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. Counterfactual model
3. Adjustment by regression modeling
4. Adjustment by “inverse weighting”
5. Doubly-robust estimator
6. Connection to missing data problems and Robins, Rotnitzky, and Zhao (JASA 1994)
7. Discussion

1. Introduction

Simplest situation:
- Observational point exposure study
- Continuous or discrete (e.g. binary) response \( Y \)
- “Treatment” (exposure) or “Control” (no exposure)

Objective: Make causal inference on the effect of treatment
- \( \text{Effect} \): comparison, i.e. difference in response between treatment and control
- Would like to be able to say that such an effect is attributable, or “caused by” treatment

More precisely: Usual “population-level” goal
- Not possible to see each subject under both treatment and control (so cannot evaluate individual effects)
- Compare mean (average) response on treatment to mean (average) response on control
- That is, compare mean response if entire population received treatment to that if entire population received control
- Difference of these means is the “average causal treatment effect”
1. Introduction

**Complication:** In observational studies, treatment exposure may be **confounded** with other factors

- The fact that a subject was exposed to treatment or not is associated with subject characteristics...
- ...that may also be associated with the way the subject might respond under treatment and control
- Treatment effects deduced from data on observed exposures will **not necessarily** have the desired **causal interpretation**...
- ...**unless** the analysis can take appropriate account of **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 the assumptions are **wrong**?
- The property of **double robustness** offers **protection** against some particular incorrect assumptions...
- ...and can lead to **more precise** inference

2. Counterfactual Model

**Define:**
- \(Z = 1\) if treatment, \(= 0\) if control
- \(X\) vector of pre-exposure **covariates**
- \(Y\) observed response
- \(Z\) is **observed** (not assigned)
- **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 **average causal treatment effect**

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

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

- \(Y_0\) response the subject would have if s/he received control
- \(Y_1\) response the subject would have if s/he received treatment

**Average causal treatment effect:**

- The **probability distribution** of \(Y_0\) represents how responses in the population would turn out if **everyone** received control, with mean \(E(Y_0) = P(Y_0 = 1)\) for binary response
- The **probability distribution** of \(Y_1\) represents this if **everyone** received treatment, with mean \(E(Y_1) = P(Y_1 = 1)\) for binary response
- **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 $Y_i = Y_{i1};$
- If $i$ was not exposed (control), $Z_i = 0$ and $Y_i = Y_{i0};$
- Do not confuse the observed response $Y$ for a subject observed to receive treatment with the counterfactual response $Y_1 \ldots$
- \ldots and similarly for unexposed (control) subjects ($Y$ vs. $Y_0$)

Challenge: Would like to estimate $\Delta$ based on the observed data
- First, a quick review \ldots

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 $\iff Z$

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 averages: Averages among those observed to receive treatment or control
\[ Y^{(1)} = n^{-1}_1 \sum_{i=1}^{n} Z_i Y_i, \quad Y^{(0)} = n^{-1}_0 \sum_{i=1}^{n} (1 - Z_i) Y_i \]
\[ n_1 = \sum_{i=1}^{n} Z_i = \# \text{ subjects observed to receive treatment} \]
\[ n_0 = \sum_{i=1}^{n} (1 - Z_i) = \# \text{ observed to receive control} \]
2. Counterfactual Model

What is being estimated? If we estimate \( \Delta \) by \( Y(1) - Y(0) \)

- \( Y(1) \) estimates \( E(Y|Z = 1) \) = population average response 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 \( E(Y_1) \) = average response if entire population received treatment

- Similarly, \( Y(0) \) 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

In contrast: Observational study

- Exposure to treatment is not controlled, so exposure may be related to the way a subject might potentially respond:
  \( (Y_0, Y_1) \perp Z \)

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

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 response:
  \( (Y_0, Y_1) \perp Z \)

- 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 \( E(Y|Z = 0) = E(Y_0) \)

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

2. Counterfactual Model

Confounders: It may be possible to identify covariates related to both potential response 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 response \( (Y_0, Y_1) \), i.e. \( (Y_0, Y_1) \) and \( Z \) are independent conditional on \( X \):
  \( (Y_0, Y_1) \perp Z | X \)

- 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 response, $E(Y|Z, X) = \alpha_0 + \alpha_2 Z + X^T \alpha_X$
- 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)

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

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

3. Adjustments by Regression Modeling

Example - continuous response: Suppose the true regression is

$$E(Y|Z, X) = \alpha_0 + \alpha_2 Z + X^T \alpha_X$$

- Can show that $E(Y|Z = 1, X) - E(Y|Z = 0, X) = \alpha_2$
- Can thus estimate $\Delta$ directly from fitting this model (e.g. by least squares)

- $\bar{\Delta} = \hat{\alpha}_2$
- Don’t even need to average

3. Adjustment by Regression Modeling

Example - binary response: Suppose the true regression is

$$E(Y|Z, X) = \frac{\exp(\alpha_0 + \alpha_2 Z + X^T \alpha_X)}{1 + \exp(\alpha_0 + \alpha_2 Z + X^T \alpha_X)}$$

- This model implies

$$E(Y|Z = 1, X) - E(Y|Z = 0, X)$$

$$= \frac{\exp(\alpha_0 + \alpha_2 Z + X^T \alpha_X)}{1 + \exp(\alpha_0 + \alpha_2 Z + X^T \alpha_X)} - \frac{\exp(\alpha_0 + X^T \alpha_X)}{1 + \exp(\alpha_0 + X^T \alpha_X)}$$

- Logistic regression yields $(\hat{\alpha}_0, \hat{\alpha}_2, \hat{\alpha}_X)$
- Estimate $\Delta$ by averaging over all observed $X_i$

$$\bar{\Delta} = n^{-1} \sum_{i=1}^n \left\{ \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_2 Z + X^T \hat{\alpha}_X)}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_2 Z + X^T \hat{\alpha}_X)} - \frac{\exp(\hat{\alpha}_0 + X^T \hat{\alpha}_X)}{1 + \exp(\hat{\alpha}_0 + X^T \hat{\alpha}_X)} \right\}$$
3. Adjustment by Regression Modeling

Critical: For the argument on slide 18 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 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.

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.

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

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

\[
E \left\{ \frac{ZY}{e(X)} \right\} = E \left\{ \frac{ZY_i}{e(X_i)} \right\} = E \left\{ \frac{Z}{e(X)} E(Y_i | X) \right\} = E(Y_i) = E \left\{ \frac{1}{e(X)} E(Y_i | X) \right\} = E(Y_i)
\]

(1)

(2)

(3)

Similarly:

\[
E \left\{ \frac{(1 - Z)Y}{1 - e(X)} \right\} = E(Y_0)
\]

Critical: For the argument on slide 25 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, \(\hat{\Delta}_{IPW,1}\) and \(\hat{\Delta}_{IPW,2}\) 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

Modified estimator:

\[
\hat{\Delta}_{DR} = n^{-1} \sum_{i=1}^{n} \left[ \frac{Z_i Y_i}{e(X_i, \beta)} - \frac{Z_i - e(X_i, \hat{\beta})}{e(X_i, \beta)} m_1(X_i, \hat{\alpha}_1) \right] \\
- n^{-1} \sum_{i=1}^{n} \left[ \frac{(1 - Z_i) Y_i}{1 - e(X_i, \beta)} + \frac{Z_i - e(X_i, \hat{\beta})}{1 - e(X_i, \beta)} m_0(X_i, \alpha_0) \right] \\
= \hat{\mu}_{DR} - \hat{\mu}_{0,DR}
\]

- \( e(X, \beta) \) 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)

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

5. Doubly Robust Estimator

First term: We showed on slide 25 that \( E \left\{ \frac{Z Y}{e(X)} \right\} = E(Y_1) \)

Second term:

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

5. Doubly Robust Estimator

Scenario 1: Postulated model \( e(X, \beta) \) is correct, but postulated model \( m_1(X, \alpha_1) \) is not
- \( e(X, \beta) = e(X) = E(Z|X) \)
- \( m_1(X, \alpha_1) \neq E(Y|Z = 1, X) \)

What does \( \hat{\mu}_{1,DR} \) estimate under these conditions?
- By 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
- So look at

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

Result: 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 model $m_1(X, \alpha_1)$ is correct, but postulated model $e(X, \beta)$ is not

- $e(X, \beta) \neq e(X) = E(Z|X)$
- $m_1(X, \alpha_1) = E(Y|Z = 1, X)$

What does $\hat{\mu}_{1,DR}$ estimate under these conditions?

- Look at

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

$$= E \left\{ \frac{ZY}{e(X, \beta)} \right\} - E \left\{ \frac{Z - e(X, \beta)}{e(X, \beta)} \right\} E(Y|Z = 1, X)$$

First term:

$$E \left\{ \frac{ZY_1}{e(X, \beta)} \right\} = E \left\{ \frac{ZY_1}{e(X, \beta)} \left| Y_1, X \right. \right\} = E \left\{ \frac{Y_1 E(Z|Y_1, X)}{e(X, \beta)} \right\}$$

using no unmeasured confounders

Second term:

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

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

$$= E \left\{ \frac{e(X) - e(X, \beta)}{e(X, \beta)} \right\} E(Y_1|X)$$

$$= E \left\{ \frac{Y_1 e(X)}{e(X, \beta)} \right\} - E(Y_1)$$

using no unmeasured confounders and

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

Result:

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

so 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 $\Delta$!
5. Doubly Robust Estimator

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); however...

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

Remarks: The doubly robust estimator

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

6. Connection to Missing Data and RRZ94

Can view this as a “missing data” problem: For each subject

- $(Y_0, Y_1, Z, X)$ are the “full data” we wish we could see
- Observe $Y = Y_1 Z + Y_0 (1 - Z)$
  - If $Z = 1$, see $(Y_1, Z, X)$, and $Y_0$ is missing; happens with probability $\epsilon(X)$ (does not depend on $Y_0$)
  - If $Z = 0$, see $(Y_0, Z, X)$, and $Y_1$ is missing; with probability $1 - \epsilon(X)$ (does not depend on $Y_0$)
- $Z$ plays the role of a missingness indicator
- $Y_0$ and $Y_1$ are missing at random; probability of missingness does not depend on the thing we don’t get to see

Result: Can exploit theory of Robins, Rotnitzky, and Zhao (1994) for data missing at random

What RRZ did: If willing to assume $\epsilon(X)$ can be modeled correctly, identified a class of estimators that are consistent when some data are missing at random

- For $\mu_1$, these estimators have form
  \[ \frac{1}{n-1} \sum_{i=1}^{n} z_i y_i \left( \frac{1 - \epsilon(x_i; \beta)}{\epsilon(x_i; \beta)} - \frac{(z_i - \epsilon(x_i; \beta))}{\epsilon(x_i; \beta)} \phi(x_i) \right) \]
  where $\phi(X)$ is any function of $X$ and $\epsilon(X, \beta) = \epsilon(X) = E(Z|X)$ (similar for $\mu_0$)
  - Second term “augments” simple IW estimator to improve precision
  - If $\phi(X) = E(Y|Z = 1, X)$, the estimator has smallest variance (in large samples) among estimators in the class
7. Discussion

- Adjustment for confounders in estimating causal effects needs unverifiable assumption of no unmeasured confounders
- Regression modeling and inverse propensity score weighting are two popular approaches
- The double robust estimator combines both and offers protection against mismodeling
- Offers gains in efficiency of estimation over inverse weighting
- Is probably not as good as regression modeling when the regression is correctly modeled, but adds protection
- Doubly robust estimators are also available for more complicated problems