

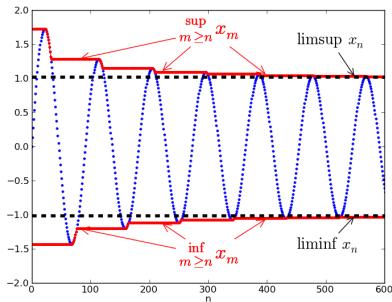
Let's begin with some definitions. Let x_n and a_n be a sequence of constants.

- $x_n = o(a_n)$ means $\frac{x_n}{a_n} \rightarrow 0$ as $n \rightarrow \infty$.

- $x_n = O(a_n)$ means $\left\| \frac{x_n}{a_n} \right\| \leq M$

- $\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \sup_{m \geq n} x_m \equiv \bar{\lim}_{n \rightarrow \infty} x_n$

$$\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \inf_{m \geq n} x_m \equiv \lim_{n \rightarrow \infty} x_n$$



(Taken from Wikipedia)

Now let x_n be a sequence of random vectors.

- $x_n = o_p(a_n)$ means $\frac{x_n}{a_n} \xrightarrow{p} 0 \equiv \forall \epsilon \lim_{n \rightarrow \infty} P \left\{ \left\| \frac{x_n}{a_n} \right\| > \epsilon \right\} = 0$

- $x_n = O_p(a_n)$ means $\forall \epsilon, \exists M_\epsilon > 0$ such that

$$\limsup_{n \rightarrow \infty} P \left\{ \left\| \frac{x_n}{a_n} \right\| > M_\epsilon \right\} < \epsilon$$

Let $H(\cdot)$ be some function from $\Theta \rightarrow \mathbb{R}$.

- Continuity at θ : $\forall \epsilon, \exists \delta$ s.t. $\sup_{\theta' \in B(\theta, \delta)} |H(\theta) - H(\theta')| < \epsilon$

- Uniform Continuity: $\forall \epsilon, \exists \delta$ s.t. $\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} |H(\theta) - H(\theta')| < \epsilon$

Now let $\{h_n(\cdot), n \geq 1\}$ be a family of functions from $\Theta \mapsto \mathbb{R}$.

(Uniform)

- Equicontinuity : $\forall \epsilon, \exists \delta \text{ s.t. } \sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} |h(\theta) - h(\theta')| < \epsilon \quad \forall h \in \{h_n(\cdot), n \geq 1\}$

Now let $\{h_n(\cdot), n \geq 1\}$ is a family of random functions from $\Theta \mapsto \mathbb{R}$.

- Stochastic Equicontinuity : $\forall \epsilon, \exists \delta \text{ s.t. } \limsup_{n \rightarrow \infty} P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} |h_n(\theta) - h_n(\theta')| > \epsilon\right) \leq \epsilon$

If $\{h_n(\cdot), n \geq 1\}$ are vector valued, i.e. $\Theta \mapsto \mathbb{R}^k$, then we use the norm

$$\forall \epsilon, \exists \delta \text{ s.t. } \limsup_{n \rightarrow \infty} P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} \|h_n(\theta) - h_n(\theta')\| > \epsilon\right) \leq \epsilon$$

Consistency

$$(1) \text{ EE: } \hat{\theta}_n \in \Theta : Q_n(\hat{\theta}_n) \leq \inf_{\theta \in \Theta} Q_n(\theta) + o_p(1)$$

$$(2) \text{ UWC: } \sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0$$

$$(3) \text{ ID: } \exists \theta_0 \in \Theta \text{ such that } \forall \epsilon > 0$$

$$\inf_{\theta \notin B(\theta_0, \epsilon)} Q(\theta) > Q(\theta_0)$$

Theorem :- (1) - (3) $\Rightarrow \hat{\theta}_n \xrightarrow{P} \theta_0$.

proof:

$$P(\hat{\theta}_n \notin B(\theta_0, \epsilon)) = P(\|\hat{\theta}_n - \theta_0\| \geq \epsilon) \leq P(Q(\hat{\theta}_n) - Q(\theta_0) \geq \delta) \quad (3)$$

for some δ

$$= P(Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n) + Q_n(\hat{\theta}_n) - Q(\theta_0) \geq \delta)$$

Why? $Q_n(\hat{\theta}_n) \leq \inf_{\theta \in \Theta} Q_n(\theta) + o_p(1)$

$$\leq Q_n(\theta_0) + o_p(1)$$

\leftarrow

$$\leq P(Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n) + Q_n(\theta_0) + o_p(1) - Q(\theta_0) \geq \delta)$$

$$\leq P(|Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n)| + |Q_n(\theta_0) - Q(\theta_0)| + o_p(1) \geq \delta)$$

$$\leq P\left(2 \sup_{\theta \in \Theta} |Q(\theta) - Q_n(\theta)| + o_p(1) \geq \delta\right)$$

$\underbrace{o_p(1)}$ by (2)

Then

$$\lim_{n \rightarrow \infty} P(\hat{\theta}_n \notin B(\theta_0, \varepsilon)) \leq \lim_{n \rightarrow \infty} P(|\text{op}(1)| \geq \delta) = 0.$$

Asymptotic Normality

- ① CF : (i) $\theta_0 \in \text{int}(\Theta)$
(ii) $Q_n(\theta)$ is twice continuously differentiable on some neighborhood $\Theta_0 \subset \Theta$ of θ_0 with probability 1.
(iii) $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, -B_0)$

$$(iv) \sup_{\theta \in \Theta_0} \left\| \frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'} - B(\theta) \right\| \xrightarrow{P} 0 \text{ for some}$$

nonstochastic $d \times d$ matrix $B(\theta)$ that is continuous at θ_0 and for which $B(\theta_0)$ is non singular.

- ② EE2 : (i) $\hat{\theta}_n \xrightarrow{P} \theta_0$
(ii) $\frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} = \text{op}\left(\frac{1}{\sqrt{n}}\right)$

Theorem .- ① and ② hold $\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B_0^{-1} - R_0 B_0^{-1})$
proof:

$$\text{op}\left(\frac{1}{\sqrt{n}}\right) = \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} = \underbrace{\frac{\partial Q_n(\theta_0)}{\partial \theta}}_{\substack{\text{mean value} \\ \theta_0}} + \underbrace{\frac{\partial^2 Q_n(\theta_n)}{\partial \theta \partial \theta'} (\hat{\theta}_n - \theta_0)}_{\substack{\text{exp around} \\ \theta_0}} \quad \text{④ } \hat{\theta}_n \xrightarrow{P} \theta_0 \text{ implies } \hat{\theta}_n \neq \theta_0$$

Multiply \sqrt{n}

$$\text{op}(1) = \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + \underbrace{\frac{\partial^2 Q_n(\theta_n)}{\partial \theta \partial \theta'} \sqrt{n} (\hat{\theta}_n - \theta_0)}_{(B(\theta_0) + \text{op}(1))}$$

$$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) = -B_0^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta_0} + \text{op}(1).$$

$$\xrightarrow{d} N(0, B_0^{-1} - R_0 B_0^{-1})$$

(*) Notice that a hidden assumption in CF (iii) is that

$$E \frac{\partial Q_n(\theta_0)}{\partial \theta} = 0 \quad \text{and} \quad E \left\| \frac{\partial Q_n(\theta_0)}{\partial \theta} \right\| < \infty$$

we may write some assumptions that ensure this. This is where CF (ii) comes into play.

Theorem.- (Dominated Convergence) Let $\{f_n\}$ be a sequence of complex valued measurable functions on a measurable space $(\Omega, \mathcal{F}, \mu)$. Suppose

- $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$ for $\forall \omega$
- $|f_n(\omega)| \leq g(\omega)$ and g is integrable

Then

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} \lim_{n \rightarrow \infty} f_n d\mu = \int_{\Omega} f d\mu.$$

Example 2: ML estimator

$$Q_n(\theta) = -\frac{1}{n} \sum_{i=1}^n \log f(w_i, \theta)$$

$$\frac{\partial Q_n(\theta)}{\partial \theta} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(w_i, \theta)$$

The true value satisfies

$$\frac{\partial}{\partial \theta} E \log f(w_i, \theta) = 0 \quad \text{if } \theta_0 \in \text{int}(\Theta)$$

We need for CF (iii) that

$$E \frac{\partial}{\partial \theta} \log f(w_i, \theta) = 0$$

these two are not the same!

We need to use the DCT to interchange differentiation with expectation.

Let $g = \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} \log f(w_i, \theta) \right\|$ be our dominating function. Then if

$$E \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} \log f(w_i, \theta) \right\| < \infty \quad \text{we can apply DCT.}$$

Thus $E \frac{\partial}{\partial \theta} \log f(w_i, \theta) = \frac{\partial}{\partial \theta} E \log f(w_i, \theta) = 0$.

Example 2: NLS

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - g(x_i, \theta))^2$$

$$\frac{\partial Q_n(\theta)}{\partial \theta} = -\frac{1}{n} \sum_{i=1}^n (y_i - g(x_i, \theta)) \frac{\partial g(x_i, \theta)}{\partial \theta}$$

The true value satisfies

$$\frac{\partial}{\partial \theta} E (y_i - g(x_i, \theta))^2 \frac{1}{2} = 0 \quad \text{if } \theta_0 \in \text{int}(\Theta)$$

We need for CF (ii) that

$$E (y_i - g(x_i, \theta)) \frac{\partial g(x_i, \theta)}{\partial \theta} = 0$$

Again, let $\beta = \sup_{\theta \in \Theta} \| (y_i - g(x_i, \theta)) \frac{\partial g(x_i, \theta)}{\partial \theta} \|$ and

assume $E \sup_{\theta \in \Theta} \| (y_i - g(x_i, \theta)) \frac{\partial g(x_i, \theta)}{\partial \theta} \| < \infty$. then, by the DCT

$$E (y_i - g(x_i, \theta)) \frac{\partial g(x_i, \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} E (y_i - g(x_i, \theta))^2 = 0.$$

Exercises

Midterm 2016 Q2

Consider the extremum estimators framework. Let $Q(\theta)$ be the nonrandom population criterion function, and suppose that it is minimized on a set $\Theta_0 \subset \Theta$, where $\Theta \subset \mathbb{R}^d$ is the parameter space:

Assumption Set-ID: For all $\theta \in \Theta_0$, $Q(\theta) = \underline{Q}$; and for all $\epsilon > 0$, $\inf_{\theta \notin B(\Theta_0, \epsilon)} Q(\theta) > \underline{Q}$,

where $B(\Theta_0, \epsilon) = \{\theta \in \Theta : d(\theta, \Theta_0) < \epsilon\}$ is the ϵ -enlargement of the set Θ_0 , and the distance between a point $b \in \mathbb{R}^d$ and a set $A \subset \mathbb{R}^d$ is defined as

$$d(b, A) = \inf_{a \in A} \|b - a\|,$$

where $\|\cdot\|$ is the Euclidean norm. Unlike in the standard extremum estimators framework, we assume that the set Θ_0 contains multiple points, i.e. the uniqueness condition fails.

Let $Q_n(\theta)$ be the random (sample) criterion function. Unlike $Q(\theta)$, we can assume that, because of random sample variation, $Q_n(\theta)$ is minimized at a unique point $\hat{\theta}_n \in \Theta$ (however, $\hat{\theta}_n$ obviously varies with n):

Assumption EE: There is a sequence $\hat{\theta}_n \in \Theta$, such that $Q_n(\hat{\theta}_n) \leq \inf_{\theta \in \Theta} Q_n(\theta) + o_p(1)$.

Lastly, assume that $Q_n(\cdot)$ converges uniformly to $Q(\cdot)$ on Θ :

Assumption U-WCON: $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \rightarrow_p 0$.

1. (5 points) How would you define the convergence in probability of the point estimator $\hat{\theta}_n$ to the set Θ_0 ?
2. (15 points) Show that under Assumptions Set-ID, EE, and U-WCON, the extremum estimator $\hat{\theta}_n$ converges in probability to the set Θ_0 .

solution :

1. Consider the case of point identification

$\forall \epsilon \lim_{n \rightarrow \infty} P(\underbrace{\|\hat{\theta}_n - \theta\|}_{d(\hat{\theta}_n, \theta)} > \epsilon) = 0$ is how we define it.
where θ is the identifying set.

now, define it as

$\forall \epsilon \lim_{n \rightarrow \infty} P(d(\hat{\theta}_n, \Theta_0) > \epsilon) = 0$.

2. We want to show that

$\forall \epsilon \lim_{n \rightarrow \infty} P(d(\hat{\theta}_n, \Theta_0) > \epsilon) = 0$
start with this measure

$$\begin{aligned} P(d(\hat{\theta}_n, \Theta_0) > \epsilon) &\stackrel{\text{ID}}{\leq} P(Q(\hat{\theta}_n) - \underline{Q} > \delta_\epsilon) \text{ for some } \delta_\epsilon \\ &= P(Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n) + Q_n(\hat{\theta}_n) - \underline{Q} + o_p(1) > \delta_\epsilon) \\ &\stackrel{\text{EE}}{\leq} P(Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n) + Q_n(\hat{\theta}_n) - \underline{Q} + o_p(1) > \delta_\epsilon) \\ &\quad \text{for any } \theta_0 \in \Theta_0 \end{aligned}$$

$$\begin{aligned}
&\leq P(|Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n)| + |Q_n(\theta_0) - Q(\theta_0)| + o_p(1) > \delta_\varepsilon) \\
&\leq P(2 \sup_{\theta \in \Theta} |Q(\theta) - Q_n(\theta)| + o_p(1) > \delta_\varepsilon) \\
&= P(o_p(1) > \delta_\varepsilon)
\end{aligned}$$

Hence

$$\begin{aligned}
P(d(\hat{\theta}_n, \theta_0) > \varepsilon) &\leq P(o_p(1) > \delta_\varepsilon) \\
\Rightarrow \lim_{n \rightarrow \infty} P(d(\hat{\theta}_n, \theta_0) > \varepsilon) &\leq \lim_{n \rightarrow \infty} P(o_p(1) > \delta_\varepsilon) = 0
\end{aligned}$$

Midterm 2017 Q2

Consider the following function

$$H_n(\theta) = n^{-1} \sum_{i=1}^n h(W_i, \theta),$$

where $h(W_i, \theta)$ is a scalar-valued function, $\{W_i : i = 1, \dots, n\}$ are iid random k -vectors, and $\theta \in \Theta \subset \mathbb{R}^d$, where Θ is bounded. Suppose that

$$Eh(W_i, \theta) = 0 \text{ for all } \theta \in \Theta,$$

$h(W_i, \theta)$ is differentiable in θ , and for some constant $K > 0$

$$E \sup_{\theta \in \Theta} \left\| \frac{\partial h(W_i, \theta)}{\partial \theta} \right\| \leq K.$$

- (a) **(30 points)** Show that $H_n(\theta)$ is stochastically equicontinuous. Hint: Use a mean-value expansion of $H_n(\theta_1) - H_n(\theta_2)$.
- (b) **(5 points)** Using the result in (a), show that $\sup_{\theta \in \Theta} |H_n(\theta)| \rightarrow_p 0$.

solution:

(a) We want to show $\forall \varepsilon > 0$

$$\limsup_{n \rightarrow \infty} P \left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} |H_n(\theta) - H_n(\theta')| > \varepsilon \right) < \varepsilon$$

$$\begin{aligned}
 H(w_i, \theta_1) - H(w_i, \theta_2) &= \frac{1}{n} \sum_{i=1}^n (h(w_i, \theta_1) - h(w_i, \theta_2)) \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{\partial h(w_i, \theta^{*})}{\partial \theta'} (\theta_1 - \theta_2)
 \end{aligned}$$

Then

$$\begin{aligned}
 |H(w_i, \theta_1) - H(w_i, \theta_2)| &\leq \frac{1}{n} \sum_{i=1}^n \left\| \frac{\partial h(w_i, \theta^{*})}{\partial \theta'} \right\| \|\theta_1 - \theta_2\| \\
 &\leq \frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \left\| \frac{\partial h(w_i, \theta)}{\partial \theta'} \right\| \|\theta_1 - \theta_2\| \\
 &\quad \underbrace{\text{E} \sup_{\theta \in \Theta} \left\| \frac{\partial h(w_i, \theta)}{\partial \theta'} \right\|}_{+ \text{op}(1)} \\
 &\leq \frac{1}{n} \sum_{i=1}^n (K + \text{op}(1)) \|\theta_1 - \theta_2\|
 \end{aligned}$$

Next,

$$\sup_{\theta_1 \in \Theta} \sup_{\theta_2 \in B(\theta_1, \delta)} |H(w_i, \theta_1) - H(w_i, \theta_2)| \leq (K + \text{op}(1)) \delta = \delta K + \text{op}(1)$$

let's put $P(\cdot)$ measure

$$\begin{aligned}
 P\left(\sup_{\theta_1 \in \Theta} \sup_{\theta_2 \in B(\theta_1, \delta)} |H(w_i, \theta_1) - H(w_i, \theta_2)| > \epsilon\right) \\
 &\leq \\
 P\left(\underbrace{\delta K + \text{op}(1)}_{\text{must be less than } \epsilon \text{ then op}(1) \text{ term will dissipate.}} > \epsilon\right)
 \end{aligned}$$

$$\text{choose } \delta K < \epsilon \Rightarrow \delta < \frac{\epsilon}{K} \text{ so for instance } \delta = \frac{\epsilon}{2K}$$

would work.

Finally,

$$\limsup_{n \rightarrow \infty} P\left(\sup_{\theta_1 \in \Theta} \sup_{\theta_2 \in B(\theta_1, \delta)} |H(w_i, \theta_1) - H(w_i, \theta_2)| > \epsilon\right) \leq \limsup_{n \rightarrow \infty} P(\text{op}(1) > \epsilon/2) = 0.$$

(b) We want to show $\sup_{\theta \in \Theta} |H_n(\theta)| \xrightarrow{P} 0$

Let $\{B(\theta_j, \delta) : j=1, \dots, J\}$ be a finite cover of Θ , with $\theta_j \in \Theta$ (we can always find such θ_j because Θ is dense). Then every $\theta \in \Theta$ is within some $B(\theta_j, \delta)$ for some j . Write

$$\cdot |H_n(\theta)| \leq \sup_{\theta' \in B(\theta_j, \delta)} |H_n(\theta')|$$

$$\leq \sup_{\theta' \in B(\theta_j, \delta)} |H_n(\theta') - H_n(\theta_j)| + |H_n(\theta_j)|$$

$$\cdot \sup_{\theta \in \Theta} |H_n(\theta)| \leq \max_{1 \leq j \leq J} \sup_{\theta' \in B(\theta_j, \delta)} |H_n(\theta') - H_n(\theta_j)| + \max_{1 \leq j \leq J} |H_n(\theta_j)|$$

$$\leq \sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} |H_n(\theta') - H_n(\theta)| + \max_{1 \leq j \leq J} |H_n(\theta_j)|$$

$$P\left(\sup_{\theta \in \Theta} |H_n(\theta)| > 2\varepsilon\right) \leq P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} |H_n(\theta') - H_n(\theta)| + \max_{1 \leq j \leq J} |H_n(\theta_j)| > 2\varepsilon\right)$$

$$\leq P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta_j, \delta)} |H_n(\theta') - H_n(\theta_j)| > \varepsilon\right)$$

$$+ \\ P\left(\max_{1 \leq j \leq J} |H_n(\theta_j)| > \varepsilon\right)$$

$$\leq P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta_j, \delta)} |H_n(\theta') - H_n(\theta_j)| > \varepsilon\right)$$

$$+ \\ \sum_{j=1}^J P(|H_n(\theta_j)| > \varepsilon)$$

$$\lim_{n \rightarrow \infty} P\left(\sup_{\theta \in \Theta} |H_n(\theta)| > 2\varepsilon\right) \leq \limsup_{n \rightarrow \infty} P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta_j, \delta)} |H_n(\theta') - H_n(\theta_j)| > \varepsilon\right)$$

$< \varepsilon$ by SE

$$+ \\ \sum_{j=1}^J \limsup_{n \rightarrow \infty} P(|H_n(\theta_j)| > \varepsilon)$$

Notice that

$H_n(\theta) \xrightarrow{P} E h(w_i, \theta) = 0 \quad \forall \theta \in \Theta$
and that includes all θ_j !

$$\leq \varepsilon + \sum_{j=1}^J \limsup_{n \rightarrow \infty} P(|H_n(\theta_j)| > \varepsilon) = \varepsilon.$$

■