Efficient estimation under non-monotone missingness

Jae-kwang Kim

March 29, 2023

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 response indicator function of Y_t.

Basic setup

- 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 (can be relaxed later).
- We are interested in estimating $\theta = E\{g(L)\}$ from the observed data.
- A direct (Inverse Propensity Weight) estimator of θ is

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

Notation

- Let $G_R(L)$ be the observed part of L based on the realization of R.
- That is,

$$G_R(L) = \begin{cases} (X, Y_1, Y_2) & \text{if } R_1 = R_2 = 1 \\ (X, Y_1) & \text{if } R_1 = 1, R_2 = 0 \\ (X, Y_2) & \text{if } R_1 = 0, R_2 = 1 \\ X & \text{if } R_1 = 0, R_2 = 0 \end{cases}$$

- Thus, any estimator is a function of observations $O_i = (R_i, G_{R_i}(L_i))$ for $i = 1, \cdots, n$.

Augmentation space

- Define
$$\Lambda_R = \{h(O); E_R\{h(O) \mid L\} = 0\}$$
be a space of functions of $O = (R, G_R(L))$ whose expectation (wrt the response mechanism) is equal to zero.
- Tsiatis (2006) called Λ_R as the augmentation space.
- For any element $\ell \in \Lambda_R$,
$$\hat{\theta}_\ell = \hat{\theta}_{\mathrm{IPW}} - \frac{1}{n}\sum_{i=1}^{n} \ell(O_i) \tag{1}$$
is unbiased for θ. Thus, $\{\hat{\theta}_\ell; \ell \in \Lambda_R\}$ is a class of unbiased estimators of θ.

Efficient estimation using projection

- Goal: Find the optimal $\ell^* \in \Lambda_R$ among the class in (1).
- Idea: Let $\varphi_0(O)$ be an initial function satisfying $E\{\varphi_0(O)\} = \theta$. If we express the class

$$\hat{\theta}_\ell = \frac{1}{n} \sum_{i=1}^{n} \{\varphi_0(O_i) - \ell(O_i)\}$$

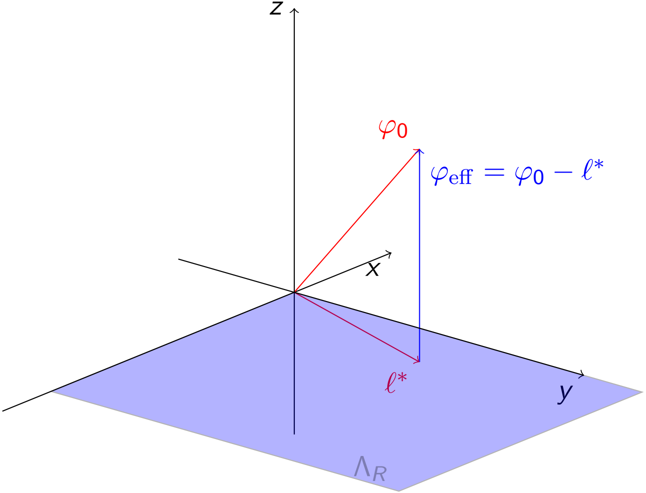
for $\ell \in \Lambda_R$, the optimal choice of ℓ is

$$\ell^* = \Pi\left(\varphi_0 \mid \Lambda_R\right)$$

where the projection ℓ^* satisfies

1. $\ell^* \in \Lambda_R$
2. $\langle \varphi_0 - \ell^*, \ell \rangle = 0$ for all $\ell \in \Lambda_R$.

Graphical illustration in $\mathbb{R}^3$: $\ell^* = \Pi(\varphi_0 \mid \Lambda_R)$



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
  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: $P(R_2 = 1 \mid R_1 = 1, X, Y_1) := \pi_{1|1}(X, Y_1)$

Characterizing the augmentation space under Case 1

- Recall

$$\Lambda_R = \{h(O); E\{h(O) \mid L\} = 0\}$$

and we have three levels of (R_1, R_2).

- We can express

$$\Lambda_R = \Lambda_{R1} \oplus \Lambda_{R2},$$

where

$$\Lambda_{R1} = \left\{ \left(\frac{R_1}{\pi_{1+}(x)} - 1 \right) b_1(x); \forall b_1(x) \right\}$$

and

$$\Lambda_{R2} = \left\{ \left(\frac{R_1 R_2}{\pi_{11}(x, y_1)} - \frac{R_1}{\pi_{1+}(x)} \right) b_2(x, y_1); \forall b_2(x, y_1) \right\}.$$

Main Result (under monotone missingness)

- For
$$\varphi_0(O) = \frac{R_1 R_2 g(X, Y_1, Y_2)}{\pi_{11}(X, Y_1)},$$
we obtain
$$\Pi\left(\varphi_0 \mid \Lambda_{R1}\right) = \left(\frac{R_1}{\pi_{1+}(X)} - 1\right) b_1^*(X)$$
and
$$\Pi\left(\varphi_0 \mid \Lambda_{R2}\right) = \left(\frac{R_1 R_2}{\pi_{11}(X, Y_1)} - \frac{R_1}{\pi_{1+}(X)}\right) b_2^*(X, Y_1),$$
where
$$\begin{aligned} b_1^*(X) &= E\{g(X, Y_1, Y_2) \mid X\} \\ b_2^*(X, Y_1) &= E\{g(X, Y_1, Y_2) \mid X, Y_1\}. \end{aligned}$$

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

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

  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 $b_2^*(x, y_1)$.

Justification

- Any element $\hat{\theta}_\ell$ in (1) can be expressed as

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

  where $g_i = g(x_i, y_{1i}, y_{2i})$.

- We can express

$$
\hat{\theta}_\ell = E_1\{E_2(\hat{\theta}_\ell)\} + \left[E_2(\hat{\theta}_\ell) - E_1\{E_2(\hat{\theta}_\ell)\}\right] + \left[\hat{\theta}_\ell - E_2(\hat{\theta}_\ell)\right] \qquad (3)
$$

  where $E_1(\cdot)$ and $E_2(\cdot)$ are the expectations with respect to the sampling mechanism associated with R_1 and R_2, respectively.

- Under the sampling mechanism wrt R_2, the third term of $\hat{\theta}_\ell$ has zero expectation. Thus, we obtain

$$E_2(\hat{\theta}_\ell) = n^{-1}\sum_{i=1}^{n}\left[b_1(x_i) + \frac{R_{1i}}{\pi_{1+}(x_i)}\{g(x_i, y_{1i}, y_{2i}) - b_1(x_i)\}\right]$$

- Also, under the sampling mechanism wrt R_1, we obtain

$$E_1\{E_2(\hat{\theta}_\ell)\} = n^{-1}\sum_{i=1}^{n} g(x_i, y_{1i}, y_{2i})$$

- Thus, we can express (3) as

$$\begin{aligned}\hat{\theta}_\ell &= \frac{1}{n}\sum_{i=1}^{n} g(x_i, y_{1i}, y_{2i}) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_{1i}}{\pi_{1+}(X_i)} - 1\right)\{g(x_i, y_{1i}, y_{2i}) - b_1(X_i)\} \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_{2i}}{\pi_{1|1}(x_i, y_{1i})} - 1\right)\frac{R_{1i}}{\pi_{1+}(x_i)}\{g(x_i, y_{1i}, y_{2i}) - b_2(x_i, y_{1i})\}.\end{aligned} \tag{4}$$

- The three terms in (3) are orthogonal:

$$
\begin{aligned}
& V\left(\hat{\theta}_\ell\right) \\
= \quad & V\left[E_1\{E_2(\hat{\theta}_\ell)\}\right] + V\left[E_2(\hat{\theta}_\ell) - E_1\{E_2(\hat{\theta}_\ell)\}\right] + V\left[\hat{\theta}_\ell - E_2(\hat{\theta}_\ell)\right] \\
:= \quad & V_1 + V_2 + V_3,
\end{aligned}
$$

where

$$
\begin{aligned}
V_1 &= n^{-1} V\{g(X, Y_1, Y_2)\} \\
V_2 &= n^{-1} E\left[\left(\frac{1}{\pi_{1+}(X)} - 1\right)\{g(X, Y_1, Y_2) - b_1(X)\}^2\right] \\
V_3 &= n^{-1} E\left[\frac{1 - \pi_{1|1}(X, Y_1)}{\pi_{11}(X, Y_1)}\{g(X, Y_1, Y_2) - b_2(X, Y_1)\}^2\right].
\end{aligned}
$$

- The variance is minimized at

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

Alternative expression 1

- We can express

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

where

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

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

Alternative expression 2

Writing $R_i = (R_{1i}, R_{2i})$, we can express

$$
\begin{aligned}
\hat{\theta}_{\text{eff}} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}\left(R_i = (1,1)\right)}{\pi_{11}(X_i, Y_{1i})} g(X_i, Y_{1i}, Y_{2i}) \qquad (6) \\
&+ \frac{1}{n}\sum_{i=1}^{n} \left\{ \mathbb{I}\left(R_i = (1,0)\right) - \pi_{10}(X_i, Y_{1i}) \frac{\mathbb{I}\left(R_i = (1,1)\right)}{\pi_{11}(X_i, Y_{1i})} \right\} B_{10}^{*}(X_i, Y_{1i}) \\
&+ \frac{1}{n}\sum_{i=1}^{n} \left\{ \mathbb{I}\left(R_i = (0,0)\right) - \pi_{00}(X_i, Y_{1i}) \frac{\mathbb{I}\left(R_i = (1,1)\right)}{\pi_{11}(X_i, Y_{1i})} \right\} B_{00}^{*}(X_i).
\end{aligned}
$$

How to find B_{10}^{*} and B_{00}^{*}?

Idea

- Apply the following equalities

$$R_{1i}R_{2i} + R_{1i}(1 - R_{2i}) + (1 - R_{1i})(1 - R_{2i}) = 1$$

and

$$R_{1i} = R_{1i}R_{2i} + R_{1i}(1 - R_{2i})$$

to the three-phase regression estimator in (5).

- That is, we can express (5) as

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

- We can rearrange the above term into three groups of the sample

$$
\begin{aligned}
& \hat{\theta}_{\mathrm{eff}} \\
= & \frac{1}{n}\sum_{i=1}^{n} R_{1i}R_{2i}\left[b_1^*(X_i) + \frac{b_2^*(X_i, Y_{1i}) - b_1^*(X_i)}{\pi_{1+}(X_i)} + \frac{g_i - b_2^*(X_i, Y_{1i})}{\pi_{11}(X_i, Y_{1i})}\right] \\
+ & \frac{1}{n}\sum_{i=1}^{n} R_{1i}(1 - R_{2i})\left\{b_1^*(X_i) + \frac{b_2^*(X_i, Y_{1i}) - b_1^*(X_i)}{\pi_{1+}(X_i)}\right\} \\
+ & \frac{1}{n}\sum_{i=1}^{n}(1 - R_{1i})(1 - R_{2i})b_1^*(X_i).
\end{aligned}
$$

- After some algebra, we obtain the expression in (6) with

$$B_{00}^*(X) = b_1^*(X) = E(g \mid X)$$

and

$$
\begin{aligned}
B_{10}^*(X, Y_1) &= b_1^*(X) + \frac{1}{\pi_{1+}(X)}\left\{b_2^*(X, Y_1) - b_1^*(X)\right\} \\
&= E(g \mid X) + \frac{1}{\pi_{1+}(X)}\left\{E(g \mid X, Y_1) - E(g \mid X)\right\}.
\end{aligned}
$$

- Now, we can express the solution $B_{00}^*(X)$ and $B_{10}^*(X, Y_1)$ as the conditional expectation of $D^*(X, Y_1, Y_2)$ given the observed variables:

$$
\begin{aligned}
B_{00}^*(X) &= E\{D^*(X, Y_1, Y_2) \mid X\} \qquad (7)\\
B_{10}^*(X, Y_1) &= E\{D^*(X, Y_1, Y_2) \mid X, Y_1\}, \qquad (8)
\end{aligned}
$$

where

$$
\begin{aligned}
D^*(X, Y_1, Y_2) &= E(g \mid X) + \frac{1}{\pi_{1+}(X)}\{E(g \mid X, Y_1) - E(g \mid X)\} \\
&+ \frac{1}{\pi_{11}(X, Y_1)}\{g(X, Y_1, Y_2) - E(g \mid X, Y_1)\}. \qquad (9)
\end{aligned}
$$

- That is, we can check (7) and (8).

- Thus, we can express

$$
\begin{aligned}
\hat{\theta}_{\text{eff}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}\left(R_i=(1,1)\right)}{\pi_{11}(L_i)}g(L_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\sum_{r\in\mathcal{R}/\{1\}}\left\{\mathbb{I}\left(R_i=r\right)-\pi_r(G_r(L_i))\frac{\mathbb{I}\left(R_i=1\right)}{\pi_{11}(L_i)}\right\}B_r^*(G_r(L_i))
\end{aligned} \tag{10}
$$

with

$$
B_r^*(G_r(L)) = E\left\{D^*(L)\mid G_r(L)\right\},
$$

where $\mathcal{R}$ is the support of R, $D^*(L)$ is defined in (9) and $\pi_r(G_r(L)) = P(R=r\mid G_r(L))$.

- The second term is the negative projection of $\hat{\theta}_{\text{IPW}}$ onto the augmentation space Λ_R.
- Thus, it is still the projection of $\hat{\theta}_{\text{IPW}}$ onto $\Lambda_R^{\perp}$, but the augmentation space Λ_R is not necessarily orthogonalized.

Remark

- Three different expressions of $\hat{\theta}_{\text{eff}}$:
  1. The original projection form in (2), which express $\hat{\theta}_{\text{eff}}$ as the projection of $\hat{\theta}_{\text{IPW}}$ onto the orthogonal complement of the augmentation space: $\Lambda_R = \Lambda_{R1} \oplus \Lambda_{R2}$.
  2. The three-phase regression estimation form in (6).
  3. The alternative projection form in (10).
- The third expression is applicable to non-monotone missing patterns.

Case 2: Non-monotone missingness

- In the non-monotone missingness, the augmentation space Λ_R is no longer orthogonal. Thus, we cannot apply the original projection in (2) or (6) based on the orthogonal decomposition of Λ_R.
- We can still apply the alternative projection formula in (10), but how to find $D^*(L)$ in (9) is not clear.
- Robins et al (1994) interpreted $D^*(L)$ in (9) as $\mathcal{M}^{-1}(g)$ where $\mathcal{M}$ is an operator defined by

$$\mathcal{M}(h) = \sum_{r \in \mathcal{R}} \pi(r, G_r(L)) E\left\{h(L) \mid G_r(L)\right\},$$

where $\pi_r(G_r(L)) = P(R = r \mid G_r(L))$.

- To apply the alternative projection formula in (10), conceptually, we have only to find $\mathcal{M}^{-1}(g)$.
- According to Robins et al (1994), $D(L) = \mathcal{M}^{-1}\{g(L)\}$ can be obtained by an iterative formula:

$$D_{t+1}(L) = (I - \mathcal{M})\, D_t(L) + g(L) \tag{11}$$

  with $D_0(L) = g(L)$.
- (Personal opinion) This approach does not seem to be promising in practice. To compute $\mathcal{M}(\cdot)$, we need two models, the response model and the outcome regression model. If we apply the iterative formula in (11), the estimation errors in the model parameters will accumulate.

Alternative approach (Idea)

- Let's consider the second expression (5) in the context of nonmonotone missingness.
- May use

$$
\begin{aligned}
\hat{\theta}_{\text{eff}} &= \frac{1}{n}\sum_{i=1}^{n} E(g_i \mid X_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}}{\pi_{1+}(X_i)} \left\{ b_2(X_i, Y_{1i}) - E(g_i \mid X_i) \right\} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{2i}}{\pi_{2+}(X_i)} \left\{ a_2(X_i, Y_{2i}) - E(g_i \mid X_i) \right\} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{R_{1i}R_{2i}}{\pi_{11}(X_i, Y_{1i})} \left\{ g_i - b_2(X_i, Y_{1i}) - a_2(X_i, Y_{2i}) + E(g_i \mid X_i) \right\}.
\end{aligned} \tag{12}
$$

- Note that (12) is design-unbiased as long as the selection probabilities are correct.
- Best choice of $a_2(\cdot)$ and $b_2(\cdot)$ will improve the efficiency.
- One choice is

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

and

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

under the outcome regression model.