Double Robustness in Estimation of Causal Treatment Effects

Marie Davidian
Department of Statistics
North Carolina State University

http://www.stat.ncsu.edu/~davidian

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*

\[ \text{mean if } \text{entire population} \text{ exposed} - \text{mean if } \text{entire population} \text{ not exposed} \]
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:

\[ Z = 1 \text{ if treatment, } = 0 \text{ if control} \]
\[ X \text{ vector of pre-exposure covariates} \]
\[ Y \text{ observed outcome} \]

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

Based on these data: Estimate the \textit{average causal treatment effect}

- To \textit{formalize} what we mean by this and what the issues are, appeal to the notion of \textit{potential outcomes} or \textit{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

\[
Y = Y_1Z + Y_0(1 - Z)
\]

- 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* \(\Delta\) 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 – $Y \parallel Z$
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 $Y \parallel Z$, then $E(Y|Z = z) = E(Y)$ for any $z$
2. Counterfactual Model

**Challenge, again:** Based on *observed data* \((Y_i, Z_i, X_i), i = 1, \ldots, n\), estimate

\[
\Delta = \mu_1 - \mu_0 = E(Y_1) - E(Y_0)
\]

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

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

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

\[
n_0 = \sum_{i=1}^{n} (1 - Z_i) = \# \text{ observed to receive control}
\]
2. Counterfactual Model

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

- $\overline{Y}^{(1)}$ estimates $E(Y|Z = 1) = \text{population mean outcome among those observed to receive treatment}$
- $E(Y|Z = 1) = E\{Y_1 Z + Y_0 (1 - Z)|Z = 1\} = E(Y_1|Z = 1)$...
- ...which is not equal to $E(Y_1) = \text{mean outcome if entire population received treatment}$
- 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
  \[
  E(Y|Z = 1) = E\{Y_1Z + Y_0(1-Z) | Z = 1\} = E(Y_1|Z = 1) = E(Y_1)
  \]
  and similarly \( E(Y|Z = 0) = E(Y_0) \)

- \( \bar{Y}^{(1)} - \bar{Y}^{(0)} \) is an *unbiased estimator* for \( \Delta \)
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:

  \[(Y_0, Y_1) \parallel Z\]

- 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 \(\Delta\)
2. Counterfactual Model

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

- If $X$ contains *all confounders*, then among subjects sharing the same $X$ 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 $X$:

  $$(Y_0, Y_1) \bot Z \mid X$$

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

- If we *believe* no unmeasured confounders, can estimate $\Delta$ by *appropriate adjustment*...
3. Adjustment by Regression Modeling

**Regression of $Y$ on $Z$ and $X$:** We can identify the regression

$$E(Y|Z, X),$$

as this depends on the observed data

- E.g., for *continuous outcome*, $E(Y|Z, X) = \alpha_0 + \alpha_Z Z + X^T \alpha_X$

- In general, $E(Y|Z = 1, X)$ is the regression among *treated*, $E(Y|Z = 0, X)$ among *control*

**Usefulness:** *Averaging* over all possible values of $X$ (*both* treatments)

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

and similarly $E\{ E(Y|Z = 0, X) \} = E(Y_0)$
3. Adjustment by Regression Modeling

Thus: Under no unmeasured confounders

\[ \Delta = E(Y_1) - E(Y_0) \]
\[ = E\{E(Y|Z = 1, X)\} - E\{E(Y|Z = 0, X)\} \]
\[ = E\{E(Y|Z = 1, X) - E(Y|Z = 0, X)\} \]

• Suggests postulating a model for the outcome regression \( E(Y|Z, X) \), fitting the model, and then and averaging the resulting estimates of

\[ E(Y|Z = 1, X) - E(Y|Z = 0, X) \]

over all observed \( X \) (both groups) to estimate \( \Delta \)
3. Adjustment by Regression Modeling

Example – continuous outcome: Suppose the true regression is

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

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

$$E(Y|Z = 1, X) - E(Y|Z = 0, X) = \alpha_0 + \alpha_Z (1) + X^T \alpha_X - \alpha_0 - \alpha_Z (0) - X^T \alpha_X$$

$$= \alpha_Z$$

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

- Can thus estimate $\Delta$ directly from fitting this model (e.g. by least squares); don’t even need to average!

- $\hat{\Delta} = \hat{\alpha}_Z$
3. Adjustment by Regression Modeling

Example – binary outcome: Suppose the true regression is

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

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

\[
E(Y|Z = 1, X) - E(Y|Z = 0, X) = \frac{\exp(\alpha_0 + \alpha_Z + X^T \alpha_X)}{1 + \exp(\alpha_0 + \alpha_Z + X^T \alpha_X)} - \frac{\exp(\alpha_0 + X^T \alpha_X)}{1 + \exp(\alpha_0 + X^T \alpha_X)}
\]

- Logistic regression yields \((\hat{\alpha}_0, \hat{\alpha}_Z, \hat{\alpha}_X)\)

- Estimate \(\Delta\) by averaging over all observed \(X_i\)

\[
\hat{\Delta} = n^{-1} \sum_{i=1}^{n} \left\{ \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_Z + X_i^T \hat{\alpha}_X)}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_Z + X_i^T \hat{\alpha}_X)} - \frac{\exp(\hat{\alpha}_0 + X_i^T \hat{\alpha}_X)}{1 + \exp(\hat{\alpha}_0 + X_i^T \hat{\alpha}_X)} \right\}
\]
3. Adjustment by Regression Modeling

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

- Thus, if we substitute estimates for $E(Y|Z = 1, X)$ and $E(Y|Z = 0, X)$ 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 $\Delta$

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

**Moral:** Estimation of $\Delta$ via regression modeling *requires* that the postulated regression model is *correct*
4. Adjustment by Inverse Weighting

**Propensity score:** Probability of treatment given covariates

\[ e(X) = P(Z = 1|X) = E\{I(Z = 1)|X\} = E(Z|X) \]

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

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

\[ e(X, \beta) \implies e(X, \hat{\beta}) \]
4. Adjustment by Inverse Weighting

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

\[
\hat{\Delta}_{\text{IPW,1}} = n^{-1} \sum_{i=1}^{n} \frac{Z_i Y_i}{e(X_i, \hat{\beta})} - n^{-1} \sum_{i=1}^{n} \frac{(1 - Z_i) Y_i}{1 - e(X_i, \hat{\beta})}
\]

- **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 \( n^{-1} \sum_{i=1}^{n} \frac{Z_i Y_i}{e(X_i, \hat{\beta})} \)

- By the law of large numbers, this should estimate the mean of a term in the sum with \( \hat{\beta} \) replaced by the quantity it estimates
4. Adjustment by Inverse Weighting

If:

\[ e(X, \beta) = e(X), \text{ the true propensity score} \]

\[
E \left\{ \frac{ZY}{e(X)} \right\} = E \left\{ \frac{ZY_1}{e(X)} \right\} = E \left[ E \left\{ \frac{ZY_1}{e(X)} \mid Y_1, X \right\} \right] 
\]

\[
= E \left\{ \frac{Y_1}{e(X)} E(Z \mid Y_1, X) \right\} = E \left\{ \frac{Y_1}{e(X)} E(Z \mid X) \right\} 
\]

\[
= E \left\{ \frac{Y_1}{e(X)} e(X) \right\} = E(Y_1) 
\]

(1) follows because \( ZY = Z\{Y_1Z + Y_0(1-Z)\} = Z^2Y_1 + Z(1-Z)Y_0 \)
and \( Z^2 = Z \) and \( Z(1-Z) = 0 \) (binary)

(2) follows because \( (Y_0, Y_1) \parallel Z \mid X \) (no unmeasured confounders)

(3) follows because \( e(X) = E(Z \mid X) \)

Similarly:

\[
E \left\{ \frac{(1-Z)Y}{1-e(X)} \right\} = E(Y_0) 
\]
4. Adjustment by Inverse Weighting

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

- Thus, if we substitute estimates for \( e(X) \) based on a *postulated propensity score model*, this *postulated model* must be identical to the *true propensity score*

- If not, \( \hat{\Delta}_{IPW,1} \) will not necessarily estimate \( \Delta \)

**Moral:** Estimation of \( \Delta \) 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 $\Delta$ based on *regression modeling* requires *correct* postulated regression model

- Estimator for $\Delta$ 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:

\[
\hat{\Delta}_{DR} = n^{-1} \sum_{i=1}^{n} \left[ \frac{Z_i Y_i}{e(X_i, \hat{\beta})} - \frac{Z_i - e(X_i, \hat{\beta})}{e(X_i, \hat{\beta})} m_1(X_i, \hat{\alpha}_1) \right]
- n^{-1} \sum_{i=1}^{n} \left[ \frac{(1 - Z_i) Y_i}{1 - e(X_i, \hat{\beta})} + \frac{Z_i - e(X_i, \hat{\beta})}{1 - e(X_i, \hat{\beta})} m_0(X_i, \hat{\alpha}_0) \right]
= \hat{\mu}_{1,DR} - \hat{\mu}_{0,DR}
\]

- \(e(X, \beta)\) is a postulated model for the true propensity score
  \(e(X) = E(Z|X)\) (fitted by logistic regression)

- \(m_0(X, \alpha_0)\) and \(m_1(X, \alpha_1)\) are postulated models for the true regressions \(E(Y|Z = 0, X)\) and \(E(Y|Z = 1, X)\) (fitted by least squares)
5. Doubly Robust Estimator

Modified estimator:

\[
\hat{\Delta}_{DR} = n^{-1} \sum_{i=1}^{n} \left[ \frac{Z_i Y_i}{e(X_i, \hat{\beta})} - \frac{Z_i - e(X_i, \hat{\beta})}{e(X_i, \hat{\beta})} m_1(X_i, \hat{\alpha}_1) \right] \\
- n^{-1} \sum_{i=1}^{n} \left[ \frac{(1 - Z_i) Y_i}{1 - e(X_i, \hat{\beta})} + \frac{Z_i - e(X_i, \hat{\beta})}{1 - e(X_i, \hat{\beta})} m_0(X_i, \hat{\alpha}_0) \right]
\]

\[
= \hat{\mu}_{1,DR} - \hat{\mu}_{0,DR}
\]

- \(\hat{\mu}_{1,DR}\) (and \(\hat{\mu}_{0,DR}\) and hence \(\hat{\Delta}_{DR}\)) may be viewed as taking the inverse weighted estimator and “augmenting” it by a second term.

What does this estimate? Consider \(\hat{\mu}_{1,DR}\) (\(\hat{\mu}_{0,DR}\) similar)
5. Doubly Robust Estimator

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

- By the law of large numbers, \( \hat{\mu}_{1,DR} \) estimates the mean of a term in the sum with \( \beta \) and \( \alpha_1 \) replaced by the quantities they estimate.

- That is, \( \hat{\mu}_{1,DR} \) estimates

\[
E \left[ \frac{ZY}{e(X, \beta)} - \frac{Z - e(X, \beta)}{e(X, \beta)} m_1(X, \alpha_1) \right] = E \left[ \frac{ZY_1}{e(X, \beta)} - \frac{Z - e(X, \beta)}{e(X, \beta)} m_1(X, \alpha_1) \right]
\]

\[
= E \left[ Y_1 + \frac{Z - e(X, \beta)}{e(X, \beta)} \{Y_1 - m_1(X, \alpha_1)\} \right] \quad \text{(by algebra)}
\]

\[
= E(Y_1) + E \left[ \frac{Z - e(X, \beta)}{e(X, \beta)} \{Y_1 - m_1(X, \alpha_1)\} \right]
\]
5. Doubly Robust Estimator

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

$$E(Y_1) + E \left[ \frac{\{Z - e(X, \beta)\}}{e(X, \beta)} \{Y_1 - m_1(X, \alpha_1)\} \right]$$

for any general functions of $X \ e(X, \beta)$ and $m_1(X, \alpha_1)$ (that may or may not be equal to the true propensity score or true regression)

- Thus, for $\hat{\mu}_{1,DR}$ 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 $e(X, \beta)$ is correct, but postulated regression model $m_1(X, \alpha_1)$ is not, i.e.,

- $e(X, \beta) = e(X) = E(Z|X) \ ( = E(Z|Y_1, X) \text{ by no unmeasured confounders})$
- $m_1(X, \alpha_1) \neq E(Y|Z = 1, X)$

*Is the second term = 0 under these conditions?*
5. Doubly Robust Estimator

\[
E \left[ \frac{\{Z - e(X)\}}{e(X)} \left\{ Y_1 - m_1(X, \alpha_1) \right\} \right] \\
= E \left( E \left[ \frac{\{Z - e(X)\}}{e(X)} \left\{ Y_1 - m_1(X, \alpha_1) \right\} \right| Y_1, X \right) \\
= E \left( \{Y_1 - m_1(X, \alpha_1)\} E \left[ \frac{\{Z - e(X)\}}{e(X)} \right| Y_1, X \right) \\
= E \left( \{Y_1 - m_1(X, \alpha_1)\} \frac{E(Z|Y_1, X) - e(X)}{e(X)} \right) \\
= E \left( \{Y_1 - m_1(X, \alpha_1)\} \frac{e(X) - e(X)}{e(X)} \right) = 0 \\
\]

(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

- $\hat{\mu}_{1,DR}$ estimates $E(Y_1)$

- Similarly, $\hat{\mu}_{0,DR}$ estimates $E(Y_0)$

- And hence $\hat{\Delta}_{DR}$ estimates $\Delta$!
5. Doubly Robust Estimator

**Scenario 2:** Postulated regression model $m_1(X, \alpha_1)$ is correct, but postulated propensity score model $e(X, \beta)$ is not

- $e(X, \beta) \neq e(X) = E(Z|X)$
- $m_1(X, \alpha_1) = E(Y|Z = 1, X) \ (= E(Y_1|X) \text{ by no unmeasured confounders})$

*Is the second term $= 0$ under these conditions?*
5. Doubly Robust Estimator

\[
E \left[ \frac{\{Z - e(X, \beta)\}}{e(X, \beta)} \{Y_1 - E(Y|Z = 1, X)\} \right]
\]

\[
= E \left( \left[ \frac{\{Z - e(X, \beta)\}}{e(X, \beta)} \{Y_1 - E(Y|Z = 1, X)\} \right] | Z, X \right) 
\]

\[
= E \left( \left[ \frac{\{Z - e(X, \beta)\}}{e(X, \beta)} \right] E \left[ \{Y_1 - E(Y|Z = 1, X)\} | Z, X \right] \right) 
\]

\[
= E \left( \left[ \frac{\{Z - e(X, \beta)\}}{e(X, \beta)} \right] \{E(Y_1|Z, X) - E(Y|Z = 1, X)\} \right) 
\]

\[
= E \left( \left[ \frac{\{Z - e(X, \beta)\}}{e(X, \beta)} \right] \{E(Y_1|X) - E(Y_1|X)\} \right) = 0 \quad (1)
\]

(1) uses \textit{no unmeasured confounders}, which says

\[
E(Y|Z = 1, X) = E(Y_1|Z = 1, X) = E(Y_1|X) = E(Y_1|Z, X)
\]
5. Doubly Robust Estimator

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

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

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

- Of course, if *both* are incorrect, $\hat{\Delta}_{DR}$ does not estimate $\Delta$ (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 $\hat{\Delta}_{DR}$ is a (*consistent*) estimator for $\Delta$!

**Definition:** This property is referred to as *double robustness*
5. Doubly Robust Estimator

Remarks: The *doubly robust* estimator

- Offers *protection* against mismodeling
- If $e(X)$ is modeled *correctly*, will have *smaller variance* than the simple *inverse weighted* estimator (in *large samples*)
- If $E(Y|Z, X)$ 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 $\hat{\Delta}_{DR}$?

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

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

$$\hat{I}_i = \frac{Z_i Y_i}{e(X_i, \hat{\beta})} - \frac{\{Z_i - e(X_i, \hat{\beta})\}}{e(X_i, \hat{\beta})} m_1(X_i, \hat{\alpha}_1)$$

$$- \left[ \frac{(1 - Z_i)Y_i}{1 - e(X_i, \hat{\beta})} + \frac{\{Z_i - e(X_i, \hat{\beta})\}}{1 - e(X_i, \hat{\beta})} m_0(X_i, \hat{\alpha}_0) \right] - \hat{\Delta}_{DR}$$

- 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 $\hat{\Delta}_{DR}$ by fitting a single *regression model* for $E(Y|Z, X)$ that *includes*

\[
\frac{Z}{e(X, \hat{\beta})} \text{ and } \frac{(1 - Z)}{\{1 - e(X, \hat{\beta})\}}
\]

as covariates.

6. Some Interesting Issues

**Issue 3:** How to select elements of $X$ 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 $\implies$ *increased precision*
  - Variable *related to exposure* but *unrelated to outcome* can be omitted $\implies$ *decreased precision*


- Best way to select for DR estimation is an *open problem*

6. Some Interesting Issues

**Issue 4:** Variants on doubly robust estimators?

6. Some Interesting Issues

Issue 5: Connection to “missing data” problems

- \((Y_0, Y_1, Z, X)\) 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 \(e(X)\); 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 \(e(X)\), \(m_0(X, \alpha_0)\), \(m_1(X, \alpha_1)\) 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