General Semiparametric Estimating Equations for Missing Data

Semiparametric estimation equations and influence functions

Estimating equation with finite-dimension parameters

- Consider the p-dimension parameter of interest to be θ.
- A p-dimension estimating function(s) for θ, denoted by $m(O;\theta)$ (O denotes individual observations), satisfies

$$E_\theta[m(O;\theta)] = 0.$$

- An estimator based on n i.i.d observations $O_1, ..., O_n$ solves

$$\sum_{i=1}^{n} m(O_i;\theta) = 0.$$

Influence function

- We denote the estimator as $\hat{\theta}$.
- Heuristically, under some regularity conditions, we can show

$$\sqrt{n}(\hat{\theta} - \theta) = n^{-1/2} \sum_{i=1}^{n} \phi(O_i; \theta) + o_p(1),$$

where

$$\phi(O; \theta) = -E[\dot{m}(O; \theta)]^{-1} m(O; \theta).$$

- $\phi(O; \theta)$ is called the influence function associated with $\hat{\theta}$.

Properties of influence function

- First, $E[\phi(O;\theta)] = 0$.
- Second, since $E_\theta[m(O;\theta)] = \int m(O;\theta)f(O;\theta)d\nu(O) = 0$ where $f(O;\theta)$ is the density function, we differentiate both sides with respect to θ to obtain

$$E[\dot{m}(O;\theta)] + E[m(O;\theta)\dot{l}_\theta(O)^T] = 0$$

  where $\dot{l}_\theta$ is the score function.
- Therefore, we have

$$E[\phi(O;\theta)\dot{l}_\theta(O)^T] = I_{p\times p}.$$

Influence function for parameter of interest

- In a semiparametric model, θ is usually decomposed into the parameter of interest, β, and the nuisance parameter, η. That is, $\theta = (\beta, \eta)$ and $\hat{\theta} = (\hat{\beta}, \hat{\eta})$. Correspondingly, the estimating function $m(O;\theta) = (m_1(O;\theta), m_2(O;\theta))$ and

$$\phi(O;\theta) = (\phi_1(O;\theta), \phi_2(O;\theta)).$$

  The score function is decomposed as

$$\dot{l}_\theta(O) = (\dot{l}_\beta(O), \dot{l}_\eta(O)).$$

- $\phi_1(O;\theta)$ is the influence function for $\hat{\beta}$.

Important properties for the influence function

- The properties of $\phi(O;\theta)$ imply

$$E[\phi_1(O;\theta)] = 0,$$

$$E[\phi_1(O;\theta)\dot{l}_\beta(O)^T] = I_{q\times q},$$

$$E[\phi_1(O;\theta)\dot{l}_\eta(O)^T] = o_{q\times(p-q)}.$$

- On the other hand, given a function ϕ_1 satisfying the above equations, we can construct the following estimating equation for β

$$\sum_{i=1}^{n}\phi_1(O_i;\beta,\hat{\eta}(\beta)) = 0,$$

  where $\hat{\eta}(\beta) - \eta_0 = O_p(n^{-1/2})$ for $|\beta - \beta_0| = O_p(n^{-1/2})$.
- The solution to the estimating equation is asymptotically linear with influence function $\phi_1(O;\theta_0)$.

What are the key messages?

- To construct all possible estimating equations for β, we need to find all functions (influence functions), $\phi_1(O;\theta)$, such that
  1. It has mean zero and finite variance;
  2. $E[\phi_1(O;\theta)\dot{l}_\beta(O)] = I_{q\times q}$;
  3. $E[\phi_1(O;\theta)\dot{l}_\eta(O)] = o_{q\times(p-q)}$.
- Of course, to actually implement it, we also need to construct a reasonable estimator for η for any given β.

A general procedure to find all influence functions

- Find a specific influence function $\phi_1^*(O;\theta)$. Usually, this is achieved through constructing a simple estimating equation for β and then obtaining its influence function.
- Find all mean zero functions, $\psi(O;\theta)$, such that

$$E[\psi(O;\theta)\dot{l}_\beta(O)] = 0, \;\; E[\psi(O;\theta)\dot{l}_\eta(O)] = 0.$$

  Equivalently, $E[\psi(O;\theta)\dot{l}_\theta(O)] = 0$.
- Then all influence functions for β are given as $\phi_1^*(O;\theta) + \psi(O;\theta)$ for some $\psi(O;\theta)$.
- In conclusion, the keys are to identify ϕ_1^* and all mean zero functions that are orthogonal to any score functions.
- A note, in the language of Hilbert space, $E[g_1 g_2] = 0$ is called g_1 and g_2 are orthogonal.

Extension to infinite-dimensional nuisance parameter

- The same conclusion applies to the situation when η is infinite-dimensional.
- The major extension includes
  - extending the definition of the score function for η;
  - characterizing the space consisting of all the score functions.

Let's practise some examples

- Example 1:

$$E[Y|X] = \beta X.$$

- Example 2: estimating $\mu = E[Y]$ in a constrained model satisfyng

$$Var(Y) = c\mu^2.$$

Efficient influence function

- From the previous derivation, the estimator corresponding to an influence function, $\phi_1(O;\theta)$, has an asymptotic variance $Var(\phi_1(O;\theta))$.
- A natural question is which influence function gives the smallest variance. Such one is called the efficient influence function, denoted as $\phi_1^{eff}(O;\theta)$.
- That is, we aim to find $\psi(O;\theta)$, which is orthogonal to all score functions, such that $Var(\phi_1^* + \psi)$ is the least.
- Conclusion: if we decompose ϕ_1^* as the projection on the space spanned by all the score functions plus the projection on its orthocomplement, then $\phi_1^{eff}(O;\theta)$ is equal to the first projection. Furthermore, $\phi_1^{eff}(O;\theta)$ is unique and doesn't depend on the choice of ϕ_1^*.
- Let's calculate the efficient influence function for the two examples.

Semiparametric estimation equations with missing data

How to apply the previous results to missing data?

- The unique challenge for the missing data is to characterize the corresponding score function and its orthogonal complement, as well as the projection on the score space.
- Let O_F be the full data with density $f(O_F;\theta)$, and let C denote the missing pattern with $C \in \{1, ..., \infty\}$, where $C = \infty$ indicates no missing data. Corresponding to each missing pattern $C = r$, we let O_r denote the observed data.
- The observed likelihood function is

$$\prod_r \left\{ \int_{O_F:G(O_F,r)=O_r} f(O_F;\theta) P(C = r|O_F) d\nu(O_F) \right\}^{I(C=r)},$$

where $G(O_F, r)$ maps O_F into the component with missing pattern r.

MAR/CAR assumption

- Under MAR (coarsening at random) where we assume $P(C = r|O_F) = P(C = r|O_r)$, the observed likelihood function becomes

$$\prod_r \left\{ \int_{O_F : G(O_F, r) = O_r} f(O_F; \theta) d\nu(O_F) P(C = r|O_r) \right\}^{I(C=r)}.$$

- The score function for θ takes form

$$\sum_r I(C = r) E[S_F(O_F; \theta)|O_r],$$

where $S_F(O_F; \theta)$ is any score function for θ in the full data likelihood.

Task 1: Determine the orthogonal space of the score space

- The function of the observed data, $\sum_r I(C=r)g_r(O_r)$, satisfies

$$E\left\{(\sum_r I(C=r)g_r(O_r))(\sum_r I(C=r)E[S_F(O_F;\theta)|O_r])\right\}=0.$$

Thus,

$$\sum_r E\left\{P(C=r|O_r)g_r(O_r)E[S_F(O_F;\theta)|O_r]\right\}=0;$$

or equivalently,

$$\sum_r P(C=r|O_r)g_r(O_r)=\psi_F(O_F;\theta)$$

for some $\psi_F(O_F;\theta)$ in the orthogonal space of the score space for the full data.

- We obtain

$$g_\infty(O_F) = \frac{1}{P(C=\infty|O_F)} \left\{ \psi_F(O_F;\theta) - \sum_{r\neq\infty} P(C=r|O_r) g_r(O_r) \right\}.$$

- In other words, any function in the orthogonal space takes the form

$$\begin{aligned}
& \frac{I(C=\infty)}{P(C=\infty|O_F)} \left\{ \psi_F(O_F;\theta) - \sum_{r\neq\infty} P(C=r|O_r) g_r(O_r) \right\} \\
& + \sum_{r\neq\infty} I(C=r) g_r(O_r) \\
= \quad & \frac{I(C=\infty)}{P(C=\infty|O_F)} \psi_F(O_F;\theta) \\
& + \sum_{r\neq\infty} \frac{I(C=r)P(C=\infty|O_F) - I(C=\infty)P(C=r|O_r)}{P(C=\infty|O_F)} g_r(O_r).
\end{aligned}$$

Task 2: Projection on the score space

- For any mean zero function of the observed data

$$\sum_r I(C=r)g_r(O_r),$$

we aim to calculate its projection on the score space.

- Equivalently, we aim to find some S_F^* that is the score function in the full data space, such that

$$E\left\{ \left(\sum_r I(C=r)g_r(O_r) - \sum_r I(C=r)E[S_F^*(O_F;\theta)|O_r]\right) \right.$$

$$\left. \times \left(\sum_r I(C=r)E[S_F(O_F;\theta)|O_r]\right) \right\} = 0$$

for any S_F in the score space of the full data.

Calculation continued

- That is,

$$\sum_r E\left\{P(C=r|O_r)(g_r(O_r) - E[S_F^*(O_F;\theta)|O_r])\right.$$

$$\left.\times E[S_F(O_F;\theta)|O_r]\right\} = 0.$$

- This gives

$$\sum_r P(C=r|O_r)(g_r(O_r) - E[S_F^*(O_F;\theta)|O_r]) = \psi_F(O_F;\theta)$$

for some $\psi_F(O_F;\theta)$ in the orthogonal space of the score space for the full data.

Calculation continued

- Assume $P(C = \infty | O_F) > 0$.
- The equation becomes a self-consistency equation to solve for $S_F^*(O_F;\theta)$:

$$S_F^*(O_F;\theta) = \frac{g_\infty(O_F) - \psi_F(O_F;\theta)}{P(C=\infty|O_F)}$$

$$+ \sum_{r\neq\infty} \frac{P(C=r|O_r)}{P(C=\infty|O_F)}(g_r(O_r) - E[S_F^*(O_F;\theta)|O_r]).$$

- It is Fredholm integral equation of the second kind and the solution exits.
- We can obtain an explicit solution if the missing patterns are monotone!

Application to some specific examples

- Regression problem: $E[Y|X] = \beta X$ with Y subject to missingness
- Longitudinal data problem: $E[Y_k|X] = \beta X, k = 1, ..., K$ with monotone missingness on Y's
- Cox model with right-censored data