STAT6061/STAT5008 – Causal Inference

Part 3-4. Doubly Robust Methods

An-Shun Tai

[1]*Department of Statistics*
National Cheng Kung University

[2]*Institute of Statistics and Data Science*
National Tsing Hua University

Current modeling strategies

➢ **Require correct modeling of the outcome variable** $\mathbb{E}\big(Y(a)\big) = \mathbb{E}\big(\mathbb{E}(Y|A = a, X; \beta_a)\big)$

1. Standardization or outcome regression

$$\hat{\tau}_o = \frac{1}{N}\sum_{i=1}^{N}\{\mu_1(X_i; \hat{\beta}_1) - \mu_0(X_i; \hat{\beta}_0)\}$$

where $\mu_a(X; \beta_a) = \mathbb{E}(Y|A = a, X; \beta_a)$

➢ **Require correct modeling of the propensity score (treatment variable)** $\mathbb{E}\big(Y(a)\big) = \mathbb{E}\left(\frac{I(A = a)}{\Pr(A = a|X; \alpha)} Y\right)$

1. IPW (Horvitz-Thompson estimator)

$$\hat{\tau}_2^{HT} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{A_i}{e(X_i; \hat{\alpha})} Y_i - \frac{1 - A_i}{1 - e(X_i; \hat{\alpha})} Y_i\right\}$$

where $e(X; \alpha) = \Pr(A = 1|X; \alpha)$

2. Propensity score matching

➢ **Nonparametric approach**

1. Mahalanobis metric matching
2. Coarsened exact matching (CEM)

Doubly robust estimator

Theorem 3.4

If the conditional exchangeability $(A \perp \{Y(1), Y(0)\}|X)$ *and positivity* $(0 < e(X) < 1)$ *hold, then*

$$\mathbb{E}\big(Y(1)\big) = \mathbb{E}\left(\frac{AY}{e(X;\alpha)} - \frac{A - e(X;\alpha)}{e(X;\alpha)}\mu_1(X;\beta_1)\right) \text{ and}$$

$$\mathbb{E}\big(Y(0)\big) = \mathbb{E}\left(\frac{(1-A)Y}{1-e(X;\alpha)} - \frac{e(X;\alpha) - A}{1-e(X;\alpha)}\mu_0(X;\beta_0)\right)$$

Moreover, if either the propensity score model or the outcome model, though not necessarily both, is correctly specified, then both equalities hold.

- The result in Theorem 3.4 motivates the following estimator of ATE

$$\hat{\tau}_{DR} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{AY}{e(X_i;\hat{\alpha})} - \frac{A - e(X_i;\hat{\alpha})}{e(X_i;\hat{\alpha})}\mu_1\big(X_i;\hat{\beta}_1\big)\right\} - \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{(1-A)Y}{1-e(X_i;\hat{\alpha})} - \frac{e(X_i;\hat{\alpha}) - A}{1-e(X_i;\hat{\alpha})}\mu_0\big(X_i;\hat{\beta}_0\big)\right\}$$

Doubly robust estimator (cont.)

- Theorem 3.4 augments the IPW estimator with an imputed outcome, leading to the augmented inverse propensity score weighting (AIPW) estimator, also known as augmented inverse probability weighting.

- Theorem 3.4 establishes that $\hat{\tau}_{DR}$ possesses the doubly robust property, remaining consistent if either the propensity score model or the outcome model is correctly specified.

$\Rightarrow$ AIPW is also referred to as the DR estimator.

- Alternative augmented estimator motivated by

$$\mathbb{E}\big(Y(1)\big) = \mathbb{E}\left(\frac{A}{e(X;\alpha)}\{Y - \mu_1(X;\beta_1)\} + \mu_1(X;\beta_1)\right) \text{ and}$$

$$\mathbb{E}\big(Y(0)\big) = \mathbb{E}\left(\frac{(1-A)}{1-e(X;\alpha)}\{Y - \mu_0(X;\beta_0)\} + \mu_0(X;\beta_0)\right)$$

- This formula improves the outcome regression estimator by incorporating weighted residuals, thereby achieving augmented robustness.

Doubly robust estimator (Proof of Theorem 3.4)

$$\mathbb{E}\left(\frac{AY}{e(X;\alpha)} - \frac{A - e(X;\alpha)}{e(X;\alpha)}\mu_1(X;\beta_1)\right) - \mathbb{E}(Y(1))$$

$$= \mathbb{E}\left(\frac{AY(1)}{e(X;\alpha)} - \frac{A - e(X;\alpha)}{e(X;\alpha)}\mu_1(X;\beta_1) - Y(1)\right)$$

$$= \mathbb{E}\left(\frac{A}{e(X;\alpha)}\{Y(1) - \mu_1(X;\beta_1)\} + \{\mu_1(X;\beta_1) - Y(1)\}\right)$$

$$= \mathbb{E}\left(\{\frac{A}{e(X;\alpha)} - 1\}\{Y(1) - \mu_1(X;\beta_1)\}\right)$$

$$= \mathbb{E}\left(\mathbb{E}\left(\{\frac{A}{e(X;\alpha)} - 1\}\{Y(1) - \mu_1(X;\beta_1)\}|X\right)\right)$$

$$= \mathbb{E}\left(\mathbb{E}\left(\{\frac{A}{e(X;\alpha)} - 1\}\,|X\right) \times \mathbb{E}(\{Y(1) - \mu_1(X;\beta_1)\}|X)\right) = \mathbb{E}(\{\frac{e_{true}(X) - e(X;\alpha)}{e(X;\alpha)}\} \times \{\mu_{1,true}(X) - \mu_1(X;\beta_1)\})$$

where $\mu_{1,true}(X) = \mathbb{E}(Y(1)|X)$ and $e_{true}(X) = \Pr(A = 1|X)$

- Therefore, the equality holds if either $\mu_{1,true}(X) = \mu_1(X;\beta_1)$ or $e_{true}(X) = e(X;\alpha)$

Simulation study from Ding (2024), Section 12.3.2

1. both the propensity score and outcome models are correct;

	reg	HT	Hajek	DR
ave.bias	0.00	0.02	0.03	0.01
true.se	0.11	0.28	0.26	0.13
est.se	0.10	0.25	0.23	0.12

2. the propensity score model is wrong but the outcome model is correct

	reg	HT	Hajek	DR
ave.bias	0.00	-0.76	-0.75	-0.01
true.se	0.12	0.59	0.47	0.18
est.se	0.13	0.50	0.38	0.18

3. the propensity score model is correct but the outcome model is wrong

	reg	HT	Hajek	DR
ave.bias	-0.05	0.00	-0.01	0.00
true.se	0.11	0.15	0.14	0.14
est.se	0.11	0.14	0.13	0.14

4. both the propensity score and outcome models are wrong.

	reg	HT	Hajek	DR
ave.bias	-0.08	0.11	-0.07	0.16
true.se	0.13	0.32	0.20	0.41
est.se	0.13	0.25	0.16	0.26

More on the DR estimator

- **Double robustness is a large-sample property.**

- **Protection against model misspecification:**

The DR estimator gives you two chances to obtain a consistent estimate — if either the propensity score model or the outcome model is correctly specified, consistency is achieved.

- **Variance comparison (semiparametric efficient):**

1. When both the models of propensity score and the outcome are correctly specified, $\hat{\tau}_{DR}$ has smaller variance than the IPW and the outcome regression estimators in large samples.
2. If only the outcome model is correctly specified, $\hat{\tau}_{DR}$ generally has larger variance than the direct outcome regression estimator in large samples.

- **Finite sample concern (Kang and Schafer, 2007):**

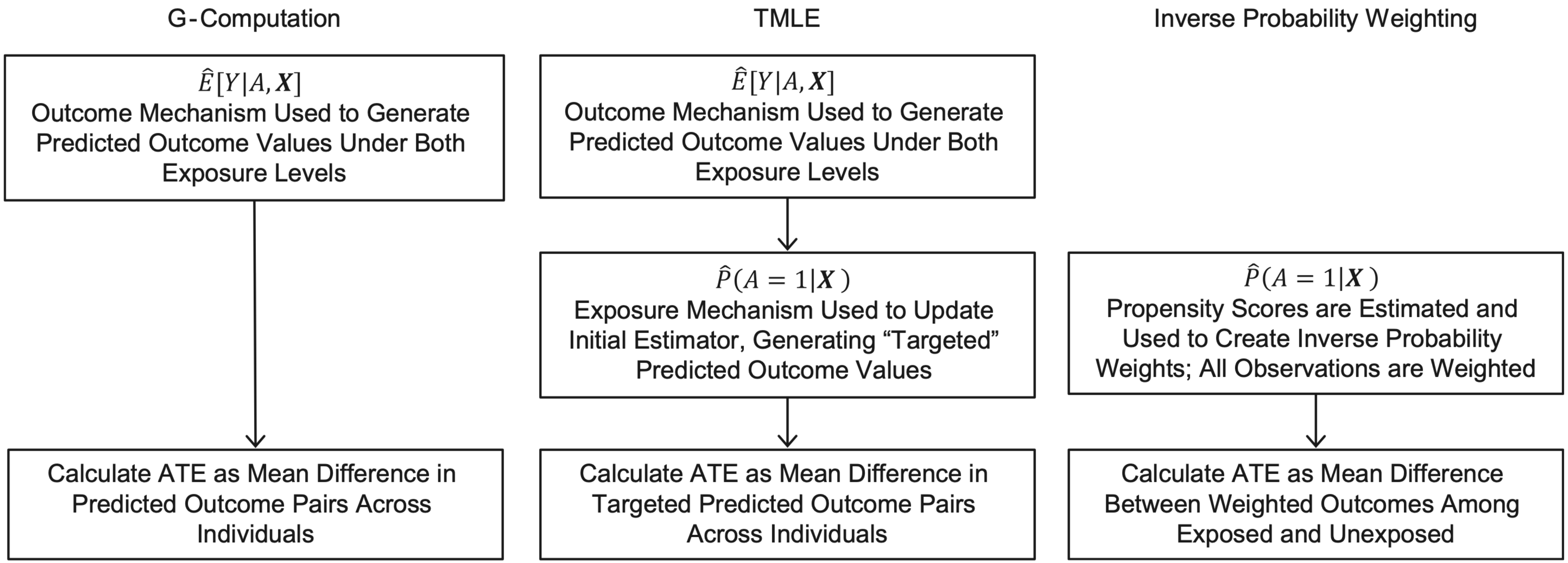
When both models are misspecified, the DR estimator can perform substantially worse than simple outcome regression or IPW in finite samples. (See Pages 5 and 6)

- **It is suggested to approximate the variance of $\hat{\tau}_{DR}$ via the nonparametric bootstrap.**

Targeted Maximum Likelihood Estimation (TMLE)

(Van Der Laan and Rubin, 2006; Schuler and Rose, 2017)

TMLE = Machine learning–friendly + Doubly robust + Efficient causal inference estimator.



Basic Steps of TMLE for ATE

(Van Der Laan and Rubin, 2006; Schuler and Rose, 2017)

$$\text{ATE: } \tau = \mathbb{E}\big(Y(1)\big) - \mathbb{E}\big(Y(0)\big)$$

Step 1: Initial outcome regression and propensity score

- Outcome Regression :
estimate the conditional outcome model $\mu_a(X;\beta_a) = \mathbb{E}(Y|A = a, X; \beta_a)$ for $a = 0, 1$.
- Propensity Score:
estimate the treatment assignment model $e(X;\alpha) = \Pr(A = 1|X;\alpha)$.

Step 2: Construct the clever covariate

Define the clever covariate $H(A,X)$, which is a specially crafted function of A and X using the propensity score. For the ATE parameter, the clever covariate is:

$$H(A,X) = \frac{A}{e(X;\hat{\alpha})} - \frac{1-A}{1-e(X;\hat{\alpha})}$$

Basic Steps of TMLE for ATE

(Van Der Laan and Rubin, 2006; Schuler and Rose, 2017)

$$\text{ATE: } \tau = \mathbb{E}\big(Y(1)\big) - \mathbb{E}\big(Y(0)\big)$$

Step 3: Update initial estimate of $\mathbb{E}(\boldsymbol{Y}|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{X}; \boldsymbol{\beta}_{\boldsymbol{a}})$ by regressing on the clever covariate.

Regress the observed outcome Y on $H(A, X)$, treating $\hat{Y} = \mathbb{E}\big(Y\big|A, X; \hat{\beta}_a\big)$ as a fixed offset, in order to estimate δ.

For a binary or bounded outcome:

$$logit\big(\mathbb{E}^*(Y|A, X; \delta)\big) = logit\big(\hat{Y}\big) + \delta \times H(A, X)$$

- This yields the fluctuation (targeting) coefficient $\hat{\delta}$. Equivalently, $\hat{\delta}$ is chosen to solve the score equation (setting the derivative of log-likelihood to zero):

$$\frac{1}{N}\sum_{i=1}^{N} H(A, X)\{Y_i - \mathbb{E}^*(Y|A, X; \delta)\}$$

(=the score equations of the GLM)

Basic Steps of TMLE for ATE

(Van Der Laan and Rubin, 2006; Schuler and Rose, 2017)

$$\text{ATE: } \tau = \mathbb{E}\big(Y(1)\big) - \mathbb{E}(Y(0))$$

Step 4: Compute the final TMLE estimate

$$\hat{\tau}_{TMLE} = \frac{1}{N}\sum_{i=1}^{N}\{\mathbb{E}^*\big(Y\big|A=1,X;\hat{\delta}\big) - \mathbb{E}^*\big(Y\big|A=0,X;\hat{\delta}\big)\}$$

➢ **Why use TMLE?**

1. Performs well even with flexible or machine learning models for nuisance function estimation.
2. Double robustness: Consistent if either the outcome model or the propensity score model is correctly specified.
3. Efficiency: Achieves the semiparametric efficiency bound when both models are correctly specified.

Simulation study

(Schuler and Rose, 2017)

Estimator	Mean ATE (SE)	Mean Bias	95% CI
Targeted Maximum Likelihood Estimation			
Super learner			
Outcome variables: A, X_1, X_2, X_3; exposure variables: X_1, X_2, X_3	–3.39 (0.35)	–0.01	–4.05, –2.64
Misspecified parametric regression			
Main-terms misspecification			
Outcome variables: A, X_1, X_2, X_3	–3.39 (0.35)	–0.01	–4.08, –2.64
Omitted-variable misspecification			
Outcome variables: A, X_1, X_2	–3.39 (0.36)	–0.01	–4.09, –2.63
Exposure variables: X_1, X_2	–3.39 (0.35)	–0.01	–4.07, –2.69
G-Computation			
Super learner			
Outcome variables: A, X_1, X_2, X_3	–3.27 (0.35)	0.11	–3.98, –2.56
Misspecified parametric regression			
Main-terms misspecification			
Outcome variables: A, X_1, X_2, X_3	–3.25 (0.33)	0.13	–3.91, –2.59
Omitted-variable misspecification			
Outcome variables: A, X_1, X_2	–4.98 (0.37)	–1.60	–5.69, –4.24[b]
Inverse Probability Weighting			
Super learner			
Exposure variables: X_1, X_2, X_3	–3.43 (0.37)	–0.05	–4.17, –2.63
Misspecified parametric regression			
Omitted-variable misspecification			
Exposure variables: X_1, X_2	–4.96 (0.37)	–1.58	–5.67, –4.21[b]

References

Kang, J. D., & Schafer, J. L. (2007). Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Statistical Science*, *22*(4), 523-539.

Schuler, M. S., & Rose, S. (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *American journal of epidemiology*, *185*(1), 65-73.

Van der Laan, M. J., & Rose, S. (2011). *Targeted learning: causal inference for observational and experimental data* (Vol. 4). New York: Springer.

Van Der Laan, M. J., & Rubin, D. (2006). Targeted maximum likelihood learning. *The international journal of biostatistics*, *2*(1).