Estimating Identifiable Causal Effects through Double Machine Learning

Yonghan Jung , Jin Tian , Elias Bareinboim

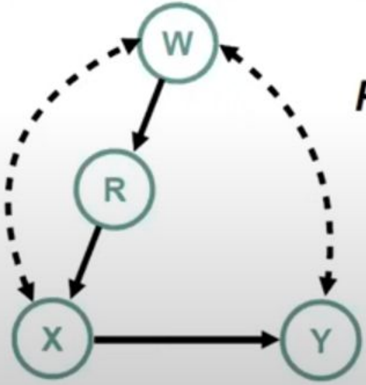
CS295 Presentation by: Chase Overcash (covercas@uci.edu)

Estimating Causal Effects

- Causal Effect P(y|do(x))
  - The effects of doing Action X = x, on Y
- Desire: Compute Causal Effects from non experimental data
- Reason: Experiments can be:
  - Costly
  - Unethical
  - Technically Infeasible
- Two scenarios:
  - Graph-based
    - Causal Graph, G
    - Observational Data, D
  - Data-driven
    - Observational Data, D
    - Learns the Markovian equivalence to causal graphs

Causal Effect Identification

- Given a causal graph - a directed acyclic graph (DAG) with bi-directed edges representing unmeasured confounders



$$P(x, y, w, r) = \sum_{u1,u2} P(y|x, u1)P(x|r, u2)P(r|w)\, P(w|u1, u2)P(u1)P(u2)$$

$$P(y, w, r|do(x)) = \sum_{u1,u2} P(y|x, u1)P(r|w)\, P(w|u1, u2)\, P(u1)P(u2)$$

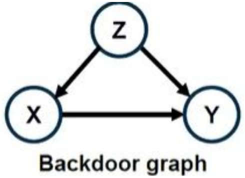
- Is P(y|do(x)) uniquely identifiable given G & an observational distribution, P(V)

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

Plug- in estimator: $P(y|do(x)) = \frac{\sum_w \hat{P}(x,y|r,w)\hat{P}(w)}{\sum_w \hat{P}(x|r,w)\hat{P}(w)}$

Backdoor (BD) Estimators

Backdoor Criterion:



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

	Inverse Probability Weight	Regression
Estimand (Q)	$\mathbb{E}\left[\frac{I_x(X)}{P(X\|Z)} I_y(Y)\right]$	$\mathbb{E}_{\mathbf{Z}}[P(y\|x,Z)]$
Estimator ($\hat{Q}$)	$\frac{1}{N}\sum_{i=1}^{N} \frac{I_x(X_{(i)})}{\hat{P}(X_{(i)}\|Z_{(i)})} I_y(Y_{(i)})$	$\frac{1}{N}\sum_{i=1}^{N} \hat{P}(y\|x,Z_{(i)})$
For correct estimation	Nuisance functional should be correctly estimated .	
For $\sqrt{N}$-consistency	Estimates for nuisances converge at $o_P(N^{-1/2})$.	

Risk of Classic Estimators

- Model misspecification => Incorrect Estimation
  - Complicated Data Generation
- Slow convergence => Not Sqrt(N) consistent
  - Machine Learning Models
- Standard Plug-in estimator for the general causal functional estimates suffer from same problems

Double/Debiased Machine Learning (DML) Estimator

Goal: For a given causal graph, develop DML style estimators for any identifiable causal effects

Assumptions: Discreteness, Positivity

Estimand Representation of Q	$\mathbb{E}\left[\frac{I_x(X)}{P(X\|Z)}(I_y(Y) - P(y\|x,Z)) + P(y\|x,Z)\right]$
Estimators Estimator $\hat{Q}$	$\frac{1}{N}\sum_{i=1}^{N}\frac{I_x(X_{(i)})}{\hat{P}(X_{(i)}\|Z_{(i)})}(I_y(Y) - \hat{P}(y\|x,Z_{(i)})) + \hat{P}(y\|x,Z_{(i)})$ where training of $\hat{P}$ and estimator evaluation 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 $o_P(N^{-1/4})$.

DML Recipe (Chernozukov 2016)

DML estimator

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

Neyman orthogonal score (NOS) ϕ

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

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

Construction of the DML estimator

- Let $\{D_0, D_1\}$ denote the randomly split halves of the dataset D. Let $\hat{\eta}_k$ denote the estimate of η from D_k for $k \in \{0,1\}$.
- Let T^k denote the solution satisfying $\mathbb{E}_{D_k}\left[\phi(\mathbf{V}; \hat{\eta}_{1-k}, T^k)\right] = o_P(N^{1/2})$ where $N \equiv |D|$, and $\mathbb{E}_{D_k}$ denote the empirical expectation over D_k.
- $T \equiv (T^0 + T^1)/2$ is a DML estimator.

Debiasedness

A DML estimator is $\sqrt{N}$-consistent whenever $\hat{\eta}$ converges to true nuisance at $N^{-1/4}$ rate.

Multi-outcome Sequential BD Criterion (mSBD)

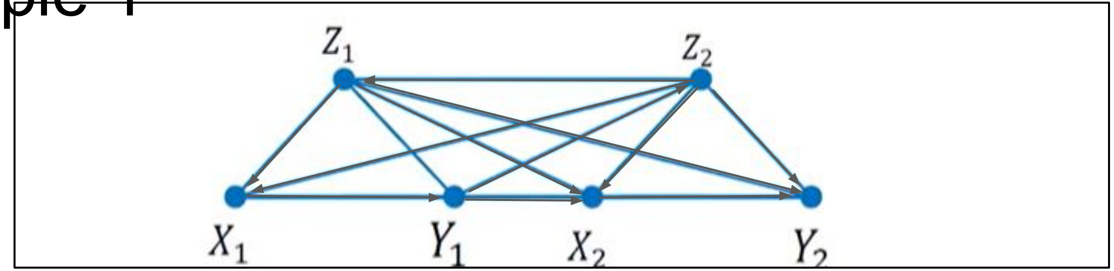
mSBD Criterion (informally): A sequential of Z variables satisfies the mSBD criterion relative to {X,Y} if a non-causal path between X_i in X and Y_i in Y are blocked by Z_i conditioned on the previous ones, $\{X^{i-1}, Y^{i-1}, Z^{i-1}\}$

mSBD adjustment (Jung et al 2020): If $\mathbf{Z} = \{Z_1, \cdots, Z_n\}$ satisfies the mSBD criterion relative to $(\mathbf{X}, \mathbf{Y})$,

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

$$P(\mathbf{y}|do(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{Y_i \in \mathbf{Y}} P(y_i|\mathbf{x}^{(i)}, \mathbf{z}^{(i)}, \mathbf{y}^{(i-1)}) \prod_{Z_i \in \mathbf{Z}} P(z_i|\mathbf{x}^{(i-1)}, \mathbf{z}^{(i-1)}, \mathbf{y}^{(i-1)}) \equiv M[\mathbf{y}|\mathbf{x};\mathbf{z}]$$

mSBD Example 1

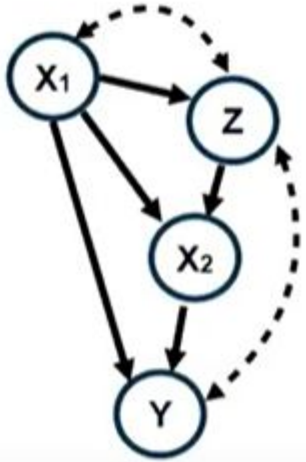


$\{Z_1, Z_2\}$ satisfies mSBD criterion relative to $\{X_1, X_2\}$, $\{Y_1, Y_2\}$

$P(y_1,y_2|\ do(x_1,x_2)) = \sum_{z1,z2}[P(z_1)P(y_1|x_1z_1)P(z_2|z_1x_1y_1)P(y_2|z_1,x_1,y_1,z_2,x_2)$

mSBD - Question

mSBD Criterion (informally): A sequential of Z variables satisfies the mSBD criterion relative to {X,Y} if a non-causal path between X_i in X and Y_i in Y are blocked by Z_i conditioned on the previous ones, $\{X^{i-1}, Y^{i-1}, Z^{i-1}\}$



Z = ?

P(y, do(x)) = ?

Result 1: mSBD Estimator

Neyman orthogonal score for mSBD

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

where

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

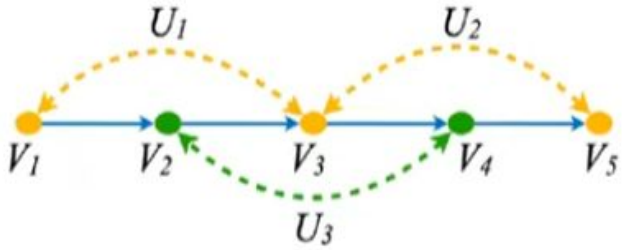
$$W_i = \prod_{p=1}^{i} \frac{I_{x_p}(X_p)}{P(x_p | \mathbf{Z}^{(p)}, \mathbf{x}^{(p-1)}, \mathbf{y}^{(p-1)})}$$

DML estimator for mSBD

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

$$P_{\mathbf{x}}(\mathbf{y}^{\geq i-1} | \mathbf{Z}^{(i-1)}, \mathbf{y}^{(i-2)}) I_{y^{(i-2)}}(Y^{(i-2)}) = I_{y^{(i-2)}}(Y^{(i-2)}) \sum_{z^{\geq i+1}} \prod_{k=i-1}^{n} P(y_k | x^{(k)}, y^{(k-1)}, z^{(k)}) P(z_k | x^{(k-1)}, y^{(k-1)}, z^{(k-1)})$$

Revisit ID Algorithm



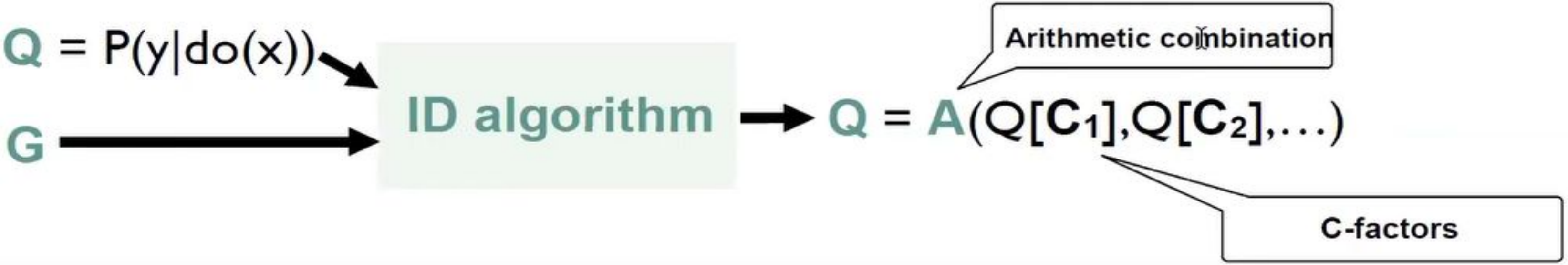
- C-componet: A set of variables connected by a bi-directed path, for example: $\{V_1,V_3,V_5\}$ and $\{V_2,V_4\}$ from the graph above.
- C-factor: Q[C]: the distribution of C under the intervention, i.e. Q[C] = P(C|do(V\C))
- The distribution can be factorized w.r.t. C-factors:
  - $P(v) = Q[V_2,V_4](v_2,v_4,v_1,v_3)Q[V_1,V_3,V_5](v_1,v_3,v_5,v_2,v_4)$

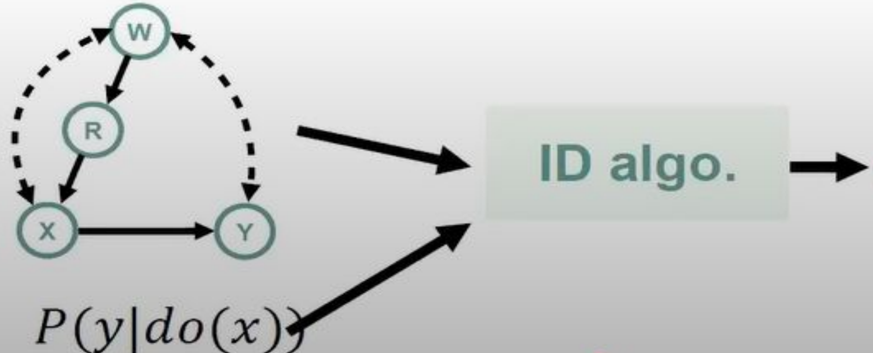
Revisit ID Algorithm

ID($\mathbf{X}, \mathbf{Y}, G$)

1. Let $\mathbf{S}_1, \mathbf{S}_2, \ldots$ be the C-components of G.
2. Let $Q[\mathbf{S}_i] = \prod_{V_k \in \mathbf{S}_i} P(v_k | v^{(k-1)})$.
3. Let $\mathbf{D}_1, \mathbf{D}_2, \ldots$ be C-components of $G(\mathbf{D})$ where $\mathbf{D} = An(\mathbf{Y})_{G(\mathbf{V} \setminus \mathbf{X})}$.
4. Identify $Q[\mathbf{D}_j]$ from $Q[\mathbf{S}]$ by recursively applying C-factor operations
5. $P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{d} \setminus \mathbf{y}} \prod_j Q[\mathbf{D}_j]$ if all $Q[\mathbf{D_j}]$ is identified, FAIL otherwise.

Revisit ID algorithm cont.

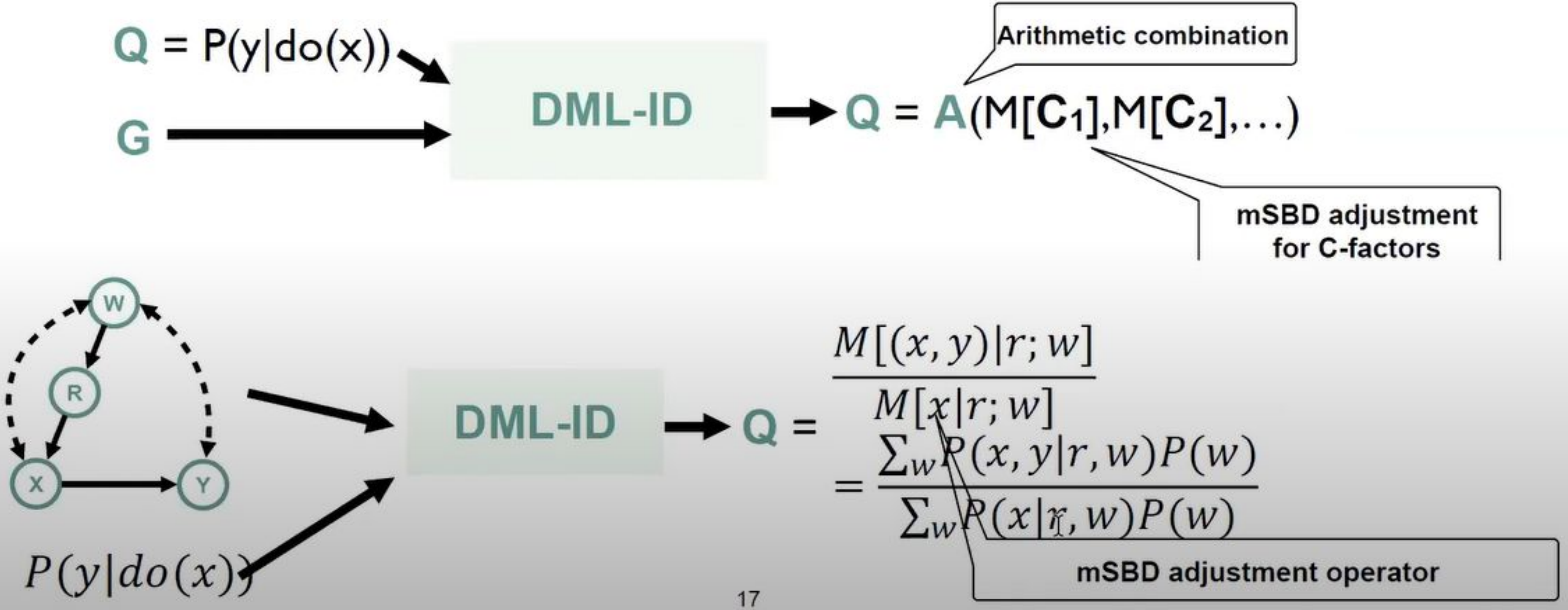




$$Q = \frac{\sum_w Q[W,X,Y]}{\sum_{w,y} Q[W,X,Y]} = \frac{\sum_w P(x,y|r,w)P(w)}{\sum_w P(x|r,w)P(w)}$$

Result 2 - DML-ID

Mechanism (C-factors are given by mSBD adjustment):



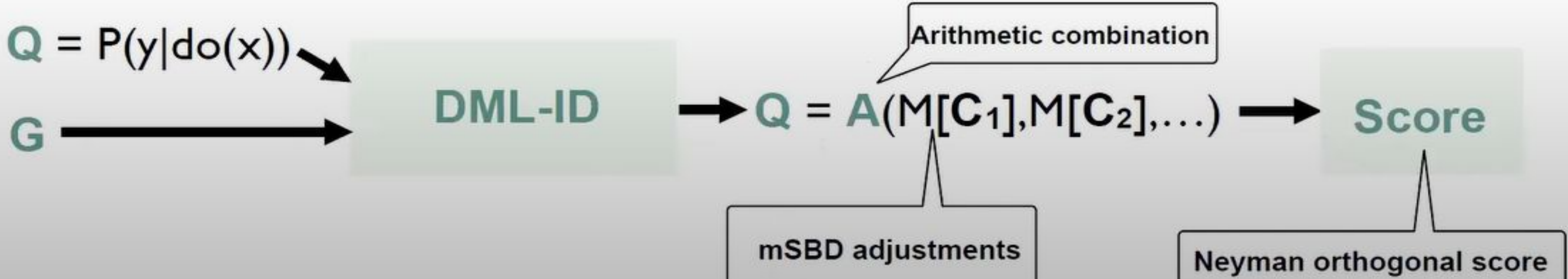
Derivation of Neyman Orthogonal Score (NOS)

A recursive algorithm for derivation of NOS

Given representation of **Q = A**(M[**C**$_1$],M[**C**$_2$],…,M[**C**$_d$]), a Neyman orthogonal score is given as

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

Neyman orthogonal score for the mSBD adjustment M_i=M[**C**$_i$]



Result 2 - DML Estimator

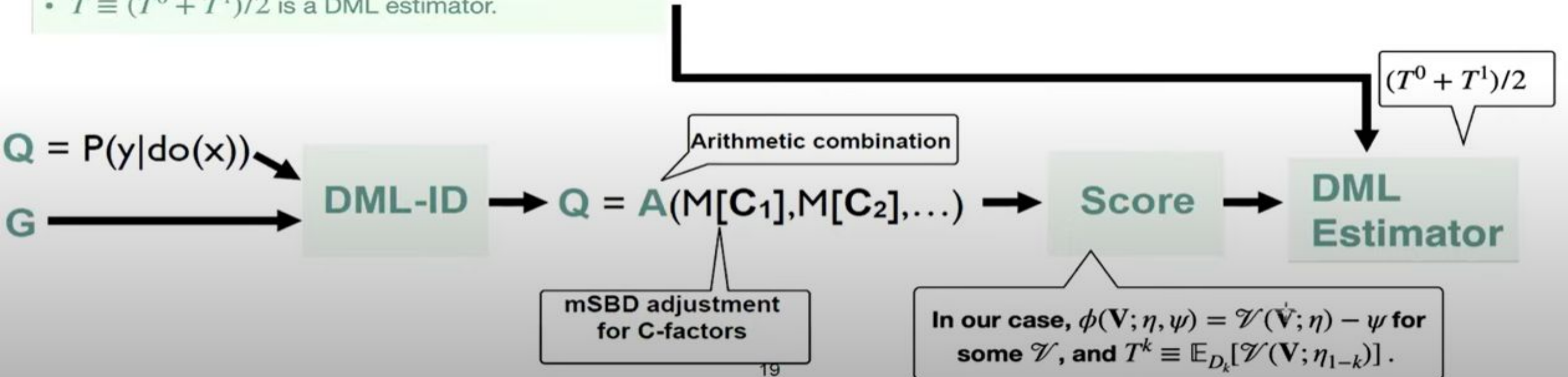
Construction of the DML estimator

- Let $\{D_0, D_1\}$ denote the randomly split halves of the dataset D. Let $\widehat{\eta}_k$ denote the estimate of η from D_k for $k \in \{0,1\}$.
- Let T^k denote the solution satisfying $\mathbb{E}_{D_k}\left[\phi(\mathbf{V}; \widehat{\eta}_{1-k}, T^k)\right] = o_P(N^{1/2})$ where $N \equiv |D|$, and $\mathbb{E}_{D_k}$ denote the empirical expectation over D_k.
- $T \equiv (T^0 + T^1)/2$ is a DML estimator.

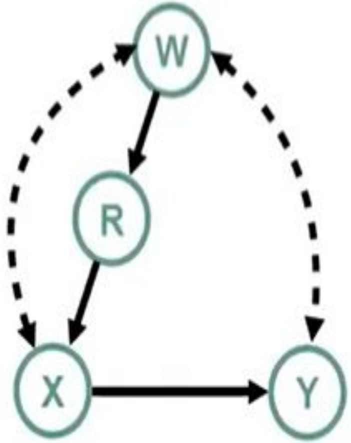
Properties

The proposed estimator is

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



DML Example



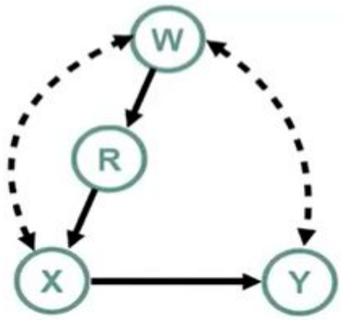
$$P(y|do(x)) = \frac{\sum_w P(x,y|r,w)P(w)}{\sum_w P(x|r,w)P(w)}$$

W is BD admissible w.r.t. $(R, \{X, Y\})$;
$M_1 \equiv M[(x,y) \mid r; w]$

W is BD admissible w.r.t. (R, X)
$M_2 \equiv M[x \mid r; w]$

$$= A(M_1, M_2) = \frac{M_1}{M_2}$$

Example: NOS



NOS ϕ_{M_i} for $M_i,\ i = 1, 2$ are given as

- $\phi_{M_i} \equiv h_{M_i} - \mu_{M_i}$
- $\mu_{M_i} \equiv E[h_{M_i}]$

$$h_{M_1} \equiv \frac{I_r(R)}{P(R|W)} \left(I_{x,y}(X,Y) - P(x,y|R,W)\right) + P(x,y|r,W)$$

$$h_{M_2} \equiv \frac{I_r(R)}{P(R|W)} \left(I_x(X) - P(x|R,W)\right) + P(x|r,W)$$

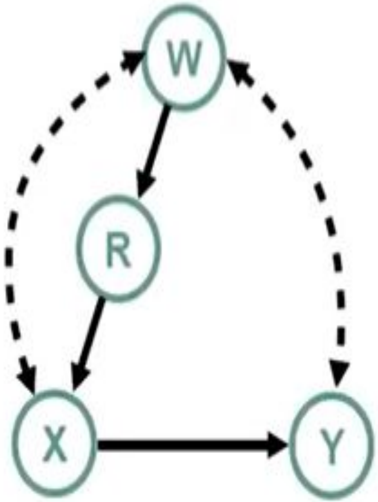
- NOS $\phi(V;\eta,\psi)$ for P(y|do(x))

$$\frac{1}{\mu_{M_2}}\left(\phi_{M_1} - \phi_{M_2}\frac{\mu_{M_1}}{\mu_{M_2}}\right)$$

- $\phi(\mathbf{V};\eta,\psi) = \mathcal{V}(\mathbf{V};\eta) - \psi$ where

$$\mathcal{V}(\mathbf{V};\eta) \equiv \frac{1}{\mu_{M_2}}\left(h_{M_1} - \phi_{M_2}\frac{\mu_{M_1}}{\mu_{M_2}}\right)$$

Example Empirical Evaluation



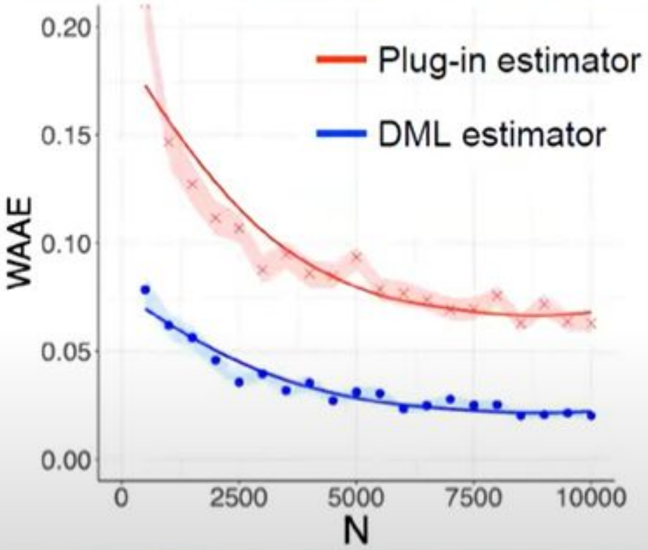
$$P(y|do(x)) = \frac{\sum_w P(x,y|r,w)P(w)}{\sum_w P(x|r,w)P(w)}$$

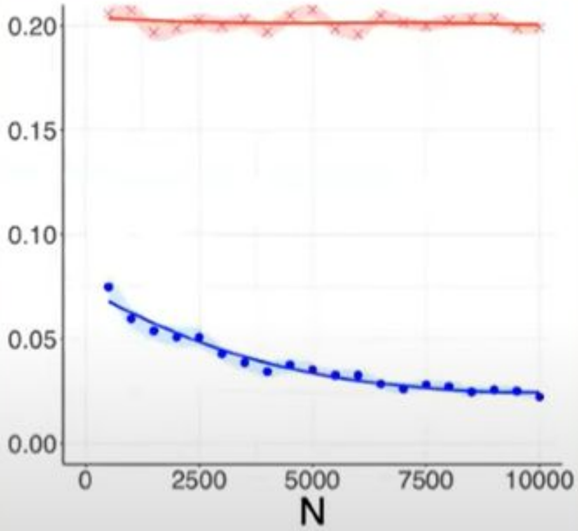
DML estimator:

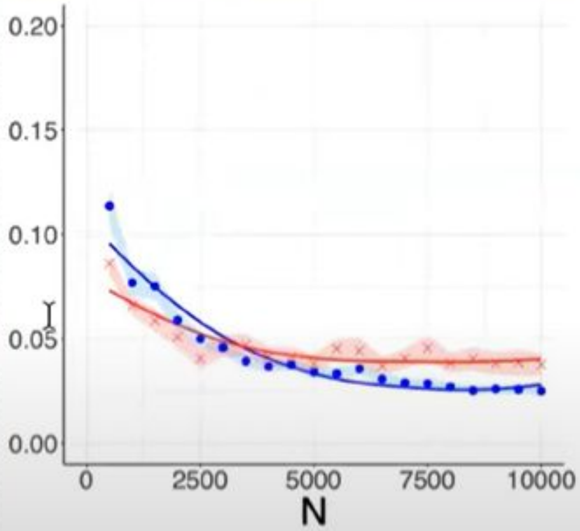
- **Doubly robustness —** Correctly estimate $P(y|do(x))$ if $\eta_1 = P(x,y|r,w)$ or $\eta_2 = P(r|w)$ are correctly estimated.
- **Debiasedness —** $\sqrt{N}$-consistent if $P(x,y|r,w)$ and $P(r|w)$ converges at $N^{-1/4}$ rate.

Empirical Evaluation

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

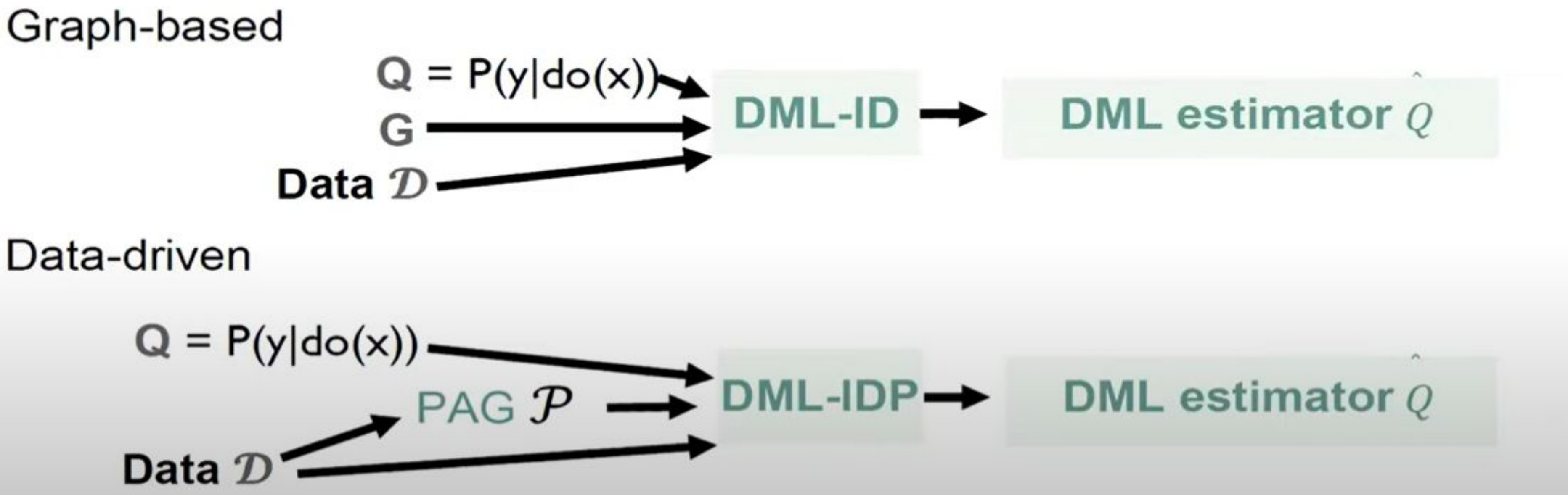






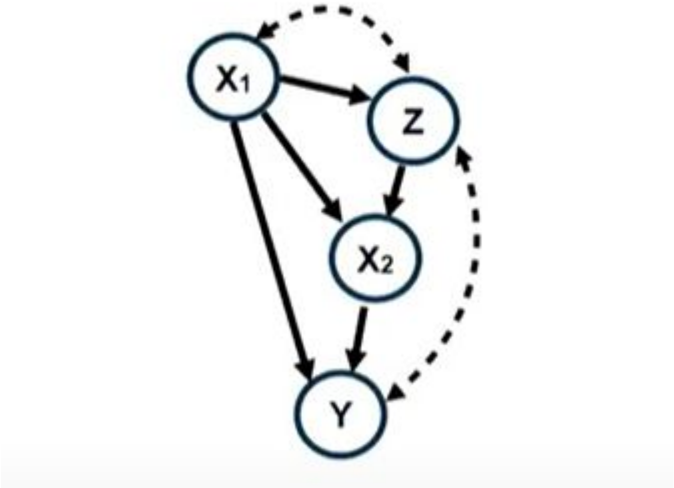
Conclusion

- DML estimators are developed for identifiable causal effects that appreciate the doubly robustness in the case of model misspecification as well as de-biasedness against biases in nuisance function estimation.



mSBD - Question
covercas@uci.edu

mSBD Criterion (informally): A sequential of Z variables satisfies the mSBD criterion relative to {X,Y} if a non-causal path between X_i in X and Y_i in Y are blocked by Z_i conditioned on the previous ones, {X^{i-1}, Y^{i-1}, Z^{i-1}}



P(y, do(x)) = ?