Basic   Chapter 7   1.

VI Convergence

(a) Convergence in Distribution (weak convergence)

Def A sequence $X_1, X_2, \ldots$ of r.v.'s converges in distribution to r.v. X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for every $x \in \mathbb{R}$ where F_X is continuous.

- **note** – we write $X_n \xrightarrow{d} X$ (sometimes called weak convergence of P_{X_n} to P_X)

  – $P_{X_n}((a,b]) = F_{X_n}(b) - F_{X_n}(a)$

$$\approx F_X(b) - F_X(a) = P_X((a,b])$$

for cty pts a, b of F_X

eg 1. $P(X_n = -\frac{1}{n}) = P(X_n = \frac{1}{n}) = \frac{1}{2}$

$P(X = 0) = 1$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

0 is a discontinuity pt of F_X

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \begin{cases} 0 & x < -\frac{1}{n} \\ \frac{1}{2} & -\frac{1}{n} \leq x < \frac{1}{n} \\ 1 & \frac{1}{n} \leq 1 \end{cases} = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases}$$

$\therefore X_n \xrightarrow{d} X$ | does $X_n(\omega) \to X(\omega)$?

Prop ① (Continuity Theorem)

Suppose $X_1, X_2, \ldots$ is a sequence of r.v.s with dist. fns $F_{X_1}, F_{X_2}, \ldots$ and c.f.'s $c_{X_1}, c_{X_2}, \ldots$.

(i) if $X_n \xrightarrow{d} X$ then $c_{X_n} \to c_X$

(ii) if $\phi = \lim c_{X_n}$ and ϕ is cont. at 0 then ϕ is the c.f. of some r.v. X and $X_n \xrightarrow{d} X$.

Proof: accept

Prop ② (Weak Law of Large Numbers)

If $X_1, X_2, \ldots$ are i.i.d. r.v.s and $EX_i = \mu \in \mathbb{R}$ then $\frac{1}{n} S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{d} \mu$.

$\boxed{\text{Proof via Markovs inequality when } EX_i^2 < \infty}$

Proof: Let X be degenerate at μ and

$$c_X(t) = E[e^{itX}] = e^{it\mu}.$$ Now

$$c_{\frac{1}{n}S_n}(t) = E[e^{it\frac{1}{n}S_n}] = \prod_{i=1}^{n} c_{X_i}(t/n) = c_{X_1}^{n}(t/n).$$

(**Lemma** $c_{X_1}(t) = 1 + i\mu t + o(t)$ (expansion of c_{X_1} to two terms)

Proof: $$\lim_{t \to 0} \frac{c_{X_1}(t) - c_{X_1}(0)}{t} = \lim_{t \to 0} E\left[\frac{e^{itX} - 1}{t}\right] = ⊛$$

and $$\left|\frac{e^{itX} - 1}{t}\right| = \left|\frac{1 + itX + o(tX) - 1}{t}\right| = O(|X|)$$ and then by

DCT $$⊛ = E\left[\lim_{t \to 0} \frac{e^{itX} - 1}{t}\right] = E[iX] = i\mu$$

Thus $c_{X_1}(t) = 1 + i\mu t + o(t)$). Therefore

$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = x + o(x)$

$$\hat{c}_{X_1}\left(\tfrac{t}{n}\right) = \left(1 + i\mu\tfrac{t}{n} + o\left(\tfrac{t}{n}\right)\right)^n \sim \left(1 + \tfrac{i\mu t}{n}\right)^n \longrightarrow e^{i\mu t}$$

and the result follows from the Continuity Theorem.

Prop 3. (Central Limit Theorem)
If $X_1, X_2, \ldots$ are iid r.v.'s with mean μ and variance σ^2 and $Z \sim N(0,1)$ then

$$Z_n = \frac{\frac{1}{n}S_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z$$

Ex $\mathbb{E}[\frac{1}{n}S_n] = \mu$, $\mathrm{Var}[\frac{1}{n}S_n] = \frac{\sigma^2}{n}$

Proof: Let $Y_n = \frac{X_n - \mu}{\sigma}$, then $\mathbb{E}[Y_n] = 0$, $\mathrm{Var}[Y_n] = 1$, $c_{Y_n}(t) = e^{-it\frac{\mu}{\sigma}} c_{X_n}\left(\frac{t}{\sigma}\right)$ and put $Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$. Then $c_{Z_n}(t)$

$$= \left(c_{Y_1}\left(\tfrac{t}{\sqrt{n}}\right)\right)^n = \left(1 + i\,\mathbb{E}[Y_1]\tfrac{t}{\sqrt{n}} + i^2\frac{\mathbb{E}[Y_1^2]}{2!}\frac{t^2}{n} + o\left(\tfrac{t^2}{n}\right)\right)^n$$

$$= \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \longrightarrow e^{-t^2/2}$$

and the result follows by the Cty Thm.

- there are numerous generalizations of the CLT

eg Simple random walk

$$S_0 = 0 \quad , \quad E[X_i] = p\cdot 1 + (1-p)(-1) = 2p-1$$

$$Var[X_i] = p\cdot 1^2 + (1-p)(-1)^2 - (2p-1)^2$$

$$= 1 - (4p^2 - 4p + 1) = 4p(1-p)$$

$$\therefore \quad \frac{\frac{1}{n}S_n - (2p-1)}{\sqrt{\frac{4p(1-p)}{n}}} \xrightarrow{D} N(0,1).$$

- **note** - convergence in distribution can be defined for a sequence of random vectors X_n to random vector $\underset{\sim}{x}$; namely $\underset{\sim}{X}_n \xrightarrow{D} \underset{\sim}{x}$ if $P_{\underset{\sim}{X}_n}(B) \to P_{\underset{\sim}{x}}(B)$ for every Borel set $B \in \mathcal{B}^k$ satisfying $P_{\underset{\sim}{x}}(\partial B) = 0$

  - also get vector versions of WLLN, CLT etc.

Prop 4 (Skorokhod Representation Theorem)
If $\{X_n\}$, X are r.v.'s on $(\Omega, \mathcal{F}, P)$ and $X_n \xrightarrow{D} X$ then $\exists$ prob. space $(\Omega', \mathcal{F}', P')$ and r.v.'s $\{X'_n\}$, X' defined on it s.t.

(i) $F_{X'_n} = F_{X_n}$, $F_{X'} = F_X$

(ii) $X'_n \xrightarrow{a.s.} X$.

Proof: see text

(b) Large Deviations

- CLT implies (for $a>0$)

$$P\left(\left|\tfrac{1}{n}S_n-\mu\right|>a\right) = P\left(\frac{\left|\tfrac{1}{n}S_n-\mu\right|}{\sigma/\sqrt{n}} > \frac{\sqrt{n}\,a}{\sigma}\right)$$

$\to 0$ as $n \to \infty$

- at what rate?

Prop ⑤ (Large Deviations) If $X_1, X_2, \ldots$ are iid with mean μ and mgf m_X and $P(X>\mu+a)>0$ then for $a>0$

$$P\left(\tfrac{1}{n}S_n-\mu>a\right) \sim \left[\inf_{t>0} \exp\{-(a+\mu)t\}\, m_X(t)\right]^n$$

- note ① $\exp\{-(a+\mu)t\}\, m_X(t) = \exp\{-at\}\, m_{X-\mu}(t)$ $\psi(t)$

$$= \left(1-at+o(t)\right)\left(1+\tfrac{1}{2}\sigma^2t^2+o(t^2)\right)$$

$$\overset{a>0}{=} 1-at+o(t) < 1 \text{ for } t \text{ sufficiently small}$$

$$\therefore 0 < \psi(a) = \inf_{t>0} \exp\{-(a+\mu)t\}\, m_X(t) < 1$$

and $P\left(\tfrac{1}{n}S_n-\mu>a\right)$ goes to 0 like $[\text{constant}<1]^n$; i.e. exponentially fast

- note ② - a similar result for $a<0$

eg Simple random walk (cont'd)

- $m_X(t) = \mathbb{E}[e^{tX}] = pe^{t} + (1-p)e^{-t}$

- put $g(t) = p\exp\{-(a+2p-1-1)t\}$

$$+ (1-p)\exp\{-(a+2p-1+1)t\}$$

and suppose $a+2p-1 < 1$, $a>0$

- then $g'(t) = -p(a+2p-2)\exp\{-(a+2p-2)t\}$

$$-(1-p)(a+2p)\exp\{-(a+2p)t\}$$

- $g'(t) = 0$

iff $p(a+2p-2)\exp\{2t\} = -(1-p)(a+2p)$

iff $t = t^* = \frac{1}{2}\ln\left[-\frac{1-p}{p}\,\frac{(a+2p)}{(a+2p-2)}\right]$

- then $P\left(\frac{1}{n}S_n - \mu > a\right) \sim (g(t_*))^n$

(a) Convergence in Probability

Def A sequence $\{X_n\}$ of r.v.'s <u>converges in probability</u> to a r.v. X, denoted $X_n \xrightarrow{P} X$, if for every $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$

Prop 6 (i) $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$

(ii) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$

Proof: (i) Put $A_{mn} = \{|X_n - X| > \frac{1}{m}\}$. Then $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_{mn}$ = set of ω s.t. $|X_n(\omega) - X(\omega)| > \frac{1}{m}$ for infinitely many n and $X_n \xrightarrow{a.s.} X$ implies that $0 = P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_{mk}\right) \overset{\text{cty of } P}{=} \lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty} A_{mn}\right)$ and $P(A_{mk}) \le P\left(\bigcup_{n=k}^{\infty} A_{mn}\right)$ implies $P(A_{mk}) \to 0$ as $k \to \infty$ which gives the result by choosing $\frac{1}{m} < \varepsilon$ and noting $P(|X_k - X| > \varepsilon) \le P(|X_k - X| > \frac{1}{m})$.

(ii) We have $F_{X_n}(x) = P(X_n \le x, X \le x+\varepsilon) + P(X_n \le x, X > x+\varepsilon)$
$\le F_X(x+\varepsilon) + P(|X_n - X| > \varepsilon)$ and $F_X(x-\varepsilon)$
$= P(X_n \le x, X \le x-\varepsilon) + P(X_n > x, X \le x-\varepsilon)$

$\leq F_{X_n}(x) + P(|X_n - X| > \varepsilon)$. Therefore,

$F_{X_n}(x) - F_X(x) \leq F_X(x+\varepsilon) - F_X(x) + P(|X_n - X| > \varepsilon)$ and

$F_X(x) - F_{X_n}(x) \leq F_X(x) - F_X(x-\varepsilon) + P(|X_n - X| > \varepsilon)$.

So $|F_X(x) - F_{X_n}(x)| \leq \max\{|F_X(x+\varepsilon) - F_X(x)|, |F_X(x) - F_X(x-\varepsilon)|\} + P(|X_n - X| > \varepsilon)$.

The result follows when we take x a cty pt of F_X.

- <u>note</u> - the converses of (i) and (ii) are false

(ii) $X_n \xrightarrow{D} X \not\Rightarrow X_n \xrightarrow{P} X$

<u>eg</u> $Z \sim N(0,1)$, $Y = -Z \sim N(0,1)$

$X_n = Z \quad \forall n$, then $X_n \xrightarrow{D} Y$

but $P(|X_n - Y| > \varepsilon) = P(2|Z| > \varepsilon) \not\to 0$

so $X_n \not\to Y$

<u>Prop</u> 7 $\quad X_n \xrightarrow{P} \mu \Leftrightarrow X_n \xrightarrow{D} \mu$

Proof: $\Rightarrow]$ Prop 5 (ii).

$\Leftarrow]$ $P(|X_n - \mu| \leq \varepsilon) = P(\mu - \varepsilon \leq X_n \leq \mu + \varepsilon)$

$= F_{X_n}(\mu+\varepsilon) - F_{X_n}(\mu - \varepsilon) + P_{X_n}(\{\mu - \varepsilon\})$

$\to 1 + \lim_{n\to\infty} P_{X_n}(\{\mu+\varepsilon\})$ and $\lim_{n\to\infty} P_{X_n}(\{\mu-\varepsilon\})$

$\leq \lim_{n\to\infty} F_{X_n}(\mu - \varepsilon) = 0$

(d) Convergence in Expectation

Def A sequence $\{X_n\}$ of r.v.'s converges in expectation of order r (≥ 1) to the r.v. X if $E|X_n|^r < \infty$ $\forall n$ and $E|X_n - X|^r \to 0$ as $n \to \infty$ and we denote this by $X \xrightarrow{r} X$.

Prop ⑧ (i) $X_n \xrightarrow{r} X$ implies $X_n \xrightarrow{s} X$ for $1 \leq s < r$
(ii) $X_n \xrightarrow{1} X$ implies $X_n \xrightarrow{P} X$

Proof: (i) $E[|X_n - X|^r] = E[|X_n - X|^{\frac{r}{s}s}]$

Jensen $\geq (E[|X_n - X|^s])^{\frac{r}{s}}$ since $x^{r/s}$ is convex.

$(x^p)'' = p(p-1)x^{p-2} \geq 0$ iff $p \geq 1$

(ii) $P(|X_n - X| > \varepsilon) \overset{\text{Markov}}{\leq} E[|X_n - X|] / \varepsilon$,

- note - the converses are false and also $X_n \xrightarrow{r} X$ does not imply $X_n \xrightarrow{a.s.} X$ or conversely

- let $\mathcal{L}_p = \{X \mid X$ a r.v. on $(\Omega, \mathcal{F}, P)$ and $E|X|^p < \infty\}$

- define $\|\cdot\|_p : \mathcal{L}_p \to [0, \infty)$ by $\|X\|_p = \{E|X|^p\}^{1/p}$ for $p \geq 1$

Prop ⑨ (Minkowski Inequality) If $X, Y \in \mathcal{L}_p$ then $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$

- clearly $\|aX\|_p = |a| \|X\|_p$ and thus $\mathcal{L}_p$ is a real vector space and $\|X\|_p$ is a (pseudo) norm on $\mathcal{L}_p$

- when $p=2$ we define $\langle\cdot,\cdot\rangle : \mathcal{L}_2 \times \mathcal{L}_2 \to \mathbb{R}$ by $\langle X, Y\rangle = \mathbb{E}XY$ $\quad (|\mathbb{E}XY| \overset{CS}{\leq} \|X\|_2 \|Y\|_2)$

- $\langle aX_1 + bX_2, Y\rangle = a\langle X_1, Y\rangle + b\langle X_2, Y\rangle$
  $\langle X, Y\rangle = \langle Y, X\rangle$
  $\langle X, X\rangle = \|X\|_2^2 \geq 0$

- thus $\langle\cdot,\cdot\rangle$ is a (pseudo) inner product on $\mathcal{L}_2$.

- hence we can talk about the angle Θ existing between r.v.s X and Y via

$$\cos\Theta = \frac{\langle X, Y\rangle}{\|X\|_2 \|Y\|_2}$$

and $X \perp Y$ iff $\langle X, Y\rangle = \mathbb{E}XY = 0$

eg - let $\mathcal{L}_2^0 \subseteq \mathcal{L}_2$ be the elements of $\mathcal{L}_2$ with 0 mean (a subspace of $\mathcal{L}_2$)

- then $$\cos\Theta = \frac{\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[X])]}{\sqrt{\mathbb{E}[(X-\mathbb{E}[X])^2]}\sqrt{\mathbb{E}[(Y-\mathbb{E}[Y])^2}}$$

$$= \mathrm{CORR}[X, Y]$$

(e) Strong Law of Large Numbers

- there are various "laws of large numbers" depending on what notion of convergence is employed.

- the WLLN uses convergence in probability

Prop ⑩ (L^2 Law of Large Numbers) If $X_1, X_2, \ldots$ is an iid sequence with mean μ and variance σ^2 then $\frac{1}{n} S_n \xrightarrow{2} \mu$.

Proof: $\mathbb{E}\left[\left(\frac{1}{n} S_n - \mu\right)^2\right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu)^2\right]$

$= \frac{\sigma^2}{n} \to 0$ as $n \to \infty$.

Prop ⑪ (Strong Law of Large Numbers) If $X_1, X_2, \ldots$ are iid with mean μ then $\frac{1}{n} S_n \xrightarrow{a.s.} \mu$.

Proof: see text.