Probability and Stochastic Processes I - Lecture 23

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/stac62/STAC622023.html

2023

IV.2 Convergence in Probability

Definition IV.2.1 The sequence X_n of r.v.'s *converges in probability* to r.v. X if

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

for any $\delta > 0$ and we write $X_n \xrightarrow{P} X$. ■

- this is different than $X_n \xrightarrow{wp1} X$ which says

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}) = 0$$

while $X_n \xrightarrow{P} X$ says for any $\delta > 0, \varepsilon > 0$ there exists $N_{\delta,\varepsilon}$ such that for all $n > N_{\delta,\varepsilon}$

$$P(\{\omega : |X_n(\omega) - X(\omega)| > \delta\}) < \varepsilon$$

Proposition IV.2.1 (i) $X_n \overset{wp1}{\to} X$ implies $X_n \overset{P}{\to} X$ and (ii) $X_n \overset{P}{\to} X$ implies $X_n \overset{d}{\to} X$.

Proof: (i) Let $A_{m,n} = \{\omega : |X_n(\omega) - X(\omega)| > 1/m\}$ so

$$\limsup_n A_{m,n} = \{\omega : |X_n(\omega) - X(\omega)| > 1/m \text{ for infinitely many } n\}.$$

By hypothesis

$$\begin{aligned} 0 &= P(\limsup_n A_{m,n}) = P(\cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_{m,n}) \\ &= \lim_{k\to\infty} P(\cup_{n=k}^{\infty} A_{m,n}) \geq \lim_{k\to\infty} P(A_{m,k}) \end{aligned}$$

so $\lim_{k\to\infty} P(A_{m,k}) = 0$ which implies $X_n \overset{P}{\to} X$.

(ii) For $\delta > 0$,

$$\begin{aligned} F_{X_n}(x) &= P(X_n \le x, X \le x+\delta) + P(X_n \le x, X > x+\delta) \\ &\le F_X(x+\delta) + P(|X_n - X| > \delta) \text{ and} \\ F_X(x-\delta) &= P(X_n \le x, X \le x-\delta) + P(X_n > x, X \le x-\delta) \\ &\le F_{X_n}(x) + P(|X_n - X| > \delta). \end{aligned}$$

Therefore

$$\begin{aligned} F_{X_n}(x) - F_X(x) &\le F_X(x+\delta) - F_X(x-\delta) + P(|X_n - X| > \delta) \\ F_X(x) - F_{X_n}(x) &\le F_X(x+\delta) - F_X(x-\delta) + P(|X_n - X| > \delta). \end{aligned}$$

Then, for $\varepsilon > 0$ there exist $N_{\delta,\varepsilon}$ s.t. for all $n > N_{\delta,\varepsilon}$, $P(|X_n - X| > \delta) < \varepsilon/2$ and so

$$|F_X(x) - F_{X_n}(x)| \le |F_X(x+\delta) - F_X(x-\delta)| + \varepsilon/2$$

If x is a cty point of F_X choose δ s.t. $|F_X(x+\delta) - F_X(x-\delta)| \le \varepsilon/2$ and so $|F_X(x) - F_{X_n}(x)| \le \varepsilon$. Since ε is arbitrary this implies the result. ■

note $X_n \xrightarrow{P} X$ does not imply $X_n \xrightarrow{wp1} X$ (example is complicated)

Example IV.2.1 $X_n \xrightarrow{d} X$ does not imply $X_n \xrightarrow{P} X$

- put $X_n = Z \sim N(0,1)$, $X = -Z \sim N(0,1)$ so $X_n \xrightarrow{d} X$ but

$$P(|X_n - X| > \delta) = P(2|Z| > \delta) = 2(1 - \Phi(\delta/2))$$

and so $X_n \overset{P}{\nrightarrow} X$ ■

Proposition IV.2.2 $X_n \xrightarrow{d} \mu$ iff $X_n \xrightarrow{P} \mu$.

Proof: Prop IV.2.1(ii) establishes that if $X_n \xrightarrow{P} \mu$, then $X_n \xrightarrow{d} \mu$. For the other direction,

$$\begin{aligned} P(|X_n - \mu| &\leq \delta) = P(\mu - \delta \leq X_n \leq \mu + \delta) \\ &= (F_{X_n}(\mu + \delta) - F_{X_n}(\mu - \delta)) + P(X_n = \mu - \delta) \\ \text{and } P(X_n &= \mu - \delta) \leq F_{X_n}(\mu - \delta) \to 0 \\ P(|X_n - \mu| &\leq \delta) \to 1 - 0 + 0 = 1 \end{aligned}$$

since $\mu \pm \delta$ are cty pts of limiting dist., which implies $X_n \xrightarrow{P} \mu$. ■

Proposition IV.2.3 (*Slutsky's Theorem*) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then (i) $X_n + Y_n \xrightarrow{d} X + c$ (ii) $X_n Y_n \xrightarrow{d} cX$ (iii) and provided $c \neq 0, X_n / Y_n \xrightarrow{d} X / c$.
Proof: Accept.

Proposition IV.2.4 If $X_n \xrightarrow{d} c$ and $h : (R^1, \mathcal{B}^1) \to (R^1, \mathcal{B}^1)$ is continuous at c, then $h(X_n) \xrightarrow{d} h(c)$.

Proof: Let $\varepsilon > 0$. Then there exists $\delta > 0$ s.t. $|h(x) - h(c)| \leq \varepsilon$ whenever $|x - c| \leq \delta$. Therefore

$$P(|h(X_n) - h(c)| > \varepsilon) \leq P(|X_n - c| > \delta) \to 0. \blacksquare$$

Example IV.2.2

- suppose $X_1, X_2, \ldots$ is an i.i.d. sequence from a distribution with mean μ and variance σ^2 so by CLT

$$\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{d} N(0,1) \text{ and if}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2\right) \text{ then}$$

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{d} \sigma^2 + \mu^2 \text{ by WLLN,}$$

$$\bar{X}^2 \xrightarrow{d} \mu^2 \text{ by Slutsky (ii) and } \frac{n}{n-1} \xrightarrow{wp1} 1, \text{ so}$$

$$S^2 \xrightarrow{d} \sigma^2 \text{ by Slutsky and } S \xrightarrow{d} \sigma \text{ by Prop. IV.2.4}$$

- therefore $T_n = \frac{\sqrt{n}(\bar{X}-\mu)}{S} = \frac{\sigma}{S}\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{d} N(0,1)$ by Slutsky
- when $X_1, X_2, \ldots$ is an i.i.d. $N(\mu, \sigma^2)$ sequence this implies

$$\text{Student}(n) \xrightarrow{d} N(0,1)$$

IV.3 Convergence in Expectation

Definition IV.3.1 The sequence X_n of r.v.'s *converges in expectation of order r* (≥ 1) to r.v. X if $E(|X_n|^r) < \infty$ for every n and

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

and we write $X_n \xrightarrow{r} X$. ■

Proposition IV.3.1 (i) If $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{s} X$ for any $1 \leq s \leq r$. (ii) If $X_n \xrightarrow{1} X$, then $X_n \xrightarrow{P} X$.

Proof: (i) Note that $d^2x^p/dx^2 = p(p-1)x^{p-2} \geq 0$ when $x \geq 0, p \geq 1$ and so $x^{r/s}$ is convex on $[0,\infty)$. Therefore,

$$E(|X_n - X|^r) = E((|X_n - X|^s)^{\frac{r}{s}}) \overset{\text{Jensen}}{\geq} (E(|X_n - X|^s))^{\frac{r}{s}}$$

which implies the result. (ii) For any $\delta > 0$

$$P(|X_n - X| > \delta) \overset{\text{Markov}}{\leq} \frac{E(|X_n - X|)}{\delta} \to 0. \ ■$$

- the converses to Prop. IV.3.1 are false

- the most important case is $r = 2$ and we let

$$L^2(P) = \{X : X \text{ is a r.v. and } E(X^2) < \infty\}$$

- define $< \cdot, \cdot >: L^2(P) \times L^2(P) \to R^1$ by $< X, Y >= E(XY)$ and note

$$(E(XY))^2 \overset{\text{Cauchy-Schwartz}}{\leq} E(X^2)E(Y^2) < \infty$$

and let $||X|| =< X, X >^{1/2}$

Proposition IV.3.2 (i) If $X, Y \in L^2(P)$, then $a + bX + cY \in L^2(P)$ for all constants a, b, c. (ii) $< \cdot, \cdot >$ is an inner product on $L^2(P)$ (iii) $|| \cdot ||$ is a norm on $L^2(P)$.
Proof: **Exercise IV.3.1.**

- this leads to a geometry of r.v.'s and the angle θ between $X - E(X), Y - E(Y) \in L^2(P)$ satisfies

$$\cos\theta = \frac{< X - E(X), Y - E(Y) >}{||X - E(X)||\,||Y - E(Y)||} = \frac{Cov(X, Y)}{Sd(x)\,Sd(Y)} = Corr(X, Y)$$

Proposition IV.3.3 *(L^2 Law of large Numbers)* If X_n is an i.i.d. sequence in $L^2(P)$ then $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{2} E(X_1)$.

Proof:

$$E\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\right)^2\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{Var(X_1)}{n} \to 0. \blacksquare$$

- in time series many s.p.'s are defined in terms of series of r.v.'s that converge in L^2

- **note** $X_n \xrightarrow{2} X$ implies $X_n \xrightarrow{1} X$ implies $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{d} X$