Semiparametric Theory and Missing Data
Chapter 7
Missing and Coarsening at Random for Semiparametric Models

Martin, Dominic and Iris

Department of Statistics, Chinese University of Hong Kong

2022 Summer Lab Meeting

July 26, 2022

Missing and Coarsened Data

Preliminary Remark and Review

- In previous chapter, we denote W_0 as auxilary variables and Z_0 as primary variable.
- In this chapter, we write $Z = (Z_1^T, Z_2^T)^T := (Z_0^T, W_0^T)^T$ as the fulldata.

Suppose $Z \sim p_Z(z, \beta, \eta)$, which is a semiparametric model, where $\eta = (\eta_1, \eta_2)$, we can write

$$p_Z(z, \beta, \eta) = p_{Z_1}(z_1, \beta, \eta_1) p_{Z_2|Z_1}(z_2|z_2, \eta_2), \tag{1}$$

- $\beta \in \mathbb{R}^q$, the finite dimensional parameter of interest is the primary interest.
- $\eta \in \mathbb{R}^r$, where $r \in \mathbb{N}^+ \cup \{\infty\}$ is important for conducting correct inference, is not of primary interest.

Notion of Coarsening

Definition 1.1 (Coarsened Data)

Let the full data be $Z_1, \cdots, Z_n$, which are iid. Let

- $G_r(Z)$ be random many-to-one function, where $r \in \{1, \ldots, l\}$.
- $G_\infty(Z)$ be the identity map, i.e. $G_\infty(Z) = Z$.
- $\mathscr{C}$ be the coarsening variable, where $\mathscr{C} \in \{1, \ldots, l\} \cup \{\infty\}$.

The data is said to be coarsened when the observed data are in form of

$$\{\mathscr{C}_i, G_{\mathscr{C}_i}(Z_i)\}_{i=1}^n \tag{2}$$

Remark 0.1 (Many-to-one function)

- *A function f is* **one-to-one** *if $f(x) = f(y)$ implies $x = y$.*
- *A function f is* **many-to-one** *if it is not one-to-one, i,e, $\exists x \neq y$ with $f(x) = f(y)$.*

Remark on Notion of Coarsening

Remark 0.2 (Coarsening index)

- *For $\mathscr{C} \neq \infty$, $G_{\mathscr{C}}(Z)$ is NOT one-to-one, i,e, there is* **information loss** *compare to the full-data set.*
- *For $\mathscr{C} = \infty$, $G_{\mathscr{C}}(Z) = Z$, i.e. no coarsening.*

Remark 0.3 (Assumption)

All theory developed later on relies on the following assumption: $\exists \epsilon > 0$ s.t. for all z,

$$\Pr(\mathscr{C} = \infty | Z = z) \geq \epsilon > 0, \tag{3}$$

i.e. there is positive probability of observing the full data.

Example of Coarsened Data

Example 1.2

An biologist want to study the relationship between X and Y, where

- X is the serum concentration.
- Y: some outcome affected by X.

The biologist also interested in the **within-person variability** in X. Hence, two blood samples of equal volume are drawn in a study for some individuals. Let X_1 and X_2 be the associated value of X of two different samples.

- For some individuals, we would observe $\{X_1, X_2, Y\}$. ($\mathscr{C} = \infty$)
- For the remaining, the biologist would simply mix the two blood samples (for saving expenses), i.e. we would observe $\{(X_1 + X_2)/2, Y\}$. ($\mathscr{C} := 1$)

We can thus define

$$G_r(X_1, X_2, Y) := \begin{cases} (X_1, X_2, Y) & r = \infty \\ \left(\frac{X_1+X_2}{2}, Y\right) & r = 1 \end{cases} \tag{4}$$

to characterize the coarsening mechanism.

Generalization of Notion of Missingness

Suppose full data is of form

$$Z = \left(Z^{(1)}, \ldots, Z^{(d)}\right)^T.$$

Having missing data is equivalent to have some $\mathcal{O} \subset \{1, \ldots, d\}$ and $\mathcal{O} \neq \{1, \ldots, d\}$ s.t.

$$G_r(Z) = \cup_{k \in \mathcal{O}} \{Z^{(k)}\}$$

being observed. Therefore, we could define G_r as a mapping function of Z to its subset to characterize the missingness.

Generalization of Notion of Missingness

Example 1.3 (Missing Data)

Let $Z = (Z^{(1)}, Z^{(2)})^T$ be vector of two random variables. Define

$\mathscr{C}$	$G_{\mathscr{C}}(Z)$
1	$Z^{(1)}$
2	$Z^{(2)}$
∞	Z

That is, if $\mathscr{C} = 1$, we only observe $Z^{(1)}$ and $Z^{(2)}$ missing; if $\mathscr{C} = 2$, we only observe $Z^{(2)}$ and $Z^{(1)}$ missing. if $\mathscr{C} = \infty$, both $Z^{(1)}$ and $Z^{(2)}$ are observed.

Remark 0.4

If we were dealing only missing data, say a l-dimensional random vector, it may be more covenient to define the missingness variable to be an l-dimensional vector of 1's and 0's to denote which element of vector is observed or missing. If it is convenient to switch to such notation, we would use R to denote such missingness indicators.

Coarsened-Data Mechanisms

Definition 1.4 (Coarsened-Data Mechanism)

- Coarsening completely at random (CCAR): $\mathscr{C} \perp\!\!\!\perp Z$, i.e.

$$\Pr(\mathscr{C} = r|Z) = \varpi(r) \quad \text{for all } r, Z.$$

- Coarsening at random (CAR):

$$\Pr(\mathscr{C} = r|Z) = \varpi\{r, G_r(Z)\},$$

  i.e. the probability of coarsening depends on Z only as a function of observed data.
- Noncoarsening at random (NCAR): if it is not CAR, i.e. $\exists z_1, Z_2$ s.t. $G_r(z_1) = G_r(z_2)$ for some r and $\mathrm{P}(\mathscr{C} = r|z_1) \neq \mathrm{P}(\mathscr{C} = r|z_2)$.

Remark 0.5 (More about Notation)

- *Full data* $Z_1, \ldots, Z_n$.
- *Observed data* $\{\mathscr{C}_i, G_{\mathscr{C}_i}(Z_i)\}_{i=1}^n$
- *Complete data* $\{Z_i : C_i = \infty\}$

The Density and Likelihood of Coarsened Data

Basics

In order to find observed-data estimator of parameter of interest β, we need to derive the likelihood of observed data. We write the density as

$$p_{\mathscr{C},Z}(r,z,\psi,\beta,\eta) = \mathrm{P}(\mathscr{C} = r|Z = z,\psi)p_Z(z,\beta,\eta), \tag{5}$$

where η and ψ are the nuisance parameter and the parameter governing coarsening mechanism respectively.

Remark 0.6

Notice $\{\mathscr{C}, Z\}$ are not fully observable, but the observable data $\{\mathscr{C}, G_{\mathscr{C}}(Z)\}$ is a known function of $\{\mathscr{C}, Z\}$. Hence studying likelihood of fulldata help study of observed data.

Remark 0.7

Let Z be a discrete random variable and v_Z be corresponding counting measure, then we can write

$$\mathrm{P}(Z \in A) = \sum_{z \in A} \mathrm{P}(Z = z) = \int_{z \in A} p_Z(z) dv_Z(z)$$

Discrete Data

For simplicity, we first consider the case when Z is discrete random vector.

We can write

$$\mathrm{P}(\mathscr{C} = r, G_{\mathscr{C}}(Z) = g_r) = \sum_{z:G_r(z)=g_r} \mathrm{P}(C = r, Z = z) = \sum_{z:G_r(z)=g_r} \mathrm{P}(C = r|Z = z)\mathrm{P}(Z = z) \tag{6}$$

and by remark 0.7, we can thus write

$$p_{\mathscr{C},G_{\mathscr{C}}(Z)}(r, g_r, \psi, \beta, \eta) = \int_{\{z:G_r(z)=g_r\}} \mathrm{P}(C = r|Z = z, \psi) p_Z(z, \beta, \eta) dv_Z(z) \tag{7}$$

Continuous Data

Consider $Z = (Z_1, \ldots, Z_l)$ is l-dimensional continuous random vector.

- Let $G_r(z)$ be l_r-dimensional, $l_r < l$ for $r \neq \infty$.
- Assume $\exists V_r(z)$ being $(l - l_r)$ dimensional so that the mapping between
$$z \leftrightarrow \{G_r^T(z), V_r^T(z)\}$$
is one-to-one for all r.

Define $z = H_r(g_r, v_r)$ as the inverse transform, then we have

$$p_{G_r,V_r}(g_r, v_r) = p_Z\{H_r(g_r, v_r)\}J(g_r, v_r), \tag{8}$$

where J is the Jacobian matrix of H_r w.r.t. (g_r, v_r). Hence we can write

$$p_{\mathscr{C},G_{\mathscr{C}}}(r, g_r) = \int p_{\mathscr{C},G_{\mathscr{C}},V_{\mathscr{C}}}(r, g_r, v_r)dv_r = \int \mathrm{P}(\mathscr{C} = r|G_r = g_r, V_r = v_r)p_{G_r,V_r}(g_r, v_r)dv_r \tag{9}$$

$$= \int \mathrm{P}(\mathscr{C} = r|Z = H_r(g_r, v_r))p_{G_r,V_r}(r, g_r, v_r)dv_r \tag{10}$$

$$\Rightarrow p_{\mathscr{C},G_{\mathscr{C}}}(r, g_r, \psi, \beta, \eta) = \int \mathrm{P}(\mathscr{C} = r|Z = H_r(g_r, v_r), \psi)p_Z\{H_r(g_r, v_r), \beta, \eta\}J(g_r, v_r)dv_r \tag{11}$$

CAR Likelihood

Remark 0.8

The difference between discrete and continuous case only differ in the Jacobian matrix.

We then derive the likelihood of the observed data under CAR assumption. For the time being, we denote such a model by

$$\mathrm{P}(\mathscr{C} = r | Z = z) = \varpi\{r, G_r(z), \psi\}, \tag{12}$$

where ψ is unknown parameter that is **functionally independent** if (β, η).

Remark 0.9

- *Sometimes, the coarsening probability ϖ is known to investigator, in such case the parameter ψ is not needed.*
- *We assumed that the model for $\varpi\{r, G_r(z), \psi\}$ is known and correctly specified.*

CAR Likelihood

Assume that Z is discrete,

$$p_{\mathscr{C},G_{\mathscr{C}}(Z)}(r,g_r,\psi,\beta,\eta)=\int_{z:G_r(z)=g_r}\mathrm{P}(\mathscr{C}=r|Z=z,\psi)p_Z(z,\beta,\eta)d\nu_Z(z)$$
$$=\int_{z:G_r(z)=g_r}\varpi\{r,g_r(z),\psi\}p_Z(z,\beta,\eta)d\nu_Z(z)=\varpi\{r,g_r(z),\psi\}\int_{z:G_r(z)=g_r}p_Z(z,\beta,\eta)d\nu_Z(z)$$

Assume that Z is continuous. Notice that $G_r(H_r(g_r,v_r))=g_r$, hence

$$\begin{aligned}p_{\mathscr{C},G_{\mathscr{C}}(Z)}(r,g_r,\psi,\beta,\eta)&=\int\varpi(r,g_r,\psi)p_Z\{H_r(g_r,v_r),\beta,\eta\}J(g_r,v_r)dv_r\\&=\varpi(r,g_r,\psi)\int p_Z\{H_r(g_r,v_r),\beta,\eta\}J(g_r,v_r)dv_r\end{aligned}$$

In both cases, we have

$$p_{\mathscr{C},G_{\mathscr{C}}(Z)}(r,g_r,\psi,\beta,\eta)=\varpi(r,g_r,\psi)p_{G_r(Z)}(g_r,\beta,\eta)\tag{13}$$

Remark on Likelihood Methods

Suppose we posit a parametric model for full data and aim to estimate β using coarsened data, We then maximize

$$\Pi_{i=1}^{n} p_{\mathscr{C},G_{\mathscr{C}}(Z)}(r_i, g_{r_i}, \psi, \beta, \eta) = \{\Pi_{i=1}^{n} \varpi(r_i, g_{r_i}, \psi)\} \left\{\Pi_{i=1}^{n} p_{G_{r_i}(Z)}(g_{r_i}, \beta, \eta)\right\} \tag{14}$$

with respect to β. Thus it is sufficient to maximize

$$\Pi_{i=1}^{n} p_{G_{r_i}(Z)}(g_{r_i}, \beta, \eta), \quad p_{G_r(Z)}(g_r, \beta, \eta) = \int_{z:G_r(z)=g_r} p_Z(z, \beta, \eta) dv_Z(z). \tag{15}$$

Hence as long as we believe the CAR assumption, we can find the MLE for β and η without specifying any model for the coarsening process.

Examples

Example 2.1 (Blood Concentration example (continued))

Assume

- $X_j = \alpha + e_j$, where $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$, $e_j \sim N(0, \sigma_e^2)$ independently.
- The individuals chosen to have their blood samples combinded chosen at random (CCAR), i.e. $\mathrm{P}(\mathscr{C} = 1|Z) = \varpi$ and $\mathrm{P}(\mathscr{C} = \infty|Z) = 1 - \varpi$.

It follows that (X_1, X_2) follows bivariate normal with covariance matrix Σ. Hence, we can write the likelihood as

$$\prod_{i=1}^{n}\Bigg\{\left(|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}\{(X_{i1}-\mu_\alpha, X_{i2}-\mu_\alpha)^T\Sigma^{-1}(X_{i1}-\mu_\alpha, X_{i2}-\mu_\alpha)\}\right]\right)^{I(C_i=\infty)} \tag{16}$$

$$\times\left((\sigma_\alpha^2+\sigma_e^2/2)^{-1/2}\exp\left[-\frac{\{(X_{i1}+X_{i2})/2-\mu_\alpha\}^2}{2(\sigma_\alpha^2+\sigma_e^2/2)}\right]\right)^{I(C_i=1)}\Bigg\} \tag{17}$$

Examples

Remark 0.10

Although MLE is preferred for obtaining estimators in finite-dimensional parametric models of full data Z, it may not be feasible for semiparametric models.

Example 2.2 (Logistic Regression)

Let Y is binary response with covariate X. Assume the logit model

$$P(Y = 1|X) = \frac{\exp(\beta^T X^*)}{1 + \exp(\beta^T X^*)},$$

where $X^* = (1, X^T)^T$. With full data, the likelihood for a single observation is

$$p_{Y|X}(y|x)p_X(x) = \left[\frac{\exp(\beta^T x^*)y}{1 + \exp(\beta^T x^*)}\right] p_X\{x, \eta(\cdot)\} \tag{18}$$

where $\eta(\cdot)$ is infinite-dimensional nuisance function allowing all nonparametric densities for marginal distribution of X. In order to find MLE for β, it suffices to maximize

$$\Pi_{i=1}^n \left[\frac{\exp(\beta^T X_i^*) Y_i}{1 + \exp(\beta^T X_i^*)}\right] \tag{19}$$

Examples

Example 2.3 (Logistic Regression (Continued))

However, if we have coarsened data under CAR, then the likelihood contribution for th part of the likelihood that involves β for a single observation is

$$\int_{(y,x):G_r(y,x)=g_r} \left\{ \frac{\exp(\beta^T x^*)y}{1+\exp(\beta^T x^*)} \right\} p_X\{x,\eta(\cdot)\} d\nu_{Y,X}(y,x). \tag{20}$$

Remark 0.11

Maximizing the above equation with respect to both β and infinite-dimensional parameter $\eta(\cdot)$ is nearly impossible. It motivates us to consider alternative to likelihood method for parameter estimation under coarsened data.

The Geometry of Semiparametric Coarsened-Data Models

In this section, we will study the class of influence functions.

Settings

Define the settings:

- 1. H^F: full data Hilbert space of all q-dimensional, mean-zero measurable functions of Z.
  2. H : for $\{C, G_C(Z)\}$.
  3. H^{CZ}: for $\{C, Z\}$.
- The full-data nuisance score vector is

$$S_\gamma^F(Z) = \frac{\partial \log p_Z(Z, \beta_0, \gamma_0)}{\partial \gamma},$$

  corresponding the the full-data parametric submodel $p_Z(z, \beta^{q\times 1}, \gamma^{r\times 1})$.
- Λ^F is the full-data nuisance tangent space, and is given by

$$\{B^{q\times r} S_\gamma^F(Z) \text{ for all } q \times r \text{ matrices } B\}$$

.

Settings

- The class of full-data influence functions are the elements $\psi^F(Z) \in H^F$ (The full-data Hilbert Space) such that
  1. $\psi^F(Z) \in \Lambda^{F\perp}$
  2. $\mathrm{E}\{\psi^F(Z)(S_\beta^F)^\intercal\} = I^{q\times q}$, where

$$S_\beta^F = \frac{\partial \log p_Z(Z, \beta_0, \eta_0)}{\partial \beta}$$

- (See Chapter 3, p.36) The efficient full-data score

$$S_{\mathit{eff}}^F(Z) = S_\beta^F(Z) - \Pi\{S_\beta^F(Z) \mid \Lambda^F\},$$

where $\Pi(S_\beta^F \mid \Lambda^F) = \mathrm{E}(S_\beta S_\eta^\intercal)\{\mathrm{E}(S_\eta S_\eta^\intercal)\}^{-1} S_\eta$.
The efficient full-data infludence function is

$$\varphi^F(Z) = \left[\mathrm{E}\left\{S_{\mathit{eff}}^F(Z) S_{\mathit{eff}}^{F\intercal}(Z)\right\}\right]^{-1} S_{\mathit{eff}}^F(Z),$$

Settings

Goal: Find the class of observed-data influence function.

Remarks:

- For discrete data and CAR coarsening, the likelihood is

$$p_{C,G_C(Z)}(r, g_r, \psi, \beta, \eta) = \varpi(r, g_r, \psi) \int_{\{G_r(z)=g_r\}} p_Z(z, \beta, \eta) d\nu_Z(z),$$

  and it can be generalized to continuous data as in (11). So we can focus only on discrete data first.

- Note that ψ and η are nuisance parameters. It will later (Chapter 8) be shown that

$$\Lambda = \Lambda_\psi \oplus \Lambda_\eta, \quad \Lambda_\psi \perp \Lambda_\eta.$$

  But when the coarsening of the data is by design (which is known), we don't need ψ, as the coarsening mechanism is given as $\varpi(r, G_r(z))$. The author suggests us to focus on this case first.

Nuisance Tangent Space

We begin by first considering the parametric submodel of the full-data Z given by $p_Z(z, \beta^{q\times 1}, \gamma^{r\times 1})$.

Lemma 3.1

The parametric submodel observed-data score vector with respect to γ is given by

$$S_\gamma(r, g_r) = \mathrm{E}\{S_\gamma^F(Z) \mid G_r(Z) = g_r\}$$

Lemma 3.2

When the coarsening mechanism is CAR, then

$$S_\gamma(r, g_r) = \mathrm{E}\{S_\gamma^F(Z) \mid G_r(Z) = g_r\} = \mathrm{E}\{S_\gamma^F(Z) \mid C = r, G_r(Z) = g_r\}.$$

The proof is skipped.

Notice that in general $\mathrm{E}\{S_\gamma^F(Z) \mid C = r, G_r(Z) = g_r\} \neq \mathrm{E}\{S_\gamma^F(Z) \mid G_r(Z) = g_r\}$, as it might not me true that $p_{Z|C,G_C(Z)}(z \mid r, g_r) = p_{Z|G_r(Z)}(z \mid g_r)$

Nuisance Tangent Space

Theorem 3.3

Under CAR. *The space Λ_η (i.e., the mean square closure of parametric submodel nuisance tangent spaces spanned by $S_\gamma\{C, G_C(Z)\}$) is the space of elements*

$$\Lambda_\eta = [\mathrm{E}\{\alpha^F \mid C, G_C(Z)\} \text{ for all } \alpha^F \in \Lambda^F],$$

where Λ^F denotes the full-data nuisance tangent space. We will also denote this space by the shorthand notation

$$\Lambda_\eta = \mathrm{E}\{\Lambda^F \mid C, G_C(Z)\}.$$

Proof of Theorem 3.3

Proof.

By Lemma 7.2 (3.3 in this slides), the nuisance score vector is

$$S_\gamma(r, g_r) = \mathrm{E}\{S_\gamma^F(Z) \mid G_r(Z) = g_r\} = \mathrm{E}\{S_\gamma^F(Z) \mid C = r, G_r(Z) = g_r\}.$$

So

$$\begin{aligned}\Lambda_\eta &= [B^{q\times r}\mathrm{E}\{S_\gamma^F(Z) \mid C, G_C(Z)\} \text{ for all } B^{q\times r}] \\ &= [\mathrm{E}\{B^{q\times r}S_\gamma^F(Z) \mid C, G_C(Z)\} \text{ for all } B^{q\times r}].\end{aligned}$$

Obviously, the space of elements $B^{q\times r}S_\gamma^F(Z)$ is the definition of the full-data nuisance tangent space Λ^F, Hence, the desired result follows. $\square$

Orthogonal Complement

Lemma 3.4

The space $\Lambda_\eta^\perp$ consists of all elements $h^{q\times 1}\{C, G_C(Z)\} \in H$ such that

$$E[h\{C, G_C(Z)\} \mid Z] \in \Lambda^{F\perp},$$

where $\Lambda^{F\perp}$ is the space orthogonal to the full-data nuisance tangent space.

Proof of Lemma 3.4

Proof.

By the definition of orthogonal complement, $\Lambda_\eta^\perp$ is the set of elements $h(\cdot) \in H$ such that

$$\mathrm{E}[h^\intercal\{C, G_C(Z)\}\mathrm{E}\{\alpha^F(Z) \mid C, G_C(Z)\}] = 0,$$

for all $\alpha^F(Z) \in \Lambda^F$. Then, by tower expectation,

$$\begin{aligned} 0 &= \mathrm{E}[h^\intercal\{C, G_C(Z)\}\mathrm{E}\{\alpha^F(Z) \mid C, G_C(Z)\}] \\ &= \mathrm{E}[h^\intercal\{C, G_C(Z)\}\alpha^F(Z)] \\ &= \mathrm{E}[\mathrm{E}(h^\intercal\{C, G_C(Z)\} \mid Z)\alpha^F(Z)]. \end{aligned}$$

So $h\{C, G_C(Z)\} \in H$ belongs to $\Lambda_\eta^\perp$ if and only if $\mathrm{E}[h\{C < G_C(Z)\} \mid Z]$ is orthogonal to every element $\alpha^F(Z) \in \Lambda^F$. In other words,

$$\mathrm{E}[h\{C, G_C(Z)\} \mid Z] \in \Lambda^{F\perp},$$

which is the desired result. □

Mapping

Definition 3.5

Let $H^{(1)}$ and $H^{(2)}$ be two Hilbert spaces. Define the followings:

- (mapping) $K : H^{(1)} \to H^{(2)}$ means that for any $h \in H^{(1)}$, $K(h) \in H^{(2)}$
- (properties of linear mapping) K is a linear map if $K(ah_1 + bh_2) = aK(h_1) + bK(h_2)$ for any two elements $h_1, h_2 \in H^{(1)}$.

Definition 3.6 (Inverse operator)

- For any element $h^F \in H^F$, $K^{-1}(h^F)$ corresponds to the set of all elements (assuming at least one exists) $h \in H$ such that $K(h) = h^F$.
- Similarly, the space $K^{-1}(\Lambda^{F\perp})$ corresponds to all elements of $h \in H$ such that $K(h) \in \Lambda^{F\perp}$.

Mapping

In our context:

- Let $K : H \to H^F$ to be
$$K(h) = \mathrm{E}[h\{C, G_C(Z)\} \mid Z],$$
for $h \in H$. Note that K is a linear mapping because it is a conditional expectation.
- Recall that $\Lambda_\eta^\perp$ consists of elements $h \in H$ that satisfies $\mathrm{E}[h\{C, G_C(Z)\} \mid Z] \in \Lambda^{F\perp}$. Then we write
$$\Lambda_\eta^\perp = K^{-1}(\Lambda^{F\perp}).$$

Mapping of influence functions

Lemma 3.7

For any $\varphi^{\star F}(Z) \in \Lambda^{F\perp}$, *let* $K^{-1}(\psi^{\star F}(Z)\}$ *denote the space of elements* $\tilde{h}(C, G_C(Z)) \in H$ *such that*

$$K[\tilde{h}\{C, G_C(Z)\}] = \mathrm{E}[\tilde{h}\{C, G_C(Z)\} \mid Z] = \varphi^{\star F}(Z).$$

If we could identify any element h *such that*

$$K(h) = \varphi^{\star F}(Z),$$

then

$$K^{-1}\{\varphi^{\star F}(Z)\} = h\{C, G_C(Z)\} + \Lambda_2,$$

where Λ_2 *is the linear subspace of* H *consisting of elements* $L_2\{C, G_C(Z)\}$ *such that*

$$\mathrm{E}[L_2\{C, G_C(Z)\} \mid Z] = 0 \qquad \textit{i.e.} \quad \Lambda_2 = K^{-1}(0).$$

Mapping of influence functions

Proof.

If $\tilde{h}\{C, G_C(Z)\}$ is an element of the space $h\{C, G_C(Z)\} + \Lambda_2$, then

$$\tilde{h}\{C, G_C(Z)\} = h\{C, G_C(Z)\} + L_2\{C, G_C(Z)\}$$

for some $L_2 \in \Lambda_2$. Then we can show that

$$K(\tilde{h}) = \mathrm{E}[h\{C, G_C(Z)\} \mid Z] + 0 = \varphi^{\star F}(Z).$$

Conversely, if $\mathrm{E}[\tilde{h}\{C, G_C(Z)\} \mid Z] = \varphi^{\star F}(Z)$, then

$$\tilde{h}\{C, G_C(Z)\} = h\{C, G_C(Z)\} + [\tilde{h}\{C, G_C(Z)\} - h\{C, G_C(Z)\}],$$

where it is observed that the second term $[\tilde{h}\{C, G_C(Z)\} - h\{C, G_C(Z)\}] \in \Lambda_2$. $\square$

Mapping of influence functions

Therefore, in order to construct $\Lambda_F^{\perp} = K^{-1}(\Lambda^{F\perp})$, then for each $\varphi^{\star F}(Z) \in \Lambda^{F\perp}$, the strategy is as follows:

1. identify one element h such that

$$\mathrm{E}[h\{C, G_C(Z)\} \mid Z] = \varphi^{\star F}(Z),$$

2. find $\Lambda_2 = K^{-1}(0)$.

Formulating $\Lambda_\eta^\perp$

Here is the last theorem from me today.

Theorem 3.8

Under the assumption that

$$E\{\mathbb{1}(C=\infty) \mid Z\} = \varpi(\infty, Z) > 0 \quad \textit{for all } Z \textit{ (a.e.)},$$

the space $\Lambda_\eta^\perp$ consists of all elements that can be written as

$$\frac{\mathbb{1}(C=\infty)\varphi^{\star F}(Z)}{\varpi(\infty,Z)} + \frac{\mathbb{1}(C=\infty)}{\varpi(\infty,Z)}\left[\sum_{r\neq\infty}\varpi\{r,G_r(Z)\}L_{2r}\{G_r(Z)\}\right] - \sum_{r\neq\infty}\mathbb{1}(C=r)L_{2r}\{G_r(Z)\},$$

where, for $r \neq \infty$, $L_{2r}\{G_r(Z)\}$ is an arbitrary $q \times 1$ measurable function of $G_r(Z)$ and $\varphi^{\star F}(Z)$ is an arbitrary element of $\Lambda^{F\perp}$.

Formulating $\Lambda_\eta^\perp$

We will proceed in two steps as mentioned before.

(i) Identify an h such that $\mathrm{E}[h\{C, G_C(Z)\} \mid Z] = \varphi^{\star F}(Z)$.

This can be motivated by the idea of an inverse probability weighted complete-case estimator. Consider

$$h\{C, G_C(Z)\} = \frac{\mathbb{1}(C = \infty)\varphi^{\star F}(Z)}{\varpi(\infty, Z)},$$

notice that denominator is not zero a.e. by assumption, one can check that

$$\mathrm{E}\left\{\frac{\mathbb{1}(C = \infty)\varphi^{\star F}(Z)}{\varpi(\infty, Z)} \mid Z\right\} = \frac{\varphi^{\star F}(Z)}{\varpi(\infty, Z)}\mathrm{E}\{\mathbb{1}(C = \infty) \mid Z\} = \varphi^{\star F}(Z).$$

Therefore, we can write $\Lambda_\eta^\perp = K^{-1}(\Lambda^{F\perp})$ as the direct sum of two linear subspaces; namely

$$\Lambda_\eta^\perp = \frac{\mathbb{1}(C = \infty)\varphi^{\star F}(Z)}{\varpi(\infty, Z)} \oplus \Lambda_2.$$

Formulating $\Lambda_\eta^\perp$

(ii) Derive the linear space Λ_2.

Assumed the coarsening variable C is discrete, we can decompose $H\{C, G_C(Z)\}$ as

$$\mathbb{1}(C = \infty)h_\infty(Z) + \sum_{r \neq \infty} \mathbb{1}(C = r)h_r\{G_r(Z)\}$$

where h_∞ and h_r are arbitrary $q \times 1$ functions. For $L_2\{C, G_C(Z)\} \in \Lambda_2$, we can express it as

$$\mathrm{E}\left[\mathbb{1}(C = \infty)L_{2\infty}(Z) + \sum_{r \neq \infty} \mathbb{1}(C = r)L_{2r}\{G_r(Z)\} \mid Z\right] = 0.$$

or equivalently,

$$\varpi(\infty, Z)L_{2\infty}(Z) + \sum_{r \neq \infty} \varpi\{r, G_r(Z)\}L_{2r}\{G_r(Z)\} = 0.$$

Formulating $\Lambda_\eta^\perp$

Therefore, for any set of $L_{2r}\{G_r(Z)\}$, $r \neq \infty$, we can define a typical element of Λ_2 as

$$\mathbb{1}(C = \infty) \underbrace{\left[\frac{1}{\varpi(\infty, Z)} \sum_{r \neq \infty} \varpi\{r, G_r(Z)\} L_{2r}\{G_r(Z)\} \right]}_{=: L_{2\infty}} - \sum_{r \neq \infty} \mathbb{1}(C = r) L_{2r}\{G_r(Z)\}.$$

Hence, the desired result follows.

Formulating $\Lambda_\eta^\perp$

Finally, wecall that in the previous proof, we write

$$\Lambda_\eta^\perp = \frac{\mathbb{1}(C=\infty)\varphi^{\star F}(Z)}{\varpi(\infty, Z)} \oplus \Lambda_2,$$

for an influence function $\varphi^{\star F}$. We shall end this section by defining the two linear subspaces.

Definition 3.9

The linear subspace contained in H consisting of elements

$$\left\{ \frac{\mathbb{1}(C=\infty)\varphi^{\star F}(Z)}{\varpi(\infty, Z)};\ \text{for all } \varphi^{\star F} \in \Lambda^{F\perp} \right\},$$

also denoted as $\mathbb{1}(C=\infty)\Lambda^{F\perp}/\varpi(\infty, Z)$ will be fined to be the inverse probability weighted copmlete-case (IPWCC) space.

Definition 3.10

The linear space $\Lambda_2 \subset H$ will be defined to be the augmentation space.

Example: Restricted Moment Model with Missing Data by Design

Example

Consider the semiparametric restricted model moment model that assumes

$$E(Y \mid X) = \mu(X, \beta),$$

where Y is the response variable and X is a vector of covariates. Here, $Z = (Y, X)$ denotes full data.

We showed in (4.48) that a typical element of $\Lambda^{F\perp}$ is

$$A(X)\{Y - \mu(X, \beta_0)\}.$$

This motivates the GEE to be

$$\sum_{i=1}^{n} A(X_i)\{Y_i - \mu(X_i, \beta)\} = 0 \tag{7.38}$$

using a sample of data (Y_i, X_i), $i = 1, \ldots, n$.

Example

Suppose, by design, we **coarsen the data at random**.

For example, let X be partitioned as $X = (X^{(1)\intercal}, X^{(2)\intercal})^\intercal$.

- The full data are $Z_i = (Y_i, X_i^{(1)}, X_i^{(2)})$, $i = 1, \dots, n$.
- Y_i and $X_i^{(1)}$ are observed on everyone, whereas $X_i^{(2)}$ may be missing.
- The complete-case binary indicator R_i takes the value 1 with probability $\pi(Y_i, X_i^{(1)})$, where the function $0 < \pi(y, x^{(1)}) < 1$ is a known function chosen by the investigator.
- The observed data are $(R_i, Y_i, X_i^{(1)}, R_i X_i^{(2)})$.

Example

The space $\Lambda_\eta^\perp$ is derived in (7.32) of theorem 7.2,

$$\left\{ \frac{R\psi^{*F}(Z)}{\pi(Y, X^{(1)})} + L_2\{\mathscr{C}, G_{\mathscr{C}}(Z)\} \, ; \psi^{*F}(Z) \in \Lambda^{F\perp}, L_2\{\mathscr{C}, G_{\mathscr{C}}(Z)\} \in \Lambda_2 \right\}. \tag{7.39}$$

After some algebra, any element $L_2\{\mathscr{C}, G_{\mathscr{C}}(Z)\} \in \Lambda_2$ can be expressed as

$$\left\{ \frac{R - \pi(Y, X^{(1)})}{\pi(Y, X^{(1)})} \right\} L(Y, X^{(1)}). \tag{7.40}$$

Since a typical element of $\psi^{*F}(Z) \in \Lambda^{F\perp}$ for the restricted moment model is

$$A(X)\{Y - \mu(X, \beta_0)\},$$

for arbitrary $A(X)$, then by (7.39) and (7.40), **a typical element of $\Lambda_\eta^\perp$** is

$$\frac{R[A(X)\{Y - \mu(X, \beta_0)\}]}{\pi(Y, X^{(1)})} + \left\{ \frac{R - \pi(Y, X^{(1)})}{\pi(Y, X^{(1)})} \right\} L(Y, X^{(1)}),$$

for arbitrary $A(X)$ and $L(Y, X^{(1)})$.

Example

We have shown that identifying elements orthogonal to the nuisance tangent space and using these as estimating functions may guide us in constructing estimating equations whose solution would yield a consistent, asymptotically normal estimator for β.

For this problem, we consider estimating β with a sample of coarsened data

$$(R_i, Y_i, X_i^{(1)}, R_i X_i^{(2)}), \qquad i = 1, \dots, n,$$

by using the m-estimator that solves

$$\sum_{i=1}^{n} \left[\frac{R_i [A(X_i)\{Y_i - \mu(X_i, \beta)\}]}{\pi(Y_i, X_i^{(1)})} + \left\{ \frac{R_i - \pi(Y_i, X_i^{(1)})}{\pi(Y_i, X_i^{(1)})} \right\} L(Y_i, X_i^{(1)}) \right] = 0. \qquad (7.41)$$

Example

If this estimator is to be **consistent**, at the least

$$\mathrm{E}\left[\frac{R[A(X)\{Y-\mu(X,\beta_0)\}]}{\pi(Y,X^{(1)})}+\left\{\frac{R-\pi(Y,X^{(1)})}{\pi(Y,X^{(1)})}\right\}L(Y,X^{(1)})\right]=0.$$

Using the the law of iterated conditioning, where we first condition on Y, X, we obtain

$$\mathrm{E}\left[\frac{A(X)\{Y-\mu(X,\beta_0)\}}{\pi(Y,X^{(1)})}\mathrm{E}(R\mid Y,X)+\left\{\frac{\mathrm{E}(R\mid Y,X)-\pi(Y,X^{(1)})}{\pi(Y,X^{(1)})}\right\}L(Y,X^{(1)})\right]. \tag{7.42}$$

Since

$$\mathrm{E}(R\mid Y,X)=\mathrm{P}(R=1\mid Y,X)=\mathrm{P}(R=1\mid Y,X^{(1)},X^{(2)})=\pi(Y,X^{(1)}),$$

the (7.42) becomes

$$\mathrm{E}[A(X)\{Y-\mu(X,\beta_0)\}+0]=0. \tag{7.43}$$

Example

Also, the usual expansion of m-estimators can be used to derive **asymptotic normality**.

That is,

$$
\begin{aligned}
0 &= \sum_{i=1}^{n} \left[\frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i)\{Y_i - \mu(X_i, \hat{\beta}_n)\} + \left\{ \frac{R_i - \pi(Y_i, X_i^{(1)})}{\pi(Y_i, X_i^{(1)})} \right\} L(Y_i, X_i^{(1)}) \right] \\
&= \sum_{i=1}^{n} \left[\frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i)\{Y_i - \mu(X_i, \beta_0)\} + \left\{ \frac{R_i - \pi(Y_i, X_i^{(1)})}{\pi(Y_i, X_i^{(1)})} \right\} L(Y_i, X_i^{(1)}) \right] \\
&\quad - \left[\sum_{i=1}^{n} \frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i) D(X_i, \beta_n^*) \right] (\hat{\beta}_n - \beta_0),
\end{aligned}
$$

where $D(X_i, \beta) = \partial \mu(X, \beta) / \partial \beta^T$ and β_n^* is an intermediate value between $\hat{\beta}_n$ and β_0.

Example

Therefore,

$$n^{1/2}(\hat{\beta}_n - \beta_0) = \left[n^{-1} \sum_{i=1}^{n} \frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i) D(X_i, \beta_n^*) \right]^{-1}$$

$$\times\, n^{-1/2} \sum_{i=1}^{n} \left[\frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i)\{Y_i - \mu(X_i, \beta_0)\} \right.$$

$$\left. + \left\{ \frac{R_i - \pi(Y_i, X_i^{(1)})}{\pi(Y_i, X_i^{(1)})} \right\} L(Y_i, X_i^{(1)}) \right].$$

Under suitable regularity conditions,

$$n^{-1} \sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i) D(X_i, \beta_n^*) \right\} \xrightarrow{\text{pr}} \mathrm{E}\left\{ \frac{R}{\pi(Y, X^{(1)})} A(X) D(X, \beta_0) \right\}.$$

Using iterated conditioning, where first we condition on Y, X, we obtain

$$n^{-1} \sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i) D(X_i, \beta_n^*) \right\} \xrightarrow{\text{pr}} \mathrm{E}\{A(X) D(X, \beta_0)\}.$$

Example

Consequently,

$$n^{1/2}(\hat{\beta}_n - \beta_0) = n^{-1/2} \sum_{i=1}^{n} [\mathrm{E}\{A(X)D(X, \beta_0)\}]^{-1}$$
$$\times \left[\frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i)\{Y_i - \mu(X_i, \beta_0)\} + \left\{ \frac{R_i - \pi(Y_i, X_i^{(1)})}{\pi(Y_i, X_i^{(1)})} \right\} L(Y_i, X_i^{(1)}) \right]$$
$$+ o_p(1).$$

Therefore, **the i-th influence function for $\hat{\beta}_n$ is**

$$[\mathrm{E}\{A(X)D(X, \beta_0)\}]^{-1} \left[\frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i)\{Y_i - \mu(X_i, \beta_0)\} \right.$$
$$\left. + \left\{ \frac{R_i - \pi(Y_i, X_i^{(1)})}{\pi(Y_i, X_i^{(1)})} \right\} L(Y_i, X_i^{(1)}) \right],$$

which we demonstrated has mean zero, in (7.42) and (7.43).

Example

- The estimator for β, given as the solution to (7.41), is referred to as an augmented inverse probability weighted complete-case (AIPWCC) estimator.
- If $L(Y, X^{(1)})$ is chosen to be identically equal to zero, then the estimating equation in (7.41) becomes

$$\sum_{i=1}^{n} \frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i)\{Y_i - \mu(X_i, \beta)\} = 0. \tag{7.45}$$

  The solution to (7.45) is referred to as an inverse probability weighted complete-case (IPWCC) estimator.
- The arbitrary function $L(Y, X^{(1)})$ allows contributions from individuals with missing data into the estimating equation. Properly chosen augmentation will result in an estimator with greater efficiency.

The Logistic Regression Model

Let us consider the likelihood for the logistic regression model if we had missing data by design as below,

- Y is an observed, binary response variable.
- X can be partitioned as $X = (X^{(1)\intercal}, X^{(2)\intercal})^\intercal$, where $X^{(1)}$ is observed on everyone, whereas $X^{(2)}$ may be missing.
- $X^{(2)}$ could be observed with probability $\pi(Y_i, X_i^{(1)})$ by design.
- To allow for an intercept term in the model, we define $X^* = (1, X^{(1)\intercal}, X^{(2)\intercal})^\intercal$ and $X^{(1*)} = X^* = (1, X^{(1)\intercal})^\intercal$.

The density of the full data (Y, X) is

$$p_{Y,X}(y, x, \beta, \eta_1, \eta_2) = p_{Y|X}(y \mid x, \beta) p_{X^{(2)}|X^{(1)}}(x^{(2)} \mid x^{(1)}, \eta_1) p_{X^{(1)}}(x^{(1)}, \eta_2)$$

$$= \left[\frac{\exp\{(\beta_1^\intercal x^{(1*)} + \beta_2^\intercal x^{(2)})y\}}{1 + \exp(\beta_1^\intercal x^{(1*)} + \beta_2^\intercal x^{(2)})} \right] p_{X^{(2)}|X^{(1)}}(x^{(2)} \mid x^{(1)}, \eta_1) p_{X^{(1)}}(x^{(1)}, \eta_2),$$

where $\beta = (\beta_1^\intercal, \beta_2^\intercal)^\intercal$.

The Logistic Regression Model

The density of the observed data $(R, Y, X^{(1)}, RX^{(2)})$ is

$$
\begin{aligned}
&\{p_{Y|X}(y \mid x, \beta) p_{X^{(2)}|X^{(1)}}(x^{(2)} \mid x^{(1)}, \eta_1)\}^r \\
&\times \left\{ \int p_{Y|X}(y \mid x, \beta) p_{X^{(2)}|X^{(1)}}(x^{(2)} \mid x^{(1)}, \eta_1) d\nu_{X^{(2)}}(x^{(2)}) \right\}^{1-r} p_{X^{(1)}}(x^{(1)}, \eta_2) \\
=& \left[\frac{\exp\{(\beta_1^\intercal x^{(1*)} + \beta_2^\intercal x^{(2)}) y\}}{1 + \exp(\beta_1^\intercal x^{(1*)} + \beta_2^\intercal x^{(2)})} \right]^r \{p_{X^{(2)}|X^{(1)}}(x^{(2)} \mid x^{(1)}, \eta_1)\}^r \qquad (7.46) \\
&\times \left\{ \int \left[\frac{\exp\{(\beta_1^\intercal x^{(1*)} + \beta_2^\intercal x^{(2)}) y\}}{1 + \exp(\beta_1^\intercal x^{(1*)} + \beta_2^\intercal x^{(2)})} \right] p_{X^{(2)}|X^{(1)}}(x^{(2)} \mid x^{(1)}, \eta_1) d\nu_{X^{(2)}}(x^{(2)}) \right\}^{1-r} \qquad (7.47) \\
&\times p_{X^{(1)}}(x^{(1)}, \eta_2).
\end{aligned}
$$

Notice that in the likelihood formulation above, nowhere do the probabilities $\pi(Y, X^{(1)})$ come into play, even though they are known to us by design.

The Logistic Regression Model

Since the logistic regression model is just a simple example of a restricted moment model, estimators for the parameter can be found easily by solving the estimating equation (7.41), where

$$\mu(X_i, \beta) = \frac{\exp(\beta^\intercal X_i^*)}{1 + \exp(\beta^\intercal X_i^*)}.$$

With no missing data, we showed in (4.65) taht the optimal choice for $A(X)$ is X^*. Consequently, one easy way of obtaining an estimator for β is by solving (7.41) using $A(X) = X^*$ and $L(Y_i, X_i^{(1)}) = 0$, leading to

$$\sum_{i=1}^{n} \frac{R_i}{\pi(Y_i, X_i^{(1)})} X_i^* \left\{ Y_i - \frac{\exp(\beta^\intercal X_i^*)}{1 + \exp(\beta^\intercal X_i^*)} \right\} = 0. \tag{7.48}$$

- This estimator is an inverse probability weighted complete case (IPWCC) estimator for β.
- Although this estimator is a consistent, asymptotically normal semiparametric estimator for β, it is by no means efficient.

The Logistic Regression Model

It is intuitively clear that additional efficiency can be gained by using the data from individuals $i : Ri = 0$. Therefore, it would be preferable to use an AIPWCC estimator given by (7.41),

$$\sum_{i=1}^{n} \left[\frac{R_i}{\pi(Y_i, X_i^{(1)})} X_i^* \left\{ Y_i - \frac{\exp(\beta^\intercal X_i^*)}{1 + \exp(\beta^\intercal X_i^*)} \right\} + \left\{ \frac{R_i - \pi(Y_i, X_i^{(1)})}{\pi(Y_i, X_i^{(1)})} \right\} L(Y_i, X_i^{(1)}) \right] = 0, \tag{7.49}$$

with some properly chosen $L(Y, X^{(1)})$.

Recap and Review of Notation

Basic ideas and notations

Full data

- Full data Z has density from a semiparametric model $p_X(z, \beta, \eta)$, where β is q-dimensional parameter of interest and η infinite-dimensional nuisance parameter.
- $\mathscr{H}^F$ denotes the full-data Hilbert space defined as all mean-zero, q-dimensional measurable functions of Z with finite variance equipped with the covariance inner product.
- Λ^F is the full-data nuisance tangent space.
- $\Lambda^{F\perp}$ is the set of elements $\psi^{*F}(Z)$ that are orthogonal to Λ^F.

Basic ideas and notations

Observed (coaresened) data

- Coarsened data are denoted by $\{\mathscr{C}, G_{\mathscr{C}}(Z)\}$, where $\mathscr{C} \in \{1, \ldots, l, \infty\}$, and the $G_\infty(Z) = Z$ is reserved to denote complete data.
- Three types of coarsening mechanisms:
  (1) Coarsening completely at random (CCAR);
  (2) Coarsening at random (CAR);
  (3) Noncoarsening at random (NCAR).
- When coarsening is CAR, the coarsening probabilities are

$$P(\mathscr{C} = r \mid Z) = \varpi\{r, G_r(Z)\}.$$

- A key assumption is

$$P(\mathscr{C} = \infty \mid Z = z) = \varpi\{\infty, Z\} > \epsilon > 0, \qquad \forall z.$$

- H denotes the observed-data Hilbert space of q-dimensional, mean-zero, finite-variance, measurable functions of $\{\mathscr{C}, G_{\mathscr{C}}(Z)\}$.

Basic ideas and notations

Observed (coaresened) data

- A typical function $h\{\mathscr{C}, G_{\mathscr{C}}(Z)\}$ can be written as

$$h\{\mathscr{C}, G_{\mathscr{C}}(Z)\} = I(\mathscr{C} = \infty)h_{\infty}(Z) + \sum_{r \neq \infty} I(\mathscr{C} = r)h_r\{G_r(Z)\}.$$

- The observed-data nuisance tangent space

$$\Lambda = \Lambda_{\psi} \oplus \Lambda_{\eta}, \qquad \Lambda_{\psi} \perp \Lambda_{\eta}.$$

  Specifically,

$$\Lambda_{\eta} = \left\{ \mathrm{E}\{\alpha^F(Z) \mid \mathscr{C}, G_{\mathscr{C}}(Z)\}\,; \alpha^F(Z) \in \Lambda^F \right\} = \mathrm{E}\{\Lambda^F \mid \mathscr{C}, G_{\mathscr{C}}(Z)\}.$$

- In this chapter, we assume the coarsening is by design, therefore, the observed-data nuisance tangent space $\Lambda = \Lambda_{\eta}$.

Basic ideas and notations

Observed (coaresened) data

- Observed data estimating equations are motivated by considering elements in the space $\Lambda_\eta^\perp$, where

$$\Lambda_\eta^\perp = \left\{ \frac{I(\mathscr{C} = \infty)\Lambda^{F\perp}}{\varpi(\infty, Z)} \oplus \Lambda_2 \right\}$$

  and

$$\Lambda_2 = \Big\{ L_2\{\mathscr{C}, G_{\mathscr{C}}(Z)\} \,; \mathrm{E}[L_2\{\mathscr{C}, G_{\mathscr{C}}(Z)\} \mid Z] \Big\} = 0.$$

- The two linear spaces that make up $\Lambda_\eta^\perp$ are the IPWCC space $\frac{I(\mathscr{C}=\infty)\Lambda^{F\perp}}{\varpi(\infty, Z})$ and the augmentation space Λ_2.
- A typical element of Λ_2 $(r \neq \infty)$ is

$$\frac{I(\mathscr{C} = \infty)\Lambda^{F\perp}}{\varpi(\infty, Z)} \left[\sum_{r \neq \infty} \varpi\{r, G_r(Z)\} L_{2r}\{G_r(Z)\} \right] - \sum_{r \neq \infty} I(\mathscr{C} = r) L_{2r}\{G_r(Z)\},$$

  where $L_{2r}\{G_r(Z)\}$ is an arbitrary function.

Thank You!