M2RI Asymptotic Statistics Lecture 6: Asymptotic Normality of Z estimators

I. Intuition

Context: z-estimator $\hat{\theta}_n$ s.t.

$$z_n(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n} z(X_i, \hat{\theta}_n) = 0$$

with $X_1, \dots$ i.i.d $\qquad z : \mathbb{R}^p \times \Theta \to \mathbb{R}$

$$\exists\, \theta_0 \in \Theta \text{ s.t. } \quad \mathbb{E}\big(z(X_1, \theta_0)\big) = 0$$

$$\hat{\theta}_n \xrightarrow[n \to +\infty]{\mathbb{P}} \theta_0 \qquad \text{(cf lectures 4 and 5)}.$$

Taylor around θ_0: (which makes sense since $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$)

$$0 = z_n(\hat{\theta}_n) \approx z_n(\theta_0) + (Jz_n)(\theta_0)\,(\hat{\theta}_n - \theta_0).$$

$$\text{"}\Rightarrow\text{"} \quad \sqrt{n}\,(\hat{\theta}_n - \theta_0) = -(Jz_n)(\theta_0)^{-1}\big(\sqrt{n}\, z_n(\theta_0)\big)$$

L.L.N $\quad (J\, z_n)(\theta_0) \xrightarrow{\mathbb{P}} \mathbb{E}\big(Jz(\theta_1, \theta_0)\big) =: J$

and if $\det J \neq 0$, C.M $\Rightarrow (Jz_n)(\theta_0)^{-1} \xrightarrow{\mathbb{P}} J^{-1}$

C.L.T. $\quad \sqrt{n}\,[z_n(\theta_0) - \underbrace{\mathbb{E}(z(X_1, \theta_0))}_{=0}] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \operatorname{cov}(z(X_1, \theta_0)))$

Slutsky

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, J^{-1} \operatorname{cov}(z(X_1, \theta_0))\, J^{-T}\big)$$

Objective of this lecture: Make this rigorous.

II. Main Theorem

Theorem (Maximal Inequality)

- $(X_i)_{i \geq 1}$ iid with distribution $\mathcal{L}(\mathbb{R}^d)$
- $\mathcal{F}$: set of functions from $\mathbb{R}^k$ to $\mathbb{R}$.
- $\exists F : \mathbb{R}^k \to \mathbb{R}$ s.t. $\forall f \in \mathcal{F}$, $|f(x)| \leq F(x)$ $\mathcal{L}$-a.e., $C_F := \mathbb{E}(F(X_1)^2) < \infty$.

Then
$$\mathbb{E}^* \sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} f(X_i) - \mathbb{E}(f(X_i)) \right| \leq C_{MI} \int_0^{\sqrt{C_F}} \sqrt{\log\left(\mathcal{N}_{[\,]}(\mathcal{F}, L^2(\mathcal{L}), \varepsilon)\right)}\, d\varepsilon$$

Proof: Long and technical, see van der Vaart 2000.

Remark: $\mathbb{E}^*$ is the "outer expectation" in case of non-measurability.

Exercice If $X_1, \dots, \overset{iid}{\sim} \mathrm{Unif}([0,1])$, $\mathcal{F} = \{ \mathbb{1}_{(-\infty, t]} ; t \in [0,1] \}$, find an upper-bound on
$$\underbrace{\mathbb{E}^* \sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} f(X_i) - \mathbb{E}(f(X_i)) \right|}_{*}.$$

Solution $\mathcal{N}_{[\,]}\left(\mathcal{F}, L^2(\mathrm{Unif}([0,1])), \varepsilon\right) \lesssim \frac{1}{\varepsilon^2}$ (cf last lecture).

$$F = \mathbb{1}_{\mathbb{R}}. \quad \mathbb{E}(F(X_1)^2) = 1 =: C_F$$

so $\quad * \lesssim \int_0^1 \sqrt{\log(1/\varepsilon^2)}\, d\varepsilon \lesssim \int_0^1 \sqrt{-\log(\varepsilon)}\, d\varepsilon$

$$\varepsilon = e \quad \Rightarrow d\varepsilon = -e \quad du$$

$$* \lesssim \int_{\infty}^{0} e^{u/2}\left(-e^{-u}\right) du = \int_{0}^{\infty} e^{-u/2}\, du \lesssim 1 .$$

Theorem (Asymptotic Normality of Z-estimators). If,

- $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z(X_i, \hat{\theta}_n) = o_{\mathbb{P}}(1)$ with $z : \mathbb{R}^k \times \Theta \to \mathbb{R}^p$ st $\mathbb{E}(\|z(X_1, \theta)\|^2) < \infty \ \forall \theta$,

  And $\exists \theta_0 \in \Theta$ st $\mathbb{E}(z(X_1, \theta_0)) = 0$, $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.

- $\exists \mathcal{N}$ neighborhood of θ_0 s.t. $\theta \mapsto \mathbb{E}(z(X_1, \theta)) \in C^1(\mathcal{N})$.
  $J\mathbb{E}(X_1, \theta_0)$ is invertible.

- $\mathcal{F}_j := \{ x \mapsto z(x, \theta)_j \ ; \theta \in A \}$.
  $\forall j, \forall \delta > 0$, $\int_0^{\delta} \sqrt{\log \mathcal{N}_{[\,]}(\mathcal{F}_j, L^2(\mathcal{L}), \varepsilon)}\, d\varepsilon < \infty$

- $\mathbb{E} \sup_{\substack{\theta \in \mathcal{N} \\ \|\theta - \theta_0\| \le \delta}} \|z(X_1, \theta) - z(X_1, \theta_0)\|^2 \xrightarrow[\delta \to 0]{} 0$.

Then
$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -\left(J\mathbb{E}(z(X_1, \theta_0))\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z(X_i, \theta_0) + o_{\mathbb{P}}(1)$$

and thus
$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, J\mathbb{E}(z(X_1, \theta_0))^{-1} \operatorname{cov}(z(X_1, \theta_0)) J\mathbb{E}(z(X_1, \theta_0))^{-T}\right).$$

Proof: $V := J\mathbb{E}(z(X_1, \theta_0))$

Taylor on $\theta \mapsto \mathbb{E}(z(X_1, \theta))$ around θ_0

$$\int_{\mathbb{R}^p} z(x,\theta)\,d\mathcal{L}(x) = \int_{\mathbb{R}^p} z(x,\theta_0)\,d\mathcal{L}(x) + V(\theta-\theta_0) + o(\|\theta-\theta_0\|)$$

And since $\hat{\theta}_n - \theta_0 = o_{\mathbb{P}}(1)$,

$$\int z(x,\hat{\theta}_n)\,d\mathcal{L}(x) = \int z(x,\theta_0)\,d\mathcal{L}(x) + \underbrace{V(\hat{\theta}_n-\theta_0) + o_{\mathbb{P}}(\|\hat{\theta}_n-\theta_0\|)}_{=(V+o_{\mathbb{P}}(1))(\hat{\theta}_n-\theta_0)}.$$

Multiplication by $\sqrt{n}$, $\int z(x,\theta_0)\,d\mathcal{L}(x) = 0$ and the first hypothesis yield

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\int z(x,\hat{\theta}_n)\,d\mathcal{L}(x) - z(X_i,\hat{\theta}_n)\right) = (V+o_{\mathbb{P}}(1))\sqrt{n}(\hat{\theta}_n-\theta_0) + o_{\mathbb{P}}(1).$$

which rewrites as

$$(V+o_{\mathbb{P}}(1))\sqrt{n}(\hat{\theta}_n-\theta_0) = o_{\mathbb{P}}(1) - \overbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(z(X_i,\theta_0) - \int_{\mathbb{R}^p} z(x,\theta_0)\,d\mathcal{L}(x)\right)}^{\text{Easy with CLT}}$$

$$+ \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\left(z(X_i,\theta_0) - z(X_i,\hat{\theta}_n)\right) - \int\left(z(x,\theta_0) - z(x,\hat{\theta}_n)\right)d\mathcal{L}(x)\right)}$$

$=: r_n$. We need to show that $r_n = o_{\mathbb{P}}(1)$.

Let's define: $\mathcal{F}_{j,\delta} := \{x \mapsto z(x,\theta)_j - z(x,\theta_0)_j \,;\, \theta \in B(\theta_0,\delta)\}$.

If $\|\hat{\theta}_n - \theta_0\| \leq \delta$, we have

$$\|r_n\| \leq \sqrt{p}\,\max_j \sup_{f\in\mathcal{F}_{j,\delta}} \frac{1}{\sqrt{n}}\left|\sum_{i=1}^n f(X_i) - \mathbb{E}(f(X_1))\right|.$$

Let's notice that $\forall \varepsilon>0$, $N_{[\,]}(\mathcal{F}_{j,\delta}, L^2(\mathcal{L}), \varepsilon) \leq N_{[\,]}(\mathcal{F}_j, L^2(\mathcal{L}), \varepsilon)$.

Then, $\forall \delta, \varepsilon > 0$ with $B(\theta_0, \delta) \subset A$,

$$\mathbb{P}(\|r_n\| \geq \varepsilon) \leq \mathbb{P}(\|\hat{\theta}_n - \theta_0\| \geq \delta) + \mathbb{P}\left(\sqrt{p} \max_j \sup_{f \in F_{j,\delta}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n f(X_i) - \mathbb{E} f(X_1) \right| \geq \varepsilon \right)$$

And since $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$,

$$\limsup_n \mathbb{P}(\|r_n\| \geq \varepsilon) \leq \mathbb{P}\left(\sqrt{p} \max_j \sup_{f \in F_{j,\delta}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n f(X_i) - \mathbb{E} f(X_1) \right| \geq \varepsilon \right).$$

$\forall f \in F_{j,\delta}$ and $x \in \mathbb{R}^b$, $|f(x)| \leq F_\delta(x)$

with $F_\delta(x) = \sup_{\substack{\theta \in A \\ \|\theta - \theta_0\| \leq \delta}} \|z(X_1, \theta) - z(X_1, \theta_0)\|$

So, by Markov and Maximal inequality,

$$\limsup_n \mathbb{P}(\|r_n\| \geq \varepsilon) \leq \sum_{j=1}^p \mathbb{P}\left(\sup_{f \in F_{j,\delta}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n f(X_i) - \mathbb{E} f(X_1) \right| \geq \frac{\varepsilon}{\sqrt{p}} \right)$$

$$\leq \sum_{j=1}^p \frac{\sqrt{p}}{\varepsilon} \mathbb{E} \sup_{f \in F_{j,\delta}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n f(X_i) - \mathbb{E} f(X_1) \right|$$

$$\leq \sum_{j=1}^p \frac{\sqrt{p}}{\varepsilon} C_{MI} \underbrace{\int_0^{\sqrt{\mathbb{E}(F_\delta(X_1)^2)}} \sqrt{\log N_{[\,]}(F_{j,\delta}, L^2(\mathcal{L}), u)}\, du}_{}$$

Can be made as small as we want because $\mathbb{E}(F_\delta(X_1)^2) \xrightarrow[\delta \to 0]{} 0$.

$\square$

III. Application to the median.

$(X_i)_{i\in\mathbb{N}}$ iid with F_{X_1} and with density f

Assumption 1: $f(x) > 0 \quad \forall x \in \mathbb{R}$ almost surely.

Population median θ_0 the only real number such that $F_{X_1}(\theta_0) = \frac{1}{2}$

Empirical median $\hat{\theta}_n$ s.t. $\sum_{i=1}^{n} \operatorname{sgn}(\hat{\theta}_n - X_i) = 0$

Remark: We might restrict n to even numbers and that all the X's are different (it happens almost surely).

$\Rightarrow$ By lecture 4, $\theta_n \xrightarrow{\mathbb{P}} \theta_0$ (The first point of the theorem is satisfied)

Assumption 2: f is continuous on a neighborhood of θ_0.

$\Rightarrow \mathbb{E}(\operatorname{sgn}(\theta - X_1)) = 2F_{X_1}(\theta) - 1$ is C^1 on a neighborhood of θ_0 with derivatives $2f(\theta_0)$ at θ_0.

(The second point of the theorem is satisfied).

Next Step: $\mathcal{F} := \{x \mapsto \operatorname{sgn}(\theta - x) ; \theta \in \mathbb{R}\}$

Control $\mathcal{N}_{[\,]}(\mathcal{F}, L^2(\mathcal{L}), \varepsilon)$

Exercice:

1) Show that $\forall N$ st $N+1 \geq \frac{1}{\varepsilon^2}$, $\exists t_1, \dots, t_N$ st

$\forall j = 0, \dots, N, \ \mathcal{L}\big((t_j, t_{j+1})\big) \leq \varepsilon^2$

with $t_0 = -\infty$ and $t_{N+1} = +\infty$.

2) By considering functions of the form $x \mapsto \mathbb{1}(t_j \leq x)$ and $x \mapsto \mathbb{1}(t_j < x)$,

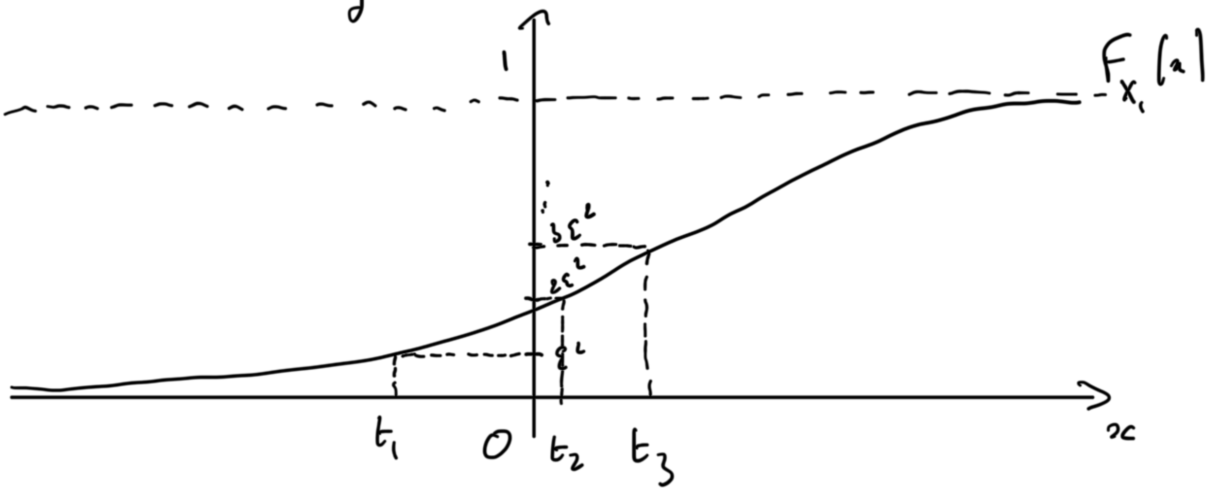
Build brackets of $\mathcal{F}_- = \{x \mapsto \mathbb{1}(\theta < x) ; \theta \in \mathbb{R}\}$.

Do the same thing for $\mathcal{F}_+ = \{x \mapsto \mathbb{1}(x < \theta) ; \theta \in \mathbb{R}\}$

3) After noticing that $\operatorname{sgn}(\theta - x) = \mathbb{1}(x < \theta) - \mathbb{1}(\theta < x)$, conclude.

Solution:

1) On a drawing



2) $\forall j = 1, \dots, N, \quad \ell_{-,j}(x) = \mathbb{1}(t_{j+1} \leq x)$

$u_{-,j}(x) = \mathbb{1}(t_j < x)$

then $\forall j \forall x, \ \ell_{-,j}(x) \leq u_{-,j}(x)$

and $\forall \theta, \exists j$ st $\forall x, \ell_{-,j}(x) \leq \mathbb{1}(\theta \leq x) \leq u_{-,j}(x)$.

Finally, $\int (u_{-,j} - \ell_{-,j})^2 d\mathcal{L} = \int \mathbb{1}(x \in (t_j, t_{j+1})) d\mathcal{L}(x)$

$$= \mathcal{L}((t_j, t_{j+1}))$$

$$\leq \varepsilon^2.$$

Same reasoning for $\mathcal{F}_+$ with $\ell_{+,j} = \mathbb{1}(x \leq t_j)$

$$u_{+,j} = \mathbb{1}(x < t_{j+1}).$$

3)

$$\underbrace{\underbrace{\ell_{+,j}(x) - u_{-,j}(x)}_{}\leq \overset{\operatorname{sgn}(\theta - x)}{\overset{\parallel}{\mathbb{1}(x<\theta) - \mathbb{1}(\theta<x)}} \leq \underbrace{u_{+,j}(x) - \ell_{-,j}(x)}_{}}_{\text{Brackets for } F.}$$

and by triangular inequality,

$$\|(\ell_{+,j} - u_{-,j}) - (u_{+,j} - \ell_{-,j})\|_{L^2(\mathcal{L})}$$

$$= \|(\ell_{+,j} - u_{+,j}) + (\ell_{-,j} - u_{-,j})\|_{L^2(\mathcal{L})}$$

$$\leq 2\varepsilon.$$

Thus, $\mathcal{N}_{[\,]}(F, L^2(\mathcal{L}), 2\varepsilon) \leq N+1$ and since we can chose $N+1 \leq \frac{1}{\varepsilon^2} + 1$,

$\mathcal{N}_{[\,]}(F, L^2(\mathcal{L}), 2\varepsilon) \leq \frac{1}{\varepsilon^2} + 1$, which gives

$$\mathcal{N}_{[\,]}\left(\mathcal{F}, L^2(\mathcal{L}), \varepsilon\right) \leq \frac{4}{\varepsilon^2} + 1$$

$\square$

Furthermore, $\mathbb{E}\left(\operatorname{sgn}\left(\theta_0 - X_1\right)^2\right) = \mathbb{E}(1) = 1$.

thus $\displaystyle\int_0^{\sqrt{\mathbb{E}(\operatorname{sgn}(\theta_0 - X_1)^2)}} \sqrt{\log\left(\mathcal{N}_{[\,]}\left(\mathcal{F}, L^2(\mathcal{L}), \varepsilon\right)\right)} < \infty.$

(The third point of the theorem is satisfied).

- $$\mathbb{E}\left(\sup_{\substack{\theta \in \mathbb{R} \\ \|\theta - \theta_0\| \leq \delta}} \left(\operatorname{sgn}(\theta - X_1) - \operatorname{sgn}(\theta_0 - X_1)\right)^2\right)$$

$$= 2\,\mathbb{P}\left(X_1 \in [\theta_0 - \delta, \theta_0 + \delta]\right) \xrightarrow[\delta \to 0]{} 0$$

(The fourth hypothesis of the theorem is satisfied).

so, $$\boxed{\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4 f^2(\theta_0)}\right)}$$