Lecture 12: L-estimators and M-estimators

L-functional and L-estimator

For a function $J(t)$ on [0,1], define the L-functional as

$$T(G) = \int xJ(G(x))dG(x), \quad G \in \mathscr{F}.$$

If $X_1,...,X_n$ are i.i.d. from F and $T(F)$ is the parameter of interest, $T(F_n)$ is called an L-estimator of $T(F)$.
$T(F_n)$ is a linear function of order statistics:

$$T(F_n) = \int xJ(F_n(x))dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} J\left(\tfrac{i}{n}\right) X_{(i)},$$

since $F_n(X_{(i)}) = i/n$, $i = 1,...,n$.

Examples

- When $J(t) \equiv 1$, $T(F_n) = \bar{X}$, the sample mean.
- When $J(t) = (1-2\alpha)^{-1} I_{(\alpha,1-\alpha)}(t)$, $T(F_n) = \bar{X}_\alpha$ is the α-trimmed sample mean.

Although the sample median is also a linear function of order statistics, it is not of the form $T(F_n)$ with an L-functional T

Asymptotic normality of L-estimators

To establish the asymptotic normality for L-estimators $T(F_n)$, we follow the following steps.

Step 1. For $x \in \mathscr{R}$, calculate

$$\phi_F(x) = \lim_{t \to 0} \frac{T(F + t(\delta_x - F)) - T(F)}{t}$$

(if it exists), where δ_x is the point mass at x.
The function ϕ_F is called the influence function of T at F.
The influence function is an important tool in the study of robuestness of estimators

Also, verify that

$$E[\phi_F(X_1)] = \int \phi_F(x) dF(x) = 0$$

Step 2. Verify that $E[\phi_F(X_1)]^2 < \infty$ and obtain

$$\sigma_F^2 = E[\phi_F(X_1)]^2 = \int [\phi_F(x)]^2 dF(x).$$

Step 3. Verify that

$$T(F_n) - T(F) = \frac{1}{n}\sum_{i=1}^{n} \phi_F(X_i) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

This holds when T is differentiable in some sense (§5.2.1).
Then

$$\sqrt{n}[T(F_n) - T(F)] \to_d N(0, \sigma_F^2).$$

Step 3 is the most difficult part.

This approach can also be applied to other functionals (§5.2).

We now apply this approach to show the asymptotic normality of the trimmed sample mean.

Step 1: Derivation of the influence function ϕ_F

$$T(G) = \int xJ(G(x))dG(x), \quad G \in \mathscr{F}$$

For F and G in $\mathscr{F}$,

$$\begin{aligned}
T(G) - T(F) &= \int xJ(G(x))dG(x) - \int xJ(F(x))dF(x) \\
&= \int_0^1 [G^{-1}(t) - F^{-1}(t)]J(t)dt \\
&= \int_0^1 \int_{F^{-1}(t)}^{G^{-1}(t)} dxJ(t)dt \\
&= \int_{-\infty}^{\infty} \int_{G(x)}^{F(x)} J(t)dtdx \\
&= \int_{-\infty}^{\infty} [F(x) - G(x)]J(F(x))dx \\
&\quad - \int_{-\infty}^{\infty} U_G(x)[G(x) - F(x)]J(F(x))dx,
\end{aligned}$$

Step 1: Derivation of the influence function ϕ_F

where

$$U_G(x) = \begin{cases} \frac{\int_{F(x)}^{G(x)} J(t)dt}{[G(x)-F(x)]J(F(x))} - 1 & G(x) \neq F(x), J(F(x)) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

and the fourth equality follows from Fubini's theorem and the fact that the region in $\mathscr{R}^2$ between curves $F(x)$ and $G(x)$ is the same as the region in $\mathscr{R}^2$ between curves $G^{-1}(t)$ and $F^{-1}(t)$.

Let $G = F + t(\delta_x - F)$, where δ_x is the degenerated distribution at x.

Since $\lim_{t\to 0} U_{F+t(\delta_x - F)}(y) = 0$, by the dominated convergence theorem,

$$\lim_{t\to 0} \int_{-\infty}^{\infty} U_{F+t(\delta_x-F)}(y)[\delta_x(y) - F(y)]J(F(y))dy = 0.$$

Hence

$$\lim_{t\to 0} \frac{T(F+t(\delta_x - F)) - T(F)}{t} = -\int_{-\infty}^{\infty} [\delta_x(y) - F(y)]J(F(y))dy,$$

which is $\phi_F(x)$, the influence function of T.

By Fubini's theorem and the fact that $\int \delta_x(y) dF(x) = F(y)$,

$$\int \phi_F(x) dF(x) = -\int_{-\infty}^{\infty} \left[\int (\delta_x - F)(y) dF(x) \right] J(F(y)) dy = 0,$$

Consider now $J(t) = (\beta - \alpha)^{-1} I_{(\alpha,\beta)}(t)$,

$$\phi_F(x) = -\frac{1}{\beta - \alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} [\delta_x(y) - F(y)] dy.$$

Assume that F is continuous at $F^{-1}(\alpha)$ and $F^{-1}(\beta)$.
$F(F^{-1}(\alpha)) = \alpha$ and $F(F^{-1}(\beta)) = \beta$.
When $x < F^{-1}(\alpha)$,

$$\begin{aligned}
\phi_F(x) &= -\frac{1}{\beta - \alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} [1 - F(y)] dy \\
&= -\frac{y[1 - F(y)]}{\beta - \alpha} \Bigg|_{F^{-1}(\alpha)}^{F^{-1}(\beta)} - \frac{1}{\beta - \alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} y dF(y) \\
&= \frac{F^{-1}(\alpha)(1 - \alpha) - F^{-1}(\beta)(1 - \beta)}{\beta - \alpha} - T(F)
\end{aligned}$$

Similarly, when $x > F^{-1}(\beta)$,

$$\begin{aligned}\phi_F(x) &= \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{F^{-1}(\beta)} F(y)]dy \\ &= \frac{F^{-1}(\beta)\beta - F^{-1}(\alpha)\alpha}{\beta-\alpha} - T(F).\end{aligned}$$

Finally, when $F^{-1}(\alpha) \le x \le F^{-1}(\beta)$,

$$\begin{aligned}\phi_F(x) &= \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{x} F(y)dy - \frac{1}{\beta-\alpha}\int_{x}^{F^{-1}(\beta)} [1-F(y)]dy \\ &= \frac{yF(y)}{\beta-\alpha}\bigg|_{F^{-1}(\alpha)}^{x} - \frac{1}{\beta-\alpha}\int_{F^{-1}(\alpha)}^{x} ydF(y) \\ &\quad + \frac{y[1-F(y)]}{\beta-\alpha}\bigg|_{x}^{F^{-1}(\beta)} - \frac{1}{\beta-\alpha}\int_{x}^{F^{-1}(\beta)} ydF(y) \\ &= \frac{x - F^{-1}(\alpha)\alpha - F^{-1}(\beta)(1-\beta)}{\beta-\alpha} - T(F).\end{aligned}$$

Step 1: Derivation of the influence function ϕ_F

Hence,

$$\phi_F(x) = \begin{cases} \frac{F^{-1}(\alpha)(1-\alpha)-F^{-1}(\beta)(1-\beta)}{\beta-\alpha} - T(F) & x < F^{-1}(\alpha) \\ \frac{x-F^{-1}(\alpha)\alpha-F^{-1}(\beta)(1-\beta)}{\beta-\alpha} - T(F) & F^{-1}(\alpha) \le x \le F^{-1}(\beta) \\ \frac{F^{-1}(\beta)\beta-F^{-1}(\alpha)\alpha}{\beta-\alpha} - T(F) & x > F^{-1}(\beta). \end{cases}$$

If F is symmetric about θ, J is symmetric about $\frac{1}{2}$ $(J(t) = J(1-t))$, and $\int_0^1 J(t)dt = 1$, then $F(x) = F_0(x-\theta)$, where F_0 is a c.d.f. that is symmetric about 0, i.e., $F_0(x) = 1 - F_0(-x)$, and

$$\begin{aligned} \int xJ(F_0(x))dF_0(x) &= \int xJ(1-F_0(-x))dF_0(x) \\ &= \int xJ(F_0(-x))dF_0(x) \\ &= -\int yJ(F_0(y))dF_0(y), \end{aligned}$$

i.e., $\int xJ(F_0(x))dF_0(x) = 0$.

Step 1: Derivation of the influence function ϕ_F

Hence,

$$\begin{aligned} T(F) &= \int xJ(F(x))dF(x) \\ &= \theta \int J(F(x))dF(x) + \int (x-\theta)J(F_0(x-\theta))dF_0(x-\theta) \\ &= \theta \int_0^1 J(t)dt + \int yJ(F_0(y))dF_0(y) \\ &= \theta. \end{aligned}$$

Assume that F is continuous at $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$.
When $\beta = 1-\alpha$, J is symmetric about $\frac{1}{2}$ and

$$\phi_F(x) = \begin{cases} \frac{F_0^{-1}(\alpha)}{1-2\alpha} & x < F^{-1}(\alpha) \\ \frac{x-\theta}{1-2\alpha} & F^{-1}(\alpha) \le x \le F^{-1}(1-\alpha) \\ \frac{F_0^{-1}(1-\alpha)}{1-2\alpha} & x > F^{-1}(1-\alpha), \end{cases}$$

where $F^{-1}(\alpha) + F^{-1}(1-\alpha) = 2\theta$, $F_0^{-1}(\alpha) = F^{-1}(\alpha) - \theta$ and $F_0^{-1}(1-\alpha) = F^{-1}(1-\alpha) - \theta$.

Step 2: Calculation of $\sigma_F^2 = E[\phi_F(X_1)]^2$

Because $F_0^{-1}(\alpha) = -F_0^{-1}(1-\alpha)$, we obtain that

$$\begin{aligned}
\int [\phi_F(x)]^2 dF(x) &= \frac{[F_0^{-1}(\alpha)]^2}{(1-2\alpha)^2}\alpha + \frac{[F_0^{-1}(1-\alpha)]^2}{(1-2\alpha)^2}\alpha \\
&\quad + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \frac{(x-\theta)^2}{(1-2\alpha)^2} dF(x) \\
&= \frac{2\alpha[F_0^{-1}(1-\alpha)]^2}{(1-2\alpha)^2} + \int_{F_0^{-1}(\alpha)}^{F_0^{-1}(1-\alpha)} \frac{x^2}{(1-2\alpha)^2} dF_0(x) \\
&= \sigma_\alpha^2.
\end{aligned}$$

Step 3: Asymptotic normality of the trimmed sample mean

It can be shown that the L-functional $T(G)$ is differentiable in some sense (see the textbook).
Hence, for the α-trimmed sample mean $\bar{X}_\alpha$,

$$\sqrt{n}(\bar{X}_\alpha - \theta) \to_d N(0, \sigma_\alpha^2).$$

M-estimators

Note that the sample mean $\bar{X}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=\min_{t\in\Theta}\frac{1}{n}\sum_{i=1}^{n}(X_i-t)^2=\min_{t\in\Theta}\int(x-t)^2dF_n$$

This idea can be generalized to get a class of estimators obtained by minimizing some functions.

Let $\rho(x,t)$ be a Borel function on $\mathscr{R}^d\times\mathscr{R}$ and $\Theta\subset\mathscr{R}$ be an open set. An *M-functional* is defined to be a solution of

$$\int\rho(x,T(G))dG(x)=\min_{t\in\Theta}\int\rho(x,t)dG(x),\qquad G\in\mathscr{F}$$

For $X_1,...,X_n$ i.i.d. from $F\in\mathscr{F}$, $T(F_n)$ is called an *M-estimator* of $T(F)$.

$$\int\rho(x,T(F_n))dF_n(x)=\min_{t\in\Theta}\int\rho(x,t)dF_n(x)$$

i.e.,

$$\frac{1}{n}\sum_{i=1}^{n}\rho(X_i,T(F_n))=\min_{t\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\rho(X_i,t)$$

Assume that $\psi(x,t) = \partial\rho(x,t)/\partial t$ exists a.e. and

$$\lambda_G(t) = \int \psi(x,t)dG(x) = \frac{\partial}{\partial t}\int \rho(x,t)dG(x).$$

Then $\lambda_G(T(G)) = 0$ and $T(F_n)$ is a solution of

$$\sum_{i=1}^{n} \psi(X_i,t) = 0.$$

Example 5.7

The following are some examples of M-estimators.

(i) If $\rho(x,t) = (x-t)^2/2$, then $T(F_n) = \bar{X}$ is the sample mean.

(ii) If $\rho(x,t) = |x-t|^p/p$, where $p \in [1,2)$, then

$$\psi(x,t) = \begin{cases} |x-t|^{p-1} & x \le t \\ -|x-t|^{p-1} & x > t. \end{cases}$$

When $p = 1$, $T(F_n)$ is the sample median. When $1 < p < 2$, $T(F_n)$ is called the pth least absolute deviations estimator or the minimum L_p distance estimator.

(iii) Let $\mathscr{F}_0 = \{f_\theta : \theta \in \Theta\}$ be a parametric family of p.d.f.'s with $\Theta \subset \mathscr{R}$ and $\rho(x,t) = -\log f_t(x)$.
Then $T(F_n)$ is an MLE.
Thus, M-estimators are extensions of MLE's in parametric models.
(iv) Let $C > 0$ be a constant.
Huber (1964) considers

$$\rho(x,t) = \begin{cases} \frac{1}{2}(x-t)^2 & |x-t| \le C \\ \frac{1}{2}C^2 & |x-t| > C \end{cases}$$

with

$$\psi(x,t) = \begin{cases} t-x & |x-t| \le C \\ 0 & |x-t| > C. \end{cases}$$

The corresponding $T(F_n)$ is a type of trimmed sample mean.
(v) Let $C > 0$ be a constant.
Huber (1964) considers

$$\rho(x,t) = \begin{cases} \frac{1}{2}(x-t)^2 & |x-t| \le C \\ C|x-t| - \frac{1}{2}C^2 & |x-t| > C \end{cases}$$

with

$$\psi(x,t) = \begin{cases} C & t-x > C \\ t-x & |x-t| \le C \\ -C & t-x < -C. \end{cases}$$

The corresponding $T(F_n)$ is a type of Winsorized sample mean.
(vi) Hampel (1974) considers $\psi(x,t) = \psi_0(t-x)$ with $\psi_0(s) = -\psi_0(-s)$ and

$$\psi_0(s) = \begin{cases} s & 0 \le s \le a \\ a & a < s \le b \\ \frac{a(c-s)}{c-b} & b < s \le c \\ 0 & s > c, \end{cases}$$

where $0 < a < b < c$ are constants.
A smoothed version of ψ_0 is

$$\psi_1(s) = \begin{cases} \sin(as) & 0 \le s < \pi/a \\ 0 & s > \pi/a. \end{cases}$$

Theorem 5.7

Let $X_1, ..., X_n$ be i.i.d. from F and T be an M-functional.
Assume that ψ is a bounded and continuous function on $\mathscr{R}^d \times \mathscr{R}$ and that $\lambda_F(t)$ is continuously differentiable at $T(F)$ and $\lambda'_F(T(F)) \neq 0$.
Then

$$\sqrt{n}[T(F_n) - T(F)] \to_d N(0, \sigma_F^2)$$

with

$$\sigma_F^2 = \frac{\int [\psi(x, T(F))]^2 dF(x)}{[\lambda'_F(T(F))]^2}.$$

Example 5.13

Consider Huber's ψ given in Example 5.7(v).
Assume that F is continuous at $\theta - C$ and $\theta + C$.
Then

$$\sigma_F^2 = \frac{\int_{\theta-C}^{\theta+C} (\theta - x)^2 dF(x) + C^2 F(\theta - C) + C^2 [1 - F(\theta + C)]}{[F(\theta + C) - F(\theta - C)]^2}$$

Asymptotic relative efficiency between Huber's M-estimator and the sample mean can be obtained.

A sketched proof of Theorem 5.7:

Let $\theta = T(F)$ and $\widehat{\theta} = T(F_n)$.
By the definition of M-estimator $\widehat{\theta}$,

$$\int \psi(x, \widehat{\theta}) dF_n(x) = 0$$

Hence

$$\begin{aligned} -\int \psi(x,\theta)dF_n(x) &= \frac{\partial}{\partial\theta}\left[\int \psi(x,\theta)dF_n(x)\right](\widehat{\theta}-\theta) + o_p(n^{-1/2}) \\ &= \frac{\partial}{\partial\theta}\left[\int \psi(x,\theta)dF(x)\right](\widehat{\theta}-\theta) + o_p(n^{-1/2}) \\ &= \lambda_F'(\theta)(\widehat{\theta}-\theta) + o_p(n^{-1/2}) \end{aligned}$$

Then

$$-\frac{1}{n}\sum_{i=1}^{n} \psi(X_i,\theta)/\lambda_F'(\theta) = (\widehat{\theta}-\theta) + o_p(n^{-1/2})$$

The result follows from the CLT since $E[\psi(X_i,\theta)] = 0$ and

$$\mathrm{Var}(\psi(X_i)) = \int [\psi(x,\theta)]^2 dF(x)$$