Ch 7. Propensity Score Approach

Jae-Kwang Kim

ISU

Basic Setup

- $z_i = (x_i, y_i), i = 1, 2, \cdots, n$: random sample
- Parameter of interest (θ_0): defined by the (unique) solution to $E\{U(\theta; Z)\} = 0$.
- Under complete response of $z_1, \cdots, z_n$, a consistent estimator of θ_0 is obtained by solving

$$\hat{U}_n(\theta) \equiv \frac{1}{n} \sum_{i=1}^{n} U(\theta; z_i) = 0$$

  for θ. We assume that the solution $\hat{\theta}_n$ is unique.
- Under some conditions, $\hat{\theta}_n$ converges in probability to θ_0.
- Note that $\hat{\theta}_n$ is asymptotically unbiased for θ_0 if $E\{U(\theta_0; Z)\} = 0$.
- What if some of y_i are missing ?

Two approaches

1. Prediction model approach: use a model for y. Solve

$$n^{-1}\sum_{i=1}^{n}\left[\delta_i U(\theta; x_i, y_i) + (1-\delta_i)E\{U(\theta; x_i, y_i) \mid x_i, \delta_i = 0\}\right] = 0$$

   for θ. Prediction model approach was discussed in Chapter 2-5.

2. Response model approach: use a model for δ_i (response indicator function). Solve

$$\sum_{i=1}^{n}\frac{\delta_i}{\pi(x_i, y_i)}U(\theta; x_i, y_i) = 0$$

   for θ, where $\pi(x, y) = P(\delta = 1 \mid x, y)$.

Complete-case (CC) method

- Definition: CC estimator of θ_0 is the solution to

$$\sum_{i=1}^{n} \delta_i U(\theta; z_i) = 0. \tag{1}$$

- Theory for CC method: If $Cov\,(\delta_i, U_i) = 0$, where $U_i = U(\theta_0; z_i)$, then the CC estimator from (1) is (approximately) unbiased.

CC method (Cont'd)

- Unless the missing mechanism is missing completely at random (MCAR), the CC method leads to biased estimation.

- Furthermore, the CC method does not make use of the observed information of x_i for $\delta_i = 0$. Thus, it is not efficient.

Weighted Complete-case (WCC) method

- Solve
$$\sum_{i=1}^{n} \delta_i \frac{1}{\pi_i} U(\theta; z_i) = 0 \tag{2}$$
for θ, where $\pi_i = Pr\left(\delta_i = 1 \mid z_i\right)$.

- In survey sampling, π_i are known and the WCC method is very popular (Horvitz-Thompson estimation) since it does not require the model assumptions about unobserved z.

Properties of WCC

- Asymptotically unbiased
- Asymptotic variance: Assuming that $Cov(\delta_i, \delta_j) = 0$ for $i \neq j$,

$$V\left(\hat{\theta}_W\right) \cong \tau^{-1} V\left\{\hat{U}_W(\theta_0)\right\}\tau^{-1'}$$

where $\tau = E\left\{\dot{U}(\theta_0; Z)\right\}$ and

$$\begin{aligned} V\left\{\hat{U}_W(\theta_0)\right\} &= V\left\{\hat{U}_n(\theta_0)\right\} + E\left\{n^{-2}\sum_{i=1}^{n}\left(\pi_i^{-1} - 1\right) U(\theta_0; z_i)^{\otimes 2}\right\} \\ &= n^{-1} E\left\{n^{-1}\sum_{i=1}^{n}\pi_i^{-1} U(\theta_0; z_i)^{\otimes 2} - \bar{U}_n(\theta_0)^{\otimes 2}\right\} \\ &\cong E\left\{n^{-2}\sum_{i=1}^{n}\pi_i^{-1} U(\theta_0; z_i)^{\otimes 2}\right\}. \qquad (3) \end{aligned}$$

Properties of WCC

- A consistent estimator for the variance of $\hat{\theta}_W$ is computed by

$$\hat{V}\left(\hat{\theta}_W\right) = \hat{\tau}^{-1}\hat{V}_u\hat{\tau}^{-1'}$$

where

$$\hat{\tau} = n^{-1}\sum_{i=1}^{n}\delta_i\pi_i^{-1}\dot{U}(\hat{\theta}_W; z_i)$$

and

$$\hat{V}_u = n^{-2}\sum_{i=1}^{n}\delta_i\pi_i^{-2}U(\hat{\theta}_W; z_i)^{\otimes 2}.$$

Example 7.1

- Let the parameter of interest be $\theta_0 = E(Y)$ and we use $U(\theta; z) = (y - \theta)$ to define θ_0. The WCC estimator of θ_0 can be written
$$\hat{\theta}_W = \frac{\sum_{i=1}^{n} \delta_i y_i / \pi_i}{\sum_{i=1}^{n} \delta_i / \pi_i}. \tag{4}$$
- The asymptotic variance of $\hat{\theta}_W$ in (4) is equal to, by (3),
$$n^{-2} \sum_{i=1}^{n} \pi_i^{-1} (y_i - \theta_0)^2$$
which is consistently estimated by
$$n^{-2} \sum_{i=1}^{n} \delta_i \pi_i^{-2} \left(y_i - \hat{\theta}_W \right)^2.$$

Full model vs Reduced model

- We now consider the case when x is decomposed into $x = (x_1, x_2)$ and MAR holds under x_1. That is,

$$P(\delta = 1 \mid x_1, y) = P(\delta = 1 \mid x_1). \tag{5}$$

- Writing $\tilde{\pi}(x_1) = P(\delta = 1 \mid x_1)$, we note that $\tilde{\pi}(x_1)$ is the conditional expectation of $\pi(x) = P(\delta = 1 \mid x)$ given x_1.
- If the parameter of interest is $\theta_0 = E(Y)$ and both $\pi(x)$ and $\tilde{\pi}(x_1)$ are known, there are two choices for deriving the WCC estimator of θ.
  1. The first one is obtained by solving

$$\hat{U}_{W1}(\theta) \equiv \sum_{i=1}^{n} \delta_i \frac{1}{\pi(x_i)}(y_i - \theta) = 0 \tag{6}$$

  2.

$$\hat{U}_{W2}(\theta) \equiv \sum_{i=1}^{n} \delta_i \frac{1}{\tilde{\pi}(x_{1i})}(y_i - \theta) = 0. \tag{7}$$

Then, a question naturally arises: which estimator is better?

Theorem 7.1 (Kim et al, 2019)

Theorem

Let $\hat{\theta}_1$ *and* $\hat{\theta}_2$ *be the solutions to (6) and (7), respectively. Under (5), both* $\hat{\theta}_1$ *and* $\hat{\theta}_2$ *are asymptotically unbiased for* $\theta_0 = E(Y)$ *and*

$$V(\hat{\theta}_1) \geq V(\hat{\theta}_2) \tag{8}$$

asymptotically.

Remark

Using (3), we have

$$V\{\hat{U}_{W1}(\theta_0)\} \cong E\left\{\sum_{i=1}^{n}\{\pi(x_i)\}^{-1}(y_i-\theta_0)^2\right\}$$

and

$$V\{\hat{U}_{W2}(\theta_0)\} \cong E\left\{\sum_{i=1}^{n}\{\tilde{\pi}(x_{1i})\}^{-1}(y_i-\theta_0)^2\right\}.$$

Note that $f(x) = 1/x$ is a convex function at $x \in (0,1)$. Thus, using Jensen's inequality, we can show that

$$E\left\{\frac{1}{\pi(x_i)} \mid x_{1i}\right\} \geq \frac{1}{E\{\pi(x_i) \mid x_{1i}\}}.$$

So, we can expect that $V\{\hat{U}_{W1}(\theta_0)\}$ is larger than $V\{\hat{U}_{W1}(\theta_0)\}$.

Density ratio expression

- Bayes formula

$$\frac{P(\delta = 0 \mid z)}{P(\delta = 1 \mid z)} = \frac{1-p}{p} \times \frac{f(z \mid \delta = 0)}{f(z \mid \delta = 1)}$$

  where $p = P(\delta = 1)$ and $z = (x, y)$.

- Thus, we can express

$$\frac{1}{\pi(z)} = 1 + \left(\frac{1}{p} - 1\right) r(z)$$

  where

$$r(z) = \frac{f(z \mid \delta = 0)}{f(z \mid \delta = 1)}$$

  is a density ratio function.

Reduced model MAR

- Now, assume that $x = (x_1, x_2)$, we can obtain

$$\frac{1}{\tilde{\pi}(x_1, y)} = 1 + \left(\frac{1}{p} - 1\right) \tilde{r}(x_1, y)$$

- Lemma 7.1:

$$\begin{aligned} E\{r(x, y) \mid x_1, y, \delta = 1\} &= \int r(x, y) f(x_2 \mid x_1, y, \delta = 1) dx_2 \\ &= \frac{\int r(x, y) f(x, y \mid \delta = 1) dx_2}{f(x_1, y \mid \delta = 1)} \\ &= \frac{f(x_1, y \mid \delta = 0)}{f(x_1, y \mid \delta = 1)} = \tilde{r}(x_1, y) \end{aligned}$$

Implication

- Two WCC estimators of $\theta = E(Y)$:

$$\hat{\theta}_1 = n^{-1} \sum_{i=1}^{n} \delta_i \{1 + (p^{-1} - 1) r(x_i, y_i)\} y_i.$$

$$\hat{\theta}_2 = n^{-1} \sum_{i=1}^{n} \delta_i \{1 + (p^{-1} - 1) \tilde{r}(x_{1i}, y_i)\} y_i.$$

- By Lemma 7.1, we have

$$E(\hat{\theta}_1 \mid x_1, y, \delta) = \hat{\theta}_2,$$

which implies that

$$E(\hat{\theta}_1) = E(\hat{\theta}_2)$$

and

$$V(\hat{\theta}_1) \geq V(\hat{\theta}_2).$$

Implication (Cont'd)

- Under MAR, we have $\pi(x, y) = \pi(x)$ and

$$\hat{\theta}_1 = n^{-1} \sum_{i=1}^{n} \delta_i \{1 + (p^{-1} - 1) r(x_i)\} y_i .$$

- Under the reduced model MAR condition (5), we have $\tilde{\pi}(x_1, y) = \tilde{\pi}(x_1)$.
- Thus, the density ratio approach gives a proof for Theorem 7.1.

Lemma 7.2: How to achieve (5)?

Lemma

If MAR condition given X holds (i.e. $Y \perp \delta \mid X$) and the reduced model for y holds $f(y \mid x) = f(y \mid x_1)$ for $x = (x_1, x_2)$, then we can obtain MAR given X_1. That is,

$$Y \perp \delta \mid X_1$$

Proof

We have only to prove that

$$f(y \mid x_1, \delta) = f(y \mid x_1).$$

Now, using Bayes formula,

$$\begin{aligned}
f(y \mid x_1, \delta) &= \frac{\int f(y \mid x, \delta) P(\delta \mid x) f(x_2 \mid x_1) f(x_1) dx_2}{\int \int f(y \mid x, \delta) P(\delta \mid x) f(x_2 \mid x_1) f(x_1) dx_2 dy} \\
&= \frac{\int f(y \mid x_1) P(\delta \mid x) f(x_2 \mid x_1) f(x_1) dx_2}{\int \int f(y \mid x_1) P(\delta \mid x) f(x_2 \mid x_1) f(x_1) dx_2 dy} \\
&= \frac{f(y \mid x_1) \int P(\delta \mid x) f(x_2 \mid x_1) dx_2}{\int f(y \mid x_1) \int P(\delta \mid x) f(x_2 \mid x_1) dx_2 dy} \\
&= \frac{f(y \mid x_1) P(\delta \mid x_1)}{\int f(y \mid x_1) P(\delta \mid x_1) dy} = f(y \mid x_1),
\end{aligned}$$

where the second equality follows by MAR assumption and the reduced model assumption.

§5.2 Regression weighting method

Motivation

- $\mathbf{x}_i$: auxiliary variables (observed throughout the sample)
- Assume that $1 = \mathbf{x}_i'\mathbf{a}$ for some $\mathbf{a}$.
- y_i: study variable (observed only when $\delta_i = 1$).
- Regression weighting technique: Use

$$w_i = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\right)' \left(\sum_{i=1}^{n} \delta_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \mathbf{x}_i$$

for the weight associated with unit i with $\delta_i = 1$.

Motivation

- Note that the regression estimator $\hat{\theta}_{reg} = \sum_{i=1}^{n} \delta_i w_i y_i$ of $\theta = E(Y)$ can be written as

$$\hat{\theta}_{reg} = \bar{\mathbf{x}}_n' \hat{\boldsymbol{\beta}}_r \tag{9}$$

  where

$$\hat{\beta}_r = \left(\sum_{i=1}^{n} \delta_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{n} \delta_i \mathbf{x}_i y_i .$$

- Under what conditions, the regression weighting method is justified (in that the resulting estimator is asymptotically unbiased under the response model) ?

Main Result (Fuller et al, 1994)

Assume that auxiliary variables $\mathbf{x}_i$ are observed throughout the sample and the response probability satisfies

$$\frac{1}{\pi_i} = \mathbf{x}_i' \boldsymbol{\lambda} \tag{10}$$

for all unit i in the sample, where $\boldsymbol{\lambda}$ is unknown. We assume that an intercept is included in xi. Under these conditions, the regression estimator defined by (9) is asymptotically unbiased for $\theta = E(Y)$.

Justification

- Because an intercept term is included in $\mathbf{x}_i$, we have

$$\hat{\theta}_n \equiv \bar{y}_n = \bar{\mathbf{x}}_n' \hat{\boldsymbol{\beta}}_n$$

where

$$\hat{\beta}_n = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i .$$

- Note that we can write

$$\hat{\theta}_{reg} - \hat{\theta}_n = \bar{\mathbf{x}}_n' \left(\sum_{i=1}^{n} \delta_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{n} \delta_i \mathbf{x}_i \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_n \right)$$

and so

$$E\left(\hat{\theta}_{reg} - \hat{\theta}_n \mid \mathbf{X}, \mathbf{Y} \right) \cong \bar{\mathbf{x}}_n' \left(\sum_{i=1}^{n} \pi_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{n} \pi_i \mathbf{x}_i \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_n \right)$$

where the expectation is taken with respect to the response mechanism.

Justification (Cont'd)

- Thus, to show that $\hat{\theta}_{reg}$ is asymptotically unbiased, we have only to show that

$$\sum_{i=1}^{n} \pi_i \mathbf{x}_i \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_n \right) = 0 \qquad (11)$$

  holds.

- By (10), we have

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}_n \right) \\ &= \sum_{i=1}^{n} \pi_i \left(\boldsymbol{\lambda}' \mathbf{x}_i \right) \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_n \right), \end{aligned}$$

  which implies that (11) holds.

Variance estimation of the regression estimator

- To discuss variance estimation of the regression estimator where the covariates $\mathbf{x}_i$ satisfy (10), note that

$$
\begin{aligned}
\bar{\mathbf{x}}_n'\hat{\boldsymbol{\beta}}_r &= \bar{\mathbf{x}}_n'\boldsymbol{\beta} + \bar{\mathbf{x}}_n'\left(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}\right) \\
&= \bar{\mathbf{x}}_n'\boldsymbol{\beta} + \bar{\mathbf{x}}_n'\left(\sum_{i=1}^{n}\delta_i\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_{i=1}^{n}\delta_i\mathbf{x}_i\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right) \\
&\cong \bar{\mathbf{x}}_n'\boldsymbol{\beta} + \bar{\mathbf{x}}_n'\left(\sum_{i=1}^{n}\pi_i\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_{i=1}^{n}\delta_i\mathbf{x}_i\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)
\end{aligned}
$$

  where $\boldsymbol{\beta}$ is the probability limit of $\hat{\boldsymbol{\beta}}_r$
- By the fact that 1 is included in $\mathbf{x}_i$ and by (10), it can be shown that

$$
\bar{\mathbf{x}}_n'\left(\sum_{i=1}^{n}\pi_i\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_{i=1}^{n}\delta_i\mathbf{x}_i\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right) \tag{12}
$$

  by some matrix algebra.

Variance estimation of the regression estimator

- Approximate variance

$$V\left(\hat{\theta}_{reg}\right) \cong V\left(\frac{1}{n}\sum_{i=1}^{n} d_i\right) \tag{13}$$

where $d_i = \mathbf{x}_i'\boldsymbol{\beta} + \delta_i \pi_i^{-1}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)$.

- Variance estimation can be implemented by using a standard variance estimation formula applied to $\hat{d}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}_r + \delta_i n w_i (y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_r)$. That is,

$$\hat{V}\left(\hat{\theta}_{reg}\right) = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{d}_i - \bar{\hat{d}}_n\right)^2$$

where $\bar{\hat{d}}_n = \sum_{i=1}^{n} \hat{d}_i / n$.