M2RI Asymptotic Statistics Lecture 5: Bracketing Number

Context With M and Z estimators, we need to show

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \sup_{\theta\in\Theta} \|Z_n(\theta) - Z(\theta)\| \xrightarrow{\mathbb{P}} 0 .$$

This lecture will explain how to do this for M-estimators, Z-estimators being handled in the same way.

Hypothesis:
- $X_1, \dots$ iid
- $M_n(\theta) = \frac{1}{n}\sum_{i=1}^n m(X_i, \theta) \qquad m : \mathbb{R}^k \times \Theta \to \mathbb{R}$
- $M(\theta) = \mathbb{E}(m(X_1, \theta))$

Notation: $F = \{ m(\cdot, \theta) ; \theta \in \Theta \}$.

Rewriting: $\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| = \sup_{f\in F} |M_n(f) - M(f)| = \sup_{f\in F} \left| \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}(f(X_1)) \right|$

💡: We need to measure the complexity of the class of functions F.

1st Idea Since we know how to handle F' finite (Law of large numbers).

So we can search F' finite such that $\forall f \in F, \exists f' \in F'$ s.t. $\|f - f'\|_\infty \leq \varepsilon$.

Then
$$\sup_{f\in F} \left| \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}(f(X_1)) \right| \leq \sup_{f'\in F'} \left| \frac{1}{n}\sum_{i=1}^n f'(X_i) - \mathbb{E} f'(X_1) \right| + 2\varepsilon$$
p.s.

But $N(F, \|\cdot\|_\infty, \varepsilon) = \inf_{N\in\mathbb{N}} \{ N : \exists F' \text{ with } |F'| = N \text{ and } \forall f \in F, \exists f' \in F', \|f - f'\|_\infty \leq \varepsilon \}$

(covering number) can be ∞ on simple examples.

Exercice What is $N(F, \|\cdot\|_\infty, \varepsilon)$ when

$$F = \left\{ \mathbb{1}_{(-\infty, t]} \; ; \; t \in [0,1] \right\} \quad ?$$

Definition (Bracketing numbers)

For l and u two functions from $\mathbb{R}^k$ to $\mathbb{R}$ and such that $\forall x,\ l(x) \leq u(x)$.

$$[l,u] := \left\{ f : \mathbb{R}^k \to \mathbb{R} \; : \; \forall x,\ l(x) \leq f(x) \leq u(x) \right\} \text{ (bracket)}.$$

We define the bracketing number $\mathcal{N}_{[\,]}(F, L^q(\mathcal{L}), \varepsilon)$ as

$$\mathcal{N}_{[\,]}\left(F, L^q(\mathcal{L}), \varepsilon\right) = \inf_{N \in \mathbb{N}} \left\{ \exists [l_1,u_1], \dots, [l_N, u_N] : \forall j,\ \|u_j - l_j\|_{L^q(\mathcal{L})} \leq \varepsilon,\ F \subset \bigcup_j [l_j, u_j] \right\}.$$

Exercice $P = \mathrm{Unif}\,[0,1]$ $\quad F = \left\{ f_t(x) = \mathbb{1}_{(-\infty,t]}(x) : t \in [0,1] \right\}$

Find (up to mult. constants) $\mathcal{N}_{[\,]}(F, L_2(P), \varepsilon)$.

Solution

- $\forall s,t \in [0,1],\ \|f_t - f_s\|_{L_2(P)} = \left(\int_0^1 \left(\mathbb{1}_{(-\infty,t)}(u) - \mathbb{1}_{(-\infty,s)}(u) \right)^2 du \right)^{1/2}$

$$= \sqrt{|t-s|}$$

- $[f_0, f_{\varepsilon^2}], [f_{\varepsilon^2}, f_{2\varepsilon^2}], \dots, [f_{\lfloor \frac{1}{\varepsilon^2} \rfloor \varepsilon^2}, f_1]$ satisfies the covering condition and the metric condition and has $\propto 1/\varepsilon^2$ brackets

- $\|f_a - f_b\|_2 = \sqrt{|a-b|}$ so given a set of brackets $[f_{a_1}, f_{b_1}] \dots [f_{a_N}, f_{b_N}]$, at most a length $\sum_i |a_i - b_i|$ is covered, and since $\forall i,\ \sqrt{|a_i - b_i|} \leq \varepsilon$,
$N\varepsilon^2$ is covered.

So, since everything must be covered, $N\varepsilon^2 \gtrsim 1 \Rightarrow N \gtrsim 1/\varepsilon^2$.

Definition ($\mathcal{L}$-Glivenko-Cantelli)

$\mathcal{F}$ is said $\mathcal{L}$-Glivenko-Cantelli i

- $\forall f \in \mathcal{F}$, $\int |f| d\mathcal{L} < \infty$
- $\forall (X_i)_{i\in\mathbb{N}}$ i.i.d with distribution $\mathcal{L}$,

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}\left(f(X_1)\right) \right| = o_{\mathbb{P}}(1).$$

Theorem: If $\mathcal{F}$ is such that

- $\forall f \in \mathcal{F}$, $\int |f| d\mathcal{L} < \infty$
- $\forall \varepsilon > 0$, $\mathcal{N}_{[\,]}(\mathcal{F}, L^1(\mathcal{L}), \varepsilon) < \infty$.

Then $\mathcal{F}$ is $\mathcal{L}$-Glivenko-Cantelli.

Proof. Let $\varepsilon > 0$ and $N = \mathcal{N}_{[\,]}(\mathcal{F}, L^1(\mathcal{L}), \varepsilon) < \infty$ and $[l_1, u_1], \dots, [l_N, u_N]$ a bracket that satisfies this bound.

$$\forall f \in \mathcal{F}, \forall j, \frac{1}{n}\sum_{i=1}^{n} l_j(X_i) \leq \frac{1}{n}\sum_{i=1}^{n} f(X_i) \leq \frac{1}{n}\sum_{i=1}^{n} u_j(X_i),$$

and since $\mathbb{E}(u_j(X_1)) - \mathbb{E}(l_j(X_1)) \leq \mathbb{E}(|l_j(X_1) - u_j(X_1)|) \leq \varepsilon$

$$\Rightarrow \mathbb{E}(l_j(X_1)) \leq \mathbb{E}(f(X_1)) \leq \mathbb{E}(u_j(X_1)) \leq \mathbb{E}(l_j(X_1)) + \varepsilon$$

So,

$$\frac{1}{n}\sum_i l_j(X_i) - \mathbb{E}(u_j(X_1)) \leq \frac{1}{n}\sum_i f(X_i) - \mathbb{E}(f(X_1)) \leq \frac{1}{n}\sum_i u_j(X_i) - \mathbb{E}(l_j(X_1))$$

And thus,

$$\frac{1}{n}\sum_{i=1}^{n} l_j(X_i) - \mathbb{E}(l_j(X_1)) - \varepsilon \leq \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}(f(X_1)) \leq \frac{1}{n}\sum_{i=1}^{n} u_j(X_i) - \mathbb{E}(u_j(X_1)) + \varepsilon$$

$$\Rightarrow \mathbb{P}\left(\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}(f(X_1)) \right| \geq 2\varepsilon \right)$$

$$\leq \mathbb{P}\left(\max_{j=1,\dots,N}\ \max\left\{\left|\frac{1}{n}\sum_{i=1}^{n} l_j(X_i)-\mathbb{E}\big(l_j(X_1)\big)\right|,\left|\frac{1}{n}\sum_{i=1}^{n} u_j(X_i)-\mathbb{E}\big(u_j(X_1)\big)\right|\right\}\geq \varepsilon\right)$$

And now, from the law of large numbers, since the $\max_{j=1,\dots,N}$ is finite, we know that the last term $\xrightarrow[n\to+\infty]{} 0$.

$$\text{So, } \sup_{f\in F}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i)-\mathbb{E}\big(f(X_1)\big)\right| = o_{\mathbb{P}}(1).$$

Proposition Let $\mathcal{L}$ be a distribution on $\mathbb{R}^k$, $F=\{g_\theta\ ;\ \theta\in\Theta\}$ with $\forall\theta$, $g_\theta:\mathbb{R}^k\to\mathbb{R}$. Furthermore, let's assume that

- Θ is a compact set of a metric space
- $\forall x$, $\theta\mapsto g_\theta(x)$ is continuous
- $\int \sup_{\theta\in\Theta} |g_\theta(x)|\, d\mathcal{L}(x) < \infty$.

Then F is $\mathcal{L}$-Glivenko-Cantelli.

Proof: Let $\varepsilon>0$. We are going to show that $\mathcal{N}_{[\,]}(F, L^1(\mathcal{L}),\varepsilon) < +\infty$

For $\theta\in\Theta$, we consider $(B_{\theta,N})_{N\in\mathbb{N}}$ with $B_{\theta,N}=\{\tilde\theta\in\Theta,\ d(\theta,\tilde\theta)<\frac{1}{N}\}$.

and define

$$\forall x,\quad \tilde l_{\theta,N}(x)=\inf_{\tilde\theta\in B_{\theta,N}} g_{\tilde\theta}(x)\ ;\ \tilde u_{\theta,N}(x)=\sup_{\tilde\theta\in B_{\theta,N}} g_{\tilde\theta}(x).$$

We have, $\forall x$, $\tilde u_{\theta,N}(x)-\tilde l_{\theta,N}(x)\xrightarrow[n\to+\infty]{} 0$ by continuity and

$$\tilde{u}_{\theta,N} - \tilde{\ell}_{\theta,N} \leq 2 \sup_{\theta \in \Theta} |g_\theta| ,$$

and thus $\int \sup_{N \in \mathbb{N}} |\tilde{u}_{\theta,N} - \tilde{\ell}_{\theta,N}| \, d\mathcal{L} < +\infty$ so by dominated convergence,

$$\int |\tilde{u}_{\theta,N} - \tilde{\ell}_{\theta,N}| \, d\mathcal{L} \xrightarrow[N \to +\infty]{} 0$$

So $\exists N_\theta$ s.t. $\int |\tilde{u}_{\theta,N_\theta} - \tilde{\ell}_{\theta,N_\theta}| \, d\mathcal{L} \leq \varepsilon$.

Now, $\Theta \subset \{ \bigcup_{\theta \in \Theta} B_{\theta, N_\theta} \}$ and by Compacity (Borel-Lebesgue),

$\exists \theta_1, \dots, \theta_m$ s.t. $\Theta \subset \bigcup_{j=1}^{m} B_{\theta_j, N_{\theta_j}}$.

We define $\forall j, x$,

$$\ell_j(x) = \tilde{\ell}_{\theta_j, N_{\theta_j}} \qquad u_j(x) = \tilde{u}_{\theta_j, N_{\theta_j}} .$$

Then we have build a <u>finite</u> sequence of brackets that works.

So $\mathcal{N}_{[\,]}(F, L^1(\mathcal{L}), \varepsilon) < \infty$. $\square$

Application to Maximum Likelihood

Theorem Statistical Model $\{\mathcal{L}_\theta \,;\, \theta \in \Theta\}$ with $\forall \theta$, $\mathcal{L}_\theta$ with density f_θ.

$X_1, \dots \overset{iid}{\sim} f_{\theta_0}$, $\hat{\theta}_n$ Maximum Likelihood Estimator.

<u>Assumptions</u>:
- Θ Compact
- $\forall \theta$, $f_{\theta}(x) > 0$ a.s.

- $\forall x$, $\theta \mapsto f_\theta(x)$ is continuous
- $\int \sup_{\theta \in \Theta} |\log(f_\theta(x))| \, f_{\theta_0}(x) \, dx < \infty$
- $\theta \Rightarrow \mathcal{L}_\theta$ is injective on Θ.

<u>Then:</u> $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.

<u>Proof:</u> By the last theorem,

$$\sup_{\theta \in \Theta} \left| \sum_{i=1}^{n} \log(f_\theta(X_i)) - \mathbb{E}(\log(f_\theta(X_1))) \right| \xrightarrow{\mathbb{P}} 0$$

We simply need to check the well-posed indentification from last lecture with $M(\theta) = \mathbb{E}(\log(f_\theta(X_1)))$.

For $\theta \neq \theta_0$,

$$\begin{aligned}
M(\theta) - M(\theta_0) &= \mathbb{E}(\log(f_\theta(X_1))) - \mathbb{E}(\log(f_{\theta_0}(X_1))) \\
&= \int \log\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right) f_{\theta_0}(x) \, dx \\
&\underset{\substack{\log(t) \leq 2(\sqrt{t}-1) \\ \text{for } t > 0}}{\leq} 2 \int \left(\sqrt{\frac{f_\theta(x)}{f_{\theta_0}(x)}} - 1 \right) f_{\theta_0}(x) \, dx \\
&= 2 \int \sqrt{f_\theta(x)} \sqrt{f_{\theta_0}(x)} \, dx - 2 \int f_{\theta_0}(x) \, dx \\
&= 2 \int \sqrt{f_\theta(x)} \sqrt{f_{\theta_0}(x)} \, dx - \int f_\theta(x) \, dx - \int f_{\theta_0}(x) \, dx \\
&= - \int \left(\sqrt{f_\theta(x)} - \sqrt{f_{\theta_0}(x)} \right)^2 dx \\
&< 0.
\end{aligned}$$

Finally, M is continuous by continuity under the integral.

Hence, $\forall \varepsilon > 0$, $\sup\limits_{\substack{\theta \in \Theta \\ \|\theta - \theta_0\| \geq \varepsilon}} M(\theta) < M(\theta_0)$.

We can apply last lecture's result

Then $\hat{\theta}_n \xrightarrow{P} \theta$.