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

- **Estimand (Q)**
  \[ \mathbb{E} \left[ \frac{I_x(X)}{P(X|Z)} I_y(Y) \right] \]

- **Estimator (Q)**
  \[ \frac{1}{N} \sum_{i=1}^{N} \frac{I_x(X_{(i)})}{P(X_{(i)}|Z_{(i)})} I_y(Y_{(i)}) \]

### Regression

- **Estimand (Q)**
  \[ \mathbb{E}_Z [P(y|x, Z)] \]

- **Estimator (Q)**
  \[ \frac{1}{N} \sum_{i=1}^{N} P(y|x, Z_{(i)}) \]

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 $\Rightarrow$ Incorrect Estimation
  - Complicated Data Generation
- Slow convergence $\Rightarrow$ 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 | $\mathbb{E} \left[ \frac{I_X(X)}{P(X|Z)} \left( I_Y(Y) - P(y|x, Z) \right) + P(y|x, Z) \right]$ |
|----------|-----------------------------------------------------------------|
| Estimators | $\frac{1}{N} \sum_{i=1}^{N} \frac{I_X(X(i))}{P(X(i)|Z(i))} \left( I_Y(Y) - \hat{P}(y|x, Z(i)) \right) + \hat{P}(y|x, Z(i))$ |

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 \( \psi \); and
2. (Cross-fitting) training and evaluating nuisances \( \hat{\eta} \) is done with two distinct sets of samples.

#### Neyman orthogonal score (NOS) \( \phi \)

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

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

#### 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 \( \eta \) from \( D_k \) for \( k \in \{0, 1\} \).
- Let \( T^k \) denote the solution satisfying
  \[ \mathbb{E}_{D_k}[\phi(V; \hat{\eta}_{1-k}, T^k)] = o_p(N^{1/2}) \]
  where \( N \equiv |D| \), and \( \mathbb{E}_{D_k} \) denote the empirical expectation over \( D_k \).
- \( T = (T^0 + T^1)/2 \) 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 \(Z = \{Z_1, \ldots, Z_n\}\) satisfies the mSBD criterion relative to \((X, Y)\),

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

**mSBD Example 1**

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

\[
P(y_1, y_2 | do(x_1, x_2)) = \sum_{z_1, z_2} [P(z_1)P(y_1 | x_1, z_1)P(z_2 | z_1, x_1, y_1)P(y_2 | z_1, x_1, y_1, z_2, x_2)]
\]
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, $\{X_i^{i-1}, Y_i^{i-1}, Z_i^{i-1}\}$

$Z = ?$

$P(y, \text{do}(x)) = ?$
Result 1: mSBD Estimator

Neyman orthogonal score for mSBD

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

where

\[ H_i = P_x(y^{z_{i-1}} | Z^{(i-1)}, y^{(i-2)}) \] and

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

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.
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. \(Q[C] = P(C|do(V \setminus C))\)
- 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

\[ \text{ID}(X, Y, G) \]

1. Let \( S_1, S_2, \ldots \) be the C-components of \( G \).

2. Let \( Q[S_i] = \prod_{v_k \in S_i} P(v_k | v^{(k-1)}) \).

3. Let \( D_1, D_2, \ldots \) be C-components of \( G(D) \) where \( D = An(Y)_{G(V \setminus X)} \).

4. Identify \( Q[D_j] \) from \( Q[S] \) by recursively applying C-factor operations.

5. \[ P_x(y) = \sum_{D \setminus Y} \prod_{j} Q[D_j] \text{ if all } Q[D_j] \text{ is identified, FAIL otherwise.} \]
Revisit ID algorithm cont.

\[ Q = P(y|do(x)) \]

**ID algorithm**

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

\[ P(y|do(x)) \]

**ID algo.**

\[
Q = \frac{\sum_w Q[W, X, Y]}{\sum_{w,y} Q[W, X, Y]} = \frac{\sum_w P(x, y|r, w)P(w)}{\sum_w P(x|r, w)P(w)}
\]
Result 2 - DML-ID

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

\[ Q = P(y|do(x)) \]

\[ G \]

DML-ID \[ \rightarrow Q = A(M[C_1], M[C_2], \ldots) \]

Arithmetic combination

mSBD adjustment for C-factors

\[ Q = \frac{M[(x, y)|r; w]}{\sum_w P(x, y|r, w)P(w)} \]

mSBD adjustment operator
Derivation of Neyman Orthogonal Score (NOS)

A recursive algorithm for derivation of NOS

Given representation of \( Q = A(M[C_1], M[C_2], \ldots, M[C_d]) \), a Neyman orthogonal score is given as

\[
\sum_{i=1}^{d} \phi_{M_i} \frac{\partial}{\partial M_i} A(M[C_1], \ldots, M[C_d])
\]

Neyman orthogonal score for the mSBD adjustment \( M_i = M[C_i] \)

\[Q = P(y|do(x))\]

DML-ID \( \rightarrow Q = A(M[C_1], M[C_2], \ldots) \rightarrow \text{Score} \)

mSBD adjustments

Arithmetic combination

Neyman orthogonal score
Result 2 - DML Estimator

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 \( \eta \) from \( D_k \) for \( k \in \{0,1\} \).
- Let \( T^k \) denote the solution satisfying
  \[ E_{D_k} \left[ \phi(V; \hat{\eta}_{1-k}, T^k) \right] = o_p(N^{1/2}) \] where \( N \equiv |D| \), and \( E_{D_k} \) 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)

\[ Q = P(y|do(x)) \]

DML-ID \[ Q = A(M[C_1], M[C_2], ...) \]

Score

DML Estimator

Arithmetic combination

msBD adjustment for C-factors

In our case, \( \phi(V; \eta, \psi) = \mathcal{V}(\hat{V}; \eta) - \psi \) for some \( \mathcal{V} \), and \( T^k \equiv E_{D_k}[\mathcal{V}(V; \eta_{1-k})]. \)
DML Example

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

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

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

\[ = A(M_1, M_2) = \frac{M_1}{M_2} \]
Example: NOS

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

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

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

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

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

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

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

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

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

\[ P(x, y | r, w), P(r | w) \] converges to true \[ P(x, y | r, w), P(r | w) \] at a rate \( N^{-1/4} \)

- Plug-in estimator
- DML estimator

Doubly robustness

\[ P(x, y | r, w) \] 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

\[ P(y, \text{do}(x)) = ? \]

**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, \( \{X_{i-1}, Y_{i-1}, Z_{i-1}\} \).