Tsiatis (2006) Chapters 10–11

Asymptotic Statistics 2025 Summer Reading Group

Naoki Eguchi

August 18, 2025

Faculty of Medicine, Kyoto University

Today's agenda

- Recap: Coarsened data, IF geometry, AIPWCC in one slide
- Chapter 10: Two-level missingness $\Rightarrow$ efficient AIPWCC & DR
- *Skip (brief mention only):* monotone coarsening details, censoring
- Generalization idea (operators $L,\ M,\ M^{-1}$) for later use
- Chapter 11: Locally efficient estimators (two representations; how to compute)
- Practical recipe: what to implement in 2h

Recap

Coarsened data and Semiparametric statistics

- Full data Z; observed data $\{C,\ G_C(Z)\}$ under CAR (coarsening at random).
  $\Rightarrow\ \forall r,\ P(C = r \mid Z) = \pi\{r,\ G_r(Z)\}$ (coarsening depends only on observed data.)
- When $C = \infty$, the data are completely observed (equal to full data).
- In semiparametric statistics, we study the influence function, an element of $\Lambda^{\perp}$.
  - Full-data tangent space: $\mathcal{H}^F$; full-data nuisance tangent space: $\Lambda_F \subset \mathcal{H}^F$.
    - (Observed) tangent space: $\mathcal{H}$; nuisance tangent space: $\Lambda = \Lambda_\psi \oplus \Lambda_\eta$ $(\Lambda_\psi \perp \Lambda_\eta)$.
- **Theorem 8.3**: all observed-data influence functions can be written as

$$\varphi\{C, G_C(Z)\} = \left[\frac{\mathbf{1}(C=\infty)}{\pi(\infty, Z, \psi_0)}\varphi^F(Z) + L_2\{C, G_C(Z)\}\right] - \Pi\big([\cdot] \mid \Lambda_\psi\big)$$

where $\varphi^F(Z)$: full-data IF, $L_2\{C, G_C(Z)\} \in \Lambda_2$ (augmentation space)

10. Improving Efficiency & Double Robustness with Coarsened Data

The optimal (variance-minimizing) observed-data IF

Theorem 10.1

- For fixed full-data IF $\varphi_F(Z) \in (\mathrm{IF})_F$, the optimal choice is

$$L_2\{C, G_C(Z)\} = -\Pi\left[\frac{\mathbf{1}(C=\infty)\,\varphi_F(Z)}{\pi(\infty, Z, \psi_0)}\,\middle|\,\Lambda_2\right].$$

- Hence the optimal observed-data influence function is given by

$$\frac{\mathbf{1}(C=\infty)\,\varphi_F(Z)}{\pi(\infty, Z, \psi_0)} - \Pi\left[\frac{\mathbf{1}(C=\infty)\,\varphi_F(Z)}{\pi(\infty, Z, \psi_0)}\,\middle|\,\Lambda_2\right].$$

- *Remark 1* : If we know ψ_0, we can choose optimal IF.
  Also, we do not have to estimate ψ because $\Lambda_\psi \subset \Lambda_2$ (already subtracted !)

Generalized by linear operator

Definition 1

- Define the linear operator $\mathcal{J} : \mathcal{H}^F \to \mathcal{H}$

$$h_F \mapsto \mathcal{J}(h_F) = \frac{\mathbf{1}(C=\infty)}{\pi(\infty, Z, \psi_0)} h_F(Z) - \Pi\left[\frac{\mathbf{1}(C=\infty)}{\pi(\infty, Z, \psi_0)} h_F(Z) \,\middle|\, \Lambda_2 \right].$$

- This operator turns full-data IF into efficient observed-data IF, so we can create efficient observed-data IF space.

Definition 2

- The class of *double-robust* observed-data influence functions is defined by

$$(\mathrm{IF})_{\mathrm{DR}} = \left\{ \mathcal{J}(\varphi_F) : \varphi_F(Z) \in (\mathrm{IF})_F \right\}.$$

DR linear space

- $(\mathrm{IF})^F = \varphi^F(Z) + \mathcal{T}^{F\perp}$ is a linear variety in $\mathcal{H}^F$.
- Since $\mathcal{J}$ is linear operator, $(\mathrm{IF})_{\mathrm{DR}} = \mathcal{J}\{(\mathrm{IF})^F\} = \mathcal{J}(\varphi^F) + \mathcal{J}(\mathcal{T}^{F\perp})$ in $\mathcal{H}$.

Definition 3

- The linear subspace
$$\mathcal{J}(\Lambda^{F\perp}) \;\subset\; \Lambda^{\perp} \;\subset\; \mathcal{H},$$
is called the **DR linear space**.
- Explicitly,
$$\mathcal{J}(\Lambda^{F\perp}) = \left\{ \mathcal{J}(\varphi^F) : \varphi^F(Z) \in \Lambda^{F\perp} \right\},$$

10.2 Improving Efficiency with Two Levels of Missingness

Set-up: Two-level missingness

- Partition $Z = (Z_1^\top, Z_2^\top)^\top$ with Z_1 always observed and Z_2 sometimes missing.
- Complete-case indicator $R \in \{0, 1\}$ with MAR:
  $P(R = 1 \mid Z) = P(R = 1 \mid Z_1) = \pi(Z_1, \psi)$ (e.g., logit model).
- Observed data: $O = (R, Z_1, RZ_2)$.
- In this case, $\dfrac{I(C = \infty)\, \varphi^{*F}(Z)}{\varpi(\infty, Z)} = \dfrac{R\varphi^{*F}(Z)}{\pi(Z_1)}$ for $\varphi^{*F}(Z) = m(Z, \beta_0) \in \Lambda^{F\perp}$, hence, we want to find its projection onto Λ_2.
- Also, $\Lambda_2 = \{\dfrac{R - \pi(Z_1)}{\pi(Z_1)}\, h_2(Z_1, \beta) : \forall h_2 : Z_1 \to \mathbb{R}^q\}$

Finding the Projection onto the Augmentation space

Theorem 10.2

- The projection of

$$\frac{R\varphi^{*F}(Z)}{\pi(Z_1)}$$

onto the augmentation space Λ_2 is the unique element

$$\left(\frac{R-\pi(Z_1)}{\pi(Z_1)}\right)h_2^0(Z_1) \; \in \; \Lambda_2,$$

where

$$h_2^0(Z_1) = \mathbb{E}\left[\varphi^{*F}(Z) \mid Z_1\right].$$

- For implementaion, we just compute $\mathbb{E}\left[\varphi^{*F}(Z) \mid Z_1\right]$.

Adaptive estimation and its algorithm

- Consider the case where we can choose an estimating function $m(Z, \beta)$.
- **Optimal** augmentation: $h_2^\star(Z_1, \beta) = \mathbb{E}[m(Z, \beta) \mid Z_1]$.
- In practice, posit a (possibly misspecified) model $p^*_{Z_2|Z_1}(\cdot \mid z_1; \xi)$ and compute

$$h_2^*(Z_1, \beta, \xi) = \int m(z_1, u, \beta)\, p^*_{Z_2|Z_1}(u \mid z_1; \xi)\, du.$$

- Estimation: fit $\hat{\psi}$ by MLE for $\pi(\cdot, \psi)$; fit $\hat{\xi}^*$ by likelihood for the posited model; solve

$$\sum_{i=1}^{n} \left[\frac{R_i}{\pi(Z_{1i}, \hat{\psi}_n)} m(Z_i, \beta) - \frac{R_i - \pi(Z_{1i}, \hat{\psi}_n)}{\pi(Z_{1i}, \hat{\psi}_n)} h_2^*(Z_{1i}, \beta, \hat{\xi}_n^*) \right] = 0.$$

- This estimation strategy is relatively easy to implement.

AIPWCC estimation which is igronrant of ξ

- While, using posit model, we can think of the limiting value $h_2^*(Z_{1i}, \beta, \xi^*)$.
- Even if $h_2^*(Z_{1i}, \beta, \xi^*) \neq h_2^0(Z_{1i}, \beta)$, it is a function of Z_{1i}, which satisfies the condition that can construct the IF.

$$-\left\{\frac{R_i - \pi(Z_{1i}, \psi_0)}{\pi(Z_{1i}, \psi_0)}\right\} h_2^*(Z_{1i}, \beta_0, \xi^*) \ \in \ \Lambda_2$$

- Therefore, the estimating equation for IF is

$$\sum_{i=1}^{n}\left[\frac{R_i\, m(Z_i, \beta)}{\pi(Z_{1i}, \hat{\psi}_n)} - \left\{\frac{R_i - \pi(Z_{1i}, \hat{\psi}_n)}{\pi(Z_{1i}, \hat{\psi}_n)}\right\} h_2^*(Z_{1i}, \beta_0, \xi^*)\right] = 0$$

- The only difference is to estimate ξ_n^* or not (AIPWCC requires integration).

Asymptotic equivalence between adaptive estimation and AIPWCC

Theorem 10.3

Let $\widehat{\beta}_n$ be the estimator obtained by the adaptive estimation, and let $\widehat{\beta}_n^*$ be the AIPWCC estimator . Under suitable regularity conditions,

$$n^{1/2}\left(\widehat{\beta}_n - \widehat{\beta}_n^*\right) \xrightarrow{P} 0.$$

Hence, the adaptive estimator is asymptotically equivalent to the AIPWCC estimator.

- Thus, we think misspecification of ξ and its estimation (=adaptive estimation) as a kind of negligible problem. (BUT no efficiency !)

Double robustness (DR)

- Under regularity, the adaptive AIPWCC estimator is **consistent** if

$$\textit{either}\ \ \pi(Z_1, \psi) \text{ is correctly specified} \quad \textbf{or} \quad p^*_{Z_2|Z_1}(\cdot \mid Z_1; \xi) \text{ is correct.}$$

- Proof sketch for monotone coarsening: algebraic identity + CAR $\Rightarrow$ decompositions where either model kills the bias term.

Takeaway

Adaptive augmentation buys efficiency; the DR structure buys protection against one misspecification.

Worked example (logistic regression; design-based missingness)

- Full-data $m(Z, \beta)$ chosen for logistic model of Y on $X = (X_1, X_2)$.
- Under two-level missingness, the AIPWCC form becomes (schematically)

$$\sum_i \left[\frac{R_i}{\pi(Y_i, X_{1i}; \hat{\psi})} X_i^* \{Y_i - \mu_\beta(X_i)\} - \frac{R_i - \pi(Y_i, X_{1i}; \hat{\psi})}{\pi(Y_i, X_{1i}; \hat{\psi})} L(Y_i, X_{1i}) \right] = 0,$$

with $L^\star(Y, X_1) = \mathbb{E}[X^* \{Y - \mu_\beta(X)\} \mid Y, X_1]$.

What we skip in Ch. 10 (for time)

- Detailed development for *monotone coarsening* and for *censoring* (survival)—we'll pick up the tools where needed in Ch. 11.
- Nonmonotone coarsening exists but is computationally heavy (iterative projections, integral operators).

Pointer

Keep the operators L and M in mind; they will be our bridge to Chapter 11.

10.5 Improving Efficiency when Coarsening is Nonmonotone

Generalization idea: operators L and M (you'll need this)

Definition 4 and 5

- Define two linear operators $\mathcal{L} : \mathcal{H}^F \to \mathcal{H}$ and $\mathcal{M} : \mathcal{H}^F \to \mathcal{H}^F$

$$\mathcal{L}\{h^F\} = \mathbb{E}\left[h^F(Z) \mid C, G_C(Z)\right], \qquad \mathcal{M}\{h^F\} = \mathbb{E}\left[\mathcal{L}\{h^F\} \mid Z\right].$$

where h^F is an arbitrary element in $\mathcal{H}^F$.

$$L\{h_F(\cdot)\} = \sum_{r=1}^{\infty} I(C = r)\,\mathbb{E}\{h_F(Z) \mid G_r(Z)\},$$

and

$$M\{h_F(\cdot)\} = \sum_{r=1}^{\infty} \#\{r, G_r(Z)\}\,\mathbb{E}\{h_F(Z) \mid G_r(Z)\}.$$

How to derive the projection : $\Pi\left[\frac{\mathbf{1}(C=\infty)h_F(Z)}{\pi(\infty,Z)}\,\middle|\,\Lambda_2\right]$

Theorem 10.6

- Inverse operator $\mathcal{M}^{-1}$ exists and is uniquely defined.
- Moreover, the projection is given by

$$\Pi\left[\frac{\mathbf{1}(C=\infty)h_F(Z)}{\pi(\infty,Z)}\,\middle|\,\Lambda_2\right]=\frac{\infty(C=\infty)h_F(Z)}{\pi(\infty,Z)}-\mathcal{L}[\mathcal{M}^{-1}\{h_F(\cdot)\}],$$

- By **Definition 1**, $\mathcal{L}[\mathcal{M}^{-1}\{\varphi^F(Z)\}]=\mathcal{J}\{\varphi^F(Z)\}$.
- Therefore, DR linear space is written by $\mathcal{J}(\Lambda^{F\perp})=\mathcal{L}[\mathcal{M}^{-1}\{\Lambda^{F\perp}\}]$.

Proof sketch for the uniqueness of $\mathcal{M}^{-1}$

- $I - \mathcal{M}$ is a contraction mapping

$$\Leftrightarrow \forall h^F \in \mathcal{H}^F, 0 \leq \exists L < 1, ||(I - \mathcal{M})h^F|| \leq L||h^F||$$

- Existence of $\mathcal{M}^{-1}$: $\sum_{k=0}^{\infty}(I - \mathcal{M})^k$ (Completeness of linear operator space)

$$\because \sum_{k=0}^{n}(I - \mathcal{M})^k \circ \mathcal{M} = I - (I - \mathcal{M})^{n+1} \to I \qquad \text{(as } n \to \infty\text{)}$$

- Uniqueness of $\mathcal{M}^{-1}$: $\mathcal{M}^{-1} = \mathcal{M}^{-1}(\mathcal{M} \circ \mathcal{M}'^{-1}) = (\mathcal{M}^{-1} \circ \mathcal{M})\mathcal{M}'^{-1} = \mathcal{M}'^{-1}$

Constructing Strategy

- we only focus on the operators $\mathcal{L}, \mathcal{M}^{-1}$ to estimate β.

$$\sum_{i=1}^{n} \mathcal{L}_i \left[\mathcal{M}_i^{-1} \{ m(Z_i, \beta) \} \right] = 0.$$

  - $\mathcal{L}$: posit model ξ for conditional probability of Z given $G_r(Z) \rightarrow \mathcal{L}(\cdot, \xi)$
  - $\mathcal{M}^{-1}$: ψ for coarsening probability, posit model ξ (same above) $\rightarrow \mathcal{M}^{-1}(\cdot, \psi, \xi)$
- To construct β, after estimating ψ for $\hat{\psi}_n$, and ξ for $\hat{\xi}_n^*$, solve the equation

$$\sum_{i=1}^{n} \mathcal{L}_i \left[\mathcal{M}_i^{-1} \{ m(Z_i, \beta), \hat{\psi}_n, \hat{\xi}_n^* \}, \hat{\xi}_n^* \right] = 0$$

Equivalence of $\mathcal{L} \circ \mathcal{M}^{-1}$ and AIPWCC

Theorem 10.7 - (i)

- Let $d^F(Z, \beta, \psi, \xi) = \mathcal{M}^{-1}\{m(Z, \beta), \psi, \xi\}$, then, the residual can separate

$$\begin{aligned}\mathcal{L}[\mathcal{M}^{-1}\{m(Z,\beta),\psi,\xi\},\xi] &= \mathcal{L}\{d_F(Z,\beta,\psi,\xi),\xi\} \\ &= \underbrace{\frac{I(C=\infty)\, m(Z,\beta)}{\pi(\infty, Z, \psi)}}_{\text{IPWCC space}} + \underbrace{L_2^*\{C, G_C(Z), \beta, \psi, \xi\}}_{\text{Augmentation space}}.\end{aligned}$$

- This means the solution to this estimating equation is an AIPWCC estimator

$$\sum_{i=1}^{n} \frac{I(C_i = \infty)\, m(Z_i, \beta)}{\varpi(\infty, Z_i, \hat{\psi}_n)} + L_2^*\{C_i, G_{C_i}(Z_i), \beta, \hat{\psi}_n, \hat{\xi}_n^*\} = 0.$$

The part of augmentation space

Theorem 10.7 - (ii)

- The element of the augmentation space is written by

$$
\begin{aligned}
& L_2^*\{C, G_C(Z), \beta, \psi, \xi\} = \\
& -\frac{I(C=\infty)}{\pi(\infty, Z, \psi)}\left(\sum_{r \neq \infty} \pi(r, G_r(Z), \psi)\, \mathbb{E}[d_F(Z, \beta, \psi, \xi) \mid G_r(Z), \xi]\right) \\
& + \sum_{r \neq \infty} I(C=r)\, \mathbb{E}[d_F(Z, \beta, \psi, \xi) \mid G_r(Z), \xi]
\end{aligned}
$$

- Let $L_{2r}\{G_r(Z)\} = -\mathbb{E}[d_F(Z, \beta, \psi, \xi) \mid G_r(Z), \xi]$, we can obtain the same representation as a typical element of Λ_2. (see (7.37))

Construct $\mathcal{M}^{-1}$ and $\mathcal{L}$

- $\mathcal{M}^{-1}$ is represented by infinite series, so be constructed by successive approximation.

$$d^F_{(j)}(Z,\beta,\psi,\xi) = \hat{\mathcal{M}}^{-1}\{m(Z,\beta),\psi,\xi\}$$

- By construction, $L_2^*\{C, G_C(Z), \beta, \psi, \xi\} \in \Lambda_2$ in any cases.
- This means that when ψ_0 (coarsening probability model) is correctly specified,

$$\frac{I(C=\infty)\, m(Z,\beta_0)}{\varpi(\infty, Z, \psi_0)} + L^*_{2(j)}\{C, G_C(Z), \beta_0, \psi_0, \xi^*\} \in \Lambda^{\perp}$$

- Hence, we can obtain an AIPWCC estimator by solving

$$\sum_{i=1}^{n} \frac{I(C_i=\infty)\, m(Z_i,\beta)}{\varpi(\infty, Z_i, \hat{\psi}_n)} + L^*_{2(j)}\{C_i, G_{C_i}(Z_i), \beta, \hat{\psi}_n, \hat{\xi}^*_n\} = 0.$$

- The estimator is RAL (i.e. consistent and asymptotically normal)

Double Robustness

- In addition to correct specification of ψ, the posit model $p^*_Z(z, \xi)$ is correctly specified, its influence function is efficient. (double-robustness)
- Also, this AIPWCC estimator is consistent and asymptotically normal even if ψ is misspecified. (Actually, we do not have to consider this pattern thanks to CAR)
- Recall this two probabilistic model
  - posit model $p^*_Z(z, \xi)$; estimate $\hat{\xi}^*_n \to \xi^*(\neq \xi_0)$ (if misspecified)
  - coarsening model $\pi\{r, G_r(Z), \psi\}$; estimate $\hat{\psi}^*_n \to \psi^*(\neq \psi_0)$ (if misspecified)

Theorem 10.8

- This theorem establishes the **double robustness** property.
- Specifically:

$$\mathbb{E}_{\xi_0,\psi_0}\left[L\big(M^{-1}\{m(Z,\beta_0),\psi_0,\xi^*\},\,\xi^*\big)\right]=0,$$

and

$$\mathbb{E}_{\xi_0,\psi_0}\left[L\big(M^{-1}\{m(Z,\beta_0),\psi^*,\xi_0\},\,\xi_0\big)\right]=0.$$

- Hence, the estimating equation remains unbiased if either the coarsening model (ψ) or the marginal model of Z (ξ) is correctly specified.

11. Locally Efficient Estimators for Coarsened-Data Semiparametric Models

Roadmap for Chapter 11

- For a full-data estimating function $m(Z, \beta) \in \Lambda^{F\perp}$, applying $\mathcal{J}$, we obtain an observed-data estimating function.

$$\sum_{i=1}^{n} \left[\frac{I(C_i = \infty) m(Z_i, \beta)}{\varpi(\infty, Z_i, \hat{\psi}_n)} - \Pi \left(\frac{I(C_i = \infty) m(Z_i, \beta)}{\varpi(\infty, Z_i)} \;\middle|\; \Lambda_2 \right) \right] = 0$$

- For computation, we need to estimate ψ in the former, ξ (posit model) inthe latter.
- We propose a locally efficient estimator, which achieves the semiparametric efficiency bound if the posited model $p_Z^*(z, \xi)$ is correct but still be consistent and asymptotically normal and RAL even if misspecified.
- Recall the essential framework : Efficient score → Efficient IF (proportional relationship)

11.1 The Observed-Data Efficient Score

Efficient score: Representation 1 (likelihood-based) : theoritical

- Basiccaly, the efficient score is given by

$$S_{\text{eff}}\{C, G_C(Z)\} = S_\beta\{C, G_C(Z)\} - \Pi\big[S_\beta\{C, G_C(Z)\} \mid \Lambda\big],$$

where $S_\beta = \mathbb{E}\big[S_\beta^F(Z) \mid C, G_C(Z)\big]$, and $\Lambda = \Lambda_\psi \oplus \Lambda_\eta \;\; (\Lambda_\psi \perp \Lambda_\eta)$.

$$\begin{aligned}
\Pi\big[S_\beta\{C, G_C(Z)\} \mid \Lambda\big] &= \Pi\big[S_\beta\{C, G_C(Z)\} \mid \Lambda_\psi\big] + \Pi\big[S_\beta\{C, G_C(Z)\} \mid \Lambda_\eta\big] \\
&= \Pi\big[S_\beta\{C, G_C(Z)\} \mid \Lambda_\eta\big] \qquad (\because S_\beta\{C, G_C(Z)\} \perp \Lambda_2)
\end{aligned}$$

- Recall $\Lambda_\eta = \{\mathbb{E}[\alpha^F(Z) \mid C, G_C(Z)] : \alpha^F(Z) \in \Lambda^F\}$, there exists $\alpha_{\text{eff}}^F(Z) \in \Lambda^F$,

$$\Pi\big[S_\beta\{C, G_C(Z)\} \mid \Lambda_\eta\big] = \mathbb{E}[\alpha_{\text{eff}}^F(Z) \mid C, G_C(Z)]$$

$$\begin{aligned}
\therefore\; S_{\text{eff}}\{C, G_C(Z)\} &= S_\beta\{C, G_C(Z)\} - \Pi\big[S_\beta\{C, G_C(Z)\} \,\big|\, \Lambda_\eta\big] \\
&= E\big[\,\{S_\beta^F(Z) - \alpha_{\text{eff}}^F(Z)\} \,\big|\, C, G_C(Z)\big]
\end{aligned}$$

Efficient score: Representation 2 (AIPWCC-based) : practical

- Since the full-data efficient score $B^F(Z) \in \Lambda^{F\perp}$, we can apply $\mathcal{J}$.

$$S_{\text{eff}}\{C, G_C(Z)\} = \mathcal{J}\left\{B_F^{\text{eff}}(Z)\right\} = I(C=\infty)\frac{B_F^{\text{eff}}(Z)}{\pi(\infty, Z)} - \Pi\left[I(C=\infty)\frac{B_F^{\text{eff}}(Z)}{\pi(\infty, Z)}\middle|\Lambda_2\right].$$

- Clearly, this element lives in DR linear space $\mathcal{J}(\Lambda^{F\perp})$
- Now, we obtain two equivalent representations about the efficient score

$$\mathbb{E}\left[\, S_\beta^F(Z) - \alpha_{\text{eff}}^F(Z) \,\middle|\, C,\, G_C(Z)\right] = \frac{I(C=\infty)\, B_{\text{eff}}^F(Z)}{\varpi(\infty, Z)} - \Pi\left[\frac{I(C=\infty)\, B_{\text{eff}}^F(Z)}{\varpi(\infty, Z)}\middle|\Lambda_2\right]$$

- For computation, this AIPWCC estimator is far more practical but still be difficult to construct. $\to$ Efficiency problem boils down to find $B_{\text{eff}}^F(Z)$.

How to derive $B_{\text{eff}}^{F}(Z)$

Theorem 11.1

- $B_{\text{eff}}^{F}(Z) \in \Lambda^{F\perp}$ is the unique solution to

$$\Pi\left[\mathcal{M}^{-1}\{B^{F}(Z)\} \,\middle|\, \Lambda^{F\perp}\right] = S_{\text{eff}}^{F}(Z),$$

- Thereofore, solving this equation, we obtain $B_{\text{eff}}^{F}(Z)$, then we can construct teh efficient score (Note : This is revisit)

$$S_{\text{eff}}\{C, G_C(Z)\} = \frac{I(C=\infty)\, B_{\text{eff}}^{F}(Z)}{\varpi(\infty, Z)} - \Pi\left[\frac{I(C=\infty)\, B_{\text{eff}}^{F}(Z)}{\varpi(\infty, Z)} \,\middle|\, \Lambda_2\right],$$

Proof sketch of Lemma 11.1 : successive approximation

- To construct $B^F_{\text{eff}}(Z)$, define $D^F_{\text{eff}}(Z) = \mathcal{M}^{-1}\{B^F_{\text{eff}}(Z)\}$.
- Then $D^F_{\text{eff}}(Z)$ is the solution to

$$S_{\text{eff}}(Z) = \Pi\big[h^F(Z) \,\big|\, \Lambda^{F\perp}\big] + \Pi\big[\mathcal{M}\{h^F(Z)\} \,\big|\, \Lambda^F\big] = (I - \mathcal{Q})\{h^F(Z)\}$$

where $\mathcal{Q}\{D^F_{\text{eff}}(Z)\} = \Pi\Big[(I - \mathcal{M})\{D^F_{\text{eff}}(Z)\} \,\Big|\, \Lambda^F\Big]$

- $\mathcal{Q}$ is a contraction mapping, so $(I - \mathcal{Q})^{-1}$ uniquely exists.
- Finally, we can successively approximate $D^F_{\text{eff}}(Z) = (\sum_{i=0}^{\infty} \mathcal{Q}^i) S^F_{\text{eff}}(Z)$

Computational focus : $\mathcal{M}^{-1}$

Theorem 11.2

- Suppose the coarsening mechanism is **monotone**.
- For any $h^F(Z) \in \mathcal{H}^F$, the inverse operator M^{-1} has the closed form

$$M^{-1}\{h^F(Z)\} = \frac{h^F(Z)}{\varpi(\infty, Z)} - \sum_{r \neq \infty} \frac{\lambda_r\{G_r(Z)\}}{K_r\{G_r(Z)\}} \mathbb{E}\left[h^F(Z) \mid G_r(Z)\right],$$

where

$$K_r(Z) = \Pr(C \geq r+1 \mid Z), \quad \lambda_r\{G_r(Z)\} = \Pr(C = r \mid C \geq r, Z).$$

- Hence, M^{-1} can be computed explicitly in terms of the hazard and survival functions of the coarsening distribution.

Lemma 11.2

- Let $\{D^{(k)}\}$ be defined recursively by

$$D^{(k+1)}(Z) = \Pi\Big((I - M)D^{(k)}(Z) \,\Big|\, \mathcal{H}^F\Big) + S_{\text{eff}}^F(Z),$$

  with $D^{(0)}$ arbitrary.
- Then

$$D^{(k)} \longrightarrow D_{\text{eff}}^F \quad \text{in } \mathcal{H}^F \text{ norm},$$

  and

$$B^{(k)}(Z) = \Pi\big(MD^{(k)}(Z) \,\big|\, \Lambda^{F\perp}\big) \longrightarrow B_{\text{eff}}^F(Z).$$

- Thus, the successive approximation scheme converges to the efficient element $B_{\text{eff}}^F(Z)$.

11.2 Strategy for Obtaining Improved Estimators

Estimation scheme

- Consider finding a full-data estimating function $m(Z, \beta)$.
- Firstly, by MLE (or likelihood methods), we estimate $\hat{\psi}_n, \hat{\xi}_n^*$.
- Secondly, we calculate full-data efficient score $S_F^{\text{eff}}(Z)$.
- By successive approximation, $\hat{B}_{\text{eff}}^F(Z) = m(Z, ^b eta, \hat{\psi}_n, \hat{\xi}_n^*)$.
- Finally, we can estimate observed-data efficient score using

$$S_{\text{eff}}\{C, G_C(Z)\} = \frac{I(C = \infty)\, B_{\text{eff}}^F(Z)}{\varpi(\infty, Z)} - \Pi\left[\frac{I(C = \infty)\, B_{\text{eff}}^F(Z)}{\varpi(\infty, Z)} \,\middle|\, \Lambda_2\right]$$

- Now, we focus on the projection $\Pi\left[\dfrac{I(C=\infty)\, m(Z,\beta)}{\varpi(\infty,Z)}\,\middle|\,\Lambda_2\right]$.

  $\rightarrow$ In some special coarsening mechanism, this projection is simplily written down.

- In actual implementation, we solve this adaptive estimating equation.

$$
\begin{aligned}
\sum_{i=1}^{n} \Bigg(& \frac{I(C_i=\infty)\, m(Z_i,\beta,\hat{\psi}_n,\hat{\xi}_n^*)}{\varpi(\infty,Z_i,\hat{\psi}_n)} \\
& - \frac{I(C_i=\infty)}{\varpi(\infty,Z_i,\hat{\psi}_n)} \left[\sum_{r\neq\infty} \varpi\{r, G_r(Z_i), \hat{\psi}_n\}\, \mathbb{E}\Big\{ D^{(j)}(Z,\beta,\hat{\psi}_n,\hat{\xi}_n^*) \,\Big|\, G_r(Z_i), \hat{\xi}_n^* \Big\} \right] \\
& + \sum_{r\neq\infty} I(C_i=r)\, \mathbb{E}\Big\{ D^{(j)}(Z,\beta,\hat{\psi}_n,\hat{\xi}_n^*) \,\Big|\, G_r(Z_i), \hat{\xi}_n^* \Big\} \Bigg) \;=\; 0
\end{aligned}
$$

locally efficient estimator for restricted moment model with monotone coarsening

- Full data: $Z = (Y, X)$, Model: $Y = \mu(X, \beta) + \varepsilon$ with $\mathbb{E}(\varepsilon \mid X) = 0$.
- Full-data efficient score: $S^F_{\text{eff}}(Z) = D^\top(X)V^{-1}(X)\varepsilon$,
  $D = \partial\mu/\partial\beta^\top, V(X) = \text{Var}(Y|X)$.
- Seek $B^{\text{eff}}_F(Z) = A(X)\varepsilon \in \mathcal{F}^\perp$ solving

$$\Pi\big[M^{-1}\{A(X)\varepsilon\} \mid \mathcal{F}^\perp\big] = S^F_{\text{eff}}(Z).$$

- Leads to an integral equation for $A(X)$; exact solution rare; use numerical/approximate A_{imp}.

Approximate locally efficient estimator (practice)

- Compute a numerically feasible $A_{\text{imp}}(X, \hat{\beta}, \hat{\psi}, \hat{\xi}^*)$.
- Plug into the AIPWCC efficient-score form:

$$\sum_{i=1}^{n} \left[I(C_i = \infty) \frac{A_{\text{imp}}(X_i, \beta, \hat{\psi}, \hat{\xi}^*)\{Y_i - \mu(X_i, \beta)\}}{\pi(\infty, Y_i, X_i; \hat{\psi})} + L_2\{C_i, G_{C_i}(Y_i, X_i), \beta, \hat{\psi}, \hat{\xi}^*\} \right] = 0.$$

- Here L_2 is the projection term (closed form under monotone coarsening).

Robustness note

The A_{imp} route targets local efficiency but may be numerically heavy; when it's too hard, fall back to the DR AIPWCC from Ch. 10.

11.3 Concluding Thoughts

Summary

- Adaptive estimation : Specified coarsening model $\rightarrow$ using posit model $\xi \rightarrow$ fit ξ_n^*
- Inverse weighted method balances between simplicity of implementation and relative efficiency.
- AIPWCC estimator is more improved one in double robustness and gaining considerable efficiency at the cost of numerical implementaion. (successive approximation)
- Locally efficient estimator aims to find optimal full-data estimating function $B^F_{\text{eff}}(Z)$ as well as optimal augmentation. (difficult to implement)

References

References

- Tsiatis, A. (2006). *Semiparametric Theory and Missing Data*. Springer (Chs. 10–11).
- A.W. van der Vaart (1998). *Asymptotic Statistics*. Cambridge University Press.