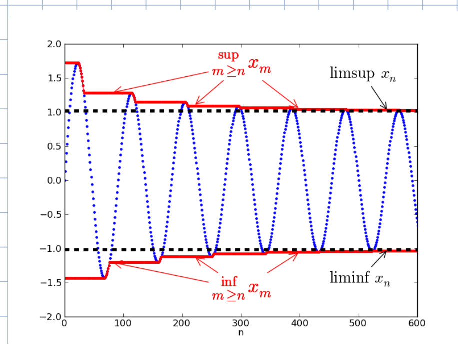
Let's begin with some definitions. Let x_n and a_n be a sequence of constants.

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

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

- $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{m\geq n} x_m \equiv \overline{\lim_{n\to\infty}} x_n$

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



(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. $\sup_{\theta' \in B(\theta,\delta)} | H(\theta) - H(\theta') | < \varepsilon$

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

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. $\sup\limits_{\theta \in \Theta} \sup\limits_{\theta' \in B(\theta,\delta)} |h(\theta) - h(\theta')| < \varepsilon \quad \forall\, h \in \{H_n(\cdot), n \geq 1\}$

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. $\limsup\limits_{n \to \infty} P\left(\sup\limits_{\theta \in \Theta} \sup\limits_{\theta' \in B(\theta,\delta)} |H_n(\theta) - H_n(\theta')| > \varepsilon \right) < \varepsilon$

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** : $\hat{\theta}_n \in \Theta : Q_n(\hat{\theta}_n) \leq \inf\limits_{\theta \in \Theta} Q_n(\theta) + o_p(1)$

(2) **UWC** : $\sup\limits_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow[p]{} 0$

(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 :

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

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

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

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

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

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

Then

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

Asymptotic Normality

① CF :

(i) $\theta_0 \in \text{int}(\Theta)$

(ii) $Q_n(\theta)$ is twice continously 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 continous at θ_0 and for which $B(\theta_0)$ is non singular.

② EE2 :

(i) $\hat{\theta_n} \xrightarrow{p} \theta_0$

(ii) $\dfrac{\partial Q_n(\hat{\theta_n})}{\partial\theta} = o_p\left(\dfrac{1}{\sqrt{n}}\right)$

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 $\sqrt{n}$

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

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

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

(*) 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.

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

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

Then

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

Example 1: 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

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

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

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

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

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

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

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

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

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

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

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

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

solution:

(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

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

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

$$\le (K + o_p(1))\ \|\theta_1-\theta_2\|$$

Next,

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

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

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

$$\le$$

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

must be less than ε 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,

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

$$= 0. \quad \blacksquare$$

(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

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

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

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

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

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 $x_n = o_p(1)$.

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

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

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

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

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

1. Let $\{X_i : i = 1, \ldots .n\}$ be iid random k-vectors such that $EX_i \neq 0$, and $Var(X_i)$ is finite and positive definite.

   (a) **(10 points)** Show that the family of random functions $\{Q_n(\theta) : n \geq 1\}$, where

$$Q_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \theta' X_i, \quad \theta \in \mathbb{R}^k,$$

   is not stochastically equicontinuous.

solution:

We want to show $\limsup_{n\to\infty} P\left(\sup_{\theta \in \mathbb{R}^k} \sup_{\tilde{\theta} \in B(\theta,\delta)} |Q_n(\theta) - Q_n(\tilde{\theta})| > \varepsilon\right) \geq \varepsilon$

Notice now we're looking inequality in the opposite direction

$$Q_n(\theta_1) - Q_n(\theta_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\theta_1 - \theta_2)'(X_i \pm EX_i)$$

$$= \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\theta_1 - \theta_2)'(X_i - EX_i)}_{\text{we can apply CLT}} + \sqrt{n}(\theta_1 - \theta_2)' EX_i$$

$\xrightarrow{d} N(0, (\theta_1 - \theta_2)' Var(X_i) (\theta_1 - \theta_2))$
$= O_p(1)$.

Now, recall that we want

$$\sup_{\theta \in \mathbb{R}^K} \sup_{\tilde{\theta} \in B(\theta,\delta)} |Q_n(\theta) - Q_n(\tilde{\theta})| \geq \sup_{\tilde{\theta} \in B(\theta,\delta)} |Q_n(\theta) - Q_n(\tilde{\theta})|$$

so choose θ_1 and θ_2 such that $\|\theta_1 - \theta_2\| \leq \delta$. We will consider that from this point onwards.

Then

$$P\left(\sup_{\theta \in \mathbb{R}^K} \sup_{\tilde{\theta} \in B(\theta,\delta)} |Q_n(\theta) - Q_n(\tilde{\theta})| > \varepsilon\right) \geq P(|Q_n(\theta_1) - Q_n(\theta_2)| > \varepsilon)$$

$$= P(|O_p(1) + \sqrt{n}(\theta_1 - \theta_2)' EX_i| > \varepsilon)$$

$$= P(O_p(1) + \sqrt{n}(\theta_1 - \theta_2)' EX_i > \varepsilon)$$

$$+$$

$$P(O_p(1) + \sqrt{n}(\theta_1 - \theta_2)' EX_i < -\varepsilon)$$

$$= P(O_p(1) > \varepsilon - \sqrt{n}(\theta_1 - \theta_2)' EX_i)$$

$$+$$

$$P(O_p(1) < -\varepsilon - \sqrt{n}(\theta_1 - \theta_2)' EX_i)$$

<u>CASE 1)</u> $(\theta_1 - \theta_2)' EX_i > 0 \Leftrightarrow -\sqrt{n}(\theta_1 - \theta_2)' EX_i \to -\infty$ as $n \to \infty$

$P(O_p(1) > \varepsilon - \sqrt{n}(\theta_1 - \theta_2)' EX_i) \to 1$

$P(O_p(1) < -\varepsilon - \sqrt{n}(\theta_1 - \theta_2)' EX_i) \to 0$

So the sum $\to 1$.

<u>CASE 2)</u> $(\theta_1 - \theta_2)' EX_i < 0 \Leftrightarrow -\sqrt{n}(\theta_1 - \theta_2)' EX_i \to \infty$ as $n \to \infty$

$P(O_p(1) > \varepsilon - \sqrt{n}(\theta_1 - \theta_2)' EX_i) \to 0$

$P(O_p(1) < -\varepsilon - \sqrt{n}(\theta_1 - \theta_2)' EX_i) \to 1$

So the sum $\to 1$.

(b) **(10 points)** Show that the family of random functions $\{H_n(\theta) : n \geq 1\}$, where

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \theta' X_i, \quad \theta \in \mathbb{R}^k,$$

is stochastically equicontinuous.

solution:

We want to show $\limsup_{n\to\infty} P\left(\sup_{\theta\in\mathbb{R}^k} \sup_{\tilde{\theta}\in B(\theta,\delta)} |Q_n(\theta) - Q_n(\tilde{\theta})| > \varepsilon\right) < \varepsilon$

$$H_n(\theta_1) - H_n(\theta_2) = \frac{1}{n}\sum_{i=1}^{n} (\theta_1-\theta_2)' X_i$$

$$|H_n(\theta_1) - H_n(\theta_2)| \leq \frac{1}{n}\sum_{i=1}^{n} \|\theta_1-\theta_2\| \, \|X_i\|$$

$$\sup_{\theta_2\in B(\theta_1,\delta)} |H_n(\theta_1) - H_n(\theta_2)| \leq \frac{1}{n}\sum_{i=1}^{n} \delta \, \|X_i\|$$

then

$$P\left(\sup_{\theta_1\in\Theta} \sup_{\theta_2\in B(\theta_1,\delta)} |H_n(\theta_1) - H_n(\theta_2)| > \varepsilon\right) \leq P\left(\frac{1}{n}\sum_{i=1}^{n} \delta\|X_i\| > \varepsilon\right)$$

$$= P\left(\frac{1}{n}\sum_{i=1}^{n} \|X_i\| > \frac{\varepsilon}{\delta}\right)$$

$$= P\left(E\|X_i\| + o_p(1) > \frac{\varepsilon}{\delta}\right)$$

$$= \underbrace{P\left(o_p(1) > \frac{\varepsilon}{\delta} - E\|X_i\|\right)}_{\to 0 \text{ as long as } \frac{\varepsilon}{\delta} - E\|X_i\| > 0}$$

$$\Rightarrow \delta < \frac{\varepsilon}{E\|X_i\|}$$

choose $\delta = \dfrac{\varepsilon}{2E\|X_i\|}$.

so the result follows.

(c) **(10 points)** Show that the family of random functions $\{G_n(\theta) : n \geq 1\}$, where

$$G_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \theta' (X_i - EX_i), \quad \theta \in \mathbb{R}^k,$$

is stochastically equicontinuous.

solution:

We want to show $\limsup_{n\to\infty} P\left(\sup_{\theta_1 \in \Theta} \sup_{\theta_2 \in B(\theta_1,\delta)} |G_n(\theta_1) - G_n(\theta_2)| > \varepsilon \right) < \varepsilon$

$$G_n(\theta_1) - G_n(\theta_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\theta_1 - \theta_2)' (X_i - EX_i)$$

$$|G_n(\theta_1) - G_n(\theta_2)| \leq \|\theta_1 - \theta_2\| \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i) \right\|$$

$$\sup_{\theta_2 \in B(\theta_1,\delta)} |G_n(\theta_1) - G_n(\theta_2)| \leq \delta \left\| \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i)}_{O_p(1)} \right\|$$

Recall that $Y_n = O_p(1)$ means that $\exists\, M_\varepsilon < \infty$ s.t.

$$\limsup_{n\to\infty} \Pr\left(\|Y_n\| > M_\varepsilon \right) < \varepsilon$$

Write

$$\limsup_{n\to\infty} P\left(\sup_{\theta_1 \in \Theta} \sup_{\theta_2 \in B(\theta_1,\delta)} |G_n(\theta_1) - G_n(\theta_2)| > \varepsilon \right) \leq \limsup_{n\to\infty} P\left(\delta \|O_p(1)\| > \varepsilon \right)$$

So we need $\frac{\varepsilon}{\delta} \geq M_\varepsilon \Rightarrow \delta \leq \frac{\varepsilon}{M_\varepsilon}$

and the result follows.

FINAL 2019 Q2

2. **(Logit Binary Choice model)** Let iid data $\{(X_i', Y_i)' : i = 1, \ldots, n\}$, where $Y_i \in \{0,1\}$ is a binary variable, be generated according to the model

$$E(Y_i \mid X_i) = P(Y_i = 1 \mid X_i) = \Lambda(X_i'\beta_0),$$

where $\Lambda(u) = e^u/(1+e^u)$ is the CDF of the Logistic distribution, X_i is the k-vector of regressors, and $\beta_0 \in \mathbb{R}^k$ is the unknown vector of parameters. Note that the conditional distribution of Y_i conditional on X_i can be described as

$$P(Y_i = y \mid X_i) = \left(\Lambda(X_i'\beta_0)\right)^y \left(1 - \Lambda(X_i'\beta_0)\right)^{1-y}, \quad y \in \{0,1\}.$$

Let $\hat{\beta}_n$ denote the maximum likelihood estimator (MLE) of β_0:

$$\hat{\beta}_n = \arg\max_{\beta \in \mathbb{R}^k} Q_n(\beta), \text{ where}$$
$$Q_n(\beta) = n^{-1} \sum_{i=1}^{n} \left\{ Y_i \ln \Lambda(X_i'\beta) + (1 - Y_i) \ln\left(1 - \Lambda(X_i'\beta)\right) \right\}.$$

Let $\tilde{\beta}_n$ denote the nonlinear least squares (NLS) estimator of β_0:

$$\tilde{\beta}_n = \arg\min_{\beta \in \mathbb{R}^k} R_n(\beta), \text{ where}$$
$$R_n(\beta) = n^{-1} \sum_{i=1}^{n} \left(Y_i - \Lambda(X_i'\beta)\right)^2 / 2.$$

(a) **(3 points)** What assumption does one need to impose on X_i to ensure identification of β_0?

solution:

First, consider the case of MLE

$$Q(\beta) = E\{ Y_i \ln \Lambda(x_i'\beta) + (1-Y_i) \ln(1-\Lambda(x_i'\beta)) \}$$

Notice that we need to show that $Q(\hat{\beta}_n) - Q(\beta_0) < 0$, i.e. $Q(\beta_0)$ is the unique maximizer of MLE.

now,

- $$Q(\hat{\beta}_n) - Q(\beta_0) = E\left\{ Y_i \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} + (1-Y_i) \ln \frac{1-\Lambda(X_i'\hat{\beta}_n)}{1-\Lambda(X_i'\beta_0)} \right\}$$

$$\leq E\left\{ Y_i \left(\frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} - 1 \right) + (1-Y_i) \frac{1-\Lambda(X_i'\hat{\beta}_n)}{1-\Lambda(X_i'\beta_0)} \right\}$$

$$= 0.$$

So when is it zero? i.e. it achieves the upper bound

- $$Q(\hat{\beta}_n) - Q(\beta_0) = E\left\{ Y_i \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} + (1-Y_i) \ln \frac{1-\Lambda(X_i'\hat{\beta}_n)}{1-\Lambda(X_i'\beta_0)} \right\}$$

$$= \underbrace{E\left\{ Y_i \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} + (1-Y_i) \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} \,\middle|\, X_i'\hat{\beta}_n = X_i'\beta_0 \right\}}_{=0} P(X_i'\hat{\beta}_n = X_i'\beta_0)$$

$$+$$

$$E\left\{ Y_i \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} + (1-Y_i) \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} \,\middle|\, X_i'\hat{\beta}_n \neq X_i'\beta_0 \right\} \underbrace{P(X_i'\hat{\beta}_n \neq X_i'\beta_0)}_{\text{if this } =0}$$

$$= 0.$$

Then, our condition is $P(X_i'\hat{\beta}_n \neq X_i'\beta_0) > 0$.

Next, for NLS we have

$$R(\beta) = \frac{1}{2} E\,(Y_i - \Lambda(X_i'\beta))^2$$

$$R_n(\beta) = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} (Y_i - \Lambda(X_i'\beta))^2$$

We want to show that $R(\hat{\beta}_n) - R(\beta_0) > 0$.

$$R(\hat{\beta}_n) - R(\beta_0) = \frac{1}{2}\left\{ E(\underbrace{Y_i - \Lambda(X_i'\beta_0)}_{u_i} + \Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))^2 - E\,u_i^2 \right\}$$

$$= \frac{1}{2}\left\{ E(\cancel{u_i^2} + 2u_i(\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n)) + (\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))^2) - \cancel{E u_i^2} \right\}$$

$$= \frac{1}{2}\left\{ E[\underbrace{E(u_i|X_i)}_{=0}(\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))] + \underbrace{E(\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))^2}_{\text{can only be zero when } P(\Lambda(X_i'\beta_0) = \Lambda(X_i'\hat{\beta}_n)) = 1} \right\} \geq 0$$

so the condition is the same.

(b) **(12 points)** Find the asymptotic variance of the MLE $\hat{\beta}_n$. Assume that EX_iX_i' is positive definite and finite. Hints: (i) Use the property

$$\frac{d\Lambda(u)}{du} = \Lambda(u)\left(1 - \Lambda(u)\right)$$

to show that

$$\frac{\partial Q_n(\beta_0)}{\partial \beta} = n^{-1} \sum_{i=1}^{n} \left(Y_i - \Lambda(X_i'\beta_0)\right) X_i.$$

(ii) Show that

$$Var(Y_i \mid X_i) = \Lambda(X_i'\beta_0)\left(1 - \Lambda(X_i'\beta_0)\right).$$

Solution:

(i)

$$\frac{\partial Q_n(\beta_0)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i \frac{\partial \Lambda(X_i'\beta_0)/\partial\beta}{\Lambda(X_i'\beta_0)} - (1-Y_i) \frac{\partial \Lambda(X_i'\beta_0)/\partial\beta}{1-\Lambda(X_i'\beta_0)} \right\}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i \frac{\cancel{\Lambda(X_i'\beta_0)}(1-\Lambda(X_i'\beta_0))X_i}{\cancel{\Lambda(X_i'\beta_0)}} - (1-Y_i) \frac{\Lambda(X_i'\beta_0)\cancel{(1-\Lambda(X_i'\beta_0))}X_i}{\cancel{1-\Lambda(X_i'\beta_0)}} \right\}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\{ (Y_i - \Lambda(X_i'\beta_0)) X_i \right\}$$

(ii)

$$Var(Y_i \mid X_i) = E\left\{ (Y_i - E(Y_i \mid X_i))^2 \mid X_i \right\}$$

$$= E\left\{ (Y_i - \Lambda(X_i'\beta_0))^2 \mid X_i \right\}$$

$$= E\left\{ \underbrace{Y_i^2}_{\text{binary}} - 2Y_i\Lambda(X_i'\beta_0) + \Lambda(X_i'\beta_0)^2 \mid X_i \right\}$$

$$= E\left\{ Y_i - 2Y_i\Lambda(X_i'\beta_0) + \Lambda(X_i'\beta_0)^2 \mid X_i \right\}$$

$$= \Lambda(X_i'\beta_0) - 2\Lambda(X_i'\beta_0)^2 + \Lambda(X_i'\beta_0)^2$$

$$= \Lambda(X_i'\beta_0)\left(1 - \Lambda(X_i'\beta_0)\right).$$

Then

$$\sqrt{n}\,\frac{\partial Q_n(\beta_0)}{\partial \beta} \to N\left(0, E(Y_i - \Lambda(X_i'\beta_0))^2 X_iX_i'\right)$$

$$= N\left(0, E\,\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))X_iX_i'\right)$$

$$= N(0, \Omega_0)$$

Recall that

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{\partial Q_n(\hat{\beta}_n)}{\partial \beta} = \frac{\partial Q_n(\beta_0)}{\partial \beta} + \frac{\partial Q_n(\beta^*)}{\partial\beta\,\partial\beta'}(\hat{\beta}_n - \beta_0)$$

where

$$\frac{\partial Q_n(\beta_0)}{\partial\beta\,\partial\beta'} = -\frac{1}{n}\sum_{i=1}^{n} \Lambda(x_i'\beta_0)(1-\Lambda(x_i'\beta_0))\, x_i x_i'$$

$$\xrightarrow{p} E[\Lambda(x_i'\beta_0)(1-\Lambda(x_i'\beta_0))\, x_i x_i'] = \Omega_0(\beta_0)$$

Finally,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = -\Omega_0^{-1}\sqrt{n}\frac{\partial Q_n(\beta_0)}{\partial\beta} + o_p(1)$$

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

(c) **(12 points)** Find the asymptotic variance of the NLS estimator $\tilde{\beta}_n$.

<u>solution:</u>

Again, we require to analyze the following objects

$$\frac{\partial R_n(\beta_0)}{\partial\beta} = -\frac{1}{n}\sum_{i=1}^{n}(y_i - \Lambda(x_i'\beta_0))\frac{\partial\Lambda(x_i'\beta_0)}{\partial\beta}$$

$$= -\frac{1}{n}\sum_{i=1}^{n}(y_i - \Lambda(x_i'\beta_0))\Lambda(x_i'\beta_0)(1-\Lambda(x_i'\beta_0))\,x_i$$

Hence,

$$\sqrt{n}\frac{\partial R_n(\beta_0)}{\partial\beta} \xrightarrow{d} N(0, E[(y_i-\Lambda(x_i'\beta_0))^2\Lambda(x_i'\beta_0)^2(1-\Lambda(x_i'\beta_0))^2 x_i x_i'])$$

$$\underset{\text{LIE}}{=} N(0, \underbrace{E[\Lambda(x_i'\beta_0)(1-\Lambda(x_i'\beta_0))]^3 x_i x_i'}_{\Sigma_0})$$

Next,

$$\frac{\partial^2 R_n(\beta_0)}{\partial\beta\,\partial\beta'} = \frac{1}{n}\sum_{i=1}^{n}\left\{-\underbrace{(y_i - \Lambda(x_i'\beta_0))}_{=o_p(1)}\frac{\partial^2\Lambda(x_i'\beta_0)}{\partial\beta\,\partial\beta'} + \frac{\partial\Lambda(x_i'\beta_0)}{\partial\beta}\frac{\partial\Lambda(x_i'\beta_0)}{\partial\beta'}\right\}$$

$$= E\frac{\partial\Lambda(x_i'\beta_0)}{\partial\beta}\frac{\partial\Lambda(x_i'\beta_0)}{\partial\beta'} + o_p(1)$$

$$= \underbrace{E\{[\Lambda(x_i'\beta_0)(1-\Lambda(x_i'\beta_0))]^2 x_i x_i'\}}_{B_0} + o_p(1)$$

Finally,

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) = -B_0 \sqrt{n} \frac{\partial R_n(\beta_0)}{\partial \beta} + o_p(1)$$

$$\xrightarrow{d} N(0, B_0^{-1} \Sigma_0 B_0^{-1})$$

(d) **(7 points)** Find the asymptotic covariance between the MLE and the NLS estimator. Using the asymptotic covariance, show that the MLE estimator is more efficient than the NLS estimator by comparing their asymptotic variances.

solution:

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = -\Omega_0^{-1} \sqrt{n} \frac{\partial Q_n(\beta_0)}{\partial \beta} + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-\Omega_0^{-1}) \{ (Y_i - \Lambda(X_i'\beta_0)) X_i \} + o_p(1) = \frac{1}{\sqrt{n}} \sum \xi_{1i} + o_p(1)$$

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) = -B_0^{-1} \sqrt{n} \frac{\partial R_n(\beta_0)}{\partial \beta} + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-B_0^{-1})(Y_i - \Lambda(X_i'\beta_0)) \Lambda(X_i'\beta_0)(1 - \Lambda(X_i'\beta_0)) X_i + o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{2i} + o_p(1)$$

So the asymptotic covariance is

$$E \underbrace{\xi_{1i}}_{k\times 1} \underbrace{\xi_{2i}'}_{1\times k} = E\{ -\Omega_0^{-1}(Y_i - \Lambda(X_i'\beta_0)) X_i (Y_i - \Lambda(X_i'\beta_0)) \Lambda(X_i'\beta_0)(1 - \Lambda(X_i'\beta_0)) X_i' (-B_0^{-1}) \}$$

$$= \Omega_0^{-1} E\{ (Y_i - \Lambda(X_i'\beta_0))^2 \Lambda(X_i'\beta_0)(1 - \Lambda(X_i'\beta_0)) X_i X_i' \} B_0^{-1}$$

$$= \Omega_0^{-1} B_0 B_0^{-1} = \Omega_0^{-1}.$$ (notice that it's symmetric $E\xi_{2i}'\xi_{1i}$.)

Recall a similar proof:

$$E((\xi_{1i} - \xi_{2i})(\xi_{1i} - \xi_{2i})') = E\xi_{1i}\xi_{1i}' + E\xi_{2i}\xi_{2i}' - \underbrace{E\xi_{1i}\xi_{2i}'}_{E\xi_{1i}\xi_{1i}'} - \underbrace{E\xi_{2i}\xi_{1i}'}_{E\xi_{1i}\xi_{1i}'} \geq 0$$

$$= E\xi_{2i}\xi_{2i}' - E\xi_{1i}\xi_{1i}' \geq 0$$

$$= \text{AsyVar}(\sqrt{n}(\tilde{\beta}_n - \beta_0)) - \text{AsyVar}(\sqrt{n}(\hat{\beta}_n - \beta_0)) \geq 0$$

so MLE is more efficient.

(e) **(6 points)** The model can also be viewed as a conditional moment restriction model: for some unique β_0,

$$E\left(Y_i - \Lambda(X_i'\beta_0) \mid X_i\right) = 0.$$

Let $\mathcal{G}$ be the set of measurable k-vector valued functions of X_i. Consider a class of estimators $\{\hat{\beta}_n^g : g \in \mathcal{G}\}$, where $\hat{\beta}_n^g$ is defined as a solution to the following sample moment condition:

$$n^{-1}\sum_{i=1}^{n}\left(Y_i - \Lambda(X_i'\hat{\beta}_n^g)\right) g(X_i) = 0.$$

Show that $\hat{\beta}_n^g$ corresponding to the optimal choice of g is the MLE.

solution:

Recall the efficient IV for a conditional moment restriction

$$E\left(m(Y_i, X_i, \beta_0) \mid Z_i \right) = 0$$

$$g^*(Z_i) = \frac{1}{E\left(m^2(Y_i, X_i, \beta_0) \mid Z_i\right)} E\left(\frac{\partial m(Y_i, X_i, \beta_0)}{\partial \beta} \mid Z_i \right)$$

And in this case: $m(Y_i, X_i, \beta_0) = Y_i - \Lambda(X_i'\beta_0)$

$Z_i = X_i$

Then,

- $E\left(\frac{\partial m(Y_i, X_i, \beta_0)}{\partial \beta} \mid X_i \right) = -\Lambda(X_i'\beta_0)\left(1 - \Lambda(X_i'\beta_0)\right) X_i$
- $E\left((Y_i - \Lambda(X_i'\beta_0))^2 \mid X_i \right) = \Lambda(X_i'\beta_0)\left(1 - \Lambda(X_i'\beta_0)\right)$

$\Rightarrow \quad g^*(X_i) = X_i$.

Therefore, the efficient IV estimator satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \Lambda(X_i'\hat{\beta}_n^*)\right) X_i = 0$$

which corresponds to the MLE estimator.

2. **(GMM under misspecification)** In this question, you will derive the asymptotic distribution of a GMM estimator when the moment conditions are misspecified. Let $\{W_i : i = 1. \ldots, n\}$ be a sequence of iid random vectors, and $g(W_i, \theta) \in \mathbb{R}^k$ be a vector-valued function continuous in θ on Θ with probability one, where $\Theta \subset \mathbb{R}^d$ is compact. Assume that $d < k$ (overidentification). Let A be a fixed $k \times k$ matrix, and consider the following GMM estimator:

$$\begin{aligned} \hat{\theta}_n &= \arg\min_{\theta\in\Theta} Q_n(\theta), \text{ where} \\ Q_n(\theta) &= \left\| A \frac{1}{n} \sum_{i=1}^{n} g(W_i, \theta) \right\|^2 /2. \end{aligned}$$

Suppose that the usual identification condition fails, i.e.

$$Eg(W_i, \theta) \neq 0 \text{ for all } \theta \in \Theta.$$

Assume however that there is θ_0 in the interior of Θ such that

$$\begin{aligned} Q(\theta_0) &< Q(\theta) \text{ for all } \theta \neq \theta_0, \theta \in \Theta, \text{ where} \\ Q(\theta) &= \|AEg(W_i, \theta)\|^2 /2. \end{aligned}$$

(a) Explain why $\hat{\theta}_n \to_p \theta_0$ (under some additional technical conditions).

solution:

Assuming (EE), (UWC) and with the assumption we are given we get consistency, regardless of $E g(W_i, \theta) = 0$ or not.

Now recall that in the Linear Algebra part we derived

$$\begin{aligned} \frac{\partial Q_n(\theta_0)}{\partial \theta} &= \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial g(W_i,\theta_0)'}{\partial\theta} - E \frac{\partial g(W_i,\theta_0)'}{\partial\theta} \right)\right] A'A \left[\frac{1}{n}\sum_{i=1}^{n} \left(g(W_i,\theta_0) - Eg(W_i,\theta_0)\right)\right] \\ &+ \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial g(W_i,\theta_0)'}{\partial\theta} - E \frac{\partial g(W_i,\theta_0)'}{\partial\theta} \right)\right] A'A \, E g(W_i,\theta_0) \\ &+ E \frac{\partial g(W_i,\theta_0)'}{\partial\theta} A'A \left[\frac{1}{n}\sum_{i=1} \left(g(W_i,\theta_0) - Eg(W_i,\theta_0)\right)\right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\,\partial\theta'} &= \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta_0)'}{\partial\theta} \right] A'A \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta_0)}{\partial\theta'} \right] \\ &+ \left[I_d \otimes \left(\frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta_0)' A'A \right) \right] \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta'} \operatorname{vec}\left(\frac{\partial g(W_i,\theta_0)}{\partial\theta'} \right) \right] \\ &= B_0 + o_p(1) \end{aligned}$$

(e) Using the results in (a)-(d), show that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal and derive the asymptotic variance-covariance matrix of $\hat{\theta}_n$. Justify any additional assumptions you have to make.

solution:

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

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-B_0^{-1}) \left\{ \left(\frac{\partial g(w_i,\theta_0)'}{\partial \theta} - E\frac{\partial g(w_i,\theta_0)'}{\partial \theta} \right) A'A\, E g(w_i,\theta_0) + E\frac{\partial g(w_i,\theta_0)'}{\partial \theta} A'A \left(g(w_i,\theta_0) - E g(w_i,\theta_0) \right) \right\}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-B_0)^{-1} \xi_i \xrightarrow{d} N(0,\ B_0^{-1} E\xi_i \xi_i' B_0^{-1})$$

provided $E\|\xi_i\|^2 < \infty$

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

(f) Describe how to obtain a robust to misspecification estimator of the asymptotic variance-covariance matrix of the GMM estimator.

solution:

$$\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i \hat{\xi}_i'$$

$$\hat{\xi}_i = \left(\frac{\partial g(w_i,\hat{\theta}_n)'}{\partial \theta} - \frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(w_i,\hat{\theta}_n)'}{\partial \theta} \right) A'A \left(\frac{1}{n} \sum_{i=1}^{n} g(w_i,\hat{\theta}_n) \right)$$
$$+ \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(w_i,\hat{\theta}_n)'}{\partial \theta} \right) A'A \left(g(w_i,\hat{\theta}_n) - \frac{1}{n}\sum_{i=1}^{n} g(w_i,\hat{\theta}_n) \right)$$

$$\hat{B}_n = \left[\frac{1}{n} \sum_{i=1} \frac{\partial g(w_i,\hat{\theta}_n)'}{\partial \theta} \right] A'A \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(w_i,\hat{\theta}_n)}{\partial \theta'} \right]$$
$$+ \left[I_d \otimes \left(\frac{1}{n} \sum_{i=1}^{n} g(w_i,\hat{\theta}_n)' A'A \right) \right] \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta'} \mathrm{vec}\left(\frac{\partial g(w_i,\hat{\theta}_n)}{\partial \theta'} \right) \right]$$

Then let $\hat{\Sigma}_n = \hat{B}_n^{-1} \hat{\Omega}_n \hat{B}_n^{-1}$.

(g) What happens to the estimator in (f) if the model is correctly specified, i.e. if $Eg(W_i, \theta_0) = 0$?

- $$\hat{\xi}_i = \left(\frac{\partial g(W_i, \hat{\theta}_n)}{\partial \theta} - E\frac{\partial g(W_i, \theta_0)'}{\partial \theta} \right) A'A \underbrace{E\, g(W_i, \theta_0)}_{=0} + o_p(1)$$
$$+ E\frac{\partial g(W_i, \theta)'}{\partial \theta} A'A \left(g(W_i, \hat{\theta}_n) - \underbrace{E\, g(W_i, \theta_0)}_{=0} \right) + o_p(1)$$

- $$\hat{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n} E\frac{\partial g(W_i, \theta)'}{\partial \theta} A'A\, g(W_i, \hat{\theta}_n)\, g(W_i, \hat{\theta}_n)' A'A\, E\frac{\partial g(W_i, \theta_0)'}{\partial \theta} + o_p(1)$$
$$= E\frac{\partial g(W_i, \theta_0)'}{\partial \theta} A'A\, E[g(W_i, \theta_0)\, g(W_i, \theta_0)']\, A'A\, E\frac{\partial g(W_i, \theta_0)'}{\partial \theta} + o_p(1)$$

which is the same as the correctly specified GMM.

- $$\hat{B}_n = E\frac{\partial g(W_i, \theta_0)'}{\partial \theta} A'A\, E\frac{\partial g(W_i, \theta_0)}{\partial \theta'} + o_p(1)$$

which is also the same under the correctly specified GMM.

(h) Is there an efficient GMM weight matrix when the model is misspecified? Explain.

solution:

$$Q(\theta) = \frac{1}{2} E g(W_i, \theta)' A'A\, E g(W_i, \theta)$$

F.O.C:

$$\frac{\partial Q(\theta)}{\partial \theta} = \underbrace{E\, \frac{\partial g(w_i,\theta)'}{\partial \theta}\, A'A\, E g(w_i,\theta)}_{\text{if correctly specified } =0 \text{ at } \theta_0 \text{ no matter } A'A} = 0$$

However if $E g(w_i,\theta) \neq 0$ for all $\theta \in \Theta$, we deal with it differently.

Misspecified:

- θ_0 is such that $E\, \frac{\partial g(w_i,\theta_0)'}{\partial \theta}\, A'A\, E\, g(w_i,\theta_0) = 0$

Correctly specified:

- θ_0 is such that $E\, g(w_i,\theta_0) = 0$.

Therefore, now there's no efficient weighting matrix since θ_0 (pseudo true value) depends on $A'A$!