Assume we face F.O.C. that are differentiable. We write

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta}$$

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \underset{K\times 1}{\frac{\partial Q_n(\theta_0)}{\partial \theta}} + \underset{K\times K}{\frac{\partial^2 Q_n(\theta_n^*)}{\partial \theta\, \partial \theta'}} \cdot \underset{K\times 1}{(\hat{\theta}_n - \theta_0)}$$ (by mean value exp)

$\|\theta_n^* - \theta_0\| \leq \underbrace{\|\hat{\theta}_n - \theta_0\|}_{= o_p(1)}$

$$o_p(1) = \sqrt{n}\, \underbrace{\frac{\partial Q_n(\theta_0)}{\partial \theta}}_{\text{Random function but non-random argument}} + \underbrace{\frac{\partial^2 Q_n(\theta_n^*)}{\partial \theta\, \partial \theta'}}_{\text{Random function with random argument.}} \sqrt{n}\,(\hat{\theta}_n - \theta_0)$$ (multiply $\sqrt{n}$)

We need uniform LLN.

If we had uniform (over θ) convergence to a non-random function $B(\cdot)$ that is continuous at θ_0, then we will have a uniform LLN (provided θ_n^* is consistent):

$$\frac{\partial^2 Q_n(\theta_n^*)}{\partial \theta\, \partial \theta'} = B(\theta_n^*) + o_p(1)$$ (by UWC)

$$= B(\theta_0) + o_p(1) \qquad (1)$$ (by $\theta_n^* \xrightarrow{p} \theta_0$ and continuity at θ_0)

To see why the 2nd line is true, write

$$P(\|B(\theta_0) - B(\theta_n^*)\| > \varepsilon) = P(\|B(\theta_0) - B(\theta_n^*)\| > \varepsilon,\ \|\theta_n^* - \theta_0\| \leq \delta)$$
$$+$$
$$P(\|B(\theta_0) - B(\theta_n^*)\| > \varepsilon,\ \|\theta_n^* - \theta_0\| > \delta)$$

Convergence in P + $B(\cdot)$ continuous at θ_0

$$\leq P(\underbrace{\|B(\theta_0) - B(\theta_n^*)\|}_{<\varepsilon \text{ by continuity}} > \varepsilon,\ \|\theta_n^* - \theta_0\| \leq \delta)$$, so the whole thing $= 0$.
$$+$$
$$\underbrace{P(\|\theta_n^* - \theta_0\| > \delta)}_{= o(1)} = o(1)$$ by consistency.

The last part of lecture 11 looks to provide Theorems to deal with this!

Write

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[m(w_i,\theta) - Em(w_i,\theta)\right]$$

and recall

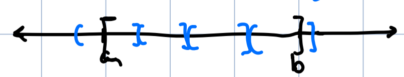
Def. (SE) We say $\{H_n(\theta) : n \geq 1\}$ is stochastically equicontinuous on Θ if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{\theta_0\in\Theta}\ \sup_{\theta_n^*\in B(\theta_0,\delta)} |H_n(\theta_n^*) - H_n(\theta_0)| > \varepsilon\right) < \varepsilon$$

- (Assumption BD) Θ is totally bounded. (There is a finite subcover that covers the entire space)

Example: $\Theta = [a,b]$

finite subcover $\supseteq [a,b]$



- (Assumption P-WCON) $H_n(\theta) \xrightarrow{p} 0$, $\forall \theta \in \Theta$.

- (Assumption SE) $\{H_n(\theta) : n \geq 1\}$ is stochastic equicontinuous on Θ.

- Property U-WCON: $\sup_{\theta\in\Theta} |H_n(\theta)| \xrightarrow{p} 0$.

Theorem 11.2 - (a) BD, P-WCON, SE $\Rightarrow$ U-WCON

(b) U-WCON $\Rightarrow$ P-WCON, SE.

proof:

(a)

$$\limsup_{n\to\infty} P\left(\sup_{\theta\in\Theta}|H_n(\theta)| > \varepsilon\right) \leq \limsup_{n\to\infty} P\left(\sup_{\theta'\in \bigcup_{j\leq J} B(\theta_j,\delta)} |H_n(\theta')| > \varepsilon\right)$$

$$= \limsup_{n\to\infty} P\left(\max_{j\leq J}\ \sup_{\theta'\in B(\theta_j,\delta)} |H_n(\theta')| > \varepsilon\right)$$

$$= \limsup_{n\to\infty} P\left(\max_{j\leq J}\ \sup_{\theta'\in B(\theta_j,\delta)} |H_n(\theta') \pm H_n(\theta_j)| > \varepsilon\right)$$

$$\le \limsup_{n\to\infty} P\Big(\max_{j\le J} \sup_{\theta' \in B(\theta_j,\delta)} |H_n(\theta') - H_n(\theta_j)| + \max_{j\le J} |H_n(\theta_j)| > \varepsilon \Big)$$

by triangle inequality

$P(A+B>\varepsilon) \le P(A>\varepsilon/2) + P(B>\varepsilon/2)$

(at least one of the two must be $>\varepsilon/2$)

$$\le \underbrace{\limsup_{n\to\infty} P\Big(\max_{j\le J} \sup_{\theta' \in B(\theta_j,\delta)} |H_n(\theta') - H_n(\theta_j)| > \varepsilon/2 \Big)}_{<\varepsilon/2 \text{ by SE}}$$

$$+$$

$$\limsup_{n\to\infty} P\Big(\max_{j\le J} |H_n(\theta_j)| > \varepsilon/2 \Big)$$

$$\le \varepsilon/2 + \limsup_{n\to\infty} P\Big(\max_{j\le J} |H_n(\theta_j)| > \varepsilon/2 \Big)$$

$$\le \varepsilon/2 + \sum_{j=1}^{J} \limsup_{n\to\infty} \underbrace{P\big(|H_n(\theta_j)| > \varepsilon/2 \big)}_{o(1) \text{ by P-WCON}}$$

$\le \varepsilon/2$, which completes the proof.

(b) • $\lim\limits_{n\to\infty} P(|H_n(\theta)| > \varepsilon) \le \lim\limits_{n\to\infty} P\Big(\sup\limits_{\theta\in\Theta} |H_n(\theta)| > \varepsilon \Big)$

$$= 0 .$$

• $\overline{\lim\limits_{n\to\infty}} P\Big(\underbrace{\sup\limits_{\theta\in\Theta} \sup\limits_{\theta'\in B(\theta,\delta)} |H_n(\theta') - H_n(\theta)|}_{\substack{\le \sup \sup |H_n(\theta')| + \sup \sup |H_n(\theta)| \\ \le \sup_{\theta\in\Theta} |H_n(\theta)| + \sup_{\theta\in\Theta} |H_n(\theta)|}} > \varepsilon \Big) \le \overline{\lim\limits_{n\to\infty}} P\Big(2\sup\limits_{\theta\in\Theta} |H_n(\theta)| > \varepsilon\Big)$

$$= 0 ,$$

which concludes the proof.

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \big[m(w_i,\theta) - E\, m(w_i,\theta) \big]$$

Theorem 11.3.- (a) Assume

(i) $\{w_i : i \ge 1\}$ are identically distributed

(ii) $m(w_i,\theta)$ is continuous in $\theta\in\Theta$, $\forall\, w \in W$.

(iii) $E \sup\limits_{\theta\in\Theta} |m(w_i,\theta)| < \infty$

(iv) Θ is compact.

Then SE, BD hold and $E\, m(w_i,\theta)$ is continuous at θ_0.

(b) If $\{w_i : i\ge 1\}$ is also independent (or stationary ergodic), then P-WCON holds and also U-WCON.

Recall from the example at the beginning all we need is U-WCON and $E\, m(w_i,\theta)$ to be continuous at θ_0 to have a uniform ULN.

$H_n(\theta) = \frac{1}{n} \Sigma\, [m(w_i,\theta) - E[\,]\,]$

proof: ⊗ Define $Y_{i\delta} := \sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta,\delta)} |m(w_i,\theta') - m(w_i,\theta)|$

(a) To show SE write

$\leq |H_n(\theta')| + |H_n(\theta)|$

$$\overline{\lim_{n\to\infty}}\; P\Big(\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |H_n(\theta') - H_n(\theta)| > \varepsilon \Big) > \varepsilon \Big)$$

$$\leq \overline{\lim_{n\to\infty}}\; P\Big(\frac{1}{n}\sum_{i=1}^{n} (Y_{i\delta} + E Y_{i\delta}) > \varepsilon \Big)$$

$$\leq \overline{\lim_{n\to\infty}}\; \frac{E\, \frac{1}{n}\sum_{i=1}^{n} (Y_{i\delta} + E Y_{i\delta})}{\varepsilon} \qquad \text{by Markov's Ineq.}$$

$$= \frac{2 E Y_{i\delta}}{\varepsilon} \qquad \text{by identically distributed } w_i$$

$$= \frac{2}{\varepsilon}\, E\Big\{ \sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |m(w_i,\theta') - m(w_i,\theta)| \Big\}$$

$$= \frac{2}{\varepsilon} \lim_{\delta\to 0} E\Big\{ \underbrace{\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |m(w_i,\theta') - m(w_i,\theta)|}_{\text{Bounded above by } 2\sup_{\theta\in\Theta}|m(w_i,\theta)| \text{ which is integrable.}} \Big\}$$

$$= \frac{2}{\varepsilon}\, E\Big\{ \underbrace{\lim_{\delta\to 0} \sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |m(w_i,\theta') - m(w_i,\theta)|}_{=0 \text{ by continuity of } m(w_i,\theta)} \Big\} \qquad \text{by DCT}$$

$$= \frac{2}{\varepsilon} \times 0$$

$$= 0.$$

A compact set is totally bounded so BD follows. To see why $E\, m(w_i,\theta)$ continuous at θ_0 holds, notice that we need to show

$$\lim_{\delta \to 0} \sup_{\theta \in B(\theta_0, \delta)} | Em(w_i, \theta) - Em(w_i, \theta_0) | = 0$$

Write

$$\lim_{\delta \to 0} \sup_{\theta \in B(\theta_0, \delta)} | Em(w_i, \theta) - Em(w_i, \theta_0) | \leq \lim_{\delta \to 0} \sup_{\theta \in B(\theta_0, \delta)} E | m(w_i, \theta) - m(w_i, \theta_0) |$$

$$= \lim_{\delta \to 0} E \underbrace{\sup_{\theta \in B(\theta_0, \delta)} | m(w_i, \theta) - m(w_i, \theta_0) |}$$

bounded by $2 \sup_{\theta \in \Theta} | m(w_i, \theta) |$
which is integrable.

$$= E \lim_{\delta \to 0} \sup_{\theta \in B(\theta_0, \delta)} | m(w_i, \theta) - m(w_i, \theta_0) |$$

$$= 0.$$

by continuity at θ_0.

(b) To see why P-WCON hold, write

$$\lim_{n \to \infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} \left[m(w_i, \theta) - Em(w_i, \theta) \right] \right| > \varepsilon \right) = \lim_{n \to \infty} P(o_p(1) > \varepsilon)$$

by WLLN with iid data.

$$= 0.$$

By Theorem 11.2 we also obtain U-WCON. ∎

Derivation of Information Matrix : Write

$$1 = \int \underbrace{f(\omega,\theta)}\, d\mu(\omega)$$

true density of data. If correctly specified this is equal to $f(\omega,\theta_0)$.

Differentiating w.r.t. θ yields

$$0 = \frac{\partial}{\partial\theta} \int f(\omega,\theta)\, d\mu(\omega)$$

$$= \int \frac{\partial}{\partial\theta} f(\omega,\theta)\, d\mu(\omega)$$

(Assume $\sup_{\theta\in\Theta} f(\omega,\theta)$ is integrable so we can always interchange)

$$= \int \left[\frac{\partial}{\partial\theta} \log f(\omega,\theta)\right] \times f(\omega,\theta)\, d\mu(\omega)$$

$$\frac{\frac{\partial}{\partial\theta} f(\omega,\theta)}{\cancel{f(\omega,\theta)}} \times \cancel{f(\omega,\theta)}$$

Differentiate again and assume $\sup_{\theta\in\Theta} |\partial/\partial\theta\, f(\omega,\theta)|$ is integrable

$$0 = \frac{\partial}{\partial\theta} \int \frac{\partial}{\partial\theta'} \log f(\omega,\theta)\ f(\omega,\theta)\, d\mu(\omega)$$

$$\left[\frac{\partial \log f(\omega,\theta)}{\partial\theta'}\right] f(\omega,\theta)$$

$$= \int \frac{\partial^2}{\partial\theta\,\partial\theta'} \log f(\omega,\theta) \times f(\omega(\theta)\, d\mu + \int \left[\frac{\partial}{\partial\theta} \log f(\omega,\theta)\right] \left[\frac{\partial}{\partial\theta'} f(\omega,\theta)\right] d\mu(\omega)$$

$$= \int \frac{\partial^2}{\partial\theta\,\partial\theta'} \log f(\omega,\theta)\ f(\omega,\theta)\, d\mu + \int \left[\frac{\partial}{\partial\theta} \log f(\omega,\theta)\right] \left[\frac{\partial}{\partial\theta'} \log f(\omega,\theta)\right] f(\omega,\theta)\, d\mu$$

If the model is correctly specified, the equation becomes

$$0 = E \frac{\partial^2}{\partial\theta\,\partial\theta'} \log f(\omega_i,\theta_0) + E \frac{\partial}{\partial\theta} \log f(\omega_i,\theta_0) \frac{\partial}{\partial\theta'} \log f(\omega_i,\theta_0)$$

$$= -B_0 + \Omega_0$$

$\Rightarrow$ our B_0 will equal the information matrix Ω_0.