Efficient estimation under non-monotone missingness by design

August 15, 2023

1 Introduction

2 Main Problem

3 Discussion

Projection: High level concept

- Let $\hat{\theta}_0$ be an unbiased estimator of θ.
- Let $\Lambda = \left\{ \hat{b}; E(\hat{b}) = 0 \right\}$ be a space of all unbiased estimators of zero.
- We consider the following class of unbiased estimators of θ:

$$\hat{\theta}_b = \hat{\theta}_0 - \hat{b} \tag{1}$$

  where $\hat{b} \in \Lambda$.
- The optimal estimator among the class in (1) is

$$\hat{\theta}_{\rm opt} = \hat{\theta}_0 - \hat{b}^*$$

  where $\hat{b}^*$ satisfies
  1. $\hat{b}^* \in \Lambda$
  2. $Cov(\hat{\theta}_0 - \hat{b}^*, \hat{b}) = 0$ for all $\hat{b} \in \Lambda$.
- The $\hat{b}^*$ satisfying the above two conditions is often called the <u>projection</u> of $\hat{\theta}_0$ onto Λ and is denoted as $\hat{b}^* = \Pi(\hat{\theta}_0 \mid \Lambda)$.

Justification

- We wish to prove

$$V\left(\hat{\theta}_0 - \hat{b}\right) \geq V\left(\hat{\theta}_0 - \hat{b}^*\right). \tag{2}$$

- Note that

$$\begin{aligned} V\left(\hat{\theta}_0 - \hat{b}\right) &= V\left(\hat{\theta}_0 - \hat{b}^*\right) + V\left(\hat{b} - \hat{b}^*\right) \\ &+ 2Cov\left(\hat{\theta}_0 - \hat{b}^*, \hat{b} - \hat{b}^*\right) \end{aligned}$$

  and the covariance term is zero by the definition of $\hat{b}^*$.

- Thus, (2) is proved.

Basic Setup

- $U = \{1, \ldots, N\}$: finite population
- Let $(x_i, y_i, \pi_i) \sim F$ for some F (unknown) for $i \in U$.
- Generate $I_i \sim \pi_i$ independently for $i \in U$
- Observe (π_i, y_i) only when $I_i = 1$
- x_i observed throughout the finite population.
- We are interested in estimating $\theta = E(Y)$, for example, from the sample.
- The Horvitz-Thompson estimator $\hat{\theta}_{\mathrm{HT}} = N^{-1} \sum_{i=1}^{N} I_i w_i y_i$ is design unbiased for θ, where $w_i = \pi_i^{-1}$, but it is not necessarily efficient.

Class of estimators

- We consider the following class of estimators

$$\hat{\theta}_b = \hat{\theta}_{\mathrm{HT}} - \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{I_i}{\pi_i} - 1 \right\} b(x_i) \tag{3}$$

  where $b \in \mathcal{L}^2 = \{b(x); \int b(x)^2 dF(x) < \infty\}$.
- Note that $\hat{\theta}_b$ is unbiased for θ for all $b \in \mathcal{L}^2$. Thus, $\{\hat{\theta}_b; b \in \mathcal{L}^2\}$ is a class of unbiased estimators of θ.
- Goal: Find the optimal $b^* \in \mathcal{L}^2$ among the class in (3).

Theorem

- The variance of $\hat{\theta}_b$ is

$$V\left(\hat{\theta}_b\right) = V\left(N^{-1}\sum_{i=1}^{N} y_i\right) + E\left\{\frac{1}{N^2}\sum_{i=1}^{N}(w_i - 1)\left(y_i - b(x_i)\right)^2\right\}$$

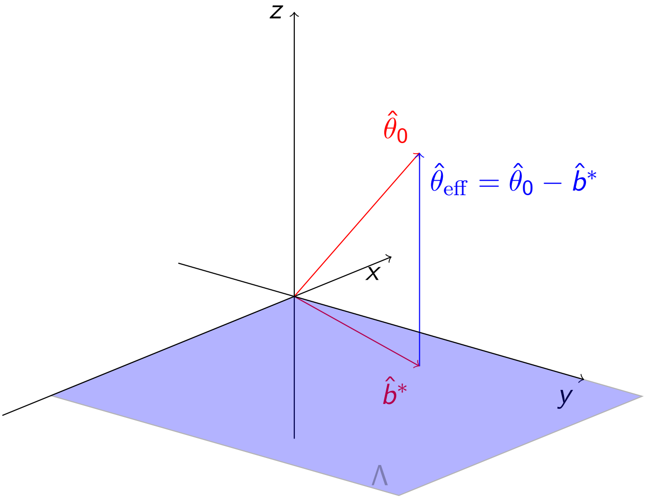
where $w_i = \pi_i^{-1}$.

- Result: The optimal b^* minimizing the variance of $\hat{\theta}_b$ is

$$b^*(x) = \frac{E\{(W-1)Y \mid x)\}}{E\{(W-1) \mid x\}}. \tag{4}$$

- $\hat{\theta}_{b^*}$ is the projection of $\hat{\theta}_{\mathrm{HT}}$ onto $\Lambda = \left\{N^{-1}\sum_{i=1}^{N}\left(\pi_i^{-1}I_i - 1\right)b(x_i)\right\}$.

Graphical illustration in $\mathbb{R}^3$: $b^* = \Pi\left(\hat{\theta}_0 \mid \Lambda\right)$



Check

- We can check
$$
Cov\left\{\hat{\theta}_{\mathrm{HT}} - N^{-1}\sum_{i=1}^{N}\left(\pi_i^{-1}I_i - 1\right)b^*(x_i), N^{-1}\sum_{i=1}^{N}\left(\pi_i^{-1}I_i - 1\right)b(x_i)\right\}
$$
is zero for all $b(x)$.
- The covariance term can be written as
$$
\begin{aligned}
(\mathrm{Cov}) &= N^{-2}E\left\{\sum_{i=1}^{N}(w_i-1)(y_i-b^*(x_i))b(x_i)\right\} \\
&= N^{-2}E\left[\sum_{i=1}^{N}E\left\{(w_i-1)y_i \mid x_i\right\}b(x_i)\right] \\
&\quad -N^{-2}E\left[E\left\{(w_i-1) \mid x_i\right\}b^*(x_i)b(x_i)\right] \\
&= 0
\end{aligned}
$$
for $b^*(x)$ in (4).

Remark

- If the sampling design is non-informative, then we have

$$E(W_i Y_i \mid x_i) = E(W_i \mid x_i) E(Y_i \mid x_i)$$

  and

$$b^*(x) = \frac{E\{(W-1)Y \mid x)\}}{E\{(W-1) \mid x\}} = E(Y \mid x)$$

- The optimality of regression estimator is based on the assumption that the sampling mechanism is non-informative.

1 Introduction

2 Main Problem

3 Discussion

Motivating Toy Example (3 items)

Table: **Case 1: Monotone missingness**

Sample	X	Y_1	Y_2	R_1	R_2
A_{11}	✓	✓	✓	1	1
A_{10}	✓	✓		1	0
A_{00}	✓			0	0

Table: **Case 2: Non-monotone missingness**

Sample	X	Y_1	Y_2	R_1	R_2
A_{11}	✓	✓	✓	1	1
A_{10}	✓	✓		1	0
A_{01}	✓		✓	0	1
A_{00}	✓			0	0

Note: R_t is the second-phase sampling indicator function of Y_t.

Basic setup

- $A = A_{11} \cup A_{10} \cup A_{01} \cup A_{00}$: index set of the original sample (representative of the population). We can view A as the index set of the first-phase sample.
- For simplicity, we will assume that A is selected by SRS.
- Define $L = (X, Y_1, Y_2)$ and

$$\pi_{ab}(L) = P\left(R_1 = a, R_2 = b \mid L\right).$$

- We assume for now that $\pi_{ab}(L)$ is known.
- We are interested in estimating $\theta = E\{g(L)\}$ from the observed data.
- A direct HT estimator of θ is

$$\hat{\theta}_{\mathrm{HT}} = \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}R_{2i}}{\pi_{11}(L_i)} g(L_i).$$

Case 1: Efficient estimation under monotone missingness

- Under monotone missingness, we have three-phase sampling structure:
  1. Baseline sample: $A_{11} \cup A_{10} \cup A_{00}$, observe x
  2. $R_1 = 1$ sample: $A_{11} \cup A_{10}$, observe (x, y_1)
  3. $R_2 = 1$ sample: A_{11}, observe (x, y_1, y_2)
- Conditional inclusion probability is non-informative sampling
  1. Baseline sample: $\pi_i = n/N$
  2. $R_1 = 1$ sample: $P(R_1 = 1 \mid X) := \pi_{1+}(X)$
  3. $R_2 = 1$ sample: $P(R_2 = 1 \mid R_1 = 1, X, Y_1) := \pi_{1|1}(X, Y_1)$

Class of estimators

- Consider

$$
\begin{aligned}
\hat{\theta}_b &= \frac{1}{n}\sum_{i=1}^{n}\frac{R_{1i}R_{2i}}{\pi_{11}(x_i, y_{1i})}g(x_i, y_{1i}, y_{2i}) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_{1i}}{\pi_{1+}(x_i)} - 1\right)b_1(x_i) \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{R_{2i}}{\pi_{1|1}(x_i, y_{1i})} - 1\right)\frac{R_{1i}}{\pi_{1+}(x_i)}b_2(x_i, y_{1i})\right\}
\end{aligned}
$$

  where $g_i = g(x_i, y_{1i}, y_{2i})$.
- Note that $\hat{\theta}_b$ is unbiased for $\theta = E\{g(X, Y_1, Y_2)\}$ regardless of the choice of b_1 and b_2.

Main Result (under monotone missingness)

- Using the projection result, the efficient estimator of $\theta = E\{g(X, Y_1, Y_2)\}$ is

$$\begin{aligned}\hat{\theta}_{\text{eff}} &= \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}R_{2i}}{\pi_{11}(X_i, Y_{1i})} g(X_i, Y_{1i}, Y_{2i}) \\ &- \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{1i}}{\pi_{1+}(X_i)} - 1 \right) b_1^*(X_i) \\ &- \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R_{1i}R_{2i}}{\pi_{11}(X_i, Y_{1i})} - \frac{R_{1i}}{\pi_{1+}(X_i)} \right) b_2^*(X_i, Y_{1i}),\end{aligned} \tag{5}$$

  where

$$\begin{aligned} b_1^*(X) &= E\{g(X, Y_1, Y_2) \mid X\} \\ b_2^*(X, Y_1) &= E\{g(X, Y_1, Y_2) \mid X, Y_1\}. \end{aligned}$$

- We need a working outcome model to compute $b_1^*(x)$ and $b_2^*(x, y_1)$.

Remark

- We can express

$$
\begin{aligned}
\hat{\theta}_{\text{eff}} &= \frac{1}{n}\sum_{i=1}^{n} b_1^*(X_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}}{\pi_{1+}(X_i)} \left\{ b_2^*(X_i, Y_{1i}) - b_1^*(X_i) \right\} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}R_{2i}}{\pi_{11}(X_i, Y_{1i})} \left\{ g(X_i, Y_{1i}, Y_{2i}) - b_2^*(X_i, Y_{1i}) \right\},
\end{aligned}
\tag{6}
$$

where

$$
\begin{aligned}
b_1^*(X) &= E\{g(X, Y_1, Y_2) \mid X\} \\
b_2^*(X, Y_1) &= E\{g(X, Y_1, Y_2) \mid X, Y_1\}.
\end{aligned}
$$

- This is the usual three-phase (sampling) regression estimator.

Case 2: Non-monotone missingness

- Idea: Let's consider a regression-type estimator in (6) in the context of nonmonotone missingness.
- May use

$$
\begin{aligned}
\hat{\theta}_{\mathrm{a,b}} &= \frac{1}{n}\sum_{i=1}^{n} E(g_i \mid X_i) \qquad (7)\\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}}{\pi_{1+}(X_i)} \left\{ b_2(X_i, Y_{1i}) - E(g_i \mid X_i) \right\}\\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{2i}}{\pi_{2+}(X_i)} \left\{ a_2(X_i, Y_{2i}) - E(g_i \mid X_i) \right\}\\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}R_{2i}}{\pi_{11}(X_i)} \left\{ g_i - b_2(X_i, Y_{1i}) - a_2(X_i, Y_{2i}) + E(g_i \mid X_i) \right\}.
\end{aligned}
$$

- Note that (7) is design-unbiased as long as the selection probabilities are correct. (We may assume that the sampling mechanism depends only on X.)
- Best choice of $a_2(\cdot)$ and $b_2(\cdot)$ will improve the efficiency.
- One choice is

$$a_2^*(X, Y_2) = E\left(g \mid X, Y_2\right) \tag{8}$$

and

$$b_2^*(X, Y_1) = E\left(g \mid X, Y_1\right) \tag{9}$$

under the outcome regression model.

1 Introduction

2 Main Problem

3 Discussion

Research Questions (to be solved by Caleb)

1. Does the choice in (8) and (9) minimize the variance among the class of unbiased estimators $\hat{\theta}_{a,b}$ in (7)?
    - If yes, prove it.
    - If no, find out the optimal choice.
2. Once the optimal estimator is found in the class in (7), can we construct mass imputation to implement the optimal estimation?
3. How can we generalize the result to multiple times ($T > 2$) and apply it to NRI?