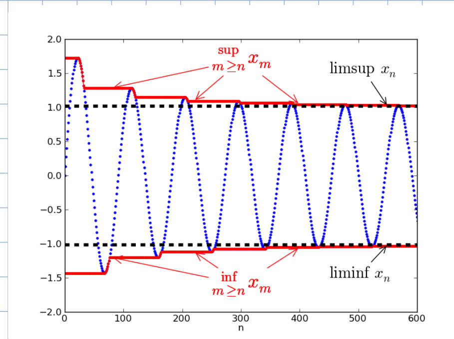
Let's begin with some definitions. Let x_n and a_n be a sequence of constants.

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

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

- $\limsup\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty} \sup\limits_{m\geq n} x_m \equiv \overline{\lim\limits_{n\to\infty}} x_n$

  $\liminf\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty} \inf\limits_{m\geq n} x_m \equiv \underline{\lim\limits_{n\to\infty}} x_n$



(Taken from Wikipedia)

Now let x_n be a sequence of random vectors.

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

- $x_n = O_p(a_n)$ means $\forall \varepsilon, \exists M_\varepsilon > 0$ such that
  $$\limsup_{n\to\infty} P\left\{ \left\| \frac{x_n}{a_n} \right\| > M_\varepsilon \right\} < \varepsilon$$

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

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

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

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

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

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

- Stochastic Equicontinuity : $\forall \varepsilon, \exists \delta$ s.t. $\limsup_{n\to\infty} P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta,\delta)} |H_n(\theta) - H_n(\theta')| > \varepsilon \right) < \varepsilon$

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

$$\forall \varepsilon, \exists \delta \text{ s.t. } \limsup_{n\to\infty} P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta,\delta)} \| H_n(\theta) - H_n(\theta')\| > \varepsilon \right) < \varepsilon$$

Consistency

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

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

(3) ID : $\exists \theta_0 \in \Theta$ such that $\forall \varepsilon > 0$

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

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

proof :

$$P(\hat{\theta}_n \notin B(\theta_0,\varepsilon)) \equiv P(\|\hat{\theta}_n - \theta_0\| \geq \varepsilon) \underset{(3)}{\leq} P(Q(\hat{\theta}_n) - Q(\theta_0) \geq \delta)$$

for some δ

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

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

Why? $Q_n(\hat{\theta}_n) \leq \inf_{\theta \in \Theta} Q_n(\theta) + o_p(1) \leq Q_n(\theta_0) + o_p(1)$

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

$$\leq P\Big(\underbrace{2\sup_{\theta \in \Theta} |Q(\theta) - Q_n(\theta)|}_{o_p(1) \text{ by } (2)} + o_p(1) \geq \delta\Big)$$

Then

$$\lim_{n\to\infty} P\left(\hat{\theta_n} \notin B(\theta_0, \varepsilon)\right) \leq \lim_{n\to\infty} P\left(o_p(1) \geq \delta\right) = 0. \quad \blacksquare$$

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) $\sqrt{n}\, \dfrac{\partial Q_n(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Omega_0)$

(iv) $\sup_{\theta \in \Theta_0} \left\| \dfrac{\partial^2 Q_n(\theta)}{\partial\theta\,\partial\theta'} - B(\theta) \right\| \xrightarrow{p} 0$ 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) $\dfrac{\partial Q_n(\hat{\theta_n})}{\partial \theta} = o_p\left(\frac{1}{\sqrt{n}}\right)$

Theorem.- ① and ② hold $\Rightarrow$ $\sqrt{n}(\hat{\theta_n} - \theta_0) \xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1})$

proof:

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{\partial Q_n(\hat{\theta_n})}{\partial \theta} \underset{\substack{\text{mean}\\\text{value}\\\text{exp around}\\\theta_0}}{=} \frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_n(\theta_n^*)}{\partial\theta\,\partial\theta'}(\hat{\theta_n} - \theta_0)$$

(*) $\hat{\theta_n} \xrightarrow{p} \theta_0$ implies $\theta_n^* \xrightarrow{p} \theta_0$

Multiply $\sqrt{n}$

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

$$\Rightarrow \quad \sqrt{n}(\hat{\theta_n} - \theta_0) = -B_0^{-1}\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta_0} + o_p(1).$$

$$\xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1})$$

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

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

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

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

- $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ for $\forall \omega$
- $|f_n(\omega)| \leq \underbrace{g(\omega)}_{\text{dominating function.}}$ and g is integrable

Then

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega \lim_{n\to\infty} f_n \, d\mu \equiv \int_\Omega f \, d\mu.$$

<u>Example 1</u>: 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 $g \equiv \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} \log f(W_i, \theta) \right\|$ 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

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

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

The true value satisfies

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

We need for CF (ii) that

$$E (Y_i - g(X_i,\theta)) \frac{\partial g(X_i,\theta)}{\partial \theta} = 0$$

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

assume $E \sup_{\theta \in \Theta} \| (Y_i - g(X_i,\theta)) \frac{\partial}{\partial \theta} g(X_i,\theta) \| < \infty$. 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: $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \to_p 0$.

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

$\forall \varepsilon \quad \lim_{n \to \infty} P(\underbrace{\|\hat{\theta}_n - \theta\|}_{d(\hat{\theta}_n, \theta)} > \varepsilon) = 0$ is how we define it.

where θ is the identifying set.

now, define it as

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

2. We want to show that

$$\forall \varepsilon \quad \lim_{n \to \infty} \underbrace{P(d(\hat{\theta}_n, \Theta_0) > \varepsilon)}_{\text{start with this measure}} = 0$$

$$P(d(\hat{\theta}_n, \Theta_0) > \varepsilon) \underset{ID}{\leq} P(Q(\hat{\theta}_n) - \underline{Q} > \delta_\varepsilon) \text{ for some } \delta_\varepsilon$$

$$= P(Q(\hat{\theta}_n) \pm Q_n(\hat{\theta}_n) - \underline{Q} > \delta_\varepsilon)$$

$$\underset{EE}{\leq} P(Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n) + Q_n(\theta_0) - \underline{Q} + o_p(1) > \delta_\varepsilon)$$

for any $\theta_0 \in \Theta_0$

$$\leq P(\ |Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n)| + |Q_n(\theta_0) - Q(\theta_0)| + o_p(1) > \delta_\varepsilon)$$

$$\leq P(2 \sup_{\theta \in \Theta} |Q(\theta) - Q_n(\theta)| + o_p(1) > \delta_\varepsilon)$$

$$= P(\ o_p(1) > \delta_\varepsilon)$$

Hence

$$P(\ d(\hat{\theta}_n, \Theta_0) > \varepsilon) \leq P(\ o_p(1) > \delta_\varepsilon)$$

$$\Rightarrow \lim_{n\to\infty} P(\ d(\hat{\theta}_n, \Theta_0) > \varepsilon) \leq \lim_{n\to\infty} P(\ o_p(1) > \delta_\varepsilon) = 0$$

<u>Midterm 2017 Q2</u>

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, $\{W_i : i = 1, \ldots, n\}$ 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 $\sup_{\theta \in \Theta} |H_n(\theta)| \to_p 0$.

<u>solution:</u>

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

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

$$H(w_i,\theta_1) - H(w_i,\theta_2) = \frac{1}{n}\sum_{i=1}^{n}\left(h(w_i,\theta_1) - h(w_i,\theta_2)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial h(w_i,\theta_i^*)}{\partial \theta'}(\theta_1 - \theta_2)$$

Then

$$|H(w_i,\theta_1) - H(w_i,\theta_2)| \leq \frac{1}{n}\sum_{i=1}^{n} \left\| \frac{\partial h(w_i,\theta_i^*)}{\partial\theta'} \right\| \|\theta_1 - \theta_2\|$$

$$\leq \frac{1}{n}\sum_{i=1}^{n} \underbrace{\sup_{\theta\in\Theta} \left\| \frac{\partial h(w_i,\theta)}{\partial\theta'}\right\|}_{E \sup_{\theta\in\Theta} \left\| \frac{\partial h(w_i,\theta)}{\partial\theta'}\right\| + o_p(1)} \|\theta_1-\theta_2\|$$

$$\leq \cancel{\frac{1}{n}\sum_{i=1}^{n}}(K + o_p(1)) \quad \|\theta_1 - \theta_2\|$$

Next,

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

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

$$P\left(\sup_{\theta_1\in\Theta} \sup_{\theta_2\in B(\theta_1,\delta)} |H(w_i,\theta_1) - H(w_i,\theta_2)| > \varepsilon\right)$$

$$\leq$$

$$P(\underbrace{\delta K}_{\text{must be less than } \varepsilon} + o_p(1) > \varepsilon)$$

then $o_p(1)$ term will dissipate.

choose $\delta K < \varepsilon \Rightarrow \delta < \frac{\varepsilon}{K}$ so for instance $\delta = \frac{\varepsilon}{2K}$ would work.

Finally,

$$\limsup_{n\to\infty} P\left(\sup_{\theta_1\in\Theta} \sup_{\theta_2\in B(\theta_1,\delta)} |H(w_i,\theta_1) - H(w_i,\theta_2)| > \varepsilon\right) \leq \limsup_{n\to\infty} P(o_p(1) > \varepsilon/2)$$

$$= 0. \quad \blacksquare$$

(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

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

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

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

$$\lim_{n \to \infty} P\left(\sup_{\theta \in \Theta} |H_n(\theta)| > 2\varepsilon\right) \leq \limsup_{n \to \infty} \underbrace{P\left(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta_j, \delta)} |H_n(\theta') - H_n(\theta_j)| > \varepsilon\right)}_{< \varepsilon \text{ by SE}}$$
$$+ \sum_{j=1}^{J} \limsup_{n \to \infty} \underbrace{P(|H_n(\theta_j)| > \varepsilon)}$$

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

$$\leq \varepsilon + \sum_{j=1}^{J} \limsup_{n \to \infty} P(o_p(1) > \varepsilon) = \varepsilon. \quad \blacksquare$$

ECON 626 PS4 Q3 (2018)

Suppose that $f : \mathbf{R} \to \mathbf{R}$ is a function that is continuous at $c \in \mathbf{R}$. Show that

$$f(c + o_P(1)) = f(c) + o_P(1).$$

solution:

Recall that continuity at c means $\theta = c + x_n$, $\sup_{\theta \in B(c,\delta)} | f(\theta) - f(c)| < \varepsilon$ (whenever $\theta \in B(c,\delta)$ or $|x_n| < \delta$ then $|f(\theta) - f(c)| < \varepsilon$)

We want to show $\lim_{n\to\infty} \Pr(|f(c + x_n) - f(c)| > \varepsilon) = 0$, where $x_n = o_p(1)$.

$$|x_n| < \delta \subseteq |f(\theta) - f(c)| < \varepsilon$$

$$1 + \Pr(|x_n| < \delta) \leq \Pr(|f(\theta) - f(c)| < \varepsilon) + 1$$

$$1 - \Pr(|f(\theta) - f(c)| < \varepsilon) \leq 1 - \Pr(|x_n| < \delta)$$

$$\Pr(|f(c + x_n) - f(c)| \geq \varepsilon) \leq \Pr(|x_n| \geq \delta)$$

$$\lim_{n\to\infty} \Pr(|f(c + x_n) - f(c)| \geq \varepsilon) \leq \lim_{n\to\infty} \Pr(o_p(1) \geq \delta) = 0. \blacksquare$$