Eco 519. Notes on Andrews (1994)

Note Title 3/13/2006

- Empirical Process Methods

- Useful for establishing uniform analogs for:

  * Laws of Large Numbers
  * Central Limit Theorems } weak
    ↳ Functional CLT } converg.

- Stochastic equicontinuity plays a crucial role here.

- Verify stochastic equicontinuity via:

  * Directly: only works for simple cases
  * Using combinatorial methods (VC conditions)
  * via bracketing

Setup

- We'll use the same notation as Andrews, which is sort of time-series oriented.

- Let $\{W_{Tt} : t \leq T, T \geq 1\}$ be a triangular array of $\mathbb{R}^k$-valued r.v.'s on a probability space $(\Omega, \mathcal{A}, P)$. Denote their range as $\mathcal{W} \subset \mathcal{B}(\mathbb{R}^k)$.

$$
\begin{array}{llllll}
W_{11} & & & & & \\
W_{21} & W_{22} & & & & \\
W_{31} & W_{32} & W_{33} & & & \\
W_{41} & W_{42} & W_{43} & W_{44} & & \\
\vdots & \vdots & \vdots & & \ddots & \\
W_{T1} & W_{T2} & \cdots & & \cdots & W_{TT}
\end{array}
$$

- Let $\mathcal{T}$ be a pseudometric space with pseudometric ρ

Example of a pseudometric space is L^p, the space of p-power integrable functions in $[0,1]$ with metric

$$\rho(\tau_1, \tau_2) = \int_0^1 |\tau_1(z) - \tau_2(z)|^p dz$$

- A pseudometric satisfies triangle inequality and nonnegativity, but $\rho(\tau_1, \tau_2) = 0 \not\Rightarrow \tau_1 = \tau_2$.

- Consider the family of functions $M : \mathcal{W} \times \mathcal{T} \longrightarrow \mathbb{R}^s$

$$\mathcal{M} = \{ m(\cdot, \tau) : \tau \in \mathcal{T} \}$$ random functions indexed by τ

- Define an $\boxed{\text{empirical process}}$ $v_T(\cdot)$ by:

$$v_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[m(W_t, \tau) - E\, m(W_t, \tau) \right]$$

with $\tau \in \mathcal{T}$.

Examples:
In parametric settings: $\mathcal{T}$ is usually a subset of $\mathbb{R}^p$; in semi-parametric settings, $\mathcal{T}$ is a class of functions or a class of subsets in $\mathbb{R}^p$.

- We now re-introduce the concept of $\boxed{\text{weak convergence}}$.

- First, focus on the metric space $(B(\mathcal{T}), d)$, where $B(\mathcal{T})$ is the space of bounded, $\mathbb{R}^s$-valued functions in $\mathcal{T}$, and d is the uniform metric:

$$d(b_1, b_2) = \sup_{\tau \in \mathcal{T}} \| b_1(\tau) - b_2(\tau) \|$$

Definition: Outer Expectation

- For each bounded, real-valued function H on Ω, we define the outer expectation E^*H as:

$$E^*(H) = \inf \left\{ E(h) : H \leq h, \text{ and } h \text{ is measurable, integrable} \right\}$$

Inner expectation would be defined analogously.

Definition: Weak Convergence

$$v_T(\cdot) \Rightarrow V(\cdot) \quad \text{if}$$

$$E^* f(v_T(\cdot)) \to E f(V(\cdot)) \quad \forall f \in \mathcal{U}(B(\mathcal{T}))$$

where $\mathcal{U}(B(\mathcal{T}))$ is the space of bdd, uniformly continuous functions on $B(\mathcal{T})$ (with respect to the uniform metric).

- We use outer expectations because requiring measurability of $V_T(\cdot)$ in $(B(T), d)$ is too restrictive:

for example, if $T = [0,1]$ and

$$m(W_t, \tau) = 1\{W_t \leq \tau\}$$

and $B(T) = \underbrace{D[0,1]}_{\text{space of cadlag functions in } [0,1]}$

then $V_T(\cdot)$ is not measurable in $(D[0,1], d, \underbrace{\mathcal{B}(D[0,1])}_{\hookrightarrow \text{ Borel sigma field in } D[0,1]})$.

<u>Cadlag</u> <u>functions</u>: Functions that are right-continuous and whose left limits always exist.

Comments on $\nu(\cdot)$:

- $\nu(\cdot)$ may or may not be defined on the same probability space as $\nu_T(\cdot)$.

- The limiting process $\nu(\cdot)$ is assumed to be measurable.

- If $\nu(\cdot)$ is uniformly ρ-continuous in τ w.p.1, then a sufficient condition for weak convergence is that the empirical process

$\{\nu_T(\cdot) : T \geq 1\}$ be stochastically equicontinuous

Make distinction between weak convergence and convergence in distribution

Stochastic Equicontinuity

- Three equivalent definitions

(i) $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicont if $\forall \varepsilon > 0$ and $\eta > 0$, $\exists \delta > 0$ s.th:

$$\overline{\lim_{T \to \infty}} P^* \left[\sup_{\rho(\tau_1, \tau_2) < \delta} \| \nu_T(\tau_1) - \nu_T(\tau_2) \| > \eta \right] < \varepsilon$$

$\overline{\lim}$ denotes upper limit:

$\overline{\lim}\, T_n = c$ if c is a term that is greater than all but a finite number of terms of $\{T_n\}$, all of which equal c.

(ii) For any sequence of constants $\delta_T \to 0$:

$$\sup_{\rho(\tau_1,\tau_2)\le\delta_T} |V_T(\tau_1) - V_T(\tau_2)| \xrightarrow{p^*} 0$$

(iii) For all sequences of random elements $\{\hat{\tau}_{1T}\}$, $\{\hat{\tau}_{2T}\}$ such that $\rho(\hat{\tau}_{1T}, \hat{\tau}_{2T}) \xrightarrow{p} 0$, we have

$$V_T(\hat{\tau}_{1T}) - V_T(\hat{\tau}_{2T}) \xrightarrow{p^*} 0$$

Example: $\longrightarrow$ Space of linear fns.

$$\mathcal{M} = \{g : g(w) = w'\tau \text{ for some } \tau \in \mathbb{R}^k\}$$

$$\|V_T(\tau_1) - V_T(\tau_2)\| = \frac{1}{\sqrt{T}} \left| \sum_{t=1}^{T} (W_t - E(W_t))'(\tau_1 - \tau_2) \right|$$

$$\le \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{T} (W_t - E(W_t)) \right\| \cdot \| \tau_1 - \tau_2 \|$$

Then ρ is Euclidean metric. Take any $\varepsilon > 0$, $\eta > 0$. Then

$$\Rightarrow \sup_{\rho(\tau_1,\tau_2)<\delta} \| V_T(\tau_1) - V_T(\tau_2)\| \leq \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{T} (W_t - E(W_t)) \right\| \cdot \delta$$

$$P\left(\sup_{\rho(\tau_1,\tau_2)<\delta} \| V_T(\tau_1) - V_T(\tau_2)\| > \eta \right)$$

$$\leq P\left(\frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{T} (W_t - E(W_t)) \right\| > \eta/\delta \right)$$

$\therefore$ stochastic equicont. follows if

$$\frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{T} (W_t - E(W_t)) \right\| = O_p(1)$$

Proposition: Weak Convergence and Stochastic Equicontinuity

Suppose:

(i) $(\mathcal{T}, \rho)$ is a totally bounded pseudo-metric space

(ii) Finite-dimensional convergence in distribution holds:
$\forall$ finite subsets $(\tau_1, \dots, \tau_J) \subset \mathcal{T}$
$(\nu_T(\tau_1)', \dots, \nu_T(\tau_J)')'$ converges in distribution

(iii) $\nu_T(\cdot)$ is stochastically equicontinuous

- Then: There exists a Borel-measurable (with respect to d) $\mathcal{B}(\mathcal{T})$-valued stochastic process $\nu(\cdot)$ with sample paths that are uniformly ρ-continuous w.p.1 such that: $\nu_T(\cdot) \Rightarrow \nu(\cdot)$

- If $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ with $\nu(\cdot)$ having these properties and (i) holds, then (ii) and (iii) hold.

Example: Sometimes it is easy to verify stochastic equicontinuity:

$$\mathcal{M} = \{ g : g(w,\tau) = w'\tau \ ; \ \tau \in \mathbb{R}^k \}$$

$$\mathcal{T} = (\mathbb{R}^k, \|\cdot\|) \Rightarrow \rho = \|\cdot\| \text{ (Euclidean metric)}$$

$$v_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (W_t'\tau - E(W_t)'\tau)$$

For any $\tau_1, \tau_2 \in \mathbb{R}^k$:

$$\|v_T(\tau_1) - v_T(\tau_2)\| = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left| (W_t - E(W_t))'(\tau_1 - \tau_2) \right|$$

$$\leq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \|W_t - E(W_t)\| \cdot \|\tau_1 - \tau_2\|$$

$$\therefore \sup_{\rho(\tau_1,\tau_2)<\delta} \|v_T(\tau_1) - v_T(\tau_2)\| \leq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \|W_t - E(W_t)\| \cdot \delta$$

$$\Rightarrow P\left(\sup_{\rho(\tau_1,\tau_2)<\delta} \|v_T(\tau_1) - v_T(\tau_2)\| > \eta \right)$$

$$\leq P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\|W_t - E(W_t)\| > \frac{\eta}{\delta}\right)$$

So, if $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\|W_t - E(W_t)\| = O_p(1)$, then for any $\varepsilon > 0$, $\eta > 0$, we can find δ small enough so that this probability is $< \varepsilon$. $\square$

- Not all classes of functions are stoch. equicontinuous: Suppose $\mathcal{T}$ is the class of all Borel sets in $\mathcal{W} \subset \mathbb{R}^p$, suppose $\{W_t\}$ are continuous r.v.'s with distribution P_1. Let:

$m(w, \tau) = 1\!\!1\{w \in \tau\}$ and the pseudometric

$$\rho(\tau_1, \tau_2) = \left(\int (m(w, \tau_1) - m(w, \tau_2))^2 \, dP_1(w)\right)^{1/2}$$

$$= \left(Pr(W \in \tau_1) + Pr(W \in \tau_2) - 2Pr(W \in \tau_1 \cap \tau_2)\right)^{1/2}$$

- For any realization $\omega \in \Omega$, there exist two finite sets τ_1, τ_2

such that for this $\omega \in \Omega$,

$$W_t \in \mathcal{T}_1 \quad \forall\, t \le T$$
$$W_t \notin \mathcal{T}_2 \quad \forall\, t \le T$$

Therefore, for any $\omega \in \Omega$, we can always find two finite sets $\mathcal{T}_1, \mathcal{T}_2$ s.th:

$$\| V_T(\mathcal{T}_1) - V_T(\mathcal{T}_2) \| = \frac{T}{\sqrt{T}} - 0 = \sqrt{T}$$

Note that for such two $\mathcal{T}_1, \mathcal{T}_2$:

$\rho(\mathcal{T}_1, \mathcal{T}_2) = 0$ because W_t is continuous

Then, we must have:
↳ for any $\delta > 0$

$$\left.\sup_{\rho(\mathcal{T}_1,\mathcal{T}_2)<\delta} \| V_T(\mathcal{T}_1) - V_T(\mathcal{T}_2) \| \ge \sqrt{T} \quad \right\} \text{ stochastic equicont. fails.}$$

The class $\mathcal{M}$ is way too rich...

M-Estimation:

- Take $\tau \in \Upsilon \subset \mathbb{R}^k$ and let:

$$\bar{m}_T(\tau) = \frac{1}{T}\sum_{t=1}^{T} m(W_t, \tau)$$

- Suppose an estimator $\hat{\tau} \in \mathbb{R}^k$ satisfies $\bar{m}_T(\hat{\tau}) = o_p(1)$. Suppose $\hat{\tau}$ is consistent for τ_0: $\rho(\hat{\tau}, \tau_0) \xrightarrow{p} 0$, where $E\bar{m}_T(\tau_0) = 0$.

- Suppose $E\bar{m}_T(\tau) \equiv \lambda(\tau)$ is smooth, differentiable. Then a Taylor approx yields:

$$0 = \lambda(\tau_0) = \lambda(\hat{\tau}) + \nabla_\tau \lambda(\tilde{\tau})(\tau_0 - \hat{\tau})$$

$$\Rightarrow \sqrt{T}(\hat{\tau} - \tau_0) = \nabla_\tau \lambda(\tilde{\tau})^{-1} \sqrt{T} \lambda(\hat{\tau})$$

↳ assume invertibility here

- So it all hinges on the asymptotic properties of $\sqrt{T}\lambda(\hat{\tau})$

$$\sqrt{T}\,\lambda(\hat{\tau}) = \sqrt{T}\,(\lambda(\hat{\tau}) - \bar{m}_T(\hat{\tau})) + \underbrace{\sqrt{T}\,\bar{m}_T(\hat{\tau})}_{=o_p(1)}$$

$$= -\sqrt{T}\,(\bar{m}_T(\hat{\tau}) - \lambda(\hat{\tau})) + o_p(1)$$

Let $\sqrt{T}(\bar{m}_T(\tau) - \lambda(\tau)) \equiv V_T(\tau)$, then

$$\sqrt{T}\,\lambda(\hat{\tau}) = V_T(\tau_0) - V_T(\hat{\tau}) + V_T(\tau_0) + o_p(1)$$

If $V_T(\tau_0) - V_T(\hat{\tau}) = o_p(1)$, then the only term that survives in the limit is $V_T(\tau_0)$, which would typically satisfy a CLT result.

<u>Claim:</u> $V_T(\tau_0) - V_T(\hat{\tau}) = o_p(1)$ if $V_T(\tau)$ is stochastically equicontinuous. (given $\hat{\tau} \xrightarrow{p} \tau_0$)

<u>Proof.</u> — Fix any $\eta > 0$ and consider:

$$\Pr(|V_T(\hat{\tau}) - V_T(\tau_0)| > \eta,\ \rho(\hat{\tau}, \tau_0) \leq \delta)$$
$$+ \Pr(\rho(\hat{\tau}, \tau_0) > \delta)$$

As $\delta \longrightarrow \infty$, this prob. is approximately $\approx \Pr(|V_T(\hat{\tau}) - V_T(\tau_0)| > \eta)$

As $\delta \to 0$, this probability is ≈ 1

- This intuition "shows" why $\forall \eta > 0$, we can find δ such that:

$$\overline{\lim_{T\to\infty}} P(|V_T(\hat{\tau}) - V_T(\tau_0)| > \eta)$$
$$\leq \overline{\lim_{T\to\infty}} P(|V_T(\hat{\tau}) - V_T(\tau_0)| > \eta, \rho(\hat{\tau}, \tau_0) \leq \delta)$$
$$+ \overline{\lim_{T\to\infty}} P(\rho(\hat{\tau}, \tau_0) > \delta)$$

- Since $\hat{\tau}$ is consistent, we can disregard the 2^{nd} term and obtain:

$$\overline{\lim_{T\to\infty}} P(|V_T(\hat{\tau}) - V_T(\tau_0)| > \eta)$$
$$\leq \overline{\lim_{T\to\infty}} P(|V_T(\hat{\tau}) - V_T(\tau_0)| > \eta, \rho(\hat{\tau}, \tau_0) \leq \delta)$$
$$\leq \overline{\lim_{T\to\infty}} P\left(\sup_{\rho(\tau,\tau_0)\leq\delta,\, \tau\in\mathcal{T}} |V_T(\tau) - V_T(\tau_0)| > \eta\right)$$
$< \varepsilon$ if $V_T(\cdot)$ is stoch. equicont.

And so, we have:

$$\sqrt{T}(\hat{\tau}-\tau_0) = \nabla_\tau \lambda(\hat{\tau})^{-1} V_T(\tau_0) + o_p(1)$$

$$\xrightarrow{d} M^{-1} \cdot N(0,S) = N(0, M^{-1} S M^{-1\prime})$$

where $M = \nabla_\tau \lambda(\tau_0)$.

Semiparametric Estimation

- Suppose $\hat{\tau}$ is a first-stage estimate for an infinite-dimensional parameter, and let $\hat{\theta}$ satisfy:

$$\sqrt{T}\, \bar{m}_T(\hat{\theta}, \hat{\tau}) = o_p(1)$$

where $\bar{m}_T(\theta, \hat{\tau}) = \frac{1}{T} \sum_{t=1}^{T} m(W_t; \theta, \hat{\tau})$

if m is smooth then,

$$o_p(1) = \bar{m}_T(\hat{\theta}, \hat{\tau}) = \bar{m}_T(\theta_0, \hat{\tau}) + \nabla_\theta \bar{m}_T(\tilde{\theta}, \hat{\tau})(\hat{\theta} - \theta_0)$$

So:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -\nabla_\theta \bar{m}_T(\tilde{\theta}, \hat{\tau})^{-1} \cdot \sqrt{T}\, \bar{m}_T(\theta_0, \hat{\tau}) + o_p(1)$$

Let $\lambda(\theta, \tau) = E\, \bar{m}_T(\theta, \tau)$. Then:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -\nabla_\theta \bar{m}_T(\tilde{\theta}, \hat{\tau})^{-1} \cdot \sqrt{T}\,(\bar{m}_T(\theta_0, \hat{\tau}) - \lambda(\theta_0, \hat{\tau}))$$
$$- \nabla_\theta \bar{m}_T(\tilde{\theta}, \hat{\tau})^{-1} \cdot \sqrt{T}\, \lambda(\theta_0, \hat{\tau}) + o_p(1)$$

- Now, we can either have:

$\sqrt{T}\, \lambda(\theta_0, \hat{\tau}) = o_p(1)$ or $\sqrt{T}\, \lambda(\theta_0, \hat{\tau}) \xrightarrow{d} N(0, A)$

Once again, let:

$$v_T(\theta, \tau) = \sqrt{T}(\bar{m}_T(\theta, \tau) - \lambda(\theta, \tau))$$

Then:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -\nabla_\theta \bar{m}_T(\tilde{\theta}, \hat{\tau})^{-1} [\underbrace{v_T(\theta_0, \hat{\tau})}_{\equiv v_T(\hat{\tau})} - \underbrace{v_T(\theta_0, \tau_0)}_{\equiv v_T(\tau_0)}]$$
$$- \nabla_\theta \bar{m}_T(\tilde{\theta}, \hat{\tau})^{-1} [v_T(\theta_0, \tau_0) + \sqrt{T}\, \lambda(\theta_0, \hat{\tau})] + o_p(1).$$

The first term is $o_p(1)$ if $V_T(\cdot)$ is stochastically equicont...