Review of Basic Asymptotic Theory

Probability Space: $(\Omega, \mathcal{F}, \mathcal{P})$, such that the following 3 conditions are satisfied:

1. $A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}$.
2. $A_1, A_2, \ldots, \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
3. $\emptyset \in \mathcal{F}$.

Question: Show that the above conditions imply $\Omega \in \mathcal{F}$ and $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.
The events lim sup and lim inf are defined as:

1. $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m = \{A_n, i.o.\}$ where $i.o.$ denotes "infinitely often".
2. $\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m = \{A_n, e.v.\}$ where $e.v.$ denotes "eventually".

Stochastic Convergence: All materials can be found in Ch3, Amemiya and Davidson Ch 18.

Almost Sure Convergence:

$$X_n \xrightarrow{a.s.} 0 \Longleftrightarrow P\left(\lim X_n(\omega) = 0\right) = 1. \Longleftrightarrow \forall \epsilon > 0, P\left(|X_n(\omega)| > \epsilon, i.o.\right) = 0.$$

$$\Longleftrightarrow \forall \epsilon > 0, P\left(\bigcap_{n=1}^{\infty} \bigcup_{m \geq n} \left(|X_m(\omega)| > \epsilon\right)\right) = 0. \Longleftrightarrow \forall \varepsilon > 0, \lim_{n\to\infty} P\left(\bigcup_{m \geq n} \{|X_m(\omega)| > \epsilon\}\right) = 0.$$

$$\Longleftrightarrow \forall \epsilon > 0, P\left(\bigcup_{n=1}^{\infty} \bigcap_{m \geq n} \{|X_m(\omega)| \leq \epsilon\}\right) = 1. \Longleftrightarrow \lim_{n\to\infty} P\left(\bigcap_{m \geq n} \{|X_m(\omega)| \leq \epsilon\}\right) = 1.$$

Question: Show the above definitions are equivalent.

L^p convergence $X_n \xrightarrow{L^p} 0$: $\lim_{n\to\infty} E|X_n|^p = 0$.

Convergence in probability: $\forall \epsilon > 0$, $\lim_{n\to\infty} P\left(|X_n(\omega)| \leq \epsilon\right) = 1$.

Convergence in distribution: $\lim_{n\to\infty} P\left(X_n \leq x\right) = P\left(X \leq x\right)$ for every continuity point x in the distribution of X.

Relations:

1. $\left(X_n \xrightarrow{a.s.} 0\right) \Longrightarrow \left(X_n \xrightarrow{p} 0\right)$: Note that $P\left(\bigcap_{m \geq n} |X_m(\omega)| \leq \epsilon\right) \leq P\left(|X_n(\omega)| \leq \epsilon\right)$.
2. $\left(X_n \xrightarrow{L^p} 0, p > 0\right) \Longrightarrow \left(X_n \xrightarrow{p} 0\right)$. Note that $P\left(|X_n| > \epsilon\right) \leq \epsilon^{-p} E|X_n|^p$ by the Markov Inequality.
3. $\left(X_n \xrightarrow{p} 0\right) \Longleftrightarrow \left(X_n \xrightarrow{d} 0\right)$: Almost by definition. Note however that $\left(X_n \xrightarrow{d} X\right) \not\Rightarrow \left(X_n - X \xrightarrow{p} 0\right)$ unless X is degenerate.

Borel-Cantelli lemma(BC): $\sum_{n=1}^{\infty} P\left(E_n\right) < \infty \Longrightarrow P\left(E_n, i.o.\right) = 0$.

Proof: $P\left(E_n, i.o.\right) = \lim_{n\to\infty} P\left(\bigcup_{n \geq m} E_m\right) \leq \lim_{n\to\infty} \sum_{m \geq n} P\left(E_m\right) \to 0$ where the last equality is by the summability assumption.

Example: $X_i \sim IIDUniform(0,1)$, show that $\min_{1\leq i\leq n} X_i \xrightarrow{a.s.} 0$.

Proof: Note that $P\left(\min_{1\leq i\leq n} X_i > \epsilon\right) = \prod_{i=1}^{n} P\left(X_i > \epsilon\right) = (1-\epsilon)^n$. And $\sum_{n=1}^{\infty} (1-\epsilon)^n < \infty$. Use Borel-Cantelli to conclude.

Look at the example in Amemiya P88: Note that need $\sum_{n=1}^{\infty} P\left(|X_n| > \epsilon\right) = \infty$ for the arguments to hold, hence $P\left(|X_n| > \epsilon\right) = \frac{1}{n}$ will do.

But Borel-Cantelli may not be neccessary for a.s. convergence: Let ω be uniformly distributed on $(0,1)$. Define $X_n(\omega) = n$ if $\omega \leq \frac{1}{n}$ and $X_n(\omega) = 0$ if $\omega > \frac{1}{n}$. Obviously $X_n(\omega) \overset{a.s.}{\to} 0$ but BC does not hold.

a.s. does not imply L^p convergence: The same example above, note $EX_n = 1$ for all n, although $X_n \xrightarrow{a.s.} 0$. So when does a.s. convergence imply convergence in distribution: need to control for the cases where things go really wrong with small probability.

Monotone Convergence Theorem(MON): If $X_n \overset{a.s.}{\to} X$ and X_n is increasing almost surely, then $\lim_{n\to\infty} EX_n = EX$.

Dominated Convergence Theorem(DOM): If $X_n \overset{a.s.}{\to} X$ and $E\left(\sup_n |X_n(\omega)|\right) < \infty$, then $\lim_{n\to\infty} EX_n = EX$. Note that this also applies to the Lebesgue measure, which is not a probablity measure, in which case we have: If $f_n(x) \overset{a.s.}{\to} f(x)$ with respect to the Lebesgue measure, and $\int |\sup_n f_n(x)| dx < \infty$, then $\lim_{n\to\infty} \int f_n(x)\, dx = \int f(x)\, dx$.

Uniform Integrability(UI):Definition X_n is U.I. if $\lim_{M\to\infty} \sup_n E\left(|X_n| 1\left(|X_n| > M\right)\right) = 0$.

Theorem: If X_n is U.I. and if $X_n \overset{a.s.}{\to} 0$ then $\lim E|X_n| = 0$. Hence $\lim EX_n = 0$.

Proof: Write $E|X_n| = E|X_n|1\left(|X_n| > M\right) + E|X_n|1\left(|X_n| < M\right)$. The first term $\to 0$ as $M \to \infty$ by U.I. Use DOM to show that given M, the second term $\to 0$ since it is dominated by M.

Stochastic Order: $X_n = o_p(1)$ if $X_n \xrightarrow{p} 0$. $X_n = O_p(1)$ if $\lim_{M\to\infty} \limsup_n P\left(|X_n| > M\right) = 0$.

Facts: $X_n = O_p(a_n)$ means $a_n^{-1} X_n = O_p(1)$. $O_p(1)\, o_p(1) = o_p(1)$. $O_p(a_n)\, O_p(b_n) = O_p(a_n b_n)$. $O_p(a_n) + O_p(b_n) = O_p(a_n + b_n) = O_p(\max(a_n, b_n))$.

Slutsky: See Amemiya thm 3.2.7. p89.

Continuous Mapping: See Amemiya thm 3.2.5. p88.

Weak Law of Large Numbers(WLLN): $X_t, t = 1, \ldots$. Let $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$, underwhat conditions does $\bar{X}_n - E\bar{X}_n \xrightarrow{p} 0$.

Sufficient to show $E|\bar{X}_n - E\bar{X}_n|^p \to 0$, say $p = 2$ but other $p > 0$ also works.

WLLN for independent nonindentically distributed X_t, Davidson p293: Let X_t be an independent sequence. If $\sum_{t=1}^{\infty} \sigma_t^2 / t^2 < \infty$ then $E\left(\bar{X}_n - E\bar{X}_n\right)^2 \to 0$.

Proof: $E\left(\bar{X}_n - E\bar{X}_n\right)^2 = Var\left(\bar{X}_n\right) = \frac{1}{n^2}\sum_{t=1}^{n} \sigma_t^2$. Use Kronecker's Lemma(See p34 Davidson), which says that for a positive sequence of numbers x_t and a sequence of numbers that monotonically increase to infinity, a_t, if $\sum_{t=1}^{\infty} x_t / a_t < \infty$, then $\frac{1}{a_n}\sum_{t=1}^{n} x_t \to 0$ as $n \to \infty$. Now take $x_t = \sigma_t^2$ and take $a_t = t^2$.

An easy WLLN: X_t uncorrelated with mean 0 and $Var(X_t) = \sigma^2$.

Strong Law of Large Numbers(SLLN): Under what conditions does $\bar{X}_n - E\bar{X}_n \xrightarrow{a.s.} 0$.

SLLN 1: Thm 3.3.1 Amemiya P90 If X_t independent, $\sum_{t=1}^{\infty} \frac{\sigma_t^2}{t^2} < \infty$, then $\bar{X}_n - EX_n \xrightarrow{a.s.} 0$.

SLLN 2: Thm 3.3.2 Amemiya p90 If X_t iid with finite mean u, then $\bar{X}_n - u \overset{a.s.}{\to} 0$.

Triangular Array: Useful for convergence in probability and in distribution $\{\{X_{nt}, t = 1, \ldots, n\}, n = 1, \ldots, \infty\}$.

For example, given $X_t, t = 1, \ldots, \infty$ independent and mean 0, let $\sigma_t^2 = Var(X_t)$, and let $C_n^2 = \sum_{t=1}^n \sigma_t^2$. Then $X_{nt} = \frac{X_t}{C_n}$ is an triangular array of random variables, for $t = 1, \ldots, n$ and for $n = 1, \ldots, \infty$.

Question: Show that $Var\left(\sum_{t=1}^n X_{nt}\right) = \sum_{t=1}^n Var(X_{nt}) = 1$.

Consistency of least square coefficient: p95 Amemiya

Model 1: $y_t = x_t'\beta + u_t$, u_t uncorrelated, mean 0, $Eu_t^2 = \sigma^2$ for all t. $\hat{\beta} = (X'X)^{-1} X'y$.

Consistency Theorem: If $\lambda_s(X'X) \to \infty$, then $\hat{\beta} \overset{p}{\to} \beta$. Use $\lambda_s(A)$ and $\lambda_l(A)$ to denote the smallest and largest eigenvalues of A.

Proof: $\lambda_s(X'X) \to \infty \Longrightarrow \lambda_l\left[(X'X)^{-1}\right] \to 0 \Longrightarrow \text{trace}\left[(X'X)^{-1}\right] \to 0$. But the trace of $(X'X)^{-1}$ is the sum of the variances for the elements in $\hat{\beta} - \beta$.

Consistency of $\hat{\sigma}^2$, Thm 3.5.2 Amemiy p96 Assume u_t iid in Model 1. Then $\hat{\sigma}^2 \overset{p}{\to} \sigma^2$, where $\hat{\sigma}^2 = T^{-1}\hat{u}'\hat{u} = T^{-1}u'u - T^{-1}u'Pu$ for $P = X(X'X)^{-1} X'$.

Proof:

1. $T^{-1}u'u \overset{a.s.}{\longrightarrow} \sigma^2$ by **SLLN 2**.
2. $P\left(T^{-1}u'Pu > \epsilon\right) \leq \epsilon^{-1}ET^{-1}u'Pu = (T\epsilon)^{-1}\sigma^2 tr(P) = \frac{\sigma^2 k}{T\epsilon} \to 0$.
3. Reminder $tr(P) = tr\left(X(X'X)^{-1}X'\right) = tr\left[(X'X)^{-1}(X'X)\right] = tr[I_k]$.

Convergence in distribution: See Thm 22.8 p353, Davidson $X_n \overset{d}{\to} X \Longrightarrow \lim_{n\to\infty} E[f(X_n)] = E[f(X)]$ for every bounded continuous f. However $EX_n \nrightarrow EX$ if X_n unbounded.

Characteristic Function: Useful in showing convergence in distribution because it is bounded and continuous.

Definition: $\phi_X(\lambda) = Ee^{i\lambda X} = E[\cos(\lambda X) + i\sin(\lambda X)]$.

Examples:

1. X Standard Normal $N(0,1)$, then $\phi_X(\lambda) = e^{-\lambda^2/2}$.
2. X Cauchy with density $f(x) = \frac{1}{\pi(1+x^2)}$, then $\phi_X(\lambda) = e^{-|\lambda|}$. (Davidson p167)

Property:

1. $\phi_{aX+b}(\lambda) = Ee^{i\lambda(aX+b)} = e^{ib\lambda}\phi_X(a\lambda)$.
2. Let X_{nt} be iid, let $S_n = \sum_{t=1}^n X_{nt}$, then $\phi_{S_n}(\lambda) = \prod_{t=1}^n \phi_{X_{nt}}(\lambda) = [\phi_{X_{nt}}(\lambda)]^n$.

Examples:

1. If X_t iid normal $N(0,1)$, let $S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t$, then

$$\phi_{S_n}(\lambda) = \left[\phi_{X_t}\left(\lambda/\sqrt{n}\right)\right]^n = \left[\exp\left(-\left(\lambda/\sqrt{n}\right)^2\right)\right]^n = \exp\left(-\frac{\lambda^2}{2}\right).$$

   Still $N(0,1)$

2. If X_t iid Cauchy, let $S_n = \frac{1}{n} \sum_{t=1}^{n} X_t$, then

$$\phi_{S_n}(\lambda) = \left[\phi_{X_t}(\lambda/n)\right]^n = \left[\exp\left(-|\lambda/n|\right)\right]^n = \exp\left(-|\lambda|\right).$$

   Still Cauchy. So No LLN for Cauchy random variables, because it has no mean.

Central Limit Theorem: CLT

Sequence Version(See Amemiya p92) Let

$$S_n = \frac{\bar{X}_n - E\bar{X}_n}{\sqrt{Var\left(\bar{X}_n\right)}}.$$

Under what conditions does $S_n \xrightarrow{d} N(0,1)$.

Triangular Array Version(See Davidson p 368.) Then $X_{nt} = \frac{X_t}{C_n}$ is an triangular array of random variables, for $t = 1, \ldots, n$ and for $n = 1, \ldots, \infty$. The CLT looks for conditions under which $S_n = \sum_{t=1}^{n} X_{nt} \xrightarrow{d} N(0,1)$. Remember that by definition $Var(S_n) = 1$.

Question: Show that the two formulations are equivalent.

Lindeberg Condition(Sufficient Condition for CLT)

Sequence Version(Thm 3.3.6. Amemiya. p92):

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{t=1}^{n} E\left[X_t^2 1\left(|X_t| > \epsilon C_n\right)\right] = 0. \quad \forall \epsilon > 0.$$

Array Version(Thm 23.6, Davidson p369):

$$\lim_{n\to\infty} \sum_{t=1}^{n} EX_{nt}^2 1\left(|X_{nt}| > \epsilon\right) = 0. \quad \forall \epsilon > 0.$$

Question: Convince yourself that the two versions are the same thing.

Liapounov Condition: implies Lindeberg condition

Sequence Version(Amemiya p92):

$$\lim_{n\to\infty} \left[C_n^2\right]^{-1/2} \left(\sum_{t=1}^{n} E|X_t|^3\right)^{1/3} = 0.$$

Array Version(Davidson p 373):

$$\exists \delta > 0, \quad s.t. \quad \lim_{n\to\infty} \sum_{t=1}^{n} E|X_{nt}|^{2+\delta} = 0.$$

Question: take $\delta = 1$ in the Array Version then you get the Sequence Version, convince yourself that they say the same thing.

Liapounov Condition implies Lindeberg Condition:

Proof(use the array version):

$$E|X_{nt}|^{2+\delta} \geq E\left(1\left(|X_{nt}| > \epsilon\right) X_{nt}^{2+\delta}\right) \geq \epsilon^{\delta} E\left(1\left(|X_{nt}| > \epsilon\right) X_{nt}^{2}\right)$$

Therefore

$$\lim_{n\to\infty} \sum_{t=1}^{n} E\left(X_{nt}^2 1\left(|X_{nt}| > \epsilon\right)\right) \leq \frac{1}{\epsilon^{\delta}} \lim_{n\to} \sum_{t=1}^{n} E|X_{nt}|^{2+\delta} = 0.$$

Question: Convince yourself that the same proof works for the sequence version for the special case of $\delta = 1$.

Lindeberg-Levy CLT: If X_t iid $\left(0, \sigma^2\right)$, then the CLT holds. Convince yourself that this implies the Lindeberg condition.

A sufficient condition for Lindeberg condition: Let $\sigma_t^2 \geq 1$ for all t and nondecreasing, If $\frac{X_t^2}{\sigma_t^2}$ is Uniformly Integrable(U.I.) and if $\sup_n n\frac{\max_{1\leq t\leq n}\sigma_t^2}{\sum_{t=1}^{n}\sigma_t^2} < \infty$, then the Lindeberg condition holds.

Proof:

$$\begin{aligned}\frac{1}{C_n^2}\sum_{t=1}^{n} E\left[X_t^2 1\left(|X_t| > \epsilon C_n\right)\right] &\leq \frac{1}{C_n^2} n \max_{1\leq t\leq n} \sigma_t^2 E\left(\frac{X_t^2}{\sigma_t^2}\right) 1\left(|\frac{X_t}{\sigma_t}| > C_n\frac{\epsilon}{\sigma_t}\right) \\ &\leq \left(\frac{1}{C_n^2} n \max_{1\leq t\leq n}\sigma_t^2\right) \sup_t E\left[\left(\frac{X_t}{\sigma_t}\right)^2 1\left(\left(\frac{X_t}{\sigma_t}\right)^2 > \frac{\varepsilon^2}{\max_{1\leq t\leq n}\sigma_t^2} C_n^2\right)\right] \longrightarrow 0.\end{aligned}$$

where the convergence follows from the first term being finite by assumption and the second term $\to 0$ by UI and that $\frac{\varepsilon^2}{\max_{1\leq t\leq n}\sigma_t^2} C_n^2 \longrightarrow \infty$

Asymptotic Normality of least square coefficients: (Thm 3.5.3 p96 Amemiya) Model 1, u_t iid $\left(0, \sigma^2\right)$. x_t is a scalar, if $\lim_T \frac{\max_{1\leq t\leq T} x_t^2}{\sum_{t=1}^{T} x_t^2} = 0$, then $\sigma^{-1}\left(x'x\right)^{1/2}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, 1\right)$.

Proof: Note first $\sigma^{-1}\left(x'x\right)^{1/2}\left(\hat{\beta} - \beta\right) = \frac{\sum_{t=1}^{T} x_t u_t}{\sigma\sqrt{\sum_{t=1}^{T} x_t^2}} = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t u_t}{\sqrt{Var\left(\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right)}}$, so sufficient to check

Lindeberg condition for $x_t u_t$, which is

$$\begin{aligned}
&\frac{1}{\sigma^2 \sum_{t=1}^T x_t^2} \sum_{t=1}^T E\left[(x_t u_t)^2 1\left(|x_t u_t| > \epsilon\sigma\sqrt{\sum_{t=1}^T x_t^2}\right)\right] \\
=&\frac{1}{\sigma^2 \sum_{t=1}^T x_t^2} \sum_{t=1}^T E\left[(x_t u_t)^2 1\left((x_t u_t)^2 > \epsilon^2\sigma^2\sum_{t=1}^T x_t^2\right)\right] \\
=&\frac{1}{\sigma^2 \sum_{t=1}^T x_t^2} \sum_{t=1}^T E x_t^2\left[u_t^2 1\left(u_t^2 > \epsilon^2\frac{\sigma^2\sum_{t=1}^T x_t^2}{x_t^2}\right)\right] \\
\leq&\frac{1}{\sigma^2} \max_{1\leq t\leq T} E\left[u_1^2 1\left(u_1^2 > \epsilon^2\frac{\sigma^2\sum_{t=1}^T x_t^2}{x_t^2}\right)\right] \\
=&\frac{1}{\sigma^2} E\left[u_1^2 1\left(u_1^2 > \epsilon^2\frac{\sigma^2\sum_{t=1}^T x_t^2}{\max_{1\leq t\leq T} x_t^2}\right)\right] \longrightarrow 0.
\end{aligned}$$

where the convergence follows from $Eu_1^2 \leq \infty$ and the stated assumption that $\frac{\sigma_t^2\sum_{t=1}^T x_t^2}{\max_{1\leq t\leq T} x_t^2} \longrightarrow \infty$.

What if x_t is a k-dimensional vector? Thm 3.5.4 p 97 Amemiya: Model 1, u_t iid $\left(0, \sigma^2\right)$, for each $i = 1, \ldots, k$, $\lim_{T\to\infty} \left(x_i' x_i\right)^{-1} \max_{1\leq t\leq T} x_{ti}^2 = 0$. Let $S = diag\left(\left(x_i' x_i\right)^{1/2}\right)$, $Z = XS^{-1}$, $\lim_{T\to\infty} Z'Z = R$ nonsingular, then $S\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 R^{-1}\right)$.

Proof: $S\left(\hat{\beta} - \beta\right) = S\left(X'X\right)^{-1} X'u = S\left(X'X\right)^{-1} SS^{-1}X'u = \left(Z'Z\right)^{-1} Z'u$. Use Cramer-Wold Device(Amemiya p93 Thm 3.3.8) sufficient to show that for $c \neq 0$, $c'\left(Z'Z\right)^{-1} Z'u \xrightarrow{d} N\left(0, \sigma^2 c' R^{-1} c\right)$. But $c'\left(Z'Z\right)^{-1} Z'u \overset{LD}{=} \gamma' Z'u$ for $\gamma' = c'R^{-1}$ by Slutsky. So take $x_t = \gamma' z_t$ in Thm 3.5.3 and check its condition:

$$\lim_{T\to\infty} \frac{\max_{1\leq t\leq T}\left(\gamma' z_t\right)^2}{\gamma' Z' Z\gamma} \leq \lim_{T\to\infty} \frac{\left(\gamma'\gamma\right)\max_{1\leq t\leq T} z_t' z_t}{\gamma' Z' Z\gamma} \leq \frac{\gamma'\gamma}{\left[\lambda_s\left(Z'Z\right)\right]\gamma'\gamma} \sum_{i=1}^k \frac{\max_{1\leq t\leq T} x_{ti}^2}{\sum_{t=1}^T x_{ti}^2} \longrightarrow 0.$$

1st inequality by Cauchy-Schwartz, 2nd inequality by definition of smallest eigenvalue and by definition of z_{it}, convergence by λ_s bounded away from 0 for large sample and the last term goes to 0 for each $i = 1, \ldots, k$ by assumption. Note that $Z'Z$ is the sample var-cov matrix for the regressors.

Uniform Convergence(in probability): Definition $\hat{Q}(\theta)$ converges in probality to $Q(\theta)$ uniformly over the compact set $\theta \in \Theta$ if $\forall \epsilon > 0$, $\lim_{T\to\infty} P\left(\sup_{\theta\in\Theta} |\hat{Q}(\theta) - Q(\theta)| > \epsilon\right) = 0$.

The concept of uniform convergence in probability is useful in two things:

1. Showing consistency of M-estimators.

2. Showing consistency of estimated variance-covariance etc.

Consistency of M-Estimators Read Thm 4.1.1 p106 Amemiya, and read Thm 2.1 p2121 Newey and McFadden.

Statement of theorem: If

1. $\lim_{T\to\infty} P\left(\sup_{\theta\in\Theta}|Q_T(\theta)-Q(\theta)|>\epsilon\right)=0$.
2. $Q(\theta)$ continuous and uniquely maximized at θ_0.
3. $\hat{\theta}=argmaxQ_T(\theta)$ over compact parameter set Θ.

plus some continuity and measurability requirements for $Q_T(\theta)$, then $\hat{\theta}\xrightarrow{p}\theta_0$.

Steps in Proof: Want $\forall\delta>0$, $P\left(|\hat{\theta}-\theta_0|\le\delta\right)\longrightarrow 1$.

1. First by condition (2), $\forall\delta>0$, $\exists\epsilon>0$ s.t. $|Q(\theta)-Q(\theta_0)|<\epsilon\Longrightarrow|\theta-\theta_0|<\delta$. Therefore the problem is translated into showing $P\left(|Q\left(\hat{\theta}\right)-Q(\theta_0)|<\epsilon\right)\longrightarrow 1$, or equivalently to $P\left(Q\left(\hat{\theta}\right)-Q(\theta_0)>-\epsilon\right)\longrightarrow 1$.
2. Since $Q_T\left(\hat{\theta}\right)>Q_T(\theta_0)$ by condition (3), then $Q\left(\hat{\theta}\right)-Q_T\left(\hat{\theta}\right)+Q_T\left(\hat{\theta}\right)-Q_T(\theta_0)+Q_T(\theta_0)-Q(\theta_0)\equiv Q\left(\hat{\theta}\right)-Q(\theta_0)>-\epsilon$ is implied by $Q\left(\hat{\theta}\right)-Q_T\left(\hat{\theta}\right)>-\epsilon/2$ and $Q_T(\theta_0)-Q(\theta_0)>-\epsilon/2$. But both of these last two events are implied by the event $\sup_{\theta\in\Theta}|Q_T(\theta)-Q(\theta)|<\epsilon/2$, the probability of which tends to 1 by condition (1).

Counter-examples: Read Ex 4.1.1-4.1.3 p108-109 in Amemiya.

Consistency of estimated var-cov matrix, Jacobian, etc: Read Thm 4.1.5 p113 Amemiya. Note the it is sufficient for uniform convergence to hold over a shrinking neighborhood of θ_0, the Newey and McFadden chapter made extensive use of this in many lemmas, i.e. only need $\forall\delta_n\longrightarrow 0$, $P\left(\sup_{|\theta-\theta_0|<\delta_n}|Q_T(\theta)-Q_T(\theta_0)|>\epsilon\right)\longrightarrow 0$, the same proof goes through. However, δ_n may goes to 0 arbitrary slowly, so essentially there is no difference.

Conditions for Uniform Convergence: Stochastic(uniform) Equicontinuity

Read Ch 21 of Davidson, skip the measurability problem part.

First think about just sequence of deterministic functions $f_n(\theta)$. When does a sequence of functions converges uniformly?

Uniform Equicontinuity for deterministic sequence of functions:

$$\lim_{\delta\to 0}\sup_n\sup_{|\theta'-\theta|<\delta}|f_n(\theta')-f_n(\theta)|=0.$$

This is not very useful, what if $f_n(\theta)$ may be discontinuous but the size of the jump goes to 0, say $\theta\in[0,1]$, $f_n(\theta)=0$ for $\theta\in[0,1/2]$, and $f_n(\theta)=\frac{1}{n}$ for $\theta\in(1/2,1]$. Still would like to say $f_n(\theta)\to 0$ uniformly in θ. So better need the following modification:

Asymptotic uniform equicontinuity for deterministic sequence of functions:

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{|\theta'-\theta|<\delta}|f_n(\theta')-f_n(\theta)|=0.$$

Convince yourself that this definition works for the previous example. The idea is to bound the fluctuation between fixed grids of Θ.

Uniform convergence of deterministic sequence of functions Read Thm 21.7 p335 Davidson

Statement of theorem: Θ compact, $\sup_{\theta\in\Theta}|f_n(\theta)| \longrightarrow 0$ if and only if $f_n(\theta) \longrightarrow 0$ for each θ and f_n is asymptotically uniformly equicontinuous.

Proof(sufficiency only): partition Θ into m balls with radius δ each, let $\tilde{\theta}_i$ be the center of the ith ball, $i = 1, \ldots, m$. Note the following

$$\sup_{\theta\in\Theta}|f_n(\theta)| \le \max_{1\le i\le m} \sup_{|\theta'-\tilde{\theta}_i|<\delta} |f_n(\theta') - f_n\left(\tilde{\theta}_i\right)| + \max_{1\le i\le m} |f_n\left(\tilde{\theta}_i\right)| \le \sup_{|\theta-\theta'|<\delta} |f_n(\theta) - f_n(\theta')| + \max_{1\le i\le m} |f_n\left(\tilde{\theta}_i\right)|$$

Take lim sup of the leftmost side and the rightmost side. "Only if" part see p336 Davidson. Now turning attention to the main problem: the stochastic case.

Stochastic uniform equicontinuity: See Davidson Ch 21, p327. You may also consult the original papers of Andrews and Newey in the reading list, if you have time and enjoy it.

Definition: A sequence of random functions $Q_n(\theta)$ is stochastic uniform equicontinuity if $\forall\epsilon > 0$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(\sup_{|\theta-\theta'|<\delta} |Q_n(\theta) - Q_n(\theta')| > \epsilon\right) = 0.$$

Alternative definition: $\forall\epsilon > 0$, $\exists\delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{|\theta-\theta'|<\delta} |Q_n(\theta) - Q_n(\theta')| > \epsilon\right) < \epsilon$$

Yet another definition: $\forall\epsilon > 0, \eta > 0$, $\exists\delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{|\theta-\theta'|<\delta} |Q_n(\theta) - Q_n(\theta')| > \epsilon\right) < \eta$$

Convince yourself that these definitions are all the same thing.

Uniform convergence in probability: Davidson Thm 21.9 p337. If $Q_n(\theta) \xrightarrow{p} 0$ for each θ, and $Q_n(\theta)$ is stochastic equicontinuous on $\theta\in\Theta$ compact, then $\sup_{\theta\in\Theta}|Q_n(\theta)| \xrightarrow{p} 0$.

Proof: Given $\delta > 0$, the parameter set Θ can always be partitioned into a finite number of m balls with radius no larger than δ, take $\tilde{\theta}_i$ to be the center of each ball, $i = 1, \ldots, m$, then

$$\begin{aligned}
P\left(\sup_{\theta\in\Theta}|\hat{Q}_n(\theta)| > \epsilon\right) &= P\left(\max_{1\le i\le m}\sup_{|\theta-\tilde{\theta}_i|}\left[|\hat{Q}_n(\theta) - \hat{Q}_n\left(\tilde{\theta}_i\right)| + |\hat{Q}_n\left(\tilde{\theta}_i\right)|\right] > \epsilon\right) \\
&\le P\left(\max_{1\le i\le m}\sup_{|\theta-\tilde{\theta}_i|<\delta} |\hat{Q}(\theta) - \hat{Q}\left(\tilde{\theta}_i\right)| > \epsilon/2\right) + P\left(\max_{1\le i\le m}|\hat{Q}_n\left(\tilde{\theta}_i\right)| > \epsilon/2\right) \\
&\le P\left(\sup_{|\theta-\theta'|<\delta}|\hat{Q}_n(\theta) - Q_n(\theta')| > \epsilon/2\right) + \sum_{i=1}^{m} P\left(|Q_n\left(\tilde{\theta}_i\right)| > \epsilon/2\right)
\end{aligned}$$

Take $\delta \to 0$ to get rid of the first term and $n \to \infty$ to get rid of the second term.

Comment: Although useful as highlevel conditions, the stochastic equicontinuity condition itself is not easy to use(may not be directly useful at all). If you are interested more in stochastic equicontinuity and the related subject of empirical processes where they are really useful in tackling nonsmooth problems, consult the Andrews chapter in Handbook as well as the two books by Pollard.

For simple problems where the objective function is smooth, differentiable, etc, uniform convergence can be verified without any of these fancy conditions, see the examples in Amemiya ex 4.3.1. and ex 4.3.2. pp131.

It is good to have some simple sufficient condition for stochastic equicontinuity.

Lipschitz condition for stochastic uniform equicontinuity: Read Lemma 2.9 p2138 Newey and McFadden, Theorem 21.10 p339 Davidson, consult the original Andrews(1992) and Newey(1991) papers only if you are interested.

Statement of condition: If $\forall \theta, \theta' \in \Theta$, if $|Q_n(\theta) - Q_n(\theta')| \leq B_n d(\theta, \theta')$ where $\lim_{\delta\to 0} \sup_{|\theta-\theta'|<\delta} d(\theta,\theta') = 0$ and $B_n = O_p(1)$, then $Q_n(\theta)$ is stochastic equicontinuous.

Proof: Note that $\sup_{\theta-\theta'|<\delta} |Q_n(\theta) - Q_n(\theta')| > \epsilon \Longrightarrow B_n \sup_{\theta-\theta'|<\delta} d(\theta,\theta') > \epsilon$. Therefore

$$\begin{aligned}
&\lim_{\delta\to 0} \limsup_{n\to\infty} P\left(\sup_{|\theta-\theta'|<\delta} |Q_n(\theta) - Q(\theta')| > \epsilon\right) \\
\leq &\lim_{\delta\to 0} \limsup_{n\to\infty} P\left(B_n \sup_{|\theta-\theta'|<\delta} d(\theta,\theta') > \epsilon\right) = \lim_{\delta\to 0} \limsup_{n\to\infty} P\left(B_n > \frac{\epsilon}{\sup_{|\theta-\theta'|<\delta} d(\theta',\theta)}\right) \\
= &\lim_{M\to\infty} \limsup_{n\to\infty} P(B_n > M) = 0. \qquad \text{where} \qquad M = \frac{\epsilon}{\sup_{|\theta-\theta'|<\delta} d(\theta',\theta)}.
\end{aligned}$$

Example: Suppose $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^n f(z_t,\theta)$, z_t iid, $f(z_t,\theta)$ differentiable with $f_\theta(z_t,\theta)$, then $|Q(\theta) - Q(\theta')| \leq \frac{1}{n}\sum_{t=1}^n |f_\theta(z_t,\bar{\theta})||\theta-\theta'|$, for $\bar{\theta} \in (\theta,\theta')$. If $b(z_t) = \sup_{\theta\in\Theta} f_\theta(z_t,\theta)$ is such that $Eb(z_t) < \infty$, then the Lipschitz condition holds with $B_n = \frac{1}{n}\sum_{t=1}^n b(z_t)$, b/o $P(B_n > M) \leq \frac{b(z)}{M}$. But what to do when the Lipschitz condition is not applicable?

An Uniform WLLN: See Theorem 4.2.1, pp116 of Amemiya, which is a slightly weaker version of lemma 2.4 p2129 of Newey and McFadden.

Statement of theorem: Θ compact, y_t iid, $g(y_t,\theta)$ continuous in θ for each y_t a.s., $Eg(y_t,\theta) = 0$, $E\sup_{\theta\in\Theta} |g(y_t,\theta)| < \infty$, then $\lim_{n\to\infty} P\left(\sup_{\theta\in\Theta} |\frac{1}{n}\sum_{t=1}^n g(y_t,\theta)| > \epsilon\right) = 0 \ \forall \epsilon > 0$.

Proof: Use pointwise convergence + stochastic equicontinuity.

1. $E\sup_{\theta\in\Theta} |g(y_t,\theta)| < \infty \Longrightarrow E|g(y_t,\theta)| > \infty$ for each θ, so use SLLN 2 to conclude $\frac{1}{n}\sum_{t=1}^n g(y_t,\theta) \stackrel{a.s.(p)}{\longrightarrow} 0$ for each θ.

2. Verify stochastic equicontinuity for $\frac{1}{n}\sum_{t=1}^n g(y_t,\theta)$:

$$\sup_{|\theta-\theta'|<\delta} |\frac{1}{n}\sum_{t=1}^n g(y_t,\theta) - g(y_t,\theta')| \leq \sup_{|\theta-\theta'|<\delta} \frac{1}{n}\sum_{t=1}^n |g(y_t,\theta) - g(y_t,\theta')| \leq \frac{1}{n}\sum_{t=1}^n \sup_{|\theta-\theta'|<\delta} |g(y_t,\theta) - g(y_t,\theta')|.$$

Therefore

$$\begin{aligned}
&\lim_{\delta\to 0} \limsup_{n\to\infty} P\left(\sup_{|\theta-\theta'|<\delta} |\frac{1}{n}\sum_{t=1}^n g(y_t,\theta) - g(y_t,\theta')| > \epsilon\right) \leq \lim_{\delta\to 0} \limsup_{n\to\infty} P\left(\frac{1}{n}\sum_{t=1}^n \sup_{|\theta-\theta'|<\delta} |g(y_t,\theta) - g(y_t,\theta')| > \epsilon\right) \\
\leq &\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{E\sum_{t=1}^n \sup_{|\theta-\theta'|<\delta} |g(y_t,\theta) - g(y_t,\theta')|}{n\epsilon} = \lim_{\delta\to 0} E \sup_{|\theta-\theta'|<\delta} |g(y_t,\theta) - g(y_t,\theta')|
\end{aligned}$$

Finally use (uniform b/o compact Θ) continuity of $g(y_t,\theta)$ and DOM. Since $\lim_{\delta\to 0} \sup_{|\theta-\theta'|<\delta} |g(y_t,\theta) - g(y_t,\theta')|$ almost surely, and $E\sup_\delta \sup_{|\theta-\theta'|<\delta} |g(y_t,\theta) - g(y_t,\theta')| < E2\sup_\theta |g(y_t,\theta)| < \infty$.