Double Robustness in Estimation of Causal Treatment Effects

Marie Davidian
Department of Statistics
North Carolina State University

http://www.stat.ncsu.edu/∼davidian

Based in part on Lunceford, J.K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine* 23:2937–2960

Outline

1. Introduction
2. Model based on potential outcomes (counterfactuals)
3. Adjustment by regression modeling
4. Adjustment by "inverse weighting"
5. Doubly-robust estimator
6. Some interesting issues
7. Discussion

1. Introduction

Simplest situation: *Observational study*

- *Point exposure study*
- Continuous or discrete (e.g. binary) *outcome* Y
- "*Treatment*" (exposure) or "*Control*" (no exposure)

Objective: Make *causal inference* on the *effect* of treatment

- Would like to be able to say that such an effect is *attributable*, or "*caused by*" treatment
- *Average causal effect* – based on *population mean outcome*

{ mean if *entire population* exposed − mean if *entire population* not exposed }

1. Introduction

Complication: *Confounding*

- The fact that a subject was exposed or not is associated with *subject characteristics* that may *also* be associated with his/her potential outcomes under treatment and control
- To estimate the *average causal effect* from *observational data* requires taking appropriate account of this *confounding*

Challenge: Estimate the *average causal treatment effect* from *observational data* , *adjusting appropriately* for confounding

- Different methods of adjustment are available
- Any method requires *assumptions* ; what if some of them are *wrong* ?
- The property of *double robustness* offers *protection* against some particular incorrect assumptions...
- ... *and* can lead to *more precise* inferences

2. Counterfactual Model

Define:

Z $= 1$ if treatment, $= 0$ if control

$\boldsymbol{X}$ vector of pre-exposure *covariates*

Y *observed* outcome

- Z is *observed* (not assigned)
- *Observed data* are i.i.d. copies $(Y_i, Z_i, \boldsymbol{X}_i)$ for each subject $i = 1, \ldots, n$

Based on these data: Estimate the *average causal treatment effect*

- To *formalize* what we mean by this and what the issues are, appeal to the notion of *potential outcomes* or *counterfactuals*

2. Counterfactual Model

Counterfactuals: Each subject has *potential outcomes* (Y_0, Y_1)

Y_0 outcome the subject would have if s/he received *control*

Y_1 outcome the subject would have if s/he received *treatment*

Average causal treatment effect:

- The *probability distribution* of Y_0 represents how outcomes in the population would turn out if *everyone* received *control*, with *mean* $E(Y_0)$ $(= P(Y_0 = 1)$ for *binary* outcome)
- The *probability distribution* of Y_1 represents this if *everyone* received *treatment*, with *mean* $E(Y_1)$ $(= P(Y_1 = 1)$ for *binary* outcome)
- *Thus*, the *average causal treatment effect* is

$$\Delta = \mu_1 - \mu_0 = E(Y_1) - E(Y_0)$$

2. Counterfactual Model

Problem: Do not see (Y_0, Y_1) for all n subjects; *instead* we only *observe*

$$Y = Y_1 Z + Y_0(1 - Z)$$

- If i was exposed to treatment, $Z_i = 1$ and $Y_i = Y_{1i}$
- If i was not exposed (control), $Z_i = 0$ and $Y_i = Y_{0i}$

Challenge: Would like to *estimate* Δ based on the *observed data*

- First, a quick *review* of some *statistical concepts*. . .

2. Counterfactual Model

Unconditional (marginal) expectation: Conceptually, the "*average*" across all possible values a *random variable* can take on in the population

Statistical independence: For two random variables Y and Z

- Y and Z are *independent* if the *probabilities* with which Y takes on its values are *the same* regardless of the value Z takes on
- *Notation* – $Y \perp\!\!\!\perp Z$

2. Counterfactual Model

Conditional expectation: For two random variables Y and Z,

$$E(Y|Z = z)$$

is the "*average*" across all values of Y for which the corresponding value of Z is equal to z

- $E(Y|Z)$ is a *function* of Z taking on values $E(Y|Z = z)$ as Z takes on values z
- $E\{\, E(Y|Z)\,\} = E(Y)$ (*"average the averages"* across all possible values of Z)
- $E\{\, Yf(Z)|Z\,\} = f(Z)E(Y|Z)$ – $f(Z)$ is *constant* for any value of Z so factors out of the average
- If $Y \perp\!\!\!\perp Z$, then $E(Y|Z = z) = E(Y)$ for any z

2. Counterfactual Model

Challenge, again: Based on *observed data* $(Y_i, Z_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, estimate

$$\Delta = \mu_1 - \mu_0 = E(Y_1) - E(Y_0)$$

Observed sample means: Averages among those *observed* to receive treatment or control

$$\overline{Y}^{(1)} = n_1^{-1} \sum_{i=1}^{n} Z_i Y_i, \quad \overline{Y}^{(0)} = n_0^{-1} \sum_{i=1}^{n} (1 - Z_i) Y_i$$

$$n_1 = \sum_{i=1}^{n} Z_i = \text{ \# subjects } \textit{observed} \text{ to receive } \textit{treatment}$$

$$n_0 = \sum_{i=1}^{n} (1 - Z_i) = \text{ \# } \textit{observed} \text{ to receive } \textit{control}$$

2. Counterfactual Model

What is being estimated? If we estimate Δ by $\overline{Y}^{(1)} - \overline{Y}^{(0)}$

- $\overline{Y}^{(1)}$ estimates $E(Y|Z=1) =$ population mean outcome *among those observed to receive treatment*
- $E(Y|Z=1) = E\{Y_1 Z + Y_0(1-Z)|Z=1\} = E(Y_1|Z=1)\ldots$
- ...which is *not equal to* $E(Y_1) =$ mean outcome if *entire population* received treatment
- Similarly, $\overline{Y}^{(0)}$ estimates $E(Y|Z=0) = E(Y_0|Z=0) \neq E(Y_0)$
- *Thus*, what is being estimated in general is

$$E(Y|Z=1) - E(Y|Z=0) \neq \Delta = E(Y_1) - E(Y_0)$$

2. Counterfactual Model

Exception: Randomized study

- Treatment is assigned with *no regard* to how a subject might respond to either treatment or control
- *Formally*, treatment received is *independent* of potential outcome:

$$(Y_0, Y_1) \perp\!\!\!\perp Z$$

- This means that

$$E(Y|Z=1) = E\{Y_1 Z + Y_0(1-Z)\,|Z=1\} = E(Y_1|Z=1) = E(Y_1)$$

  and similarly $E(Y|Z=0) = E(Y_0)$

- $\overline{Y}^{(1)} - \overline{Y}^{(0)}$ is an *unbiased estimator* for Δ

2. Counterfactual Model

In contrast: *Observational study*

- Exposure to treatment is *not controlled*, so exposure may be *related* to the way a subject might potentially respond:

$$(Y_0, Y_1) \not\!\perp\!\!\!\perp Z$$

- And, indeed, $E(Y|Z=1) \neq E(Y_1)$ and $E(Y|Z=0) \neq E(Y_0)$, and $\overline{Y}^{(1)} - \overline{Y}^{(0)}$ is *not* an unbiased estimator for Δ

2. Counterfactual Model

Confounders: It may be possible to identify *covariates* related to *both* potential outcome and treatment exposure

- If $\boldsymbol{X}$ contains *all confounders*, then among subjects sharing the same $\boldsymbol{X}$ there will be *no association* between exposure Z and potential outcome (Y_0, Y_1), i.e. (Y_0, Y_1) and Z are *independent conditional* on $\boldsymbol{X}$:

$$(Y_0, Y_1) \perp\!\!\!\perp Z \,|\, \boldsymbol{X}$$

- *No unmeasured confounders* is an *unverifiable assumption*

- If we *believe* no unmeasured confounders, can estimate Δ by *appropriate adjustment*...

3. Adjustment by Regression Modeling

Regression of Y on Z and $\boldsymbol{X}$: We can *identify* the *regression*

$$E(Y|Z, \boldsymbol{X}),$$

as this depends on the *observed data*

- E.g., for *continuous outcome*, $E(Y|Z, \boldsymbol{X}) = \alpha_0 + \alpha_Z Z + \boldsymbol{X}^T \alpha_X$
- In general, $E(Y|Z = 1, \boldsymbol{X})$ is the regression among *treated*, $E(Y|Z = 0, \boldsymbol{X})$ among *control*

Usefulness: *Averaging* over *all* possible values of $\boldsymbol{X}$ (*both* treatments)

$$E\{\, E(Y|Z = 1, \boldsymbol{X})\,\} = E\{\, E(Y_1|Z = 1, \boldsymbol{X})\,\} = E\{\, E(Y_1|\boldsymbol{X})\,\} = E(Y_1)$$

and similarly $E\{\, E(Y|Z = 0, \boldsymbol{X})\,\} = E(Y_0)$

3. Adjustment by Regression Modeling

Thus: Under *no unmeasured confounders*

$$
\begin{aligned}
\Delta &= E(Y_1) - E(Y_0) \\
&= E\{\, E(Y|Z=1,\boldsymbol{X})\,\} - E\{\, E(Y|Z=0,\boldsymbol{X})\,\} \\
&= E\{\, E(Y|Z=1,\boldsymbol{X}) - E(Y|Z=0,\boldsymbol{X})\,\}
\end{aligned}
$$

- Suggests postulating a *model* for the *outcome regression* $E(Y|Z,\boldsymbol{X})$, *fitting* the model, and then and *averaging* the resulting estimates of

$$E(Y|Z=1,\boldsymbol{X}) - E(Y|Z=0,\boldsymbol{X})$$

over *all* observed $\boldsymbol{X}$ (*both* groups) to estimate Δ

3. Adjustment by Regression Modeling

Example – continuous outcome: Suppose the *true* regression is

$$E(Y|Z, \boldsymbol{X}) = \alpha_0 + \alpha_Z Z + \boldsymbol{X}^T \alpha_X$$

- If this *really is* the true outcome regression model, then

$$\begin{aligned} E(Y|Z=1, \boldsymbol{X}) - E(Y|Z=0, \boldsymbol{X}) &= \alpha_0 + \alpha_Z(1) + \boldsymbol{X}^T\alpha_X - \alpha_0 - \alpha_Z(0) - \boldsymbol{X}^T\alpha_X \\ &= \alpha_Z \end{aligned}$$

- So $\Delta = E\{E(Y|Z=1, \boldsymbol{X}) - E(Y|Z=0, \boldsymbol{X})\} = \alpha_Z$
- Can thus estimate Δ *directly* from fitting this model (e.g. by *least squares*); don't even need to average!
- $\widehat{\Delta} = \widehat{\alpha}_Z$

3. Adjustment by Regression Modeling

Example – binary outcome: Suppose the *true* regression is

$$E(Y|Z,\boldsymbol{X}) = \frac{\exp(\alpha_0 + \alpha_Z Z + \boldsymbol{X}^T\alpha_X)}{1+\exp(\alpha_0 + \alpha_Z Z + \boldsymbol{X}^T\alpha_X)}$$

- If this *really is* the true outcome regression model, then

$$\begin{aligned} E(Y|Z=1,\boldsymbol{X}) - E(Y|Z=0,\boldsymbol{X}) \\ = \frac{\exp(\alpha_0 + \alpha_Z + \boldsymbol{X}^T\alpha_X)}{1+\exp(\alpha_0 + \alpha_Z + \boldsymbol{X}^T\alpha_X)} - \frac{\exp(\alpha_0 + \boldsymbol{X}^T\alpha_X)}{1+\exp(\alpha_0 + \boldsymbol{X}^T\alpha_X)} \end{aligned}$$

- *Logistic regression* yields $(\widehat{\alpha}_0, \widehat{\alpha}_Z, \widehat{\alpha}_X)$

- Estimate Δ by *averaging* over *all observed* $\boldsymbol{X}_i$

$$\widehat{\Delta} = n^{-1}\sum_{i=1}^{n}\left\{\frac{\exp(\widehat{\alpha}_0 + \widehat{\alpha}_Z + \boldsymbol{X}_i^T\widehat{\alpha}_X)}{1+\exp(\widehat{\alpha}_0 + \widehat{\alpha}_Z + \boldsymbol{X}_i^T\widehat{\alpha}_X)} - \frac{\exp(\widehat{\alpha}_0 + \boldsymbol{X}_i^T\widehat{\alpha}_X)}{1+\exp(\widehat{\alpha}_0 + \boldsymbol{X}_i^T\widehat{\alpha}_X)}\right\}$$

3. Adjustment by Regression Modeling

Critical: For the argument on slide 16 to go through, $E(Y|Z, \boldsymbol{X})$ must be the *true regression* of Y on Z and $\boldsymbol{X}$

- Thus, if we substitute estimates for $E(Y|Z = 1, \boldsymbol{X})$ and $E(Y|Z = 0, \boldsymbol{X})$ based on a *postulated outcome regression model*, this *postulated model* must be identical to the *true regression*
- If not, average of the difference *will not necessarily* estimate Δ

Result: Estimator for Δ obtained from regression adjustment will be *biased* (*inconsistent*) if the regression model used is *incorrectly specified*!

Moral: Estimation of Δ via regression modeling *requires* that the postulated regression model is *correct*

4. Adjustment by Inverse Weighting

Propensity score: Probability of treatment given covariates

$$e(\boldsymbol{X}) = P(Z = 1|\boldsymbol{X}) = E\{I(Z = 1)|\boldsymbol{X}\} = E(Z|\boldsymbol{X})$$

- $\boldsymbol{X} \perp\!\!\!\perp Z | e(\boldsymbol{X})$
- Under *no unmeasured confounders*, $(Y_0, Y_1) \perp\!\!\!\perp Z | e(\boldsymbol{X})$
- Customary to *estimate* by *postulating* and *fitting* a *logistic regression* model, e.g.

$$P(Z = 1|\boldsymbol{X}) = e(\boldsymbol{X}, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \boldsymbol{X}^T \beta_1)}{1 + \exp(\beta_0 + \boldsymbol{X}^T \beta_1)}$$

$$e(\boldsymbol{X}, \boldsymbol{\beta}) \Longrightarrow e(\boldsymbol{X}, \widehat{\boldsymbol{\beta}})$$

4. Adjustment by Inverse Weighting

One idea: Rather than use the difference of *simple averages* $\overline{Y}^{(1)} - \overline{Y}^{(0)}$, estimate Δ by the difference of *inverse propensity score weighted averages*, e.g.,

$$\widehat{\Delta}_{IPW,1} = n^{-1} \sum_{i=1}^{n} \frac{Z_i Y_i}{e(\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}})} - n^{-1} \sum_{i=1}^{n} \frac{(1 - Z_i) Y_i}{1 - e(\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}})}$$

- *Interpretation*: Inverse weighting creates a *pseudo-population* in which there is *no confounding*, so that the *weighted averages* reflect averages in the true population

Why does this work? Consider $n^{-1} \sum_{i=1}^{n} \frac{Z_i Y_i}{e(\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}})}$

- By the *law of large numbers*, this should estimate the *mean* of a term in the sum with $\widehat{\boldsymbol{\beta}}$ replaced by the quantity it estimates

4. Adjustment by Inverse Weighting

If: $e(\boldsymbol{X}, \boldsymbol{\beta}) = e(\boldsymbol{X})$, the *true propensity score*

$$
\begin{aligned}
E\left\{\frac{ZY}{e(\boldsymbol{X})}\right\} &= E\left\{\frac{ZY_1}{e(\boldsymbol{X})}\right\} = E\left[E\left\{\frac{ZY_1}{e(\boldsymbol{X})}\middle| Y_1, \boldsymbol{X}\right\}\right] && (1)\\
&= E\left\{\frac{Y_1}{e(\boldsymbol{X})}E(Z|Y_1, \boldsymbol{X})\right\} = E\left\{\frac{Y_1}{e(\boldsymbol{X})}E(Z|\boldsymbol{X})\right\} && (2)\\
&= E\left\{\frac{Y_1}{e(\boldsymbol{X})}e(\boldsymbol{X})\right\} = E(Y_1) && (3)
\end{aligned}
$$

(1) follows because $ZY = Z\{Y_1 Z + Y_0(1-Z)\} = Z^2 Y_1 + Z(1-Z)Y_0$ and $Z^2 = Z$ and $Z(1-Z) = 0$ (*binary*)

(2) follows because $(Y_0, Y_1) \perp\!\!\!\perp Z|\boldsymbol{X}$ (*no unmeasured confounders*)

(3) follows because $e(\boldsymbol{X}) = E(Z|\boldsymbol{X})$

Similarly: $E\left\{\dfrac{(1-Z)Y}{1-e(\boldsymbol{X})}\right\} = E(Y_0)$

4. Adjustment by Inverse Weighting

Critical: For the argument on slide 22 to go through, $e(\boldsymbol{X})$ must be the *true propensity score*

- Thus, if we substitute estimates for $e(\boldsymbol{X})$ based on a *postulated propensity score model*, this *postulated model* must be identical to the *true propensity score*
- If not, $\widehat{\Delta}_{IPW,1}$ *will not necessarily* estimate Δ

Moral: Estimation of Δ via inverse weighted *requires* that the *postulated propensity score model* is *correct*

5. Doubly Robust Estimator

Recap: $\Delta = E(Y_1) - E(Y_0)$

- Estimator for Δ based on *regression modeling* requires *correct* postulated regression model
- Estimator for Δ based on *inverse propensity score weighting* requires *correct* postulated propensity model

Modified estimator: Combine both approaches in a *fortuitous* way

5. Doubly Robust Estimator

Modified estimator:

$$\widehat{\Delta}_{DR} = n^{-1}\sum_{i=1}^{n}\left[\frac{Z_iY_i}{e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} - \frac{\{Z_i - e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})\}}{e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} m_1(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}}_1)\right]$$

$$-n^{-1}\sum_{i=1}^{n}\left[\frac{(1-Z_i)Y_i}{1-e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} + \frac{\{Z_i - e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})\}}{1-e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} m_0(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}}_0)\right]$$

$$= \widehat{\mu}_{1,DR} - \widehat{\mu}_{0,DR}$$

- $e(\boldsymbol{X},\boldsymbol{\beta})$ is a *postulated* model for the *true propensity score* $e(\boldsymbol{X}) = E(Z|\boldsymbol{X})$ (fitted by *logistic regression*)
- $m_0(\boldsymbol{X},\boldsymbol{\alpha}_0)$ and $m_1(\boldsymbol{X},\boldsymbol{\alpha}_1)$ are *postulated* models for the *true regressions* $E(Y|Z=0,\boldsymbol{X})$ and $E(Y|Z=1,\boldsymbol{X})$ (fitted by *least squares*)

5. Doubly Robust Estimator

Modified estimator:

$$
\begin{aligned}
\widehat{\Delta}_{DR} &= n^{-1}\sum_{i=1}^{n}\left[\frac{Z_i Y_i}{e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} - \frac{\{Z_i - e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})\}}{e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} m_1(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}}_1)\right] \\
&\quad - n^{-1}\sum_{i=1}^{n}\left[\frac{(1-Z_i)Y_i}{1-e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} + \frac{\{Z_i - e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})\}}{1-e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} m_0(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}}_0)\right] \\
&= \widehat{\mu}_{1,DR} - \widehat{\mu}_{0,DR}
\end{aligned}
$$

- $\widehat{\mu}_{1,DR}$ (and $\widehat{\mu}_{0,DR}$ and hence $\widehat{\Delta}_{DR}$) may be viewed as taking the *inverse weighted* estimator and "*augmenting*" it by a second term

What does this estimate? Consider $\widehat{\mu}_{1,DR}$ ($\widehat{\mu}_{0,DR}$ similar)

5. Doubly Robust Estimator

$$\widehat{\mu}_{1,DR} = n^{-1} \sum_{i=1}^{n} \left[\frac{Z_i Y_i}{e(\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}})} - \frac{\{Z_i - e(\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}})\}}{e(\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}})} m_1(\boldsymbol{X}_i, \widehat{\boldsymbol{\alpha}}_1) \right]$$

- By the *law of large numbers*, $\widehat{\mu}_{1,DR}$ estimates the *mean* of a term in the sum with with $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_1$ replaced by the quantities they estimate

- That is, $\widehat{\mu}_{1,DR}$ estimates

$$\begin{aligned} &E\left[\frac{ZY}{e(\boldsymbol{X}, \boldsymbol{\beta})} - \frac{\{Z - e(\boldsymbol{X}, \boldsymbol{\beta})\}}{e(\boldsymbol{X}, \boldsymbol{\beta})} m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1) \right] \\ &\quad = E\left[\frac{ZY_1}{e(\boldsymbol{X}, \boldsymbol{\beta})} - \frac{\{Z - e(\boldsymbol{X}, \boldsymbol{\beta})\}}{e(\boldsymbol{X}, \boldsymbol{\beta})} m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1) \right] \\ &\quad = E\left[Y_1 + \frac{\{Z - e(\boldsymbol{X}, \boldsymbol{\beta})\}}{e(\boldsymbol{X}, \boldsymbol{\beta})} \{Y_1 - m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1)\} \right] \quad \text{(by } \textit{algebra}\text{)} \\ &\quad = E(Y_1) + E\left[\frac{\{Z - e(\boldsymbol{X}, \boldsymbol{\beta})\}}{e(\boldsymbol{X}, \boldsymbol{\beta})} \{Y_1 - m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1)\} \right] \end{aligned}$$

5. Doubly Robust Estimator

Thus: $\widehat{\mu}_{1,DR}$ estimates

$$E(Y_1) + E\left[\frac{\{Z - e(\boldsymbol{X}, \boldsymbol{\beta})\}}{e(\boldsymbol{X}, \boldsymbol{\beta})}\{Y_1 - m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1)\}\right] \quad (1)$$

for *any general functions* of $\boldsymbol{X}$ $e(\boldsymbol{X}, \boldsymbol{\beta})$ and $m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1)$ (that *may or may not* be equal to the *true* propensity score or *true* regression)

- Thus, for $\widehat{\mu}_{1,DR}$ to estimate $E(Y_1)$, the *second term* in (1) must $= 0$!
- *When does the second term $= 0$*?

5. Doubly Robust Estimator

Scenario 1: *Postulated propensity score model* $e(\boldsymbol{X}, \boldsymbol{\beta})$ is *correct*, but *postulated regression model* $m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1)$ is *not*, i.e.,

- $e(\boldsymbol{X}, \boldsymbol{\beta}) = e(\boldsymbol{X}) = E(Z|\boldsymbol{X})$ ($= E(Z|Y_1, \boldsymbol{X})$ by *no unmeasured confounders*)
- $m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1) \neq E(Y|Z = 1, \boldsymbol{X})$

Is the second term = 0 under these conditions?

5. Doubly Robust Estimator

$$
\begin{aligned}
& E\left[\frac{\{Z-e(\boldsymbol{X})\}}{e(\boldsymbol{X})}\{Y_1-m_1(\boldsymbol{X},\boldsymbol{\alpha}_1)\}\right] \\
&\qquad = E\left(E\left[\frac{\{Z-e(\boldsymbol{X})\}}{e(\boldsymbol{X})}\{Y_1-m_1(\boldsymbol{X},\boldsymbol{\alpha}_1)\}\middle| Y_1,\boldsymbol{X}\right]\right) \\
&\qquad = E\left(\{Y_1-m_1(\boldsymbol{X},\boldsymbol{\alpha}_1)\}\, E\left[\frac{\{Z-e(\boldsymbol{X})\}}{e(\boldsymbol{X})}\middle| Y_1,\boldsymbol{X}\right]\right) \\
&\qquad = E\left(\{Y_1-m_1(\boldsymbol{X},\boldsymbol{\alpha}_1)\}\frac{\{E(Z|Y_1,\boldsymbol{X})-e(\boldsymbol{X})\}}{e(\boldsymbol{X})}\right) \\
&\qquad = E\left(\{Y_1-m_1(\boldsymbol{X},\boldsymbol{\alpha}_1)\}\frac{\{E(Z|\boldsymbol{X})-e(\boldsymbol{X})\}}{e(\boldsymbol{X})}\right) \qquad (1) \\
&\qquad = E\left(\{Y_1-m_1(\boldsymbol{X},\boldsymbol{\alpha}_1)\}\frac{\{e(\boldsymbol{X})-e(\boldsymbol{X})\}}{e(\boldsymbol{X})}\right) = 0
\end{aligned}
$$

(1) uses *no unmeasured confounders*

5. Doubly Robust Estimator

Result: *As long as* the propensity score model is *correct*, even if the *postulated regression model* is *incorrect*

- $\widehat{\mu}_{1,DR}$ estimates $E(Y_1)$
- *Similarly*, $\widehat{\mu}_{0,DR}$ estimates $E(Y_0)$
- And hence $\widehat{\Delta}_{DR}$ estimates Δ!

5. Doubly Robust Estimator

Scenario 2: *Postulated regression model* $m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1)$ is *correct*, but *postulated propensity score model* $e(\boldsymbol{X}, \boldsymbol{\beta})$ is *not*

- $e(\boldsymbol{X}, \boldsymbol{\beta}) \neq e(\boldsymbol{X}) = E(Z|\boldsymbol{X})$
- $m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1) = E(Y|Z = 1, \boldsymbol{X})$ ($= E(Y_1|\boldsymbol{X})$ by *no unmeasured confounders*)

Is the second term = 0 under these conditions?

5. Doubly Robust Estimator

$$
\begin{aligned}
E&\left[\frac{\{Z-e(\boldsymbol{X},\boldsymbol{\beta})\}}{e(\boldsymbol{X},\boldsymbol{\beta})}\{Y_1 - E(Y|Z=1,\boldsymbol{X})\}\right] \\
&= E\left(\left[\frac{\{Z-e(\boldsymbol{X},\boldsymbol{\beta})\}}{e(\boldsymbol{X},\boldsymbol{\beta})}\{Y_1 - E(Y|Z=1,\boldsymbol{X})\}\middle| Z,\boldsymbol{X}\right]\right) \\
&= E\left(\frac{\{Z-e(\boldsymbol{X},\boldsymbol{\beta})\}}{e(\boldsymbol{X},\boldsymbol{\beta})}\; E\left[\{Y_1 - E(Y|Z=1,\boldsymbol{X})\}| Z,\boldsymbol{X}\right]\right) \\
&= E\left(\frac{\{Z-e(\boldsymbol{X},\boldsymbol{\beta})\}}{e(\boldsymbol{X},\boldsymbol{\beta})}\{E(Y_1|Z,\boldsymbol{X}) - E(Y|Z=1,\boldsymbol{X})\}\right) \\
&= E\left(\frac{\{Z-e(\boldsymbol{X},\boldsymbol{\beta})\}}{e(\boldsymbol{X},\boldsymbol{\beta})}\{E(Y_1|\boldsymbol{X}) - E(Y_1|\boldsymbol{X})\}\right) = 0 \qquad (1)
\end{aligned}
$$

(1) uses *no unmeasured confounders*, which says

$$
E(Y|Z=1,\boldsymbol{X}) = E(Y_1|Z=1,\boldsymbol{X}) = E(Y_1|\boldsymbol{X}) = E(Y_1|Z,\boldsymbol{X})
$$

5. Doubly Robust Estimator

Result: *As long as* the regression model is *correct*, even if the *postulated propensity model* is *incorrect*

- $\widehat{\mu}_{1,DR}$ estimates $E(Y_1)$
- *Similarly*, $\widehat{\mu}_{0,DR}$ estimates $E(Y_0)$
- And hence $\widehat{\Delta}_{DR}$ estimates Δ!

Obviously: From these calculations if *both* models are correct, $\widehat{\Delta}_{DR}$ estimates Δ!

- Of course, if *both* are *incorrect*, $\widehat{\Delta}_{DR}$ does not estimate Δ (not *consistent*)

5. Doubly Robust Estimator

Summary: If

- The *regression model* is *incorrect* but the *propensity model* is *correct*

OR

- The *propensity model* is *incorrect* but the *regression model* is *correct*

then $\widehat{\Delta}_{DR}$ is a (*consistent*) estimator for Δ!

Definition: This property is referred to as *double robustness*

5. Doubly Robust Estimator

Remarks: The *doubly robust* estimator

- Offers *protection* against mismodeling
- If $e(\boldsymbol{X})$ is modeled *correctly*, will have *smaller variance* than the simple *inverse weighted* estimator (in *large samples*)
- If $E(Y|Z, \boldsymbol{X})$ is modeled correctly, may have *larger variance* (in *large samples*) than the *regression* estimator ...
- ... but gives *protection* in the event it is *not*

6. Some Interesting Issues

Issue 1: How do we get *standard errors* for $\widehat{\Delta}_{DR}$?

- One way: use standard *large sample theory*, which leads to the so-called *sandwich estimator*

$$\sqrt{n^{-2}\sum_{i=1}^{n}\widehat{I}_i^2}$$

$$\widehat{I}_i = \frac{Z_iY_i}{e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} - \frac{\{Z_i - e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})\}}{e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} m_1(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}}_1) - \left[\frac{(1-Z_i)Y_i}{1-e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} + \frac{\{Z_i - e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})\}}{1-e(\boldsymbol{X}_i,\widehat{\boldsymbol{\beta}})} m_0(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}}_0)\right] - \widehat{\Delta}_{DR}$$

- Use the *bootstrap* (i.e., *resample* B data sets of size n)

6. Some Interesting Issues

Issue 2: Computation?

- Is there *software*? A *SAS* procedure is *coming soon* . . .
- In many situations, it is possible to find $\widehat{\Delta}_{DR}$ by fitting a single *regression model* for $E(Y|Z,\boldsymbol{X})$ that *includes*

$$\frac{Z}{e(\boldsymbol{X},\widehat{\boldsymbol{\beta}})} \text{ and } \frac{(1-Z)}{\{1-e(\boldsymbol{X},\widehat{\boldsymbol{\beta}})\}}$$

  as covariates.

- See Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* 61, 962–972.

6. Some Interesting Issues

Issue 3: How to select elements of $\boldsymbol{X}$ to include in the models?

- For the inverse weighted estimators:
  - Variables *unrelated to exposure* but *related to outcome* should *always* be included in the propensity score model $\Longrightarrow$ *increased precision*
  - Variable *related to exposure* but *unrelated to outcome* can be *omitted* $\Longrightarrow$ *decreased precision*
- See Brookhart, M. A. et al. (2006). Variable selection for propensity score models. *American Journal of Epidemiology* 163, 1149–1156.
- Best way to select for DR estimation is an *open problem*
- See Brookhart, M. A. and van der Laan, M. J. (2006). A semiparametric model selection criterion with applications to the marginal structural model. *Computational Statistics and Data Analysis* 50, 475–498.

6. Some Interesting Issues

Issue 4: Variants on doubly robust estimators?

- See Tan, Z. (2006) A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association* 101, 1619–1637.

6. Some Interesting Issues

Issue 5: Connection to "missing data" problems

- $(Y_0, Y_1, Z, \boldsymbol{X})$ are the "*full data*" we wish we could see, but we only *observe* $Y = Y_1 Z + Y_0(1 - Z)$
- $Z = 1$ means Y_1 is observed but Y_0 is *missing*; happens with probability $e(\boldsymbol{X})$; vice versa for $Z = 0$
- Missing data theory of *Robins, Rotnitzky, and Zhao (1994)* applies and leads to the *doubly robust estimator*
- The theory shows that the *doubly robust estimator* with *all of* $e(\boldsymbol{X})$, $m_0(\boldsymbol{X}, \boldsymbol{\alpha}_0)$, $m_1(\boldsymbol{X}, \boldsymbol{\alpha}_1)$ *correctly specified* has *smallest variance* among all estimators that require one to model the propensity score *correctly* (but make *no further assumptions* about anything)

7. Discussion

- *Regression modeling* and *inverse propensity score weighting* are two popular approaches when one is willing to assume *no unmeasured confounders*
- The *double robust estimator* combines both and offers *protection* against *mismodeling*
- Offers gains in *precision* of estimation over simple inverse weighting
- May not be as precise as *regression modeling* when the *regression* is *correctly modeled*, but adds protection, and modifications are available
- Doubly robust estimators are also available for *more complicated* problems