Various Modes of Convergence

Definitions

- (convergence in probability) A sequence of random variables $\{X_n\}$ is said to converge **in probability** to a random variable X as $n \to \infty$ if for any $\varepsilon > 0$ we have
$$\lim_{n\to\infty} P\left[\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\right] = 0.$$
We write $X_n \xrightarrow{p} X$ or $\text{plim}X_n = X$.

- (convergence in distribution) Let F and F_n be the distribution functions of X and X_n, respectively. The sequence of random variables $\{X_n\}$ is said to converge **in distribution** to a random variable X as $n \to \infty$ if
$$\lim_{n\to\infty} F_n(z) = F(z)$$
for all $z \in R$ and z is a continuity points of F. We write $X_n \xrightarrow{d} X$ or $F_n \xrightarrow{d} F$.

- (almost sure convergence) We say that a sequence of random variables $\{X_n\}$ converges almost surely or with probability 1 to a random variable X as $n \to \infty$ if
$$P\left[\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right] = 1.$$
We write $X_n \xrightarrow{a.s.} X$.

- (L^r convergence) A sequence of random variables $\{X_n\}$ is said to converge **in $\mathbf{L}^r$ norm** to a random variable X as $n \to \infty$ if for some $r > 0$
$$\lim_{n\to\infty} E\left[|X_n - X|^r\right] = 0.$$
We denote as $X_n \xrightarrow{L^r} X$. If $r = 2$, it is called mean square convergence and denoted as $X_n \xrightarrow{m.s.} X$.

Relationship among various modes of convergence

$$\begin{array}{c}\text{[almost sure convergence]} \Rightarrow \text{[convergence in probability]} \Rightarrow \text{[convergence in distribution]} \\ \Uparrow \\ \text{[convergence in } L^r \text{ norm]}\end{array}$$

Example 1 *Convergence in distribution does not imply convergence in probability.*

$\Rightarrow$ Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Define the random variables X_n and X such that
$$\begin{array}{c} X_n(\omega_1) = X_n(\omega_2) = 1, X_n(\omega_3) = X_n(\omega_4) = 0 \text{ for all } n \\ X(\omega_1) = X(\omega_2) = 0, X(\omega_3) = X(\omega_4) = 1 \end{array}$$
Moreover, we assign equal probability to each event. Then,
$$F(x) = \left\{\begin{array}{l} 0, \text{ if } x < 0 \\ \frac{1}{2}, \text{ if } 0 \leq x < 1 \\ 1, \text{ if } x \geq 1 \end{array}\right\} \qquad F_n(x) = \left\{\begin{array}{l} 0, \text{ if } x < 0 \\ \frac{1}{2}, \text{ if } 0 \leq x < 1 \\ 1, \text{ if } x \geq 1 \end{array}\right\}$$
Since $F_n(x) = F(x)$ for all n, it is trivial that $X_n \xrightarrow{d} X$. However,
$$\lim_{n\to\infty} P\left[\omega : |X_n(\omega) - X(\omega)| \geq \frac{1}{2}\right] = 1$$
Note that $|X_n(\omega) - X(\omega)| = 1$ for all n and ω. Hence, $X_n \stackrel{p}{\nrightarrow} X$.

Example 2 *Convergence in probability does not imply almost sure convergence.*

$\Rightarrow$ Consider the sequence of independent random variables $\{X_n\}$ such that

$$P[X_n = 1] = \frac{1}{n}, \qquad P[X_n = 0] = 1 - \frac{1}{n} \qquad\qquad n \geq 1$$

Obviously for any $0 < \varepsilon < 1$, we have

$$P[|X_n - X| > \varepsilon] = P[X_n = 1] = \frac{1}{n} \to 0$$

Hence, $X_n \xrightarrow{p} X$. In order to show $X_n \overset{a.s.}{\nrightarrow} X$, we need the following lemma.

Lemma 3 *$X_n \overset{a.s.}{\to} X \Leftrightarrow P(B_m(\varepsilon)) \to 0$ as $m \to \infty$ for all $\varepsilon > 0$ where $B_m(\varepsilon) = \bigcup_{n=m}^{\infty} A_n(\varepsilon)$ and $A_n(\varepsilon) = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$.*

Proof. Let $C = \{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}, A(\varepsilon) = \{\omega : \omega \in A_n(\varepsilon) \ i.o.\}$.
Then, $P(C) = 1$ if and only if $P(A(\varepsilon)) = 0$ for all $\varepsilon > 0$. However, $B_m(\varepsilon)$ is a decreasing sequence of events, $B_m(\varepsilon) \downarrow A(\varepsilon)$ as $m \to \infty$ and so $P(A(\varepsilon)) = 0$ if and only if $P(B_m(\varepsilon)) \to \infty$ as $m \to \infty$. ■

Continuing the counter-example, we have

$$\begin{aligned} P(B_m(\varepsilon)) &= 1 - \lim_{M \to \infty} P[X_n = 0 \text{ for all } n \text{ such that } m \leq n \leq M] \\ &= 1 - \left(1 - \frac{1}{m}\right)\left(1 - \frac{1}{m+1}\right)\cdots \\ &= 1 \end{aligned}$$

Hence, $X_n \overset{a.s.}{\nrightarrow} X$.

Example 4 *Convergence in probability does not imply convergence in L^r − norm.*

$\Rightarrow$ Let $\{X_n\}$ be a random variable such that

$$P[X_n = e^n] = \frac{1}{n}, P[X_n = 0] = 1 - \frac{1}{n}$$

Then, for any $\varepsilon > 0$ we have

$$P[|X_n| < \varepsilon] = 1 - \frac{1}{n} \to 1 \text{ as } n \to \infty$$

Hence, $X_n \xrightarrow{p} 0$. However, for each $r > 0$,

$$E[|X_n - 0|^r] = E[X_n^r] = e^{rn}\frac{1}{n} \to \infty \text{ as } n \to \infty$$

Hence, $X_n \overset{L^r}{\nrightarrow} 0$.

Some useful theorems

Theorem 5 *Let $\{X_n\}$ be a random vector with a fixed finite number of elements. Let g be a real-valued function* ***continuous*** *at a constant vector point α. Then $X_n \overset{p(a.s.)}{\to} \alpha$ implies $g(X_n) \overset{p(a.s.)}{\to} g(\alpha)$.*

$\Rightarrow$ By continuity of g at α, for any $\epsilon > 0$ we can find δ such that $\|X_n - \alpha\| < \delta$ implies $|g(X_n) - g(\alpha)| < \epsilon$. Therefore,

$$P[\|X_n - \alpha\| < \delta] \leq P[|g(X_n) - g(\alpha)| < \epsilon] \to 0 \text{ as } n \to \infty.$$

Theorem 6 *Suppose that* $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} \alpha$ *where* α *is* ***non-stochastic****. Then*

$$(i)\ X_n + Y_n \xrightarrow{d} X + \alpha$$
$$(ii)\ X_n Y_n \xrightarrow{d} \alpha X$$
$$(iii)\ \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{\alpha} \text{ provided } \alpha \text{ is not zero.}$$

- Note the condition that $Y_n \xrightarrow{p} \alpha$ where α is non-stochastic. If α is also a random vector, $X_n + Y_n \xrightarrow{d} X + \alpha$ is not necessarily true. A counter-example is given by

$$P[X_n = 0] = P[X_n = 1] = \frac{1}{2} \text{ for all } n$$
$$P[Y_n = 0] = P[Y_n = 1] = \frac{1}{2} \text{ for all } n$$

Then,

$$X_n \xrightarrow{d} Z \text{ and } Y_n \xrightarrow{d} Z$$

where $P[Z = 0] = P[Z = 1] = \frac{1}{2}$. However,

$$X_n + Y_n \xrightarrow{d} W$$

where $P[W = 0] = P[W = 2] = \frac{1}{4}$ and $P[W = 1] = \frac{1}{2}$. Hence, $W \neq 2Z$.

Theorem 7 *Let* $\{X_n\}$ *be a random vector with a fixed finite number of elements. Let* g *be a* ***continuous*** *real-valued function . Then* $X_n \xrightarrow{d} X$ *implies* $g(X_n) \xrightarrow{d} g(X)$*.*

Theorem 8 *Suppose* $X_n \xrightarrow{d} X$ *and* $X_n - Y_n \xrightarrow{p} 0$*, then* $Y_n \xrightarrow{d} X$*.*

Inequalities frequently used in large sample theory

Proposition 9 *(Chebychev's inequality) For* $\varepsilon > 0$

$$P[|X| \geq \varepsilon] \leq \frac{E(X^2)}{\varepsilon^2}$$

Proposition 10 *(Markov's inequality) For* $\varepsilon > 0$ *and* $p > 0$

$$P[|X| \geq \varepsilon] \leq \frac{E(X^p)}{\varepsilon^p}$$

Proposition 11 *(Jensen's inequality) If a function* ϕ *is convex on an interval* I *containing the support of a random variable* X*, then*

$$\phi(E(X)) \leq E(\phi(X))$$

Proposition 12 *(Cauchy-Schwartz inequality) For random variables* X *and* Y

$$E(XY)^2 \leq E(X^2)\, E(Y^2)$$

Proposition 13 *(Hölder's inequality) For any* $p \geq 1$

$$E|XY| \leq (E|X|^p)^{\frac{1}{p}} (E|Y|^q)^{\frac{1}{q}}$$

where $q = \frac{p}{p-1}$ *if* $p > 1$*, and* $q = \infty$ *if* $p = 1$*.*

Proposition 14 *(Lianpunov's inequality) If* $r > p > 0$*,*

$$(E|X|^r)^{\frac{1}{r}} \geq (E|X|^p)^{\frac{1}{p}}$$

Proposition 15 *(Minkowski's inequality) For $r \geq 1$,*

$$\left(E\left|X+Y\right|^{r}\right)^{\frac{1}{r}} \leq \left(E\left|X\right|^{r}\right)^{\frac{1}{r}} + \left(E\left|Y\right|^{r}\right)^{\frac{1}{r}}$$

Proposition 16 *(Loève's c_r inequality) For $r > 0$,*

$$E\left|\sum_{i=1}^{m} X_i\right|^{r} \leq c_r \sum_{i=1}^{m} E\left|X_i\right|^{r}$$

where $c_r = 1$ when $0 < r \leq 1$, and $c_r = m^{r-1}$when $r > 1$.

Laws of Large Numbers

- Suppose we have a set of observation $X_1, X_2, \cdots, X_n$. A law of large numbers basically gives us the behavior of sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ when the number of observations n goes to infinity. It is needless to say that we need some restrictions(assumptions) on the behavior of each individual random variable X_i and on the relationship among $X_i's$. There are many versions of law of large numbers depending on what kind of restriction we are wiling to impose. The most generic version can be stated as

  > Given restrictions on the dependence, heterogeniety, and moments of a sequence of random variables $\{X_i\}$, $\overline{X}_n$ converges in some mode to a parameter value.

  When the convergence is in probability sense, we call it a weak law of large numbers. When in almost sure sense, it is called a strong law of large numbers.

- We will have a kind of trade-off between dependence or heterogeneity and existence of higher moments. As we want to allow for more dependence and heterogeneity, we have to accept the existence of higher moment, in general.

Theorem 17 *(Komolgorov SLLN I) Let $\{X_i\}$ be a sequence of independently and identically distributed random variables. Then $\overline{X}_n \overset{a.s.}{\to} \mu$ if and only if $E(X_i) = \mu < \infty$.*

Remark 18 *The above theorem requires the existence of the first moment only. However, the restriction on dependence and heterogeneity is quite severe. The theorem requires i.i.d.(random sample), which is rarely the case in econometrics. Note that the theorem is stated in necessary and sufficient form. Since almost sure convergence always implies convergence in probability, the theorem can be stated as $\overline{X}_n \overset{p}{\to} \mu$. Then it is a weak law of large numbers.*

Theorem 19 *(Komolgorov SLLN II) Let $\{X_i\}$ be a sequence of independently distributed random variables with finite variances $Var(X_i) = \sigma_i^2$. If $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty$, then $\overline{X}_n - \overline{\mu}_n \overset{a.s.}{\to} 0$ where $\overline{\mu}_n = E\left(\overline{X}_n\right) = \frac{1}{n}\sum_{i=1}^{n} \mu_i$.*

Remark 20 *Here we allow for the heterogeneity of distributions in exchange for existence of the second moment. Still, they have to be independent. Intuitive explanation for the summation condition is that we should not have variances grow too fast so that we have a shrinking variance for the sample mean.*

- The existence of the second moment is too strict in some sense. The following theorem might be a theoretical purism. But we can obtain a SLLN with milder restriction on the moments.

Theorem 21 *(Markov SLLN) Let $\{X_i\}$ be a sequence of independently distributed random variables with finite means $E(X_i) = \mu_i < \infty$. If for some $\delta > 0$,*

$$\sum_{i=1}^{\infty} \frac{E\left|X_i - \mu_i\right|^{1+\delta}}{i^{1+\delta}} < \infty$$

then, $\overline{X}_n - \overline{\mu}_n \overset{a.s.}{\to} 0$.

Remark 22 *When $\delta = 1$, the theorem collapses to Komolgorov SLLN II. Here we don't need the existence of the second moment. All we need is the existence of the moment of order $(1+\delta)$ where $\delta > 0$.*

- We now want to allow some dependence among $X_i's$. This modification is especially important when we are dealing with time series data which has a lot of dependence structure in it.

Theorem 23 *(Ergodic theorem) Let $\{X_i\}$ be a (weakly) stationary and ergodic sequence with $E|X_i| < \infty$. Then, $\overline{X}_n - \mu \overset{a.s.}{\to} 0$ where $\mu = E(X_i)$.*

Remark 24 *By stationarity, we have $E(X_i) = \mu$ for all i. And ergodicity enables us to have, roughly speaking, an estimate of μ as a sample mean of $X_i's$. Both stationarity and ergodicity are restrictions on dependence structure - which sometimes seem quite severe for econometric data.*

- In order to allow both dependence and heterogeneity we need more specific structure on the dependence of the data series called strong mixing and uniform mixing. The LLN's in case of mixing requires some technical discussion. Anyway, one of the most important SLLN's in econometrics is McLeish's.

Theorem 25 *(McLeish) Let $\{X_i\}$ be a sequence with a uniform mixing of size $\frac{r}{2r-1}$ or a strong mixing of size $\frac{r}{r-1}$, $r > 1$, with finite means $E(X_i) = \mu_i$. If for some $\delta, 0 < \delta \leq r, \sum_{i=1}^{\infty}\left(\frac{E|X_i - \mu_i|^{r+\delta}}{t^{r+\delta}}\right)^{\frac{1}{r}} < \infty$, then $\overline{X}_n - \overline{\mu}_n \overset{a.s.}{\to} 0$.*

- Another form of SLLN important in econometric application is SLLN for a martingale difference sequence. A stochastic process X_t is called a martingale difference sequence if

$$E(X_t \mid \mathcal{F}_{t-1}) = 0 \text{ for all } t$$

where $\mathcal{F}_{t-1} = \sigma(X_{t-1}, X_{t-2}, \cdots)$ i.e., information up to time $(t-1)$.

Theorem 26 *(Chow) Let $\{X_i\}$ be a martingale difference sequence. If for some $r \geq 1$, $\sum_{i=1}^{\infty} \frac{E|X_i|^{2r}}{t^{1+r}} < \infty$, then $\overline{X}_n \overset{a.s.}{\to} 0$.*

Central Limit Theorems

- All CLT's are meant to derive the distribution of sample mean as $n \to \infty$ when appropriately scaled. We have many versions of CLT depending on our assumptions on the data. The easiest and most frequently cited CLT is

Theorem 27 *(Lindeberg-Levy CLT) Let $\{X_i\}$ be a sequence of independently and identically distributed random variables. If $Var(X_i) = \sigma^2 < \infty$, then*

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(X_i - \mu)}{\sigma} \overset{d}{\to} N(0,1).$$

Remark 28 *The conclusion of the theorem can be also written as $\sqrt{n}\left(\overline{X}_n - \mu\right) \overset{d}{\to} N\left(0, \sigma^2\right)$. We requires the existence of the second moment even if we have i.i.d. sample. (Compare this with LLN).*

Theorem 29 *(Lindeberg-Feller CLT) Let $\{X_i\}$ be a sequence of independently distributed random variables with $E(X_i) = \mu_i, Var(X_i) = \sigma_i^2 < \infty$ and distribution function $F_i(x)$. Then*

$$\frac{\sqrt{n}\left(\overline{X}_n - \overline{\mu}_n\right)}{\overline{\sigma}_n} \overset{d}{\to} N(0,1)$$

and

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \frac{1}{n}\left(\frac{\sigma_i^2}{\overline{\sigma}_n}\right) = 0$$

if and only if

$$\lim_{n\to\infty} \overline{\sigma}_n^{-2} n^{-1} \sum_{i=1}^{n} \int_{(x-\mu_i)^2 > \epsilon n \overline{\sigma}_n^2} (x-\mu_i)^2 \, dF_i(x) = 0.$$

Remark 30 *The condition is called "Lindeberg condition". The condition controls the tail behavior of X_i so that we have a proper distribution for scaled sample mean. We do not need identical distribution here. The condition is difficult to verify in practice. A search for a sufficient condition for the Lindeberg condition leads to the following CLT.*

Theorem 31 *(Liapounov CLT) Let $\{X_i\}$ be a sequence of independently distributed random variables with $E(X_i) = \mu_i, Var(X_i) = \sigma_i^2 < \infty$, and $E|X_i - \mu_i|^{2+\delta} < \infty$ for some $\delta > 0$ and all i. If $\overline{\sigma}_n^2 > \gamma > 0$ for all n sufficiently large, then $\frac{\sqrt{n}(\overline{X}_n - \overline{\mu}_n)}{\overline{\sigma}_n} \xrightarrow{d} N(0,1)$.*

Remark 32 *We can show that the moment restrictions in the theorem are enough to obtain the Lindeberg condition.*

Theorem 33 *Let $\{X_i\}$ be a (weakly) stationary and ergodic sequence with $E(X_i^2) = \sigma^2 < \infty$. Suppose that $E(X_0 \mid \mathcal{F}_{-m}) \xrightarrow{L^2} 0$ as $m \to \infty$ and $\sum_{j=0}^{\infty} (Var[E(X_0 \mid \mathcal{F}_{-j}) - E(X_0 \mid \mathcal{F}_{-j-1})])^{\frac{1}{2}} < \infty$, where $\mathcal{F}_{-m} = \sigma(\cdots, X_{-m-2}, X_{-m-1}.X_{-m})$. Then, $\overline{\sigma}_n^2 \to \overline{\sigma}^2$ as $n \to \infty$, and if $\overline{\sigma}^2 > 0$, then $\frac{\sqrt{n}\overline{X}_n}{\overline{\sigma}^2} \xrightarrow{d} N(0,1)$.*

Remark 34 *The above theorem allows some dependence structure but retains homogeneity through stationarity and ergodicity.*

Theorem 35 *(White and Domowitz) Let $\{X_i\}$ be a sequence of mixing random variables such that either uniform mixing or strong mixing is if size $\frac{r}{r-1}, r > 1$, with $E(X_i) = \mu_i, Var(X_i) = \sigma_i^2 < \infty$, and $E|X_i|^{2r} < \infty$ for all i. Define $\overline{\sigma}_{a,n}^2 = Var\left[n^{-\frac{1}{2}} \sum_{i=a+1}^{a+n} X_i\right]$. If there exists $0 < \overline{\sigma}^2 < \infty$, such that $\overline{\sigma}_{a,n}^2 \to \overline{\sigma}^2$ as $n \to \infty$ uniformly in a, then $\frac{\sqrt{n}(\overline{X}_n - \overline{\mu}_n)}{\overline{\sigma}_n} \xrightarrow{d} N(0,1)$ where $\overline{\sigma}_n = \overline{\sigma}_{0,n}^2$.*

Remark 36 *The above CLT is quite general in the sense that we can allow reasonable dependence and heterogeneity structures to be applied to econometric data. However, as shown in the statement of the theorem, it is impractical to check the conditions of the theorem in practice.*

- Finally, we will have a CLT which can be applied to a martingale difference sequence.

Theorem 37 *Let $\{X_i\}$ be a martingale difference sequence such that $E(X_i^2) = \sigma_i^2$ and $E|X_i|^{2+\delta} < \infty$ for some $\delta > 0$ and all i. If $\overline{\sigma}_n^2 > \gamma > 0$ for all n sufficiently large and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{\sigma}_n^2 \xrightarrow{p} 0$, then $\frac{\sqrt{n}\overline{X}_n}{\overline{\sigma}^2} \xrightarrow{d} N(0,1)$ where $\overline{\sigma}^2 = \lim_{n\to\infty} \overline{\sigma}_n^2$.*