Semiparametric Theory and Missing Data

Chapter 3: The Geometry of Influence Functions

Martin, Ma Ting Tin

Department of Statistics, The Chinese University of Hong Kong

June 16, 2022

Agenda

Basic Notations and Definitions

Basic Notations (To be generalized)

- Parameter: $\theta = (\beta^T, \eta^T)^T \in \Omega$ (Parameter space), where $\beta \in \mathbb{R}^q$ is the **parameter of interest** and $\eta \in \mathbb{R}^r$ is the nuisance parameter.
- Truth parameter: $\theta_0 = (\beta_0^T, \eta_0^T)^T$.
- Data: $Z_1, \ldots, Z_n \overset{iid}{\sim} p_Z(z; \theta)$ for some $\theta \in \Omega$ w.r.t dominating measure v_Z.

Definition (Asymptotically linear estimator & Influence function)

An estimator $\widehat{\beta}_n = \widehat{\beta}_n(Z_1, \ldots, Z_n)$ of β is said to be **asymptotically linear** if $\exists$ measurable q-dimensional function $\varphi(Z)$ s.t.

1. $\mathbb{E}\{\varphi(Z)\} = 0_q$.
2. $\sqrt{n}(\widehat{\beta}_n - \beta_0) = \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} \varphi(Z_i) + o_p(1)$.
3. The matrix $\mathbb{E}(\varphi\varphi^T)$ is finite and non-singular.

The random vector $\varphi(Z_i)$ is said to be the ***i*-th influence function of $\widehat{\beta}_n$**

Example of Influence Function

Example 1 (Influence function for MLE of Normal Data)

Let $Z_1, \cdots, Z_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Recall the MLE of μ and σ^2 are $\widehat{\mu}_n = \overline{Z}_n$ and $\widehat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (Z_i - \widehat{\mu}_n)^2$ respectively. We proceed to find their influence function.

- $(\widehat{\boldsymbol{\mu}}_{\boldsymbol{n}})$: $$\sqrt{n}(\widehat{\mu}_n - \mu_0) = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \right) - \sqrt{n}\mu_0 = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \mu_0).$$

  Hence the influence function associated to μ is $\varphi_\mu(Z_i) = Z_i - \mu_0$.

- $(\widehat{\boldsymbol{\sigma}}_{\boldsymbol{n}}^{\boldsymbol{2}})$: $$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (Z_i - \mu_0 + \mu_0 - \widehat{\mu}_n)^2 = \frac{1}{n} \sum_{i=1}^n (Z_i - \mu_0)^2 + (\mu_0 - \widehat{\mu}_n)^2$$

  as cross-term $2n^{-1} \sum_{i=1}^n (Z_i - \mu_0)(\mu_0 - \widehat{\mu}_n) = 0$.

  $$(\Rightarrow) \quad \sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ (Z_i - \mu_0)^2 - \sigma_0^2 \right\} + \underbrace{\overbrace{\sqrt{n}(\widehat{\mu}_n - \mu_0)}^{\xrightarrow{\text{d}} N(0,\sigma^2)} \overbrace{(\widehat{\mu}_n - \mu_0)}^{\xrightarrow{\text{pr}} 0}}_{= o_p(1)}$$

  Hence the influence function associated to σ^2 is $\varphi_{\sigma^2}(Z_i) = (Z_i - \mu_0)^2 - \sigma_0^2$.

Uniqueness of Influence function

Theorem 3.1 (Uniqueness Theorem)

An asymptotically linear estimator $\widehat{\beta}_n$ of β has a unique influence function a.s.

Proof of Theorem 3.1

Suppose the contrary that the influence function is not unqiue a.s., i.e. $\exists \varphi, \varphi^*$ s.t.

1. $\mathbb{P}_{\theta_0}\{\varphi(Z) \neq \varphi^*(Z)\} > 0$ and $\mathbb{E}_{\theta_0}\{\varphi(Z)\} = \mathbb{E}_{\theta_0}\{\varphi^*(Z)\} = 0$
2. $\sqrt{n}(\widehat{\beta}_n - \beta_0) = n^{-1/2} \sum_{i=1}^n \varphi(Z_i) + o_p(1) = n^{-1/2} \sum_{i=1}^n \varphi^*(Z_i) + o_p(1)$.

From subtracting quantities in (2), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\varphi(Z_i) - \varphi^*(Z_i)] = o_p(1) \tag{3.1.1}$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\varphi(Z_i) - \varphi^*(Z_i)] \xrightarrow{\mathsf{d}} N\left(0, \mathbb{E}_{\theta_0}\left\{(\varphi - \varphi^*)(\varphi - \varphi^*)^T\right\}\right) \tag{3.1.2}$$

The only possibility for both (3.1.1) and (3.1.2) hold is to force $\varphi(Z) = \varphi^*(Z)$ a.s, which contradicts (1).

Importance of Influence function

We have just shown the uniqueness of influence function for a given asymptotically linear estimator $\widehat{\beta}_n$. Recall that by assumption and CLT, we can write

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(Z_i) + o_p(1) \xrightarrow{\text{d}} N\Big(0_q, \mathbb{E}_{\theta_0}\{\varphi\varphi^*\}\Big)$$

Takeaway

- The influence function φ fully characterise the asymptotic properties of $\widehat{\beta}_n$.
- Value of $\mathbb{E}_{\theta_0}\{\varphi\varphi^*\}$ can be used to compare performance of different asymptotically linear estimator of β.

Agenda

Definition of efficiency

Definition (Efficiency)

Let $\widehat{\theta}_n$ be estimator of θ. Suppose that $\lim_{n\to\infty} \text{Bias}(\widehat{\theta}_n) = 0$. Then $\widehat{\theta}_n$ is said to be

1. **asymptotically efficient** if $\lim_{n\to\infty} \mathbb{V}ar(\widehat{\theta}_n) = \text{CRLB}(\theta)$ for all $\theta \in \Theta$.
2. **super efficient** if
   (i). $\lim_{n\to\infty} \mathbb{V}ar_{\theta}(\widehat{\theta}_n) \le \text{CRLB}(\theta)$ for all $\theta \in \Theta$.
   (ii). $\lim_{n\to\infty} \mathbb{V}ar_{\theta_0}(\widehat{\theta}_n) < \text{CRLB}(\theta_0)$ for some $\theta_0 \in \Theta$.

Remark

Under regularity conditions, we have $\lim_{n\to\infty} \mathbb{V}ar_{\theta}(\widehat{\theta}_{\text{MLE}}) = \text{CRLB}(\theta)$ for all $\theta \in \Theta$, which means MLE is a typical example of asymptotically efficient estimator.

We then provide an example of super efficient estimator, which is known as the Hodges' estimator.

Hodges' Estimator

Remark on Notations

We have to carefully distinguish the DGP in handling technical details. Denote

- $\mathbb{E}_\theta\{\cdot\}$ as the expectation taken w.r.t the density $p_Z(;\theta)$
- $A_n \overset{\mathcal{D}(\mu)}{\to} B$ and $A_n \overset{P_\mu}{\to} B$ as "A_n converges in distribution and probability to B respectively while μ is parameter of DGP that generate A_n".

Definition (Hodges' Estimator)

Let $Z_1, \cdots, Z_n \overset{iid}{\sim} N(\mu, 1)$. Then the MLE of μ is given by $\overline{Z}_n$. Define the **Hodges' estimator** of μ as

$$\widehat{\mu}_n = \overline{Z}_n \mathbb{1}(|\overline{Z}_n| > n^{-1/4}).$$

In this case, CRLB$(\mu) = 1$ for all μ.

Efficiency of $\overline{Z}_n$

By CLT, $\sqrt{n}(\overline{Z}_n - \mu) \overset{\mathcal{D}(\mu)}{\to} N(0,1)$ and hence $\overline{Z}_n$ is asymptotically efficient.

Hodges' Estimator

Theorem (Efficiency of Hodges' Estimator)

Hodges' Estimator $\widehat{\mu}_n = \overline{Z}_n \mathbb{1}(|\overline{Z}_n| > n^{-1/4})$ is super-efficient.

Proof

- $(\boldsymbol{\mu \neq 0})$: As $|\overline{Z}_n| \overset{\text{pr}}{\to} |\mu| > 0$, $\sqrt{n}\mathbb{1}(|\overline{Z}_n| > n^{-1/4}) = \sqrt{n} + o_p(1)$ and $\sqrt{n}(\overline{Z}_n - \widehat{\mu}_n) = o_p(1)$
$$(\Rightarrow) \quad \overbrace{\sqrt{n}(\overline{Z}_n - \mu)}^{\overset{\mathcal{D}(\mu)}{\to} N(0,1)} = \sqrt{n}(\widehat{\mu}_n - \mu) + \overbrace{o_p(1)}^{\overset{P_\mu}{\to} 0}$$
implies that $\sqrt{n}(\widehat{\mu}_n - \mu) \overset{\mathcal{D}(\mu)}{\to} N(0,1)$.
- $(\boldsymbol{\mu = 0})$: By CLT, $\sqrt{n}\overline{Z}_n \overset{\mathcal{D}(0)}{\to} N(0,1)$, we have $n^{1/4}\overline{Z}_n \overset{P_0}{\to} 0$. It follows that
$$\mathbb{P}_0(\sqrt{n}\widehat{\mu}_n = 0) \geq \mathbb{P}_0(|\overline{Z}_n| < n^{-1/4}) = \mathbb{P}_0(n^{1/4}|\overline{Z}_n| < 1) \overset{P_\mu}{\to} 1$$
Therefore, $\lim_{n\to\infty} \sqrt{n}\widehat{\mu}_n \to 0$

Hence, $\lim_{n\to\infty} \mathbb{V}ar_\mu(\widehat{\mu}_n) =$ CRLB(μ) for $\mu \neq 0$ and $\lim_{n\to\infty} \mathbb{V}ar_0(\widehat{\mu}_n) <$ CRLB(0). So $\widehat{\mu}_n$ is super-efficient.

Trade-off of Super Efficiency

Although $\widehat{\mu}_n$ is super efficient, it shows poor performance in neighbourhood of zero. Consider

$$Z_{1n}, \cdots, Z_{nn} \overset{iid}{\sim} N(\mu_n, 1), \quad \text{where } \mu_n = n^{-1/3}$$

- For MLE $\overline{Z}_n$, we have $Z_{in} - \mu_n \overset{iid}{\sim} N(0,1)$ for all $i = 1, \cdots, n$. Hence by CLT,
$$\sqrt{n}(\overline{Z}_n - \mu_n) \overset{\mathcal{D}(\mu_n)}{\to} N(0,1)$$
- Recall that $\widehat{\mu}_n = \overline{Z}_n \mathbb{1}(|\overline{Z}_n| > n^{-1/4})$ From the last bullet, we know
$$\bar{Z}_n = \mu_n + n^{-1/2} O_p(1) = n^{-1/3} + n^{-1/2} O_p(1) = O_p(n^{-1/2})$$
and hence $\mathbb{P}_{\mu_n}(\sqrt{n}\widehat{\mu}_n = 0) \to 1$. It follows that
$$\mathbb{P}_{\mu_n}(\sqrt{n}(\widehat{\mu}_n - \mu_n) = -\sqrt{n}\mu_n) \to 1$$
while we have $-\sqrt{n}\mu_n = -n^{1/6} \to \infty$ and $\sqrt{n}(\widehat{\mu}_n - \mu_n)$ diverges to $-\infty$.

Comment

As the parameter μ_n get closer to 0, the Hodges' estimator give worse performance.

Agenda

Definitions

Definition (Local Data Generating Process - LDGP)

Consider a triangular array of random variable $\{Z_{1n}, \cdots, Z_{nn}\} \overset{iid}{\sim} p(z, \theta_n)$, where

$$\sqrt{n}(\theta_n - \theta_0) \to \tau,$$

for some $\theta_0 \in \Theta$ and $\tau \in \mathbb{R}^p$. The above data generating process is said to be LDGP.

Definition (Regular Estimator)

Assume $\{Z_{1n}, \cdots, Z_{nn}\} \overset{iid}{\sim} p(z, \theta_n)$. $\widehat{\beta}_n$ is **regular** if whenever $\{\theta_n\}$ satisfies $\sqrt{n}(\theta_n - \theta_0) \to \tau$ for some $\theta_0 \in \Theta$ and $\tau \in \mathbb{R}^p$,

$$\sqrt{n}(\widehat{\beta}_n - \beta_n) \overset{\mathcal{D}(\theta_n)}{\to} Q$$

for some distribution Q which is free of value of τ.

Remark

Intuitively, regularity is equivalent to require the limiting distribution Q does not change with small pertubation from the true parameter θ_0.

Implication of Regularity in our purpose

Theorem (Implication of Regularity)

Recall that $$\theta_n = (\beta_n^T, \eta_n^T)^T \quad \text{and} \quad \theta_0 = (\beta_0^T, \eta_0^T)^T$$

Assume $\{Z_{1n}, \cdots, Z_{nn}\} \stackrel{iid}{\sim} p(z, \theta_n)$. $\widehat{\beta}_n = \widehat{\beta}_n(Z_{1n}, \ldots, Z_{nn})$ is regular, where $\sqrt{n}(\theta_n - \theta_0) \to \tau$. Then we have

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \stackrel{\mathcal{D}(\theta_0)}{\to} N(0, \Sigma^*) \quad \Rightarrow \quad \sqrt{n}\left(\widehat{\beta}_n - \beta_n\right) \stackrel{\mathcal{D}(\theta_n)}{\to} N(0, \Sigma^*)$$

Proof

Let $\theta_n = \theta_0$ for all $n \in \mathbb{N}$. Then the regularity assumption implies that

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_n\right) \stackrel{\mathcal{D}(\theta_n)}{\to} Q$$

for some distribution Q. As $\theta_n = \theta_0$, we have $\beta_n = \beta_0$ and hence

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \stackrel{\mathcal{D}(\theta_0)}{\to} Q$$

By assumption, we have $Q = N(0, \Sigma^*)$ and hence the result follows.

Importance of Regularity

In the example concerning Hodges' estimator, $(Z_{1n}, \ldots, Z_{nn}) \overset{iid}{\sim} N(\mu_n, 1)$ and

- MLE $\overline{Z}_n$ is a regular estimator of μ as

$$\sqrt{n}(\overline{Z}_n - \mu_0) \overset{\mathcal{D}(\mu_0)}{\to} N(0,1) \quad \Rightarrow \quad \sqrt{n}(\overline{Z}_n - \mu_n) \overset{\mathcal{D}(\mu_n)}{\to} N(0,1)$$

  for arbitrary sequence of $\{\mu_n\}$ satisfying $\sqrt{n}(\mu_n - \mu_0) \to \tau$ for some constant $\tau \in \mathbb{R}$.

- Hodges' estimator $\widehat{\mu}_n$ is NOT a regular estimator of μ as for $\mu_0 = 0$ and $\mu_n = n^{-1/2}$,

$$\sqrt{n}(\widehat{\mu}_n - \mu_0) \overset{P_{\mu_0}}{\to} 0 \quad \text{but} \quad \sqrt{n}(\widehat{\mu}_n - \mu_n) \overset{P_{\mu_n}}{\to} -1 \neq 0,$$

Assumption (RAL)

The class of estimators to be considered are **regular** and **asymptotically linear**.

- Regularity prevents the undesirable phenomenon in previous example.
- (Hajek's Representation theorem): Most efficient regular estimator is AL.

Generalized Setting

Let $Z \sim p_Z(z, \theta)$, where $\theta = (\beta^T, \eta^T)^T$ and θ_0 is the true value of θ. Then the score function is defined as

$$S_\theta(Z, \theta_0) = \frac{\partial}{\partial \theta} \log p_Z(z, \theta)\bigg|_{\theta=\theta_0}$$

Also, we can partition the score function as $S_\theta(Z, \theta_0) = \{S_\beta^T(Z, \theta_0), S_\eta^T(Z, \theta_0)\}^T$.

However, sometimes we may not be able to partition the parameter θ as above. We then give a more general representation. Define the parameter of interest $\beta : \mathbb{R}^p \to \mathbb{R}^q$ as a smooth function of θ, i.e. $\beta = \beta(\theta)$.

- (Parametric model): $\exists \eta : \mathbb{R}^p \to \mathbb{R}^r$ being smooth s.t. $\{\beta(\theta), \eta(\theta)\}$ and θ could be linked through a bijection.
- (Semi-parametric/Non-parametric model): The new setup is sometimes a more natural generalization.

Geometry of Influence Function

Theorem 3.2

For parameter $\theta \in \mathbb{R}^p$, define $\beta(\theta) \in \mathbb{R}^q$ as the parameter of interest $(q < p)$. Suppose

1. $\Gamma^{q\times p}(\theta) := \partial\beta(\theta)/\partial\theta^T$ exists, is of rank q, and cts in neighbourhood of θ_0.
2. $\widehat{\beta_n}$ ia asymptotically linear with influence function $\varphi(Z)$, i.e. $\sqrt{n}(\widehat{\beta_n} - \beta_n) = n^{-1/2}\sum_{i=1}^n \varphi(Z_i) + o_p(1)$. Also $\mathbb{E}_\theta(\varphi^T\varphi)$ is continuous in neighbourhood of θ_0.
3. $\widehat{\beta_n}$ is a regular estimator.

Then

$$E_{\theta_0}\left\{\varphi(Z)S_\theta^T(Z,\theta_0)\right\} = \Gamma(\theta_0)$$

Corollary 1

If θ can be partitioned as $(\beta^T, \eta^T)^T$, then

$$E_{\theta_0}\left\{\varphi(Z)S_\beta^T(Z,\theta_0)\right\} = I^{q\times q} \quad \text{and} \quad E_{\theta_0}\left\{\varphi(Z)S_\beta^T(Z,\theta_0)\right\} = \mathcal{O}^{q\times r}$$

Agenda

Definition of m-estimator

Definition (m-Estimator)

Let $m : (Z, \theta) \to m(Z, \theta) \in \mathbb{R}^p$ be a function s.t. for all $\theta \in \Theta$, $\mathbb{E}_\theta\{m(Z, \theta)\} = 0^{p \times 1}$, $\mathbb{E}_\theta\{m^T(Z, \theta) m(Z, \theta)\}$ being finite and positive definite. Suppose for $Z_1, \ldots, Z_n \overset{iid}{\sim} p_Z(z, \theta)$, $\exists \widehat{\theta}_n = \widehat{\theta}_n(Z_1, \cdots, Z_n)$ s.t.

$$\sum_{i=1}^n m(Z_i, \widehat{\theta}_n) = 0,$$

then $\widehat{\theta}_n$ is said to be m-estimator of θ.

Example 2 (MLE as a m-estimator)

Recall that $\widehat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} \prod_{i=1}^n p_Z(Z_i, \theta) = \arg\max_{\theta \in \Theta} \sum_{i=1}^n \log p_Z(Z_i, \theta)$.

Under regularity conditions, it is equivalent to solve for $\widehat{\theta}_n$ s.t.

$$\sum_{i=1}^n S_\theta(Z_i, \widehat{\theta}_n) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log p_Z(Z_i, \theta) = 0$$

Recall that $\mathbb{E}_\theta\{S_\theta(Z, \theta)\} = 0$, so we can conclude that MLE is a m-estimator.

Consistency of m-estimator

Theorem (Consistency of m-estimator)

Let $Z_1, \cdots, Z_n \overset{iid}{\sim} p_Z(z, \theta)$. Let θ_0 be the true parameter value. Assume that

1. $\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial \theta^T} m(Z, \theta_0)\right\}$ is non-singular.
2. $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \theta^T} m(Z_i, \theta) \overset{P_{\theta_0}}{\to} \mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial \theta^T} m(Z, \theta)\right\}$ **uniformly** in θ in nbhd of θ_0.
3. m admits unique solution at nbhd of θ_0.

Then $\widehat{\theta}_n \overset{\text{pr}}{\to} \theta_0$. *Proof: (Omitted)*

Sufficient condition for (2)

Condition (2) would be satisfied if both of the following holds: Let $\mathcal{N}(\theta_0)$ be some neighbourhood of θ_0

- $\partial m(Z, \theta)/\partial \theta^T$ is continuous in θ in $\mathcal{N}(\theta_0)$ a.s.
- $\sup_{\theta \in \mathcal{N}(\theta_0)} \left| \frac{\partial}{\partial \theta^T} m(Z, \theta) \right| \leq g(Z)$ for some g s.t. $\mathbb{E}_{\theta_0}\{g(Z)\} < \infty$.

Influence function of m-estimator

Assuming the regularity condition in the previous page. By Taylor's theorem and definition of m-estimator, we have

$$0 = \sum_{i=1}^{n} m(Z_i, \widehat{\theta}_n) = \sum_{i=1}^{n} m(Z_i, \theta_0) + \left\{ \sum_{i=1}^{n} \frac{\partial}{\partial \theta^T} m(Z_i, \theta_n^*) \right\} (\widehat{\theta}_n - \theta_0)$$

for some θ_n^* lies between $\widehat{\theta}_n$ and θ_0. As $\widehat{\theta}_n \stackrel{\text{pr}}{\to} \theta_0$ and by RC(2) [locally uniform convergence],

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta^T} m(Z_i, \theta_n^*) \right\} \stackrel{P_{\theta_0}}{\to} \mathbb{E}_{\theta_0} \left\{ \frac{\partial}{\partial \theta^T} m(Z, \theta_0) \right\}.$$

It follows from RC1 [non-singularity] that the above convergences also holds when they are replaced by their inverse, i.e.

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta^T} m(Z_i, \theta_n^*) \right\}^{-1} \stackrel{P_{\theta_0}}{\to} \mathbb{E}_{\theta_0} \left\{ \frac{\partial}{\partial \theta^T} m(Z, \theta_0) \right\}^{-1}.$$

Influence function of m-estimator (Continued)

Therefore, we have

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_n - \theta_0) &= -\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta^T}m(Z_i,\theta_n^*)\right]^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}m(Z_i,\theta_0)\right\} \\
&= -\left[\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z_i,\theta_0)\right\}\right]^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{m(Z_i,\theta_0)}_{\mathbb{E}_{\theta_0}(\cdot)=0}\right\} + o_p(1)
\end{aligned}
$$

Therefore, the i-th influence function for $\widehat{\theta}_n$ is given by

$$
\varphi(Z_i) = -\left[\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z_i,\theta_0)\right\}\right]^{-1} m(Z_i,\theta_0).
$$

It follows that $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \mathbb{E}_{\theta_0}(\varphi\varphi^T)\right)$, where

$$
\mathbb{E}_{\theta_0}(\varphi\varphi^T) = \left[\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}\right]^{-1}\mathbb{V}ar_{\theta_0}\{m(Z,\theta_0)\}\left[\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}\right]^{-1^T}
$$

Influence function of MLE

Again, we take MLE as an example for illustration.

Example 3

Under regularity conditions, the fisher information is given by

$$I(\theta_0) = \mathbb{E}_{\theta_0}\{-S_{\theta\theta}(Z, \theta_0)\} = \mathbb{E}_{\theta_0}\{S_\theta(Z, \theta_0) S_\theta^T(Z, \theta_0)\}$$

Recall that for MLE, its m-function is $m(Z, \theta) = S_\theta(Z, \theta)$ and hence $\partial m(Z, \theta)/\partial \theta^T = S_{\theta\theta}(Z, \theta_0)$. Therefore, the influence function for MLE $\widehat{\theta}_n$ is given by

$$\varphi(Z_i) = I(\theta_0)^{-1} S_\theta(Z_i, \theta_0)$$

Hence the asymptotical variance of MLE is given by

$$\mathbb{E}_{\theta_0}\{\varphi\varphi^T\} = I(\theta_0)^{-1}\mathbb{E}_{\theta_0}\{S_\theta(Z, \theta_0) S_\theta^T(Z, \theta_0)\}\left(I(\theta_0)^{-1}\right)^T = \left(I(\theta_0)^{-1}\right)^T = I(\theta_0)^{-1}$$

Estimation of Asymptotic Variance

Recall that

$$\mathbb{E}_{\theta_0}(\varphi\varphi^T) = \left[\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}\right]^{-1} \mathbb{V}ar_{\theta_0}\left\{m(Z,\theta_0)\right\} \left[\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}\right]^{-1^T}$$

It suffices to estimate $\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}$ and $\mathbb{V}ar_{\theta_0}\left\{m(Z,\theta_0)\right\}$.

Case 1: θ_0 is known

We estimate the mentioned quantities by

- $\displaystyle \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta^T}m(Z_i,\theta_0) \xrightarrow{\text{pr}} \mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}$
- $\displaystyle \frac{1}{n}\sum_{i=1}^{n}m(Z_i,\theta_0)m^T(Z_i,\theta_0) \xrightarrow{\text{pr}} \mathbb{V}ar_{\theta_0}\{m(Z,\theta_0)\}$

Case 2: θ_0 is unknown

Under RCs, $\widehat{\theta}_n \xrightarrow{\text{pr}} \theta_0$ and it suffices to replace θ_0 by $\widehat{\theta}_n$ in estimators in Case 1.

Proof of Theorem 3.2 (Particular Case)

We are going to prove theorem 3.2 for the class of m-estimator. Recall the setting $\theta = (\beta^T, \eta^T)^T$ with estimator $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{\eta}_n^T)^T$. It suffices to prove

Corollary 1

If θ can be partitioned as $(\beta^T, \eta^T)^T$, then

$$E_{\theta_0}\left\{\varphi(Z)S_\beta^T(Z,\theta_0)\right\} = I^{q\times q} \quad \text{and} \quad E_{\theta_0}\left\{\varphi(Z)S_\beta^T(Z,\theta_0)\right\} = \mathcal{O}^{q\times r}$$

By definition, as for all $\theta \in \Theta$,

$$0^{p\times 1} = \mathbb{E}_\theta\{m(Z,\theta)\} = \int m(z,\theta)p_Z(z,\theta)dv(z).$$

It follows that

$$0 = \frac{\partial}{\partial\theta^T}\int m(z,\theta)p(z,\theta)dv(z)$$

Under suitable RCs, the order of differentiation and integration can be swapped, i.e. we have

$$\int\left(\frac{\partial}{\partial\theta^T}m(z,\theta)\right)p(z,\theta)dv(z) + \int m(z,\theta)\overbrace{\left(\frac{\frac{\partial}{\partial\theta^T}p(z,\theta)}{p(z,\theta)}\right)}^{=\frac{\partial}{\partial\theta^T}\log p(z,\theta)=S_\theta^T(z,\theta)}p(z,\theta)dv(z) = 0$$

Evaluate the last identity at $\theta = \theta_0$ gives

$$\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\} + \mathbb{E}_{\theta_0}\left\{m(Z,\theta_0)S_\theta^T(Z,\theta_0)\right\} = 0$$

and by rearranging it, we have

$$I^{p\times p} = -\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}^{-1}\mathbb{E}_{\theta_0}\left\{m(Z,\theta_0)S_\theta^T(Z,\theta_0)\right\}$$

Recall that influence function of $\widehat{\theta}_n$ is given by

$$\varphi_{\widehat{\theta}_n}(Z_i) = -\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}^{-1} m(Z,\theta_0)$$

and hence

$$\mathbb{E}_{\theta_0}\left\{\varphi_{\widehat{\theta}_n}(Z_i)S_\theta^T(Z_i,\theta_0)\right\} = -\mathbb{E}_{\theta_0}\left\{\frac{\partial}{\partial\theta^T}m(Z,\theta_0)\right\}^{-1}\mathbb{E}_{\theta_0}\left\{m(Z,\theta_0)S_\theta^T(Z,\theta_0)\right\} = I^{p\times p}$$

and we could partition $\varphi_{\widehat{\theta}_n}(Z_i) = \{\varphi^T_{\widehat{\beta}_n}(Z_i), \varphi^T_{\widehat{\eta}_n}(Z_i)\}^T$. Therefore,

$$\mathbb{E}_{\theta_0}\begin{bmatrix} \varphi_{\widehat{\beta}_n}(Z_i)S_\beta^T(Z_i,\theta_0) & \varphi_{\widehat{\beta}_n}(Z_i)S_\eta^T(Z_i,\theta_0) \\ \varphi_{\widehat{\eta}_n}(Z_i)S_\beta^T(Z_i,\theta_0) & \varphi_{\widehat{\eta}_n}(Z_i)S_\eta^T(Z_i,\theta_0) \end{bmatrix} = I^{q\times q}.$$

and the result follows immediately.

Agenda

Contiguity

Definition (Contiguity)

- Let $\{V_n\}_{n=1}^{\infty}$ be a sequence of random vectors.
- Let P_{1n} and P_{0n} be sequences of probability measure with densities $p_{1n}(v_n)$ and $p_{0n}(v_n)$ respectively.

Then $\{P_{1n}\}$ is **contiguous** to $\{P_{0n}\}$ if

$$\forall A_n \in \sigma(V_n), \quad "P_{0n}(A_n) \to 0" \quad \Rightarrow \quad "P_{1n}(A_n) \to 0"$$

Remark: It generalize the concept of absolute continuity between measures.

Importance of Contiguity

Setting

In our setting , then random vector $V_n = (Z_{1n}, \ldots, Z_{nn})$ and

$$p_{0n}(v_n) = \prod_{i=1}^{n} p(Z_{in}, \theta_0) \quad \text{and} \quad p_{1n}(v_n) = \prod_{i=1}^{n} p(Z_{in}, \theta_n),$$

where $\sqrt{n}(\theta_n - \theta_0) \to \tau \in \mathbb{R}^p$ for some constant τ.

Implication of Contiguity

If (to be shown) P_{1n} is contiguous to P_{0n}, then for any sequence of function $\{T_n\}$,

$$\forall \epsilon > 0, \quad P_{0n}(|T_n(V_n)| > \epsilon) \to 0 \quad \Rightarrow \quad \forall \epsilon > 0, \quad P_{1n}(|T_n(V_n)| > \epsilon) \to 0.$$

Therefore, if $T_n(V_n) \overset{P_{0n}}{\to} 0$, we have $T_n(V_n) \overset{P_{1n}}{\to} 0$. While it is sometimes relatively easier to show the convergence under P_{0n}.

LeCam's Lemma

If

$$\log\left\{\frac{p_{1n}(v_n)}{p_{0n}(v_n)}\right\} \overset{\mathcal{D}(P_{0n})}{\to} N\left(-\frac{\sigma^2}{2}, \sigma^2\right)$$

for some $\sigma^2 > 0$, then $\{P_{1n}\}$ is contiguous to $\{P_{0n}\}$. (*Proof: Omitted.*)

Recall of Theorem 3.2

Theorem 3.2

For parameter $\theta \in \mathbb{R}^p$, define $\beta(\theta) \in \mathbb{R}^q$ as the parameter of interest $(q < p)$. Suppose

1. $\Gamma^{q \times p}(\theta) := \partial\beta(\theta)/\partial\theta^T$ exists, is of rank q, and cts in neighbourhood of θ_0.
2. $\widehat{\beta_n}$ ia asymptotically linear with influence function $\varphi(Z)$, i.e. $\sqrt{n}(\widehat{\beta}_n - \beta_n) = n^{-1/2} \sum_{i=1}^n \varphi(Z_i) + o_p(1)$. Also $\mathbb{E}_\theta(\varphi^T \varphi)$ is continuous in neighbourhood of θ_0.
3. $\widehat{\beta}_n$ is a regular estimator.

Then

$$E_{\theta_0}\left\{\varphi(Z) S_\theta^T(Z, \theta_0)\right\} = \Gamma(\theta_0)$$

In order to show Theorem 3.2, we first proceed to show that P_{1n} is contiguous to P_{0n} through LeCam's Lemma.

Proof of Theorem 3.2

Define $L_n(V_n) = \log\{p_{1n}(V_n)/p_{0n}(V_n)\}$. $\exists \theta_n^*$ between θ_0 and θ_n s.t.

$$L_n(V_n) = \log\left\{\prod_{i=1}^n \frac{p(Z_{in},\theta_n)}{p(Z_{in},\theta_0)}\right\} = \sum_{i=1}^n \{\log p(Z_{in},\theta_n) - \log p(Z_{in},\theta_0)\}$$

$$=(\theta_n-\theta_0)^T\sum_{i=1}^n S_\theta(Z_{in},\theta_0) + \frac{1}{2}(\theta_n-\theta_0)^T\left(\sum_{i=1}^n S_{\theta\theta}(Z_{in},\theta_n^*)\right)(\theta_n-\theta_0)$$

$$=\overbrace{\sqrt{n}(\theta_n-\theta_0)^T}^{\to\tau^T}\overbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n S_\theta(Z_{in},\theta_0)}^{\overset{\mathcal{D}(P_{0n})}{\to}N(0,I(\theta))}$$

$$+\frac{1}{2}\underbrace{\sqrt{n}(\theta_n-\theta_0)^T}_{\to\tau^T}\underbrace{\left[\frac{1}{n}\sum_{i=1}^n S_{\theta\theta}(Z_{in},\theta_n^*)\right]}_{(\text{RC2 }\&\ \widehat{\theta}_n\overset{\text{pr}}{\to}\theta_0)\approx n^{-1}\sum_{i=1}^n S_{\theta\theta}(Z_{in},\theta_0)\overset{P_{0n}}{\to}-I(\theta_0)}\underbrace{\sqrt{n}(\theta_n-\theta_0)}_{\to\tau}$$

$$\overset{\mathcal{D}(P_{0n})}{\to}N\left(-\frac{\tau^T I(\theta_0)\tau}{2},\tau^T I(\theta_0)\tau\right)$$

It follows from LeCam's Lemma that P_{1n} is contiguous to P_{0n}.

Proof of Theorem 3.2 (Continued)

Before proceeding, consider the following quantity: $\exists \theta_n^*$ between θ_n and θ_0

$$\sqrt{n}\mathbb{E}_{\theta_n}\{\varphi(Z)\} = \sqrt{n}\int \varphi(z)p(z,\theta_n)dv(z)$$

$$= \underbrace{\sqrt{n}\int \varphi(z)p(z,\theta_0)dv(z)}_{=0(\text{By definition})} + \underbrace{\sqrt{n}\int \varphi(z)\left\{\frac{\partial}{\partial\theta}p(z,\theta_n^*)\right\}^T(\theta_n-\theta_0)dv(z)}_{\theta_n^*\to\theta_0 \text{ and } \sqrt{n}(\theta_n-\theta_0)=\tau}$$

$$\to \tau\int \varphi(z)\left\{\frac{\partial}{\partial\theta}p(z,\theta_0)\Big/p(z,\theta_0)\right\}^T p(z,\theta_0)dv(z)$$

$$= \tau\mathbb{E}_{\theta_0}\left\{\varphi(Z)S_\theta^T(Z,\theta_0)\right\}$$

By assumption, $\widehat{\beta}_n$ is asymptotically linear and regular, i.e.

$$\sqrt{n}(\widehat{\beta}_n - \beta(\theta_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varphi(Z_{in}) + o_{P_{0n}}(1),$$

where $o_{P_{in}}(1)$ is a sequence of random vectors that converges to 0 in probability with respect to P_{in} for $i = 0, 1$. As P_{1n} is contiguous to the sequence P_{0n}, we have

$$\sqrt{n}(\widehat{\beta}_n - \beta(\theta_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varphi(Z_{in}) + o_{P_{1n}}(1).$$

Proof of Theorem 3.2 (Continued)

Hence, we have

$$
\begin{aligned}
&\overbrace{\sqrt{n}(\widehat{\beta}_n - \beta(\theta_n))}^{\text{(Regularity)}: \overset{\mathcal{D}(P_{1n})}{\rightarrow} N(0,\mathbb{E}_{\theta_0}\{\varphi\varphi^T\})} \\
=&\sqrt{n}(\widehat{\beta}_n - \beta(\theta_0)) + \sqrt{n}(\beta(\theta_0) - \beta(\theta_n)) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(Z_{in}) + o_{P_{0n}}(1) - \sqrt{n}(\beta(\theta_n) - \beta(\theta_0)) \\
=&\overbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\varphi(Z_{in}) - \mathbb{E}_{\theta_n}\{\varphi(Z)\}\right]}^{\sim[0,\mathbb{E}_{\theta_n}(\varphi\varphi^T)-\mathbb{E}_{\theta_n}(\varphi)\mathbb{E}_{\theta_n}(\varphi^T)]\overset{\mathcal{D}(P_{1n})}{\rightarrow} N(0,\mathbb{E}_{\theta_0}\{\varphi\varphi^T\})} + \overbrace{\sqrt{n}\mathbb{E}_{\theta_n}\{\varphi(Z)\}}^{\rightarrow\tau\mathbb{E}_{\theta_0}\left\{\varphi(Z)S_\theta^T(Z,\theta_0)\right\}} \\
&- \underbrace{\sqrt{n}\Big(\beta(\theta_n) - \beta(\theta_0)\Big)}_{\approx\sqrt{n}\Gamma(\theta_0)(\theta_n-\theta_0)\rightarrow\tau\Gamma(\theta_0)} + o_{P_{0n}}(1)
\end{aligned}
$$

It follows that

$$\tau\mathbb{E}_{\theta_0}\left\{\varphi(Z)S_\theta^T(Z,\theta_0)\right\} = \tau\Gamma(\theta_0)$$

as the choice of τ is arbitrary due to regular assumption. The result follows.

Agenda

Remark

Theorem (Additional)

Suppose that

1. $\widehat{\theta}_n \xrightarrow{\text{pr}} \theta_0$.
2. $\sup\limits_{\theta \in \mathcal{N}(\theta_0)} |\widehat{T}_n(\theta) - T(\theta)| \xrightarrow{\text{pr}} 0$ for some neighbourhood $\mathcal{N}(\theta_0)$ of θ_0.
3. $T = T(\theta)$ is continuous in $\mathcal{N}(\theta_0)$.

Then $\widehat{T}_n(\widehat{\theta}_n) \xrightarrow{\text{pr}} T(\theta)$.

Proof

Noticing that

$$|\widehat{T}_n(\widehat{\theta}_n) - T(\theta)| \le \overbrace{|\widehat{T}_n(\widehat{\theta}_n) - T(\widehat{\theta}_n)|}^{\text{(A)}} + \overbrace{|T(\widehat{\theta}_n) - T(\theta)|}^{\text{(B)}}$$

- (A): As $\widehat{\theta}_n \xrightarrow{\text{pr}} \theta_0$, $\mathbb{P}(\widehat{\theta}_n \in \mathcal{N}(\theta_0)) \to 1$. Therefore, for sufficiently large n

$$\text{(A)} = |\widehat{T}_n(\widehat{\theta}_n) - T(\widehat{\theta}_n)| \le \sup_{\theta \in \mathcal{N}(\theta_0)} |\widehat{T}_n(\theta) - T(\theta)| \xrightarrow{\text{pr}} 0$$

- (B): As T is continuous and $\widehat{\theta}_n \xrightarrow{\text{pr}} \theta_0$, (B) $\xrightarrow{\text{pr}} 0$.

The END

Thank You!