, DML estimator achieves debiasedness property:
  - (**Debiasedness**) $\sqrt{N}$-consistent whenever $\hat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

For achieving $\sqrt{N}$-consistency

- Classic BD estimators requires $\hat{\eta}$ to be in the "***Donsker***" function class, which restricts smoothness / complexity of estimators.
- DML estimator doesn't have such constraints in achieving $\sqrt{N}$-consistency

	Classic	DML
Model class	Donsker	Arbitrary
Permitted class	• Linear regression, • Logistic regression	• Any regressors in Donsker • Neural networks
Forbidden class	Random forest, Lasso, deep neural networks.	x

Preliminary — DML estimator (Chernozhukov, 2016)

Neyman orthogonal score ϕ

For the target estimand ψ (e.g., $\psi = P(y \mid do(x))$) and nuisances η (e.g., $\eta = \{P(y \mid x, z), P(x \mid z)\}$), a function $\phi(\mathbf{V}; \psi, \eta)$ is called a ***Neyman orthogonal score*** if

1. (**Moment condition**) $\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta_0)\right] = 0$ where η_0 a true nuisance, and
2. (**Orthogonality**) $(\partial / \partial \eta)\mathbb{E}_P\left[\phi(\mathbf{V}; \psi, \eta)\right] |_{\eta = \eta_0} = 0.$

DML estimator

An estimator for ψ that is composed of $\hat{\eta}$, such that

1. (Neyman orthogonal score) the estimand

Debiasedness

- *Without any smoothness & complexity constraints on the model for* $\hat{\eta}$, DML estimator achieves debiasedness property:
  - (**Debiasedness**) $\sqrt{N}$-consistent whenever $\hat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

For achieving $\sqrt{N}$-consistency

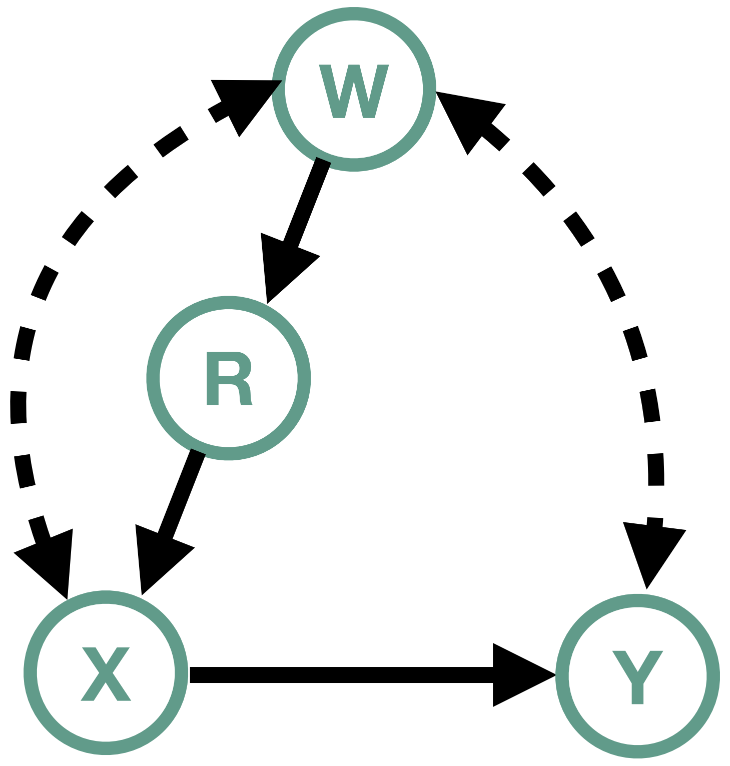
- Classic BD estimators requires $\hat{\eta}$ to be in the "***Donsker***" function class, which restricts smoothness / complexity of estimators.
- DML estimator doesn't have such constraints in achieving $\sqrt{N}$-consistency

💪 Power of DML

One can employ *any* complicated ML model and enjoy debiasedness

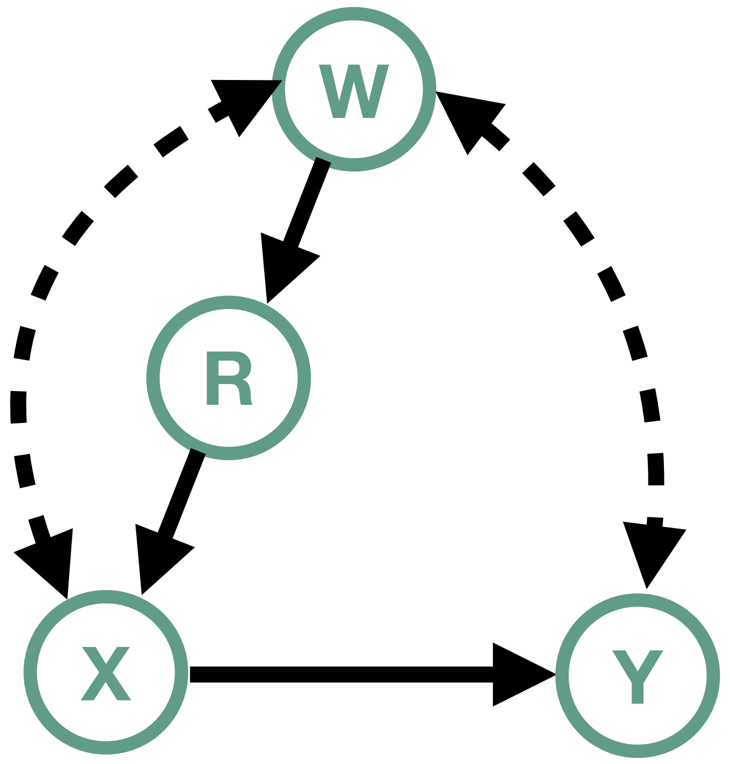
Problem Setup

Problem Setup

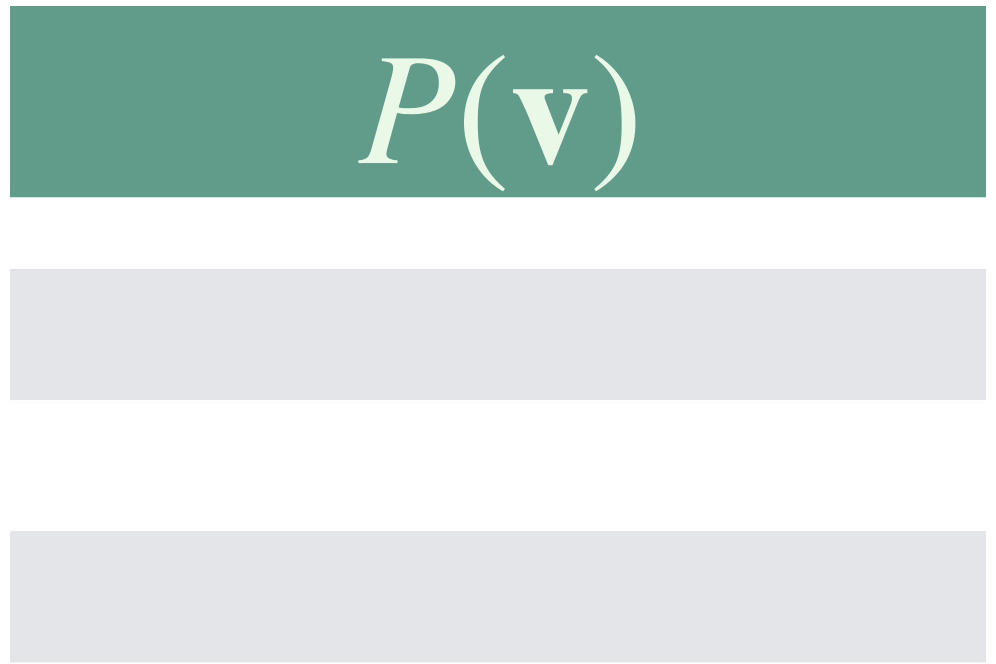


Graph G

Problem Setup

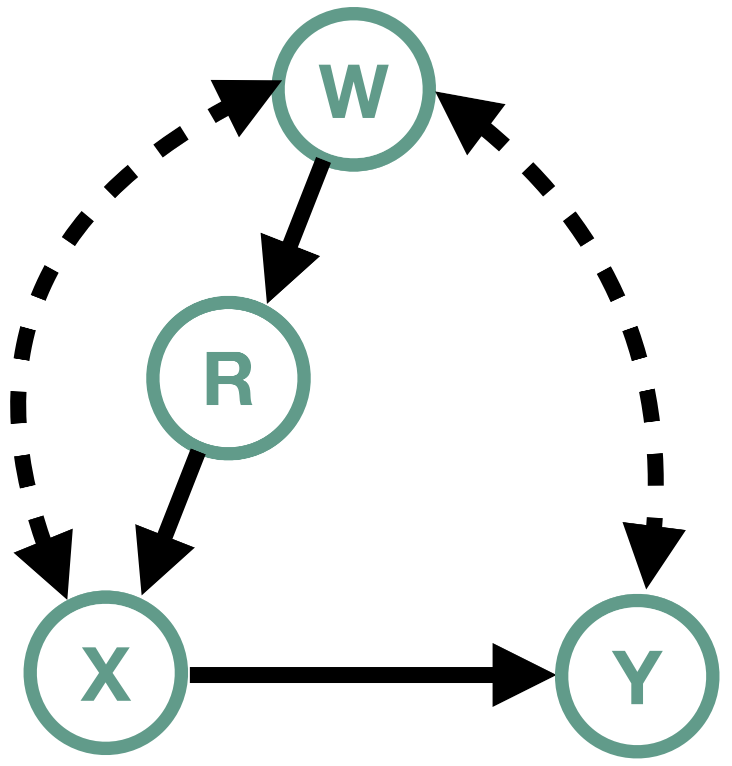


Graph G

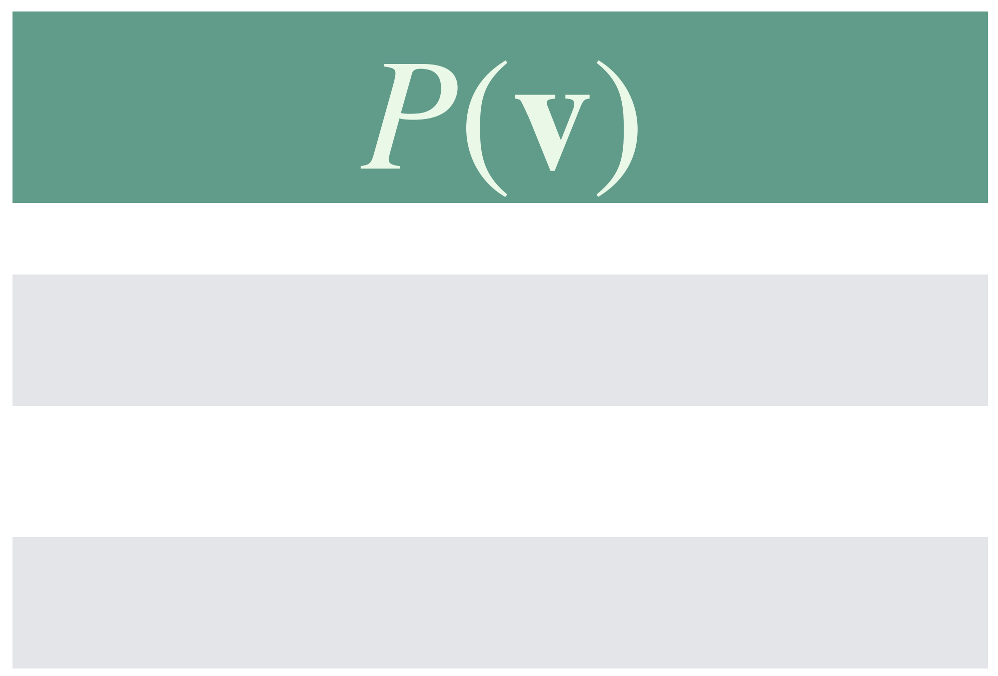


Samples D from joint distribution $P(\mathbf{v})$

Problem Setup



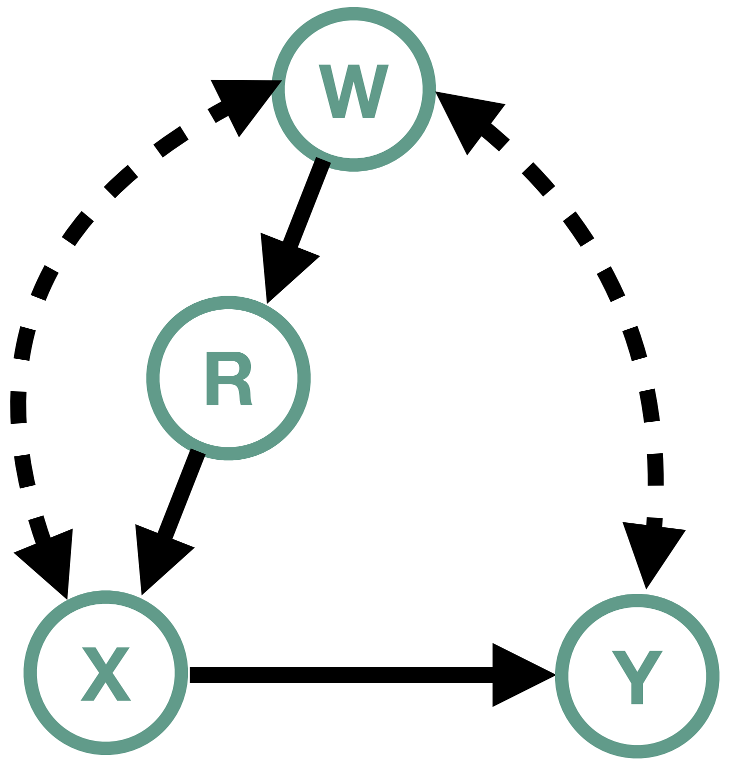
Graph G



Samples D from joint distribution $P(\mathbf{v})$

Query (Q) = $P(y \mid do(x))$

Problem Setup



Graph G

$P(\mathbf{v})$

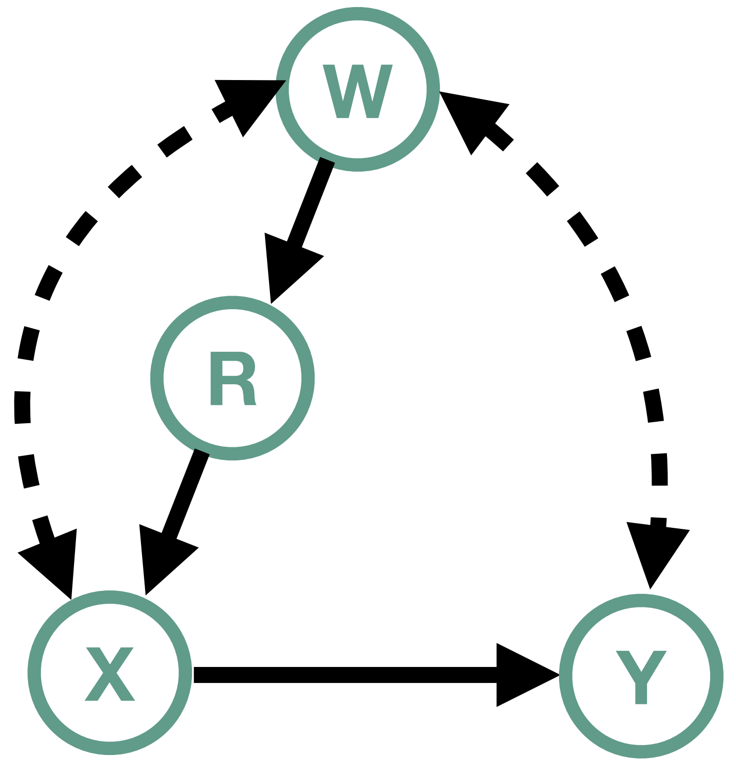
Samples D from joint distribution $P(\mathbf{v})$

Query (Q) = $P(y \mid do(x))$

Research question

Construct a DML estimator $\widehat{Q}$ that is

Problem Setup



Graph G

$P(\mathbf{v})$

Samples D from joint distribution $P(\mathbf{v})$

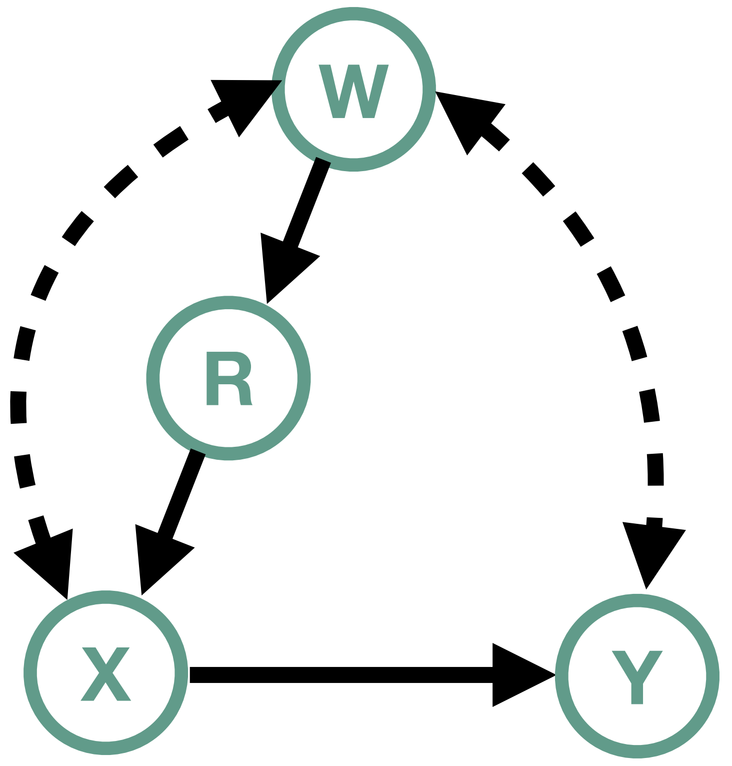
Query (Q) = $P(y \mid do(x))$

Research question

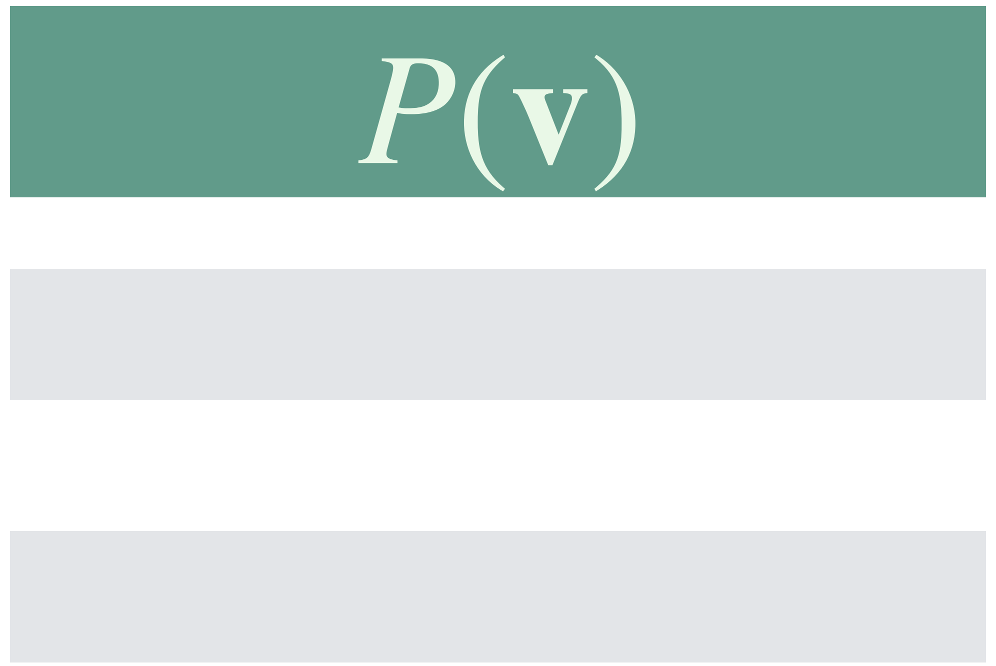
Construct a DML estimator $\widehat{Q}$ that is

- robust against model misspecification (**doubly robust**) and slow convergence (**debiased**); and

Problem Setup



Graph G



Samples D from joint distribution $P(\mathbf{v})$

Query (Q) = $P(y \mid do(x))$

Research question

Construct a DML estimator $\widehat{Q}$ that is

- robust against model misspecification (**doubly robust**) and slow convergence (**debiased**); and
- working for any identifiable causal functional. (**Complete**)

Related works

	Statistical property		Causal property	
	Doubly robustness	Debiasedness	Beyond BD	Any identifiable

Related works

	Statistical property		Causal property	
	Doubly robustness	Debiasedness	Beyond BD	Any identifiable
Jung, Tian, Bareinboim (2020a)	✗	✗	✔	✗

Related works

	Statistical property		Causal property	
	Doubly robustness	Debiasedness	Beyond BD	Any identifiable
Jung, Tian, Bareinboim (2020a)	✗	✗	✓	✗
Fulcher et al (2019), Bhattacharya et al (2020)	✓	✗	✓	✗

Related works

	Statistical property		Causal property	
	Doubly robustness	Debiasedness	Beyond BD	Any identifiable
Jung, Tian, Bareinboim (2020a)	✗	✗	✓	✗
Fulcher et al (2019), Bhattacharya et al (2020)	✓	✗	✓	✗
Jung, Tian, Bareinboim (2020b)	✗	✗	✓	✓

Related works

	Statistical property		Causal property	
	Doubly robustness	**Debiasedness**	**Beyond BD**	**Any identifiable**
Jung, Tian, Bareinboim (2020a)	✗	✗	✓	✗
Fulcher et al (2019), Bhattacharya et al (2020)	✓	✗	✓	✗
Jung, Tian, Bareinboim (2020b)	✗	✗	✓	✓
DML-ID (Jung, Tian, Bareinboim, 2021)	✓	✓	✓	✓

Our Approach for marrying DML + Identification (DML-ID)

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

- Construct a DML estimator working for any identifiable causal functional.

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

- Construct a DML estimator working for any identifiable causal functional.

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

- Construct a DML estimator working for any identifiable causal functional.

Result 0. mSBD — Notion

Result 0. mSBD — Notion

(Informal) **Multi-outcome Sequential Back-door (mSBD)**

$\mathbf{Z} = \{\mathbf{Z}_1, \cdots, \mathbf{Z}_n\}$ satisfies *mSBD criterion relative to* $\{\mathbf{X}, \mathbf{Y}\}$ if non-causal path b/w $X_i \in \mathbf{X}$ and $\mathbf{Y}_i \in \mathbf{Y}$ are blocked by $\mathbf{Z}_i$ conditioned on $\{\mathbf{X}^{(i-1)}, \mathbf{Y}^{(i-1)}, \mathbf{Z}^{(i-1)}\}$.

(**Implication**): No unmeasured confounders b/w $\mathbf{X}$ and $\mathbf{Y}$.

Result 0. mSBD — Notion

(Informal) **Multi-outcome Sequential Back-door (mSBD)**

$\mathbf{Z} = \{\mathbf{Z}_1, \cdots, \mathbf{Z}_n\}$ satisfies *mSBD criterion relative to* $\{\mathbf{X}, \mathbf{Y}\}$ if non-causal path b/w $X_i \in \mathbf{X}$ and $\mathbf{Y}_i \in \mathbf{Y}$ are blocked by $\mathbf{Z}_i$ conditioned on $\{\mathbf{X}^{(i-1)}, \mathbf{Y}^{(i-1)}, \mathbf{Z}^{(i-1)}\}$.

(**Implication**): No unmeasured confounders b/w $\mathbf{X}$ and $\mathbf{Y}$.

mSBD adjustment

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{\mathbf{V}_i \in \{\mathbf{Y}, \mathbf{Z}\}} P(\mathbf{v}_i \mid \mathbf{v}^{(i-1)}),$$

- for $\mathbf{V}_i = \mathbf{Y}_i, \mathbf{V}^{(i-1)} = \{\mathbf{X}^{(i)}, \mathbf{Z}^{(i)}, \mathbf{Y}^{(i-1)}\}$; and
- for $\mathbf{V}_i = \mathbf{Z}_i, \mathbf{V}^{(i-1)} = \{\mathbf{Y}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{Z}^{(i-1)}\}$.

Result 0. mSBD — Notion

(Informal) **Multi-outcome Sequential Back-door (mSBD)**

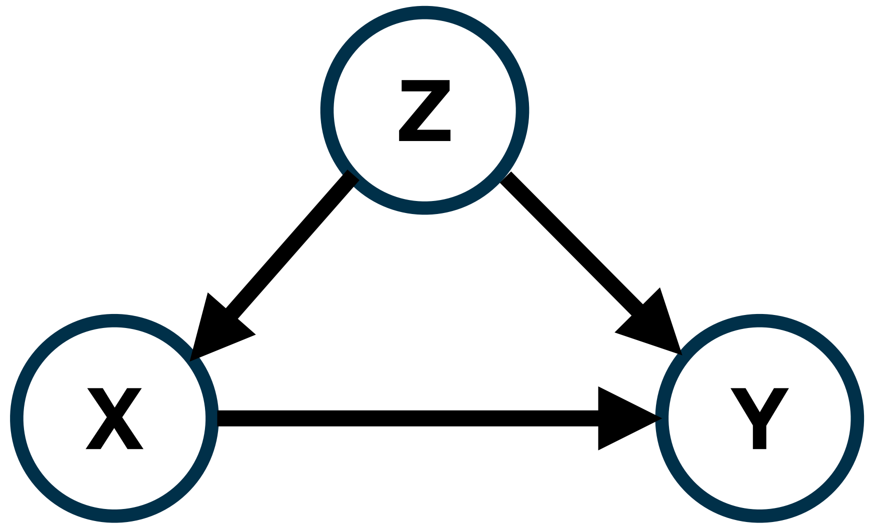
$\mathbf{Z} = \{\mathbf{Z}_1, \cdots, \mathbf{Z}_n\}$ satisfies *mSBD criterion relative to* $\{\mathbf{X}, \mathbf{Y}\}$ if non-causal path b/w $X_i \in \mathbf{X}$ and $\mathbf{Y}_i \in \mathbf{Y}$ are blocked by $\mathbf{Z}_i$ conditioned on $\{\mathbf{X}^{(i-1)}, \mathbf{Y}^{(i-1)}, \mathbf{Z}^{(i-1)}\}$.

(**Implication**): No unmeasured confounders b/w $\mathbf{X}$ and $\mathbf{Y}$.

mSBD adjustment

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{\mathbf{V}_i \in \{\mathbf{Y}, \mathbf{Z}\}} P(\mathbf{v}_i \mid \mathbf{v}^{(i-1)}),$$

- for $\mathbf{V}_i = \mathbf{Y}_i, \mathbf{V}^{(i-1)} = \{\mathbf{X}^{(i)}, \mathbf{Z}^{(i)}, \mathbf{Y}^{(i-1)}\}$; and
- for $\mathbf{V}_i = \mathbf{Z}_i, \mathbf{V}^{(i-1)} = \{\mathbf{Y}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{Z}^{(i-1)}\}$.



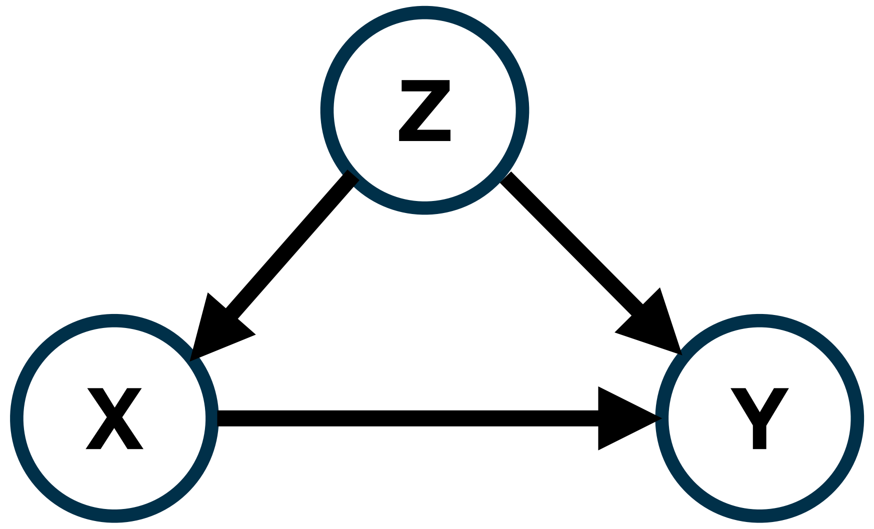
$$P(y \mid do(x)) = \sum_{z} P(y \mid x, z) P(z)$$

Result 0. mSBD — Notion

(Informal) **Multi-outcome Sequential Back-door (mSBD)**

$\mathbf{Z} = \{\mathbf{Z}_1, \cdots, \mathbf{Z}_n\}$ satisfies *mSBD criterion relative to* $\{\mathbf{X}, \mathbf{Y}\}$ if non-causal path b/w $X_i \in \mathbf{X}$ and $\mathbf{Y}_i \in \mathbf{Y}$ are blocked by $\mathbf{Z}_i$ conditioned on $\{\mathbf{X}^{(i-1)}, \mathbf{Y}^{(i-1)}, \mathbf{Z}^{(i-1)}\}$.

(**Implication**): No unmeasured confounders b/w $\mathbf{X}$ and $\mathbf{Y}$.

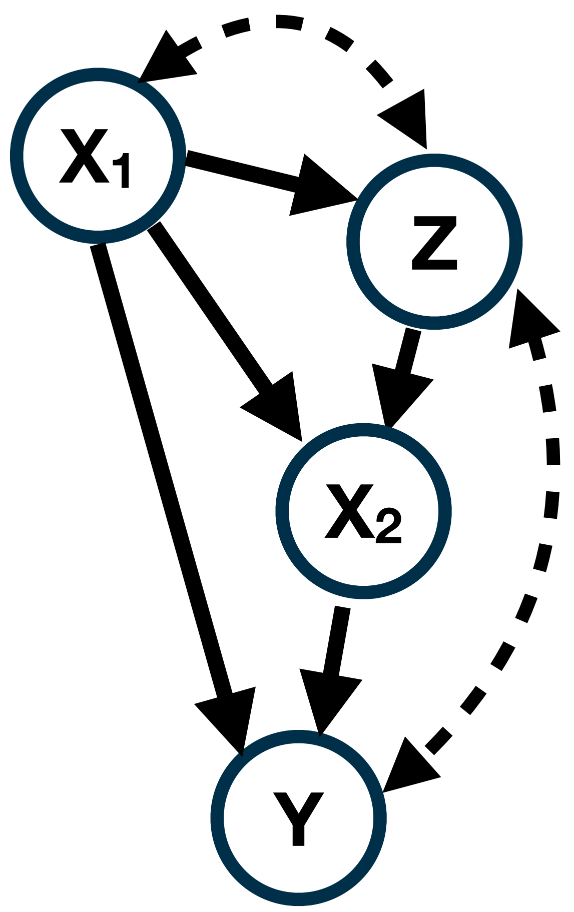


$$P(y \mid do(x)) = \sum_z P(y \mid x, z) P(z)$$

mSBD adjustment

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{V_i \in \{\mathbf{Y}, \mathbf{Z}\}} P(\mathbf{v}_i \mid \mathbf{v}^{(i-1)}),$$

- for $\mathbf{V}_i = \mathbf{Y}_i, \mathbf{V}^{(i-1)} = \{\mathbf{X}^{(i)}, \mathbf{Z}^{(i)}, \mathbf{Y}^{(i-1)}\}$; and
- for $\mathbf{V}_i = \mathbf{Z}_i, \mathbf{V}^{(i-1)} = \{\mathbf{Y}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{Z}^{(i-1)}\}$.



$$P(y \mid do(x)) = \sum_z P(y \mid x_2, z, x_1) P(z \mid x_1)$$

Result 0. mSBD — Neyman Orthogonal score & DML estimator

Result 0. mSBD — Neyman Orthogonal score & DML estimator

Neyman orthogonal score for mSBD

$$\phi(\mathbf{V};\psi,\eta) = \sum_{i=1}^{n} W_i \left(H_{i+1} - H_i\right), \text{ where}$$

$$H_i = P_{\mathbf{x}}(\mathbf{y}^{\geq i-1} \mid \mathbf{Z}^{(i-1)}, \mathbf{y}^{(i-2)}) \text{ and}$$

$$W_i = \prod_{p=1}^{i} \frac{I_{x_p}(X_p)}{P(x_p \mid \mathbf{Z}^{(p)}, \mathbf{x}^{(p-1)}, \mathbf{y}^{(p-1)})}$$

Result 0. mSBD — Neyman Orthogonal score & DML estimator

Neyman orthogonal score for mSBD

$$\phi(\mathbf{V};\psi,\eta) = \sum_{i=1}^{n} W_i \left(H_{i+1} - H_i\right), \text{ where}$$

$$H_i = P_{\mathbf{x}}(\mathbf{y}^{\geq i-1} \mid \mathbf{Z}^{(i-1)}, \mathbf{y}^{(i-2)}) \text{ and}$$

$$W_i = \prod_{p=1}^{i} \frac{I_{x_p}(X_p)}{P(x_p \mid \mathbf{Z}^{(p)}, \mathbf{x}^{(p-1)}, \mathbf{y}^{(p-1)})}$$

DML estimator for mSBD

An estimator based on the above **Neyman orthogonal score** with **cross-fitting** is a DML estimator that is

Result 0. mSBD — Neyman Orthogonal score & DML estimator

Neyman orthogonal score for mSBD

$$\phi(\mathbf{V};\psi,\eta) = \sum_{i=1}^{n} W_i \left(H_{i+1} - H_i\right), \text{ where}$$

$$H_i = P_{\mathbf{x}}(\mathbf{y}^{\geq i-1} \mid \mathbf{Z}^{(i-1)}, \mathbf{y}^{(i-2)}) \text{ and}$$

$$W_i = \prod_{p=1}^{i} \frac{I_{x_p}(X_p)}{P(x_p \mid \mathbf{Z}^{(p)}, \mathbf{x}^{(p-1)}, \mathbf{y}^{(p-1)})}$$

DML estimator for mSBD

An estimator based on the above **Neyman orthogonal score** with **cross-fitting** is a DML estimator that is

- (**Doubly robust**) consistent whenever nuisances H_i or W_i are correctly estimated; and

Result 0. mSBD — Neyman Orthogonal score & DML estimator

Neyman orthogonal score for mSBD

$$\phi(\mathbf{V};\psi,\eta) = \sum_{i=1}^{n} W_i \left(H_{i+1} - H_i\right),$$ where

$$H_i = P_{\mathbf{x}}(\mathbf{y}^{\geq i-1} \mid \mathbf{Z}^{(i-1)}, \mathbf{y}^{(i-2)})$$ and

$$W_i = \prod_{p=1}^{i} \frac{I_{x_p}(X_p)}{P(x_p \mid \mathbf{Z}^{(p)}, \mathbf{x}^{(p-1)}, \mathbf{y}^{(p-1)})}$$

DML estimator for mSBD

An estimator based on the above **Neyman orthogonal score** with **cross-fitting** is a DML estimator that is

- (**Doubly robust**) consistent whenever nuisances H_i or W_i are correctly estimated; and
- (**Debiased**) $\sqrt{N}$-consistent whenever nuisances H_i and W_i converges at $N^{-1/4}$ rate.

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

- Construct a DML estimator working for any identifiable causal functional.

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

- Construct a DML estimator working for any identifiable causal functional.

Result 1. Revisit ID algorithm

Result 1. Revisit ID algorithm

C-component & C-factor

- **C-component $\mathbf{C}$**: A set of variables s.t. connected by bidirected path.
- **C-factor $Q[\mathbf{C}]$**: A interventional distribution of $\mathbf{C}$ under the intervention $\mathbf{V} \backslash \mathbf{C}$; i.e., $Q[\mathbf{C}] \equiv P(\mathbf{C} \mid do(\mathbf{V} \backslash \mathbf{C}))$

Result 1. Revisit ID algorithm

C-component & C-factor

- **C-component** $\mathbf{C}$: A set of variables s.t. connected by bidirected path.
- **C-factor** $Q[\mathbf{C}]$: A interventional distribution of $\mathbf{C}$ under the intervention $\mathbf{V}\backslash\mathbf{C}$; i.e., $Q[\mathbf{C}] \equiv P(\mathbf{C} \mid do(\mathbf{V}\backslash\mathbf{C}))$

Complete ID algorithm (Tian and Pearl, 2003).

- A causal functional $P(\mathbf{y} \mid do(\mathbf{x}))$ is identifiable if and only if it is represented as an arithmetic of C-factors of C-component $\mathbf{C}_i$ in G (Tian and Pearl, 2003, Huang and Valtorta 2006).

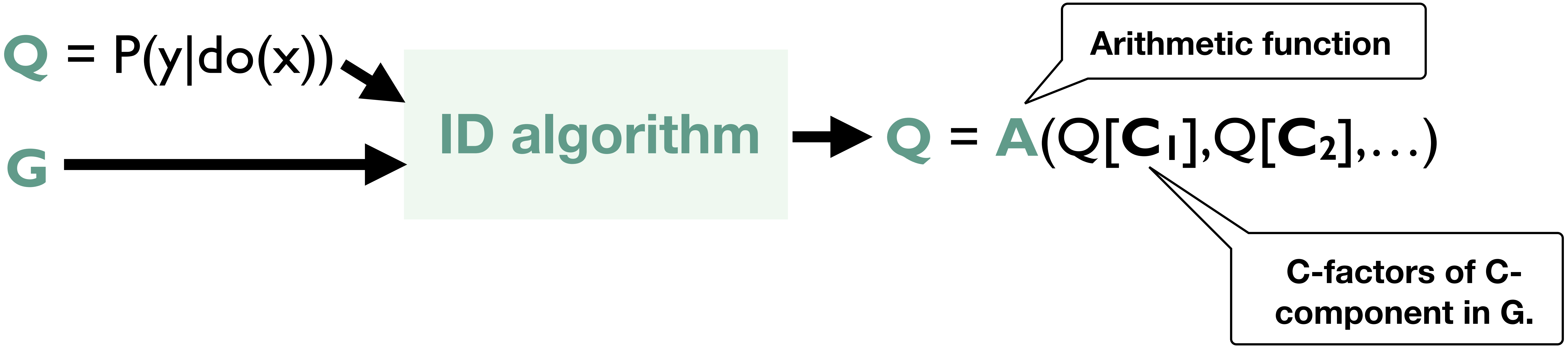
Result 1. Revisit ID algorithm

C-component & C-factor

- **C-component** $\mathbf{C}$: A set of variables s.t. connected by bidirected path.
- **C-factor** $Q[\mathbf{C}]$: A interventional distribution of $\mathbf{C}$ under the intervention $\mathbf{V} \setminus \mathbf{C}$; i.e., $Q[\mathbf{C}] \equiv P(\mathbf{C} \mid do(\mathbf{V} \setminus \mathbf{C}))$

Complete ID algorithm (Tian and Pearl, 2003).

- A causal functional $P(\mathbf{y} \mid do(\mathbf{x}))$ is identifiable if and only if it is represented as an arithmetic of C-factors of C-component $\mathbf{C}_i$ in G (Tian and Pearl, 2003, Huang and Valtorta 2006).



Result 1. DML-ID—Mechanism

Result 1. DML-ID—Mechanism

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

G

ID algorithm

$Q = A(Q[C_1], Q[C_2], \ldots)$

Arithmetic function

C-factors in (G,P)

Result 1. DML-ID—Mechanism

Q = P(y|do(x))

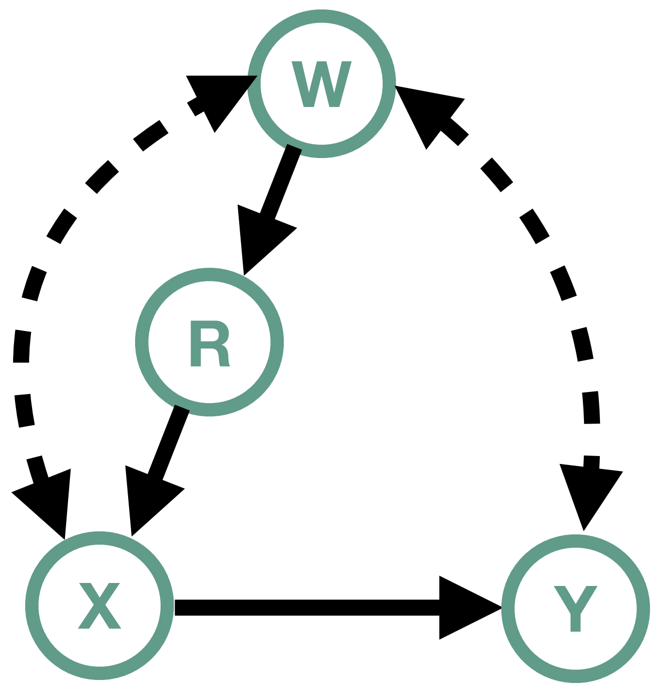
G

ID algorithm

$\mathbf{Q} = \mathbf{A}(Q[\mathbf{C}_1], Q[\mathbf{C}_2], \ldots)$

Arithmetic function

C-factors in (G,P)



$P(y \mid do(x))$

ID algo.

$$\mathbf{Q} = \frac{\sum_w Q[W, X, Y]}{\sum_{w,y} Q[W, X, Y]} = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

Result 1. DML-ID—Mechanism

Lemma: C-factor as mSBD

C-factors of C-component is identifiable and given by mSBD adjustment.

Q = P(y|do(x))

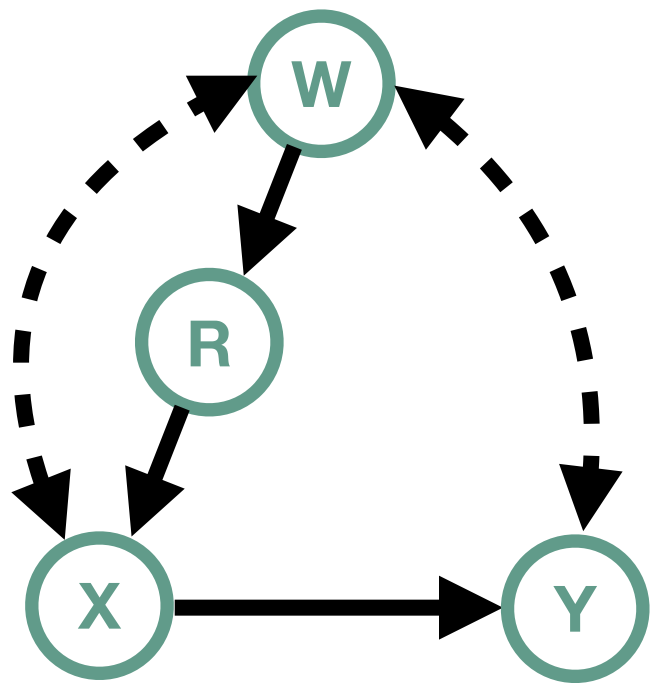
G

ID algorithm

Arithmetic function

$\mathbf{Q} = \mathbf{A}(Q[\mathbf{C}_1], Q[\mathbf{C}_2], \ldots)$

C-factors in (G,P)



$P(y \mid do(x))$

ID algo.

$$\mathbf{Q} = \frac{\sum_w Q[W, X, Y]}{\sum_{w,y} Q[W, X, Y]} = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

Result 1. DML-ID—Mechanism

Lemma: C-factor as mSBD

C-factors of C-component is identifiable and given by mSBD adjustment.

Q = P(y|do(x))

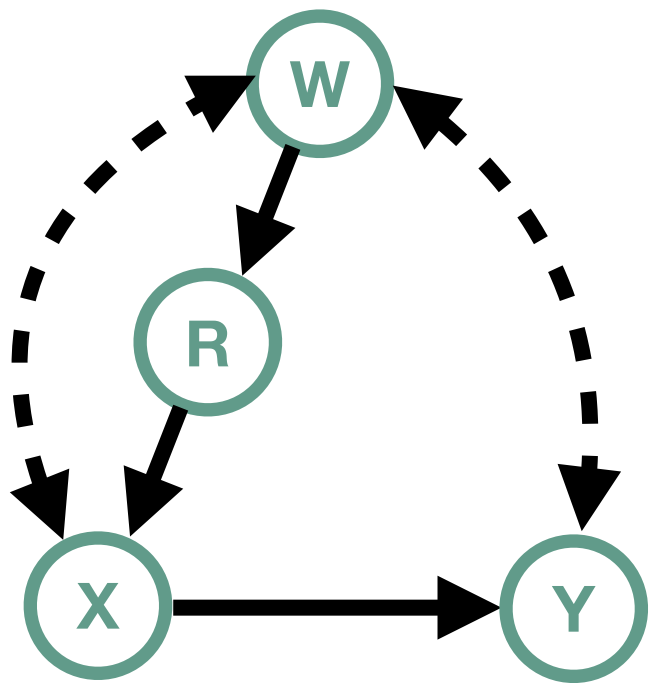
G

DML-ID

Arithmetic function

Q = **A**(M[**C**$_1$],M[**C**$_2$],…)

mSBD adjustment.



$P(y \mid do(x))$

ID algo.

$$\mathbf{Q} = \frac{\sum_w Q[W,X,Y]}{\sum_{w,y} Q[W,X,Y]} = \frac{\sum_w P(x,y \mid r,w)P(w)}{\sum_w P(x \mid r,w)P(w)}$$

Result 1. DML-ID—Mechanism

Lemma: C-factor as mSBD

C-factors of C-component is identifiable and given by mSBD adjustment.

Q = P(y|do(x))

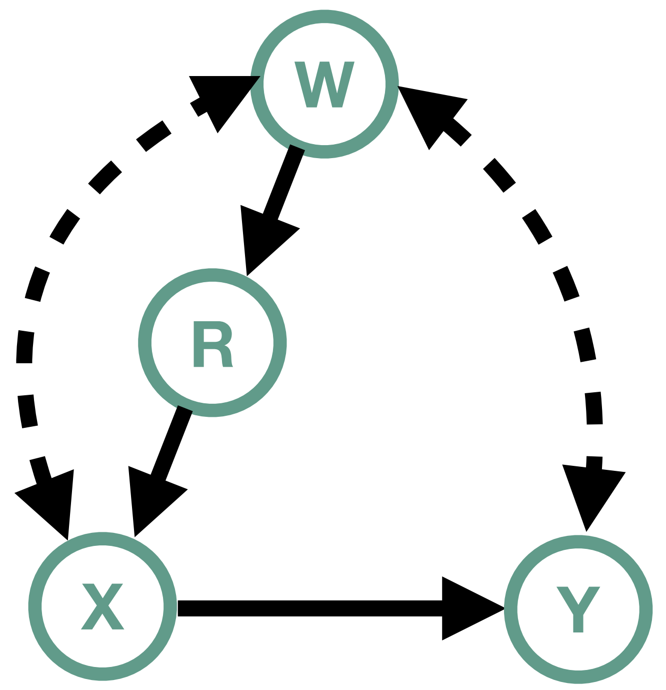
G

DML-ID

→ **Q** = **A**(M[**C**$_1$],M[**C**$_2$],…)

Arithmetic function

mSBD adjustment.



$P(y \mid do(x))$

DML-ID

→ **Q** $= \dfrac{M[(x,y) \mid r; w]}{M[x \mid r; w]} = \dfrac{\sum_w P(x,y \mid r,w)P(w)}{\sum_w P(x \mid r,w)P(w)}$

mSBD adjustment of $P(x,y \mid r,w)$ **over W**

Result 1. Derivation of Neyman Orthogonal Score

Result 1. Derivation of Neyman Orthogonal Score

Thm. 2: Derivation of Neyman Orthogonal score

Given representation of $\mathbf{Q} = \mathbf{A}(M[\mathbf{C}_1], M[\mathbf{C}_2], \ldots, M[\mathbf{C}_d])$, a Neyman orthogonal score is given as

$$\sum_{i=1}^{d} \phi_{M_i} \frac{\partial}{\partial M_i} A\left(M[\mathbf{C}_1], \cdots, M[\mathbf{C}_d]\right)$$

Neyman orthogonal score for the mSBD adjustment $M_i = M[\mathbf{C}_i]$

Result 1. Derivation of Neyman Orthogonal Score

Thm. 2: Derivation of Neyman Orthogonal score

Given representation of $\mathbf{Q} = \mathbf{A}(M[\mathbf{C}_1], M[\mathbf{C}_2], \ldots, M[\mathbf{C}_d])$, a Neyman orthogonal score is given as

$$\sum_{i=1}^{d} \phi_{M_i} \frac{\partial}{\partial M_i} A\left(M[\mathbf{C}_1], \cdots, M[\mathbf{C}_d]\right)$$

Neyman orthogonal score for the mSBD adjustment $M_i = M[\mathbf{C}_i]$

$\mathbf{Q} = P(y|do(x))$

$\mathbf{G}$

DML-ID

$\mathbf{Q} = \mathbf{A}(M[\mathbf{C}_1], M[\mathbf{C}_2], \ldots)$

Arithmetic function

mSBD adjustments

Score

Neyman orthogonal score

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

- Construct a DML estimator working for any identifiable causal functional.

Our Approach for marrying DML + Identification (DML-ID)

Result 0. (mSBD)

- Derive the Neyman orthogonal score when there are no unmeasured confounders b/w (**X**, **Y**) (called *multi-outcome sequential back-door (mSBD)*).

Result 1. (DML-ID)

- Represent any identifiable causal estimand Q=F(P) as an arithmetic (multiplication, ratio, marginalization) of mSBDs.
- Derive the Neyman orthogonal score for any arbitrary causal functional in a form of an arithmetic of scores of mSBDs.

Result 2. (DML estimator)

- Construct a DML estimator working for any identifiable causal functional.

Result 2 — DML Estimator

Result 2 — DML Estimator

Construction of DML estimator

Given derived Neyman orthogonal score, an estimator for $P(y \mid do(x))$ is constructed by

Result 2 — DML Estimator

Construction of DML estimator

Given derived Neyman orthogonal score, an estimator for $P(y \mid do(x))$ is constructed by

1. (**Neyman orthogonal score)** using the derived Neyman orthogonal score for estimating the estimand $P(y \mid do(x))$, where

Result 2 — DML Estimator

Construction of DML estimator

Given derived Neyman orthogonal score, an estimator for $P(y \mid do(x))$ is constructed by

1. (**Neyman orthogonal score)** using the derived Neyman orthogonal score for estimating the estimand $P(y \mid do(x))$, where
2. (**Cross-fitting**) training and evaluating $\hat{\eta}$ are done with two distinct sets of samples.

Result 2 — DML Estimator

Construction of DML estimator

Given derived Neyman orthogonal score, an estimator for $P(y \mid do(x))$ is constructed by

1. (**Neyman orthogonal score)** using the derived Neyman orthogonal score for estimating the estimand $P(y \mid do(x))$, where
2. (**Cross-fitting**) training and evaluating $\widehat{\eta}$ are done with two distinct sets of samples.

Properties

A proposed estimator is

Result 2 — DML Estimator

Construction of DML estimator

Given derived Neyman orthogonal score, an estimator for $P(y \mid do(x))$ is constructed by

1. (**Neyman orthogonal score)** using the derived Neyman orthogonal score for estimating the estimand $P(y \mid do(x))$, where
2. (**Cross-fitting**) training and evaluating $\widehat{\eta}$ are done with two distinct sets of samples.

Properties

A proposed estimator is

- robust against model misspecification (**doubly robust**) and slow convergence (**debiased**); and

Result 2 — DML Estimator

Construction of DML estimator

Given derived Neyman orthogonal score, an estimator for $P(y \mid do(x))$ is constructed by

1. (**Neyman orthogonal score)** using the derived Neyman orthogonal score for estimating the estimand $P(y \mid do(x))$, where
2. (**Cross-fitting**) training and evaluating $\widehat{\eta}$ are done with two distinct sets of samples.

Properties

A proposed estimator is

- robust against model misspecification (**doubly robust**) and slow convergence (**debiased**); and
- working for any identifiable causal functional. (**Complete**)

Result 2 — DML Estimator

Construction of DML estimator

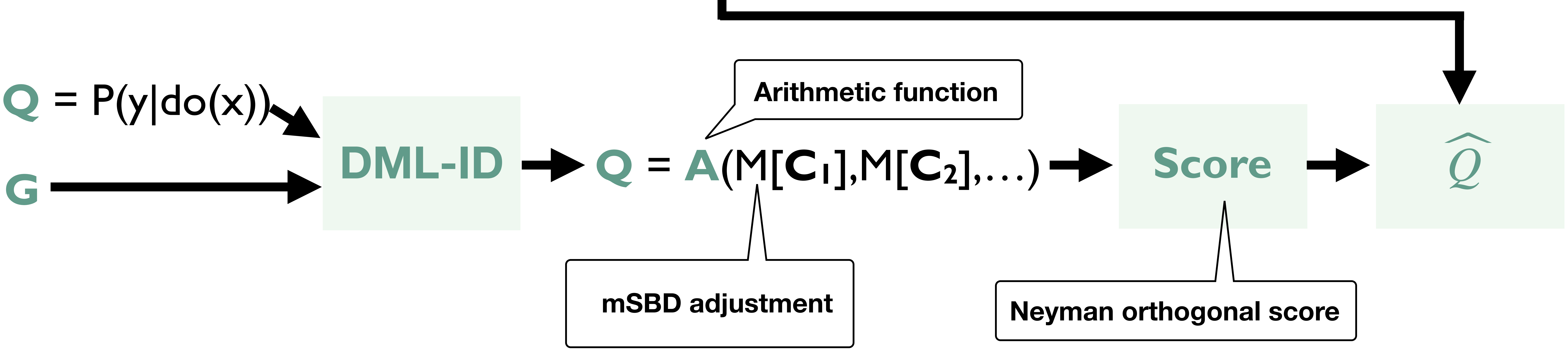
Given derived Neyman orthogonal score, an estimator for $P(y \mid do(x))$ is constructed by

1. (**Neyman orthogonal score)** using the derived Neyman orthogonal score for estimating the estimand $P(y \mid do(x))$, where
2. (**Cross-fitting**) training and evaluating $\widehat{\eta}$ are done with two distinct sets of samples.

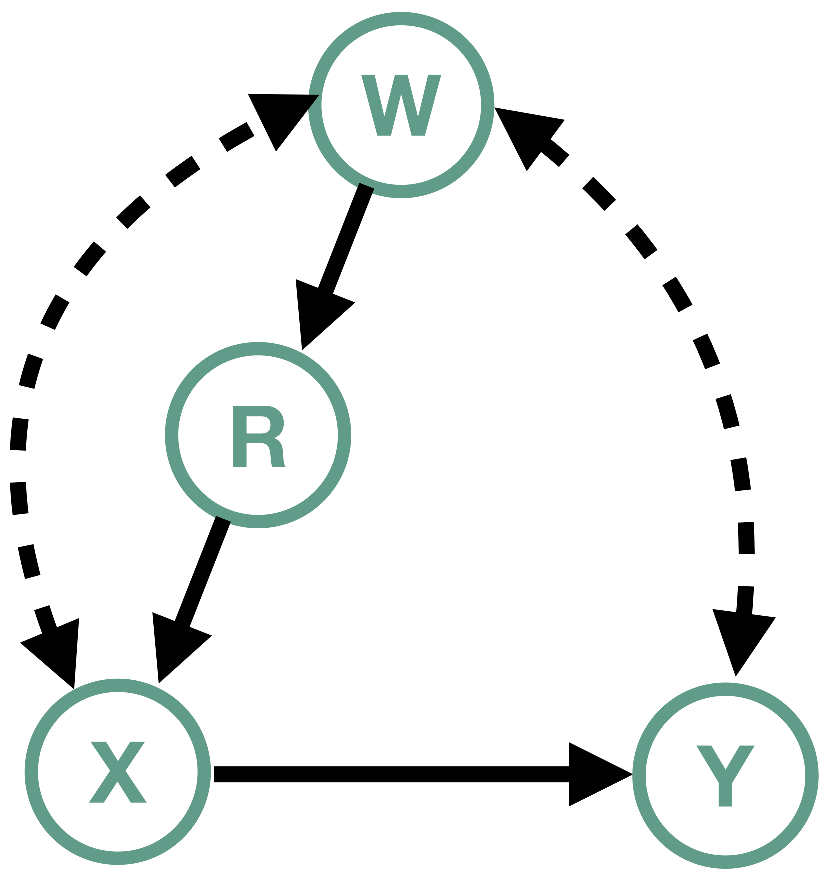
Properties

A proposed estimator is

- robust against model misspecification (**doubly robust**) and slow convergence (**debiased**); and
- working for any identifiable causal functional. (**Complete**)

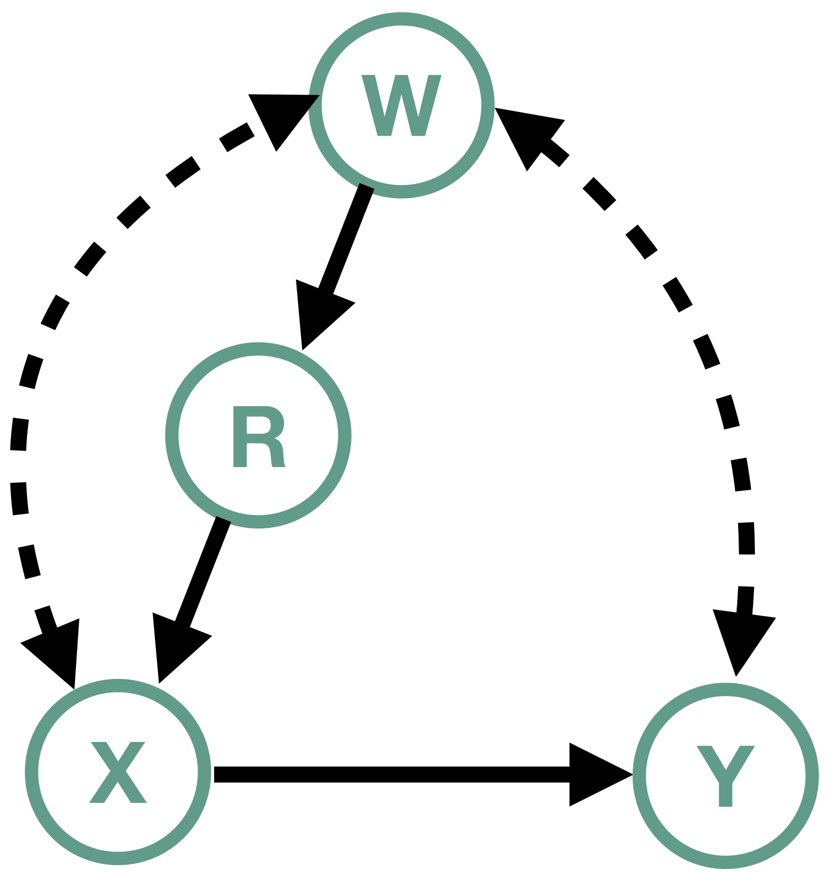


Result 2 — Empirical evidence



$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w) P(w)}{\sum_w P(x \mid r, w) P(w)}$$

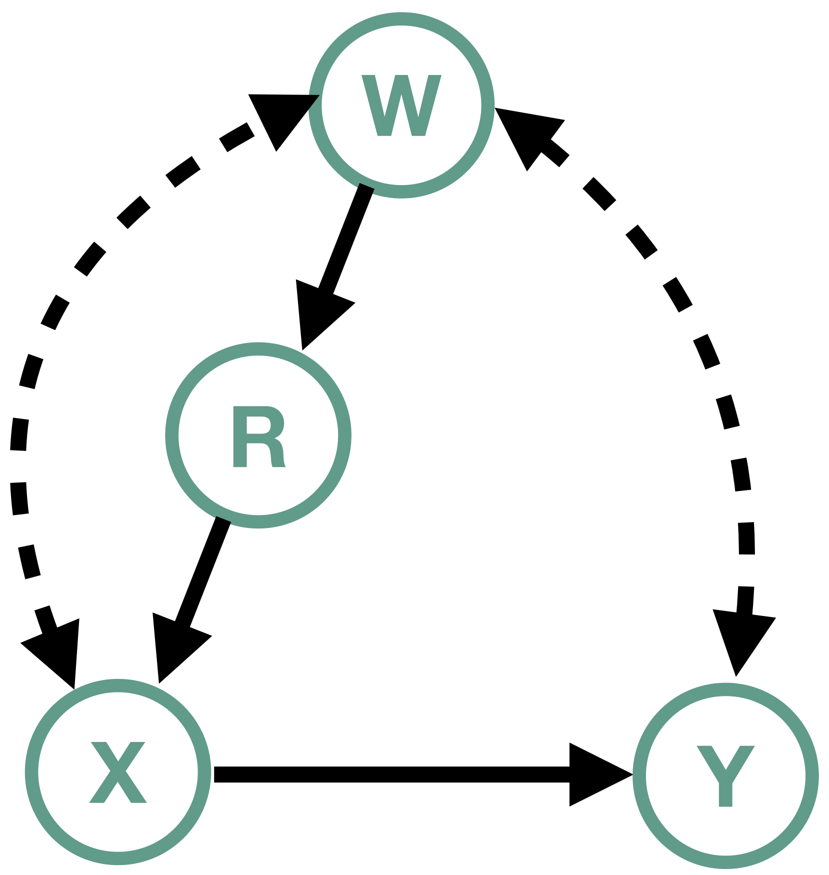
Result 2 — Empirical evidence



$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)}$$

- We compare this estimator with **Plug-in estimator**, which estimates cond.prob $\{P(x, y \mid r, w), P(x \mid r, w), P(w)\}$ and plugs those in for computing the estimand.

Result 2 — Empirical evidence



$$P(y \mid do(x)) = \frac{\sum_w P(x, y \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)}$$

- We compare this estimator with **Plug-in estimator**, which estimates cond.prob $\{P(x, y \mid r, w), P(x \mid r, w), P(w)\}$ and plugs those in for computing the estimand.
- A reason that we compare with the plug-in estimator is because this is only viable estimator that is working for arbitrary causal functional.

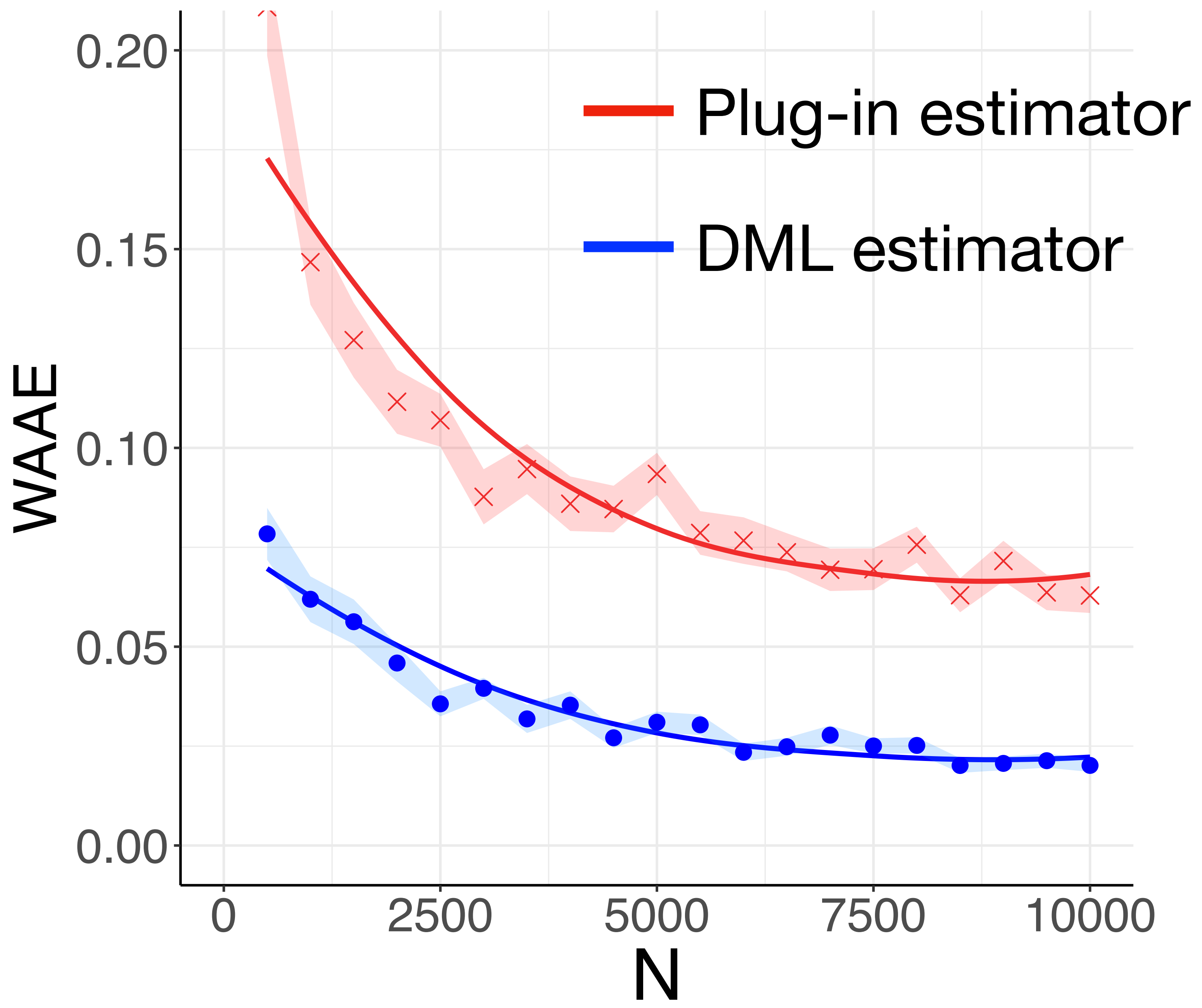
Result 2 — Empirical evidence

Debiasedness	Doubly robustness	

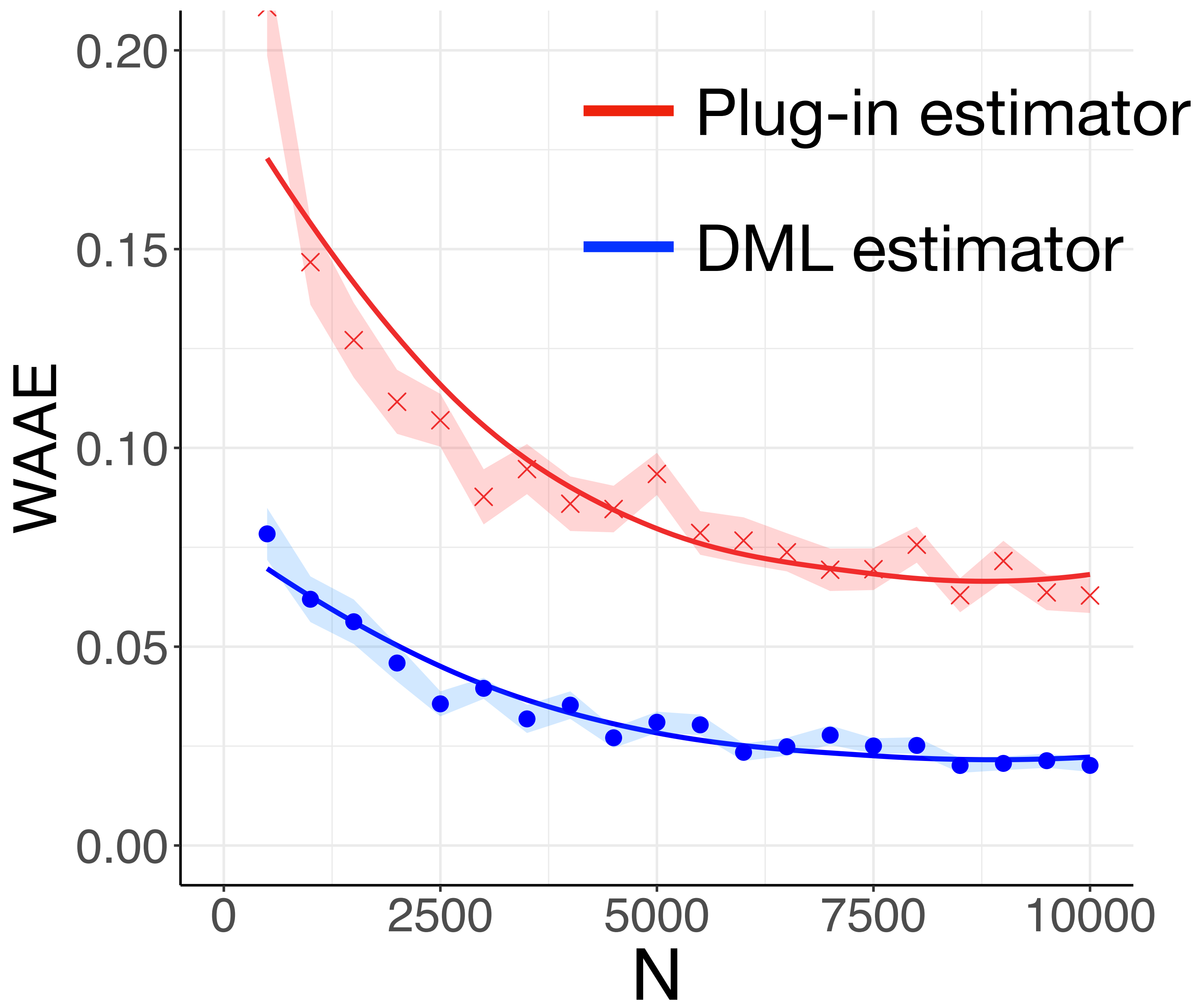
Result 2 — Empirical evidence

Debiasedness	Doubly robustness	
$\widehat{P}(x, y \mid r, w),\ \widehat{P}(r \mid w)$ converges to true $P(x, y \mid r, w), P(r \mid w)$ at a rate $N^{-1/4}$		

Result 2 — Empirical evidence

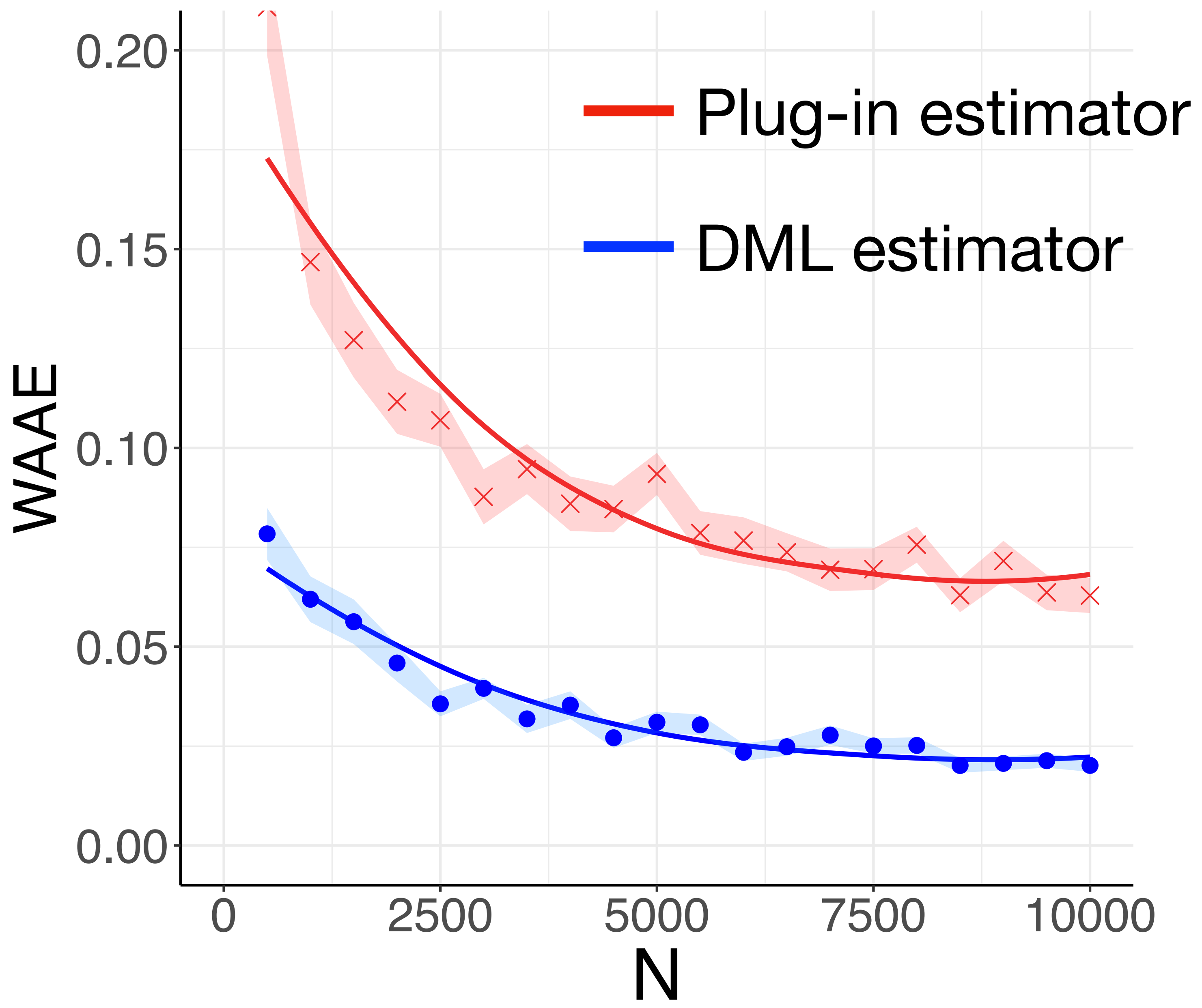
Debiasedness	Doubly robustness	
$\widehat{P}(x, y \mid r, w)$, $\widehat{P}(r \mid w)$ converges to true $P(x, y \mid r, w), P(r \mid w)$ at a rate $N^{-1/4}$		
		

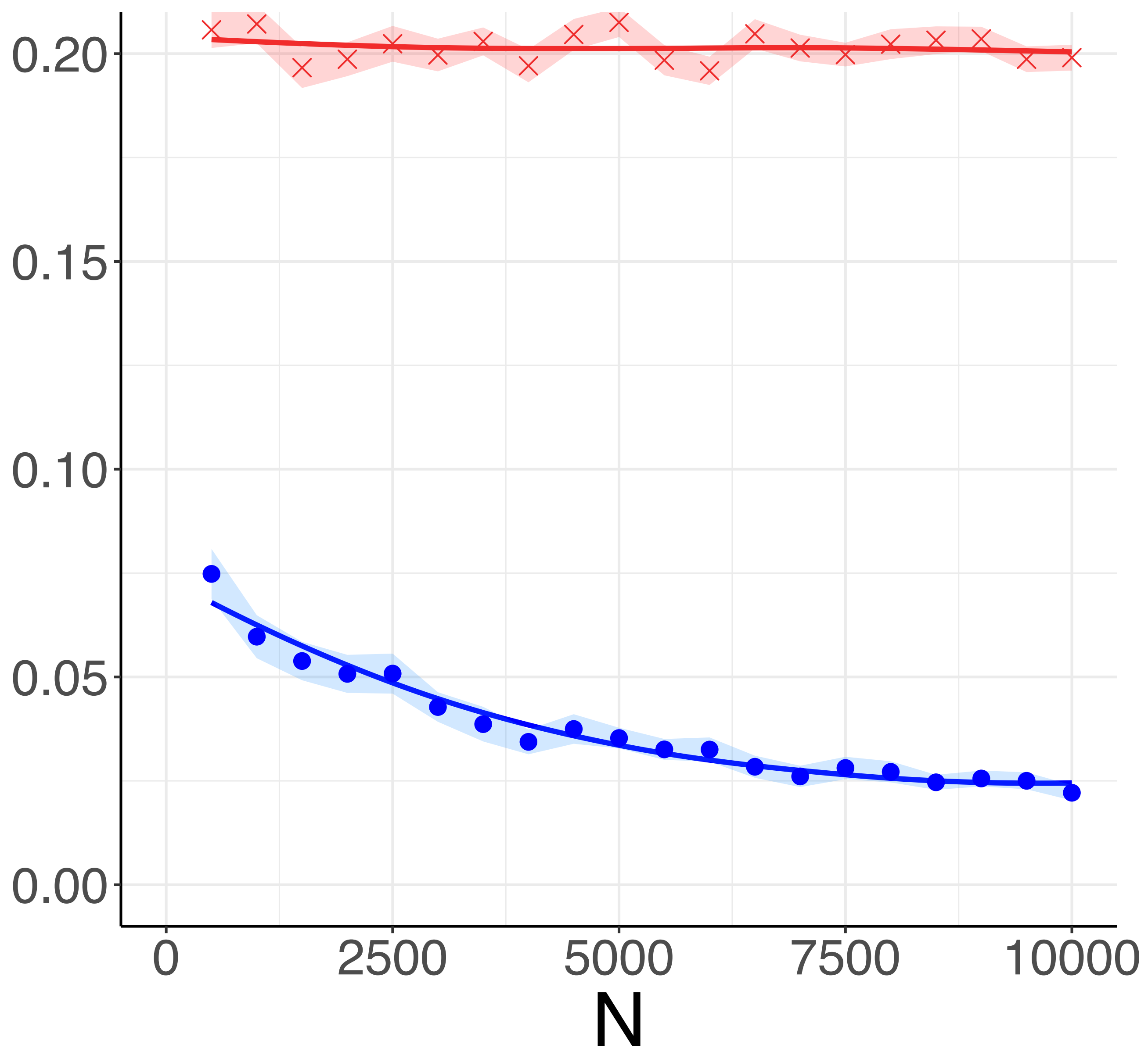
Result 2 — Empirical evidence

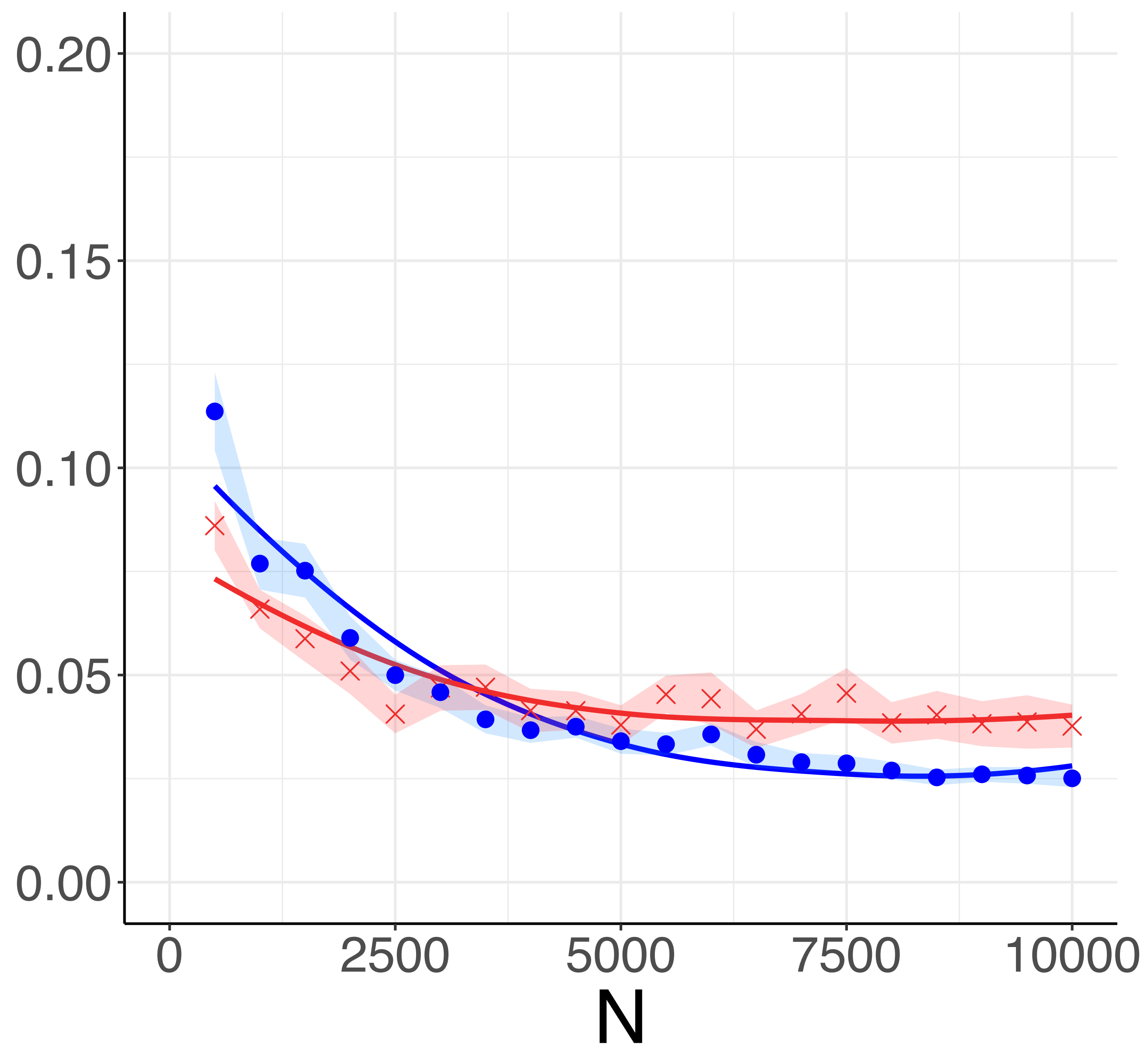
Debiasedness	Doubly robustness	
$\widehat{P}(x, y \mid r, w)$, $\widehat{P}(r \mid w)$ converges to true $P(x, y \mid r, w), P(r \mid w)$ at a rate $N^{-1/4}$	$P(x, y \mid r, w)$ misspecified.	$P(r \mid w)$ misspecified.
		

Result 2 — Empirical evidence

Debiasedness	Doubly robustness	
$\widehat{P}(x, y \mid r, w)$, $\widehat{P}(r \mid w)$ converges to true $P(x, y \mid r, w), P(r \mid w)$ at a rate $N^{-1/4}$	$P(x, y \mid r, w)$ misspecified.	$P(r \mid w)$ misspecified.



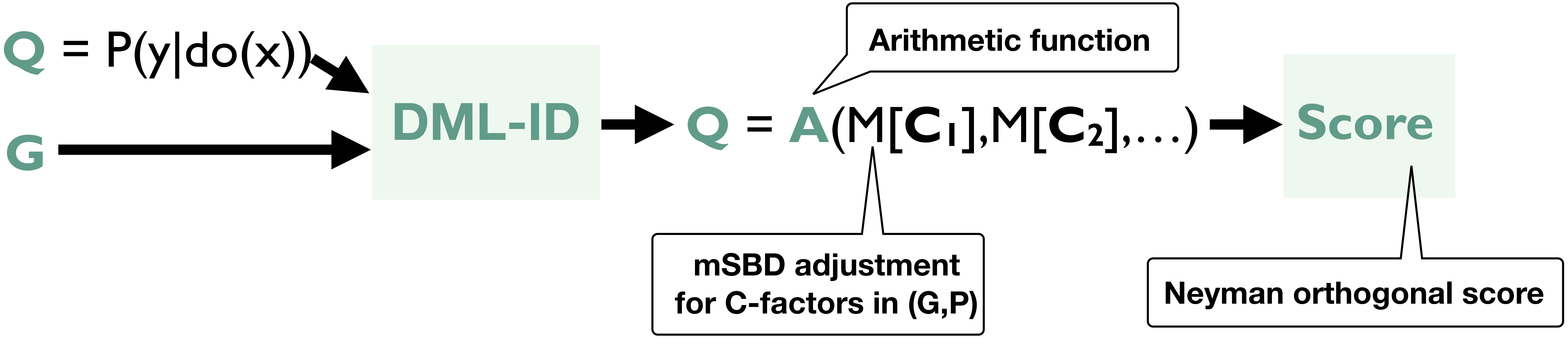




Conclusion

Conclusion

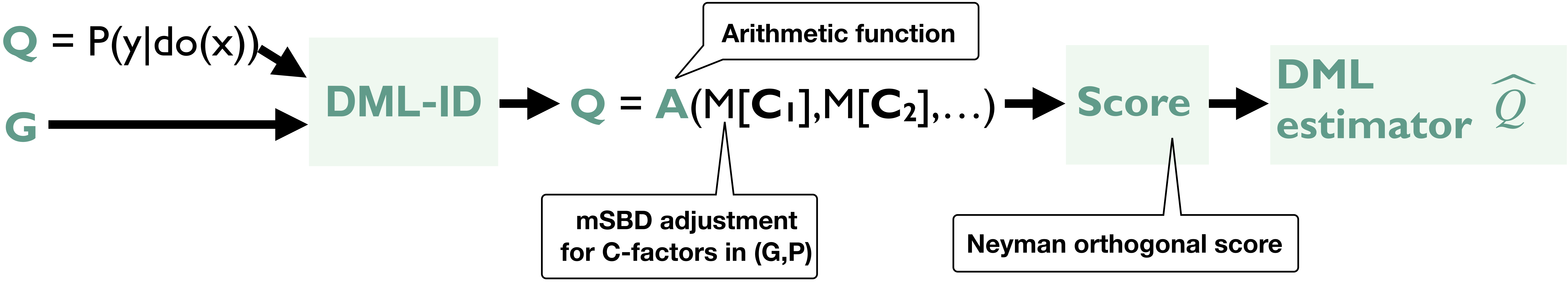
- **Result 1** — We develop a systematic procedure for deriving Neyman orthogonal score for estimands of any identifiable causal effects.



Conclusion

- **Result 1** — We develop a systematic procedure for deriving Neyman orthogonal score for estimands of any identifiable causal effects.

- **Result 2** — We develop DML estimators for any identifiable causal effect, which enjoy debiasedness and doubly robustness against model misspecification and bias.



Thank you for listening!