ECON 7710

Econometrics I

Lecture notes 2.

Convergence of random sequences:

- **Definition:** Suppose that X_n is the sequence of r.v. on $(\Omega, \mathcal{F}, P)$. This sequence converges in probability (in measure) to r.v. X if $\forall \epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

- Notation $X_n \xrightarrow{p} X$
- **Definition:** X_n converges almost surely (almost everywhere) to r.v. X if

$$P\left(\{\omega \,:\, X_n(\omega) \not\to X(\omega)\}\right) = 0.$$

- Notation $X_n \xrightarrow{a.s.} X$
- $X_n \xrightarrow{a.s.} X \;\Rightarrow\; X_n \xrightarrow{p} X$ but not the opposite
- Exception: if X_n is monotone increasing or decreasing, then $X_n \xrightarrow{a.s.} X \;\Leftrightarrow\; X_n \xrightarrow{p} X$.
  - By contradiction, suppose that $X_n \xrightarrow{p} X$, $X_n \downarrow$ and X_n does not converge a.s.
  - Then $\exists\ \epsilon > 0$ and set A, s.t. $P(A) \geq \delta > 0$, $\sup_{k\geq n} |X_k(\omega) - X(\omega)| > \epsilon$ for $\omega \in A$.
  - Since $X_n \downarrow$, then $\sup_{k\geq n} |X_k(\omega) - X(\omega)| = |X_n(\omega) - X(\omega)|$. And thus $P(|X_n - X| > \epsilon) \geq \delta$, which contradicts convergence in probability.
- **Theorem:** $X_n \xrightarrow{a.s.} X$ iff $Y_n = \sup\limits_{k\geq n} |X_k - X| \xrightarrow{p} 0$. In other words, $\forall \epsilon > 0$

$$P(\sup_{k\geq n} |X_k - X| > \epsilon) \to 0$$

- *Proof:* $X_n \xrightarrow{a.s.} X$ iff $Y_n \xrightarrow{a.s.} 0$ and sequence Y_n is monotone.

- **Theorem:** If the series $\sum_{k=1}^{\infty} P(|X_n - X| > \epsilon)$ converges for any $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$.
- *Proof:* $P\left(\cup_{k\geq n}\{|X_n - X| > \epsilon\}\right) \leq \sum\limits_{k=n}^{\infty} P(|X_n - X| > \epsilon)$.
- **Corollary:** If $X_n \xrightarrow{p} X$ then it has a subsequence X_{n_k} such that $X_{n_k} \xrightarrow{a.s.} X$
- *Proof:* Choose k such that $P(|X_{n_k} - X| > \epsilon) \leq a_k$ where $\sum_k a_k < \infty$, e.g. $a_k = 1/k^2$
- **Theorem:** If $Y_n = \sum_{k=1}^{n} X_k$ and X_k are independent, then $Y_n \xrightarrow{p} Y \;\Rightarrow\; Y_n \xrightarrow{a.s.} Y$.
- **Definition:** X_n converges to X in mean of order r (mean square if $r = 2$) or convergence in L_r if

$$\lim_{n\to\infty} E[|X_n - X|^r] = 0$$

- Notation $X_n \xrightarrow{(r)} X$
- $X_n \xrightarrow{(r)} X \;\Rightarrow\; X_n \xrightarrow{p} X$ (due to Chebychev's inequality)
- Neither convergence in probability nor a.s. convergence lead to convergence in mean.
- **Definition:** Sequence X_n is *a Cauchy sequence in probability* (a.s., in mean) if for any $\epsilon > 0$

$$\lim_{n,m\to\infty} P(|X_n - X_m| > \epsilon) = 0$$

$(\lim\limits_{n\to\infty} P(\sup\limits_{n\geq m} |X_n - X_m| > \epsilon) = 0,\; \sup\limits_{n\geq m} E[|X_n - X_m|^r] = 0)$

- **Lemma (Borel-Cantelli)** Let $\{A_n\}_{n=1}^{\infty}$ be sequence of events on $(\Omega, \mathcal{F}, P)$ and let $A = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k\geq n} A_k$. Then if $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$ then $P(A) = 0$
- *Proof:* $P(A) = \lim\limits_{n\to\infty} P\left(\bigcup_{k\geq n} A_k\right) \leq \lim\limits_{n\to\infty} \sum_{k\geq n} P(A_k) = 0$.
- **Theorem:** X_n converges in probability (a.s., in mean) iff it is a Cahuchy sequence in probability (a.s. in mean).

(Not super imp for the class → didn't go over in the class)

- *Proof:* Let X_n be Cauchy sequence in probability

- Take n_k, s.t.

$$P(|X_n - X_m| > 2^{-k}) < 2^{-k},\ n, m \geq n_k.$$

- Let $X'_k = X_{n_k}$ and $A_k = \{|X'_k - X'_{k+1}| > 2^{-k}\}$
- Then $P(A_k) \leq 2^{-k}$ and by Borel-Cantelli lemma w.p. 1 the number of occurrences of A_k is finite
- This means that for each $\omega \setminus A$ we can find $k_0(\omega)$, s.t. $|X'_k(\omega) - X'_{k+1}(\omega)| \leq 2^{-k}$.
- Thus, for all k, $l \geq k_0(\omega)$ $|X'_k(\omega) - X'_l(\omega)| \leq 2^{-k+1}$
- This means that $X'_k(\omega)$ is a Cauchy numeric sequence and $\exists\, X(\omega)$, s.t. $\lim\limits_{k\to\infty} |X'_k(\omega) - X(\omega)| = 0$
- Therefore

$$P(|X_n - X| \geq \epsilon) \leq P(|X_n - X_{n_k}| \geq \frac{\epsilon}{2}) + P(|X_{n_k} - X| \geq \frac{\epsilon}{2}) \to 0.$$

- **Theorem (continuous mapping):** Suppose that $X_n \xrightarrow{a.s.} X$ ($X_n \xrightarrow{p} X$) and $g(\cdot)$ is function continuous a.e. relative to r.v. X (i.e. continuous at each point of the set B s.t. $P(X \in B) = 1$). Then

$$g(X_n) \xrightarrow{a.s.} g(X)\ (g(X_n) \xrightarrow{p} g(X)).$$

Convergence of distributions:

- **Definition:** F_n weakly converges to F and denote $F_n \Rightarrow F$ if for all bounded continuous function $f(\cdot)$

$$\int f(x)\, dF_n(x) \to \int f(x)\, dF(x)$$

- **Theorem:** $F_n \Rightarrow F$ iff $F_n(x) \to F(x)$ at all x where $F(\cdot)$ is continuous.
- If $F(x)$ is continuous, then $F_n \Rightarrow F$ implies uniform convergence $\sup_x |F_n(x) - F(x)| \to 0$
- **Definition:** If for the distribution function of r.v. X_n and X $P_n \Rightarrow P$, then we say that X_n converges to X in distribution and denote it $X_n \xrightarrow{d} X$
- $X_n \xrightarrow{p} X$ implies $X_n \xrightarrow{d} X$ but not the other way around.

Limits of sequences of distributions:

- General definition of weak convergence requires verification of convergence for all bounded continuous functions $f(\cdot)$ which may be impractical; we need to see if instead we can only look at some smaller class

- Consider generalization of class of all cdf's $\mathcal{F}$ to class $\mathcal{G}$ such that for all $G \in \mathcal{G}$

  1. $\lim_{x_n \to +\infty} G_{X_1 \ldots X_n}(x_1, \ldots, x_n) = G_{X_1 \ldots X_{n-1}}(x_1, \ldots, x_{n-1})$
  2. $\lim_{x_n \to -\infty} G_{X_1 \ldots X_n}(x_1, \ldots, x_n) = 0$
  3. $G(-\infty) \geq 0$ and $G(+\infty) \leq 1$

  We call these the generalized distributions

- **Theorem (Helly-Bray):** The class $\mathcal{G}$ is a compact w.r.t. weak convergence $\Rightarrow$, i.e. from any sequence $\{G_n \in \mathcal{G}\}$ one can extract a converging subsequence $\mathcal{G}_{n_k} \Rightarrow G$

- The reason why we need extension of $\mathcal{F}$ is because it is not compact, i.e. $F_n \Rightarrow G$ with $F_n \in \mathcal{F}$ does not mean that $G \in \mathcal{F}$

  - Sequence of distribution functions

$$F_n(x) = \begin{cases} 0, \ x \leq -n, \\ \frac{1}{2}, \ -n < x \leq n, \\ 1, \ x > n. \end{cases}$$

    converges to $G(x) \equiv \frac{1}{2}$

- **Definition:** Sequence of distributions F_n is asymptotically tight if for any $\epsilon > 0$ we can find N such that $\inf_n(F_n(N) - F_n(-N)) > 1 - \epsilon$

- **Definition:** Class $\mathcal{L}$ of continuous and bounded functions defines the distribution if

$$\int f(x)\, dF(x) = \int f(x)\, dG(x), \quad F \in \mathcal{F}, \, G \in \mathcal{G},$$

  for all $f \in \mathcal{L}$ implies $F = G$

- **Theorem:** Suppose that $\mathcal{L}$ defines the distribution. Then $F \in \mathcal{F}$ with $F_n \Rightarrow F$ exists iff

  1. Sequence $\{F_n\}$ is asymptotically tight
  2. $\lim\limits_{n\to\infty} \int f \, dF_n$ exists for all $f \in \mathcal{L}$

- **Corollary:** Let $\mathcal{L}$ define the distributions and

$$\int f \, dF_n \to \int f \, dF, \quad F \in \mathcal{G}$$

  for any $f \in \mathcal{L}$. In addition at least one of the following three conditions hold

  1. $\{F_n\}$ is asymptotically tight
  2. $F \in \mathcal{F}$
  3. $f \equiv 1 \in \mathcal{L}$

  Then $F \in \mathcal{F}$ and $F_n \Rightarrow F$

- *Example:* Class of functions $\mathcal{L} = \{e^{itx}, \ t \in \mathbb{R}\}$ defines distributions

- **Theorem:** $F_n \Rightarrow F$ iff $\phi_n(t) \to \phi(t)$ for each t

- **Theorem:** Suppose that $\phi_n(t) = \int e^{itx} \, dF_n(x)$ is a sequence of characteristic functions and $\phi_n(t) \to \phi(t)$ for each t as $n \to \infty$. The following conditions are equivalent:

  1. $\phi(t)$ is a characteristic function
  2. $\phi(t)$ is continuous at $t = 0$
  3. Sequence $\{F_n\}$ is asymptotically tight

- Thus if convergence of c.f. occurs and one of the conditions is satisfied then there exists a distribution F which corresponds to the limit of $\{F_n\}$

Sequences of identically distributed independent r.v.:

- **Theorem (Khinchin's Law of Large Numbers):** Suppose that $\{X_n\}_{n=1}^{\infty}$ is sequence of independent identically distributed (i.i.d.) r.v. with $E[X_n] = a$ and let $S_n = \sum_{k=1}^{n} X_k$. Then $S_n/n \xrightarrow{p} a$

- *Proof:* C.f. of X_k for some neighborhood of 0

$$|\phi(t) - 1| < \frac{1}{2}.$$

Define $l(t) = \log\,\phi(t)$ in that neighborhood and given that the expectation exists

$$l'(0) = \frac{\phi'(0)}{\phi(0)} = ia.$$

Then for each t exists n such that $l(t/n)$ is well-defined

$$\phi_{S_n/n}(t) = \phi^n(t/n) = e^{l(t/n)\,n},$$

given that $l(0) = 0$, then for $n \to \infty$

$$e^{l(t/n)\,n} = exp\left(t\frac{l(t/n) - l(0)}{t/n}\right) \to e^{l'(0)t} = e^{iat}$$

This is c.f. of degenerate distribution at a. Thus $S_n/n \xrightarrow{p} a$.

- Sample average $\bar{X} = S_n/n$ allows to re-write LLN as $\bar{X} \xrightarrow{p} E[X_n]$

- Suppose that $V[X_n] = \sigma^2 < \infty$. Define the new sequence

$$Z_n = \frac{S_n - a\,n}{\sigma\sqrt{n}},$$

and let $\Phi(x)$ be the standard normal cdf.

- **Theorem (the Central Limit Theorem):** If $0 < \sigma^2 < \infty$, then

$$\boxed{\lim_{n\to\infty} \sup_x |P(Z_n < x) - \Phi(x)| = 0}$$

(i.e. uniformly in x)

- *Proof:* Uniform convergence follows from continuity of $\Phi(\cdot)$ and weak convergence. w.l.o.g. set $a = 0$. Since $V[X]$ exists then so does $\phi''(t)$ and thus

$$\phi(t) = \phi(0) + t\,\phi'(0) + \frac{t^2}{2}\phi''(0) + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2)$$

Thus for $n \to \infty$

$$\log\,\phi_{Z_n}(t) = n\,\log\left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\frac{t^2}{n}\right)\right) = n\,\left(-\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right) = -\frac{t^2}{2} + o(1) \to -\frac{t^2}{2}$$

Thus $\phi_{Z_n}(t)$ converges to the c.f. of standard normal random variables.

- CLT implies that $Z_n = \frac{S_n - an}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$

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

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

Properties of convergence:

- **Theorem:** Let X_n, X and Y_n be random vectors. Then
  - (i) $X_n \xrightarrow{a.s.} X$ implies $X_n \xrightarrow{p} X$;
  - (ii) $X_n \xrightarrow{p} X$ implies $X_n \xrightarrow{d} X$;
  - (iii) $X_n \xrightarrow{p} c$ (c is a constant) iff $X_n \xrightarrow{d} c$;
  - (iv) if $X_n \xrightarrow{d} X$ and $\|X_n - Y_n\| \xrightarrow{p} 0$, then $Y_n \xrightarrow{d} X$;
  - (v) if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ (c is a constant), then $(X_n, Y_n) \xrightarrow{d} (X, c)$;
  - (vi) if $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $(X_n, Y_n) \xrightarrow{p} (X, Y)$
- Note: (vi) is not true for convergence in distribution (see problem set)
- **Theorem (Slutsky):** Let X_n, X and Y_n be random vectors. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ (for a constant c) then
  - (i) $X_n + Y_n \xrightarrow{d} X + c$;
  - (ii) $Y_n X_n \xrightarrow{d} cX$;
  - (iii) $X_n / Y_n \xrightarrow{d} X/c$ (provided $c \neq 0$)
- *Example:* For i.i.d. $Y_1, \dots, Y_n$ with $E[Y_i^2] < \infty$ construct sample variance

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2, \text{ with } S_n^2 \xrightarrow{p} V[Y_i]$$

  by Continuous Mapping Theorem theorem. We then construct the t-statistic as

$$t_n = \frac{\sqrt{n}(\bar{Y} - E[Y_n])}{S_n}.$$

  Then by CLT, continuous mapping theorem and Slutsky's theorem $t_n \xrightarrow{d} N(0,1)$

Convergence in non-i.i.d settings:

- **Theorem (Lindeberg-Feller CLT):** Let $Y_{n,1}, \ldots, Y_{n,k_n}$ be independent r.v. with finite variances such that
$$\sum_{i=1}^{k_n} E[\|Y_{n,i}\|^2]\mathbf{1}\{\|Y_{n,i}\| > \epsilon\} \to 0, \quad \forall \epsilon > 0,$$
$\sum_{i=1}^{k_n} E[(Y_{n,i} - E[Y_{n,i}])'(Y_{n,i} - E[Y_{n,i}])] \to \Sigma$. Then
$$\sum_{i=1}^{k_n} (Y_{n,i} - E[Y_{n,i}]) \xrightarrow{d} N(0, \Sigma).$$

- *Example:* Linear regression expressed as
$$Y = X\beta + e,$$
with known $n \times p$ full rank matrix X, unobserved vector of errors e with i.i.d. components with mean zero and variance σ^2. We need to estimate β from observation (Y, X). Least squares estimator is
$$\widehat{\beta} = (X'X)^{-1}X'Y$$
which is unbiased with covariance matrix $\sigma^2(X'X)^{-1}$. Define matrix $A^{1/2}$ such that for PSD A, $A = A^{1/2'}A^{1/2}$. Then
$$(X'X)^{1/2}(\widehat{\beta} - \beta) = (X'X)^{-1/2}X'e$$
Write $(X'X)^{-1/2}X'e = \sum_{i=1}^n a_{ni}e_i$, where a_{ni} is the ith column of matrix $A = (X'X)^{-1/2}X'$. Given that $E[(X'X)^{-1/2}X'ee'X(X'X)^{-1/2'}] = \sigma^2 I$, $\Sigma = \sigma^2 I$ in the statement of Lindeberg-Feller theorem. Next we need to ensure that
$$\sum_{i=1}^n \|a_{ni}\|^2 E[e_i^2 \mathbf{1}\{\|a_{ni}\||e_i| > \epsilon\}] \to 0.$$
Note that $\sum_{i=1}^n \|a_{ni}\|^2 = \text{trace}(AA') = p$. Thus
$$\sum_{i=1}^n \|a_{ni}\|^2 E[e_i^2 \mathbf{1}\{\|a_{ni}\||e_i| > \epsilon\}] \le \sum_{i=1}^n \|a_{ni}\|^2 \max_{1 \le i \le n} E[e_i^2 \mathbf{1}\{\|a_{ni}\||e_i| > \epsilon\}]$$
$$\le pE[e_i^2] \max_{1 \le i \le n} E[\mathbf{1}\{\|a_{ni}\||e_i| > \epsilon\}] \le p\sigma^2 \max_{1 \le i \le n} \frac{\|a_{ni}\|E[|e_i|]}{\epsilon} = \frac{p\sigma^2 E[|e_i|]}{\epsilon} \max_{1 \le i \le n} \|a_{ni}\|.$$
Lideberg-Feller CLT holds for least squares estimator if $\max_{1 \le i \le n} \|a_{ni}\| \to 0$.

Delta method:

- **Theorem:** If $g(\cdot)$ satisfies the continuous mapping theorem, then $g(X_n) \xrightarrow{d} g(X)$
- **Theorem:** Suppose that $X_n \xrightarrow{d} X$ and $g(\cdot)$ is differentiable at a, $b_n \to 0$. Then

$$\frac{g(a + b_n X_n) - g(a)}{b_n} \xrightarrow{d} X\, g'(a).$$

  If $g'(a) = 0$ and $g''(a)$ exists, then

$$\frac{g(a + b_n X_n) - g(a)}{b_n^2} \xrightarrow{d} \frac{1}{2} X^2\, g''(a).$$

- *Example:* Suppose that for i.i.d. $X_1, \ldots, X_n$, $E[X_i] = 0$ and $E[X_i^2] = 1$. In this case $\sqrt{n}\bar{X} \xrightarrow{d} N(0,1)$ by CLT and $n\bar{X} \xrightarrow{d} \chi_1^2$ by CMT. Then

$$\sqrt{n}(\cos(\bar{X}) - 1) \xrightarrow{d} 0.$$

  However

$$\cos\bar{X} - \cos 0 = (\bar{X} - 0)0 + \frac{1}{2}(\bar{X} - 0)^2 (\cos x)''|_{x=0} + \ldots$$

  Thus $-2n(\cos\bar{X} - 1) \xrightarrow{d} \chi_1^2$