A course in large sample theory	Winter 2022

Modes of convergence with their relationship

Lecturer: Peirong Xu — *Bowen Zhou*

1 Modes of convergence

We begin by studying four modes of convergence of a sequence of random vectors to a limit. For a random vector $\boldsymbol{X} = (X_1, \cdots, X_d) \in \mathbb{R}^d$, the distribution function of $\boldsymbol{X}$ for $\boldsymbol{x} = (x_1, \cdots, x_d)$, is denoted by $F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}\{X_1 \le x_1, \cdots, X_d \le x_d\}$. Let $\|\boldsymbol{x}\| = \left(x_1^2 + \cdots + x_d^2\right)^{1/2}$.

Definition 1. *$\boldsymbol{X}_n$ converges in distribution to $\boldsymbol{X}$, $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$, if $F_{\boldsymbol{X}_n}(\boldsymbol{x}) \to F_{\boldsymbol{X}}(\boldsymbol{x})$, for all points $\boldsymbol{x}$ at which $F_{\boldsymbol{X}}(\boldsymbol{x})$ is continuous.*

Definition 2. *$\boldsymbol{X}_n$ converges in probability to $\boldsymbol{X}$, $\boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}$, if for every $\varepsilon > 0$, $\mathbb{P}\{\|\boldsymbol{X}_n - \boldsymbol{X}\| > \varepsilon\} \to 0$.*

Definition 3. *For a real number $r > 0$, $\boldsymbol{X}_n$ converges in the r-th mean to $\boldsymbol{X}$, $\boldsymbol{X}_n \xrightarrow{r} \boldsymbol{X}$, if $\mathbb{E}\|\boldsymbol{X}_n - \boldsymbol{X}\|^r \to 0$.*

Definition 4. *$\boldsymbol{X}_n$ converges almost surely to $\boldsymbol{X}$, $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$, if $\mathbb{P}\{\lim_{n\to\infty} \boldsymbol{X}_n = \boldsymbol{X}\} = 1$.*

Remark. When $r = 2$, convergence in the 2-th mean is also called convergence in quadratic mean, and is written $\boldsymbol{X}_n \xrightarrow{qm} \boldsymbol{X}$.

The basic relationships of the above four modes of convergence are as follows.

Theorem 5. *We have*

(a) $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}$;

(b) $\boldsymbol{X}_n \xrightarrow{r} \boldsymbol{X}$ *for some* $r > 0 \Rightarrow \boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}$;

(c) $\boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$.

Proof.

(a) If $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$,

$$\mathbb{P}\left\{\bigcup_{n=1}^{+\infty}\bigcap_{k=n}^{+\infty}\|\boldsymbol{X}_k - \boldsymbol{X}\| \le \varepsilon\right\} = 1 \quad \text{or} \quad \mathbb{P}\left\{\bigcap_{n=1}^{+\infty}\bigcup_{k=n}^{+\infty}\|\mathrm{X}_k - \mathrm{X}\| > \varepsilon\right\} = 0$$

for all $\varepsilon > 0$. Hence

$$\lim_{n\to\infty}\mathbb{P}\{\|\boldsymbol{X}_n - \boldsymbol{X}\| > \varepsilon\} \le \lim_{n\to\infty}\mathbb{P}\left\{\bigcup_{k=n}^{+\infty}\|\boldsymbol{X}_k - \boldsymbol{X}\| > \varepsilon\right\} = \mathbb{P}\left\{\bigcap_{n=1}^{+\infty}\bigcup_{k=n}^{+\infty}\|\boldsymbol{X}_k - \boldsymbol{X}\| > \varepsilon\right\} = 0.$$

(b) By Markov inequality, $\mathbb{P}\{\|\boldsymbol{X}_n - \boldsymbol{X}\| > \varepsilon\} \le \mathbb{E}\|\boldsymbol{X}_n - \boldsymbol{X}\|^r/\varepsilon^r \to 0$.

(c) Let $\varepsilon > 0$, and $\mathbf{1}_d$ represent the vector with 1 in every component. Then

$$\begin{aligned} F_{\boldsymbol{X}_n}(\boldsymbol{x}) &= \mathbb{P}\{\boldsymbol{X}_n \le \boldsymbol{x}\} \\ &= \mathbb{P}\{\boldsymbol{X}_n \le \boldsymbol{x}, \boldsymbol{X} \le \boldsymbol{x} + \varepsilon \mathbf{1}_d\} + \mathbb{P}\{\boldsymbol{X}_n \le \boldsymbol{x}, \boldsymbol{X} > \boldsymbol{x} + \varepsilon \mathbf{1}_d\} \\ &\le \mathbb{P}\{\boldsymbol{X} \le \boldsymbol{x} + \varepsilon \mathbf{1}_d\} + \mathbb{P}\{\|\boldsymbol{X}_n - \boldsymbol{X}\| > \varepsilon\} \\ &= F_{\boldsymbol{X}}(\boldsymbol{x} + \varepsilon \mathbf{1}_d) + \mathbb{P}\{\|\boldsymbol{X}_n - \boldsymbol{X}\| > \varepsilon\}, \end{aligned}$$

as $n \to \infty$, we have $\limsup F_{\boldsymbol{X}_n}(\boldsymbol{x}) \le F_{\boldsymbol{X}}(\boldsymbol{x} + \varepsilon \mathbf{1}_d)$. Similarly,

$$\begin{aligned} &\mathbb{P}\{\boldsymbol{X} \le \boldsymbol{x} - \varepsilon \mathbf{1}_d\} \\ =&\mathbb{P}\{\boldsymbol{X} \le \boldsymbol{x} - \varepsilon \mathbf{1}_d, \boldsymbol{X}_n \le \boldsymbol{x}\} + \mathbb{P}\{\boldsymbol{X} \le \boldsymbol{x} - \varepsilon \mathbf{1}_d, \boldsymbol{X}_n > \boldsymbol{x}\} \\ \le&\mathbb{P}\{\boldsymbol{X}_n \le \boldsymbol{x}\} + \mathbb{P}\{\|\boldsymbol{X}_n - \boldsymbol{X}\| > \varepsilon\} \\ =&F_{\boldsymbol{X}_n}(\boldsymbol{x}) + \mathbb{P}\{\|\boldsymbol{X}_n - \boldsymbol{X}\| > \varepsilon\}, \end{aligned}$$

hence as $n \to \infty$, $\liminf F_{\boldsymbol{X}_n}(\boldsymbol{x}) \ge F_{\boldsymbol{X}}(\boldsymbol{x} - \varepsilon \mathbf{1}_d)$. Consequently,

$$F_{\boldsymbol{X}}(\boldsymbol{x} - \varepsilon \mathbf{1}_d) \le \liminf F_{\boldsymbol{X}_n}(\boldsymbol{x}) \le \limsup F_{\boldsymbol{X}_n}(\boldsymbol{x}) \le F_{\boldsymbol{X}}(\boldsymbol{x} + \varepsilon \mathbf{1}_d).$$

If F_X is continuous at $\boldsymbol{x}$, then $\varepsilon \to 0$ imply that $F_{\boldsymbol{X}_n}(\boldsymbol{x}) \to F_{\boldsymbol{X}}(\boldsymbol{x})$.

$\square$

Remark. The converses are not hold for all $(a) - (c)$ in Theorem 5.

(a) Let $X_1, \cdots, X_n, \cdots, X \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$, and denote $\Phi(x)$ as the distribution function of $\mathsf{N}(0,1)$. Then obviously $X_n \xrightarrow{d} X$, but

$$\mathbb{P}\{|X_n - X| > \varepsilon\} = \mathbb{P}\{X_n - X > \varepsilon\} + \mathbb{P}\{X_n - X < -\varepsilon\} = 2 - 2\Phi(\varepsilon/\sqrt{2})$$

does not converge to zero for small ε and hence $X_n \overset{\mathbb{P}}{\nrightarrow} X$.

(b) Let $Z \sim \text{Unif}(0,1)$, and construct a sequence of random variables as follow: $X_1 = 1, X_2 = I_{[0,1/2)}(Z), X_3 = I_{[1/2,1)}(Z), X_4 = I_{[0,1/4)}(Z), X_5 = I_{[1/4,1/2)}(Z), \cdots, X = 0$. Obviously $X_n \overset{a.s.}{\nrightarrow} X$. For $n = 2^k + m, 0 \le m < 2^k$, $\mathbb{E}|X_n|^r = 1/k \to 0$ and $\mathbb{P}\{|X_n| > \varepsilon\} \le 1/2^k$ as $n \to 0$, so $X_n \xrightarrow{r} 0$ and $X_n \xrightarrow{\mathbb{P}} 0$.

(c) Let $X_n = 2^n I_{[0,1/n)}(Z)$ where $Z \sim \text{Unif}(0,1)$. Then $\mathbb{E}|X_n|^r = 2^{nr}/n \to \infty$ for any $r > 0$. Yet $X_n \xrightarrow{a.s.} 0$ and so $X_n \xrightarrow{\mathbb{P}} 0$.

The following are some useful tools of probability.

Lemma 6 (*Fatou*)**.** *$X_n, n = 1, 2, \cdots$ are positive random variables, then $\mathbb{E} \liminf X_n \le \liminf \mathbb{E} X_n$.*

Theorem 7 (*Monotone convergence*)**.** *If $0 \le X_1 \le X_2 \le \cdots$, and $X_n \xrightarrow{a.s.} X$, then $\mathbb{E}X_n \to \mathbb{E}X$.*

Theorem 8 (*Dominated convergence*)**.** *If $X_n \xrightarrow{a.s.} X$ and $|X_n| \le Y$ for some random variable Y with $\mathbb{E}Y < \infty$, then $\mathbb{E}X_n \to \mathbb{E}X$.*

Proposition 9 (*Hölder's inequality*)**.** *For nonnegative random variables X and Y with finite means, $\mathbb{E}XY \le (\mathbb{E}X^p)^{1/p}(\mathbb{E}Y^q)^{1/q}$ for all $p, q > 0$ with $1/p + 1/q = 1$.*

Lemma 10 (*Borel Cantelli*)**.** *$\{A_n\}$ is a sequence of events. If $\sum_{n=1}^{\infty} \mathbb{P}\{A_n\} < \infty$, then $\mathbb{P}\{A_n \ i.o.\} = 0$. Conversely, if $\{A_n\}$ are independent and $\sum_{n=1}^{\infty} \mathbb{P}\{A_n\} = \infty$, then $\mathbb{P}\{A_n \ i.o.\} = 1$.*

Remark. $\mathbb{P}\{A_n \ i.o.\} = \mathbb{P}\{A_n \text{ occur infinitely often}\} = \mathbb{P}\{\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k\}$.

2 Partial converses

As we will see, under certain additional conditions, some important partial converses of theorem 5 holds.

Theorem 11. *$\boldsymbol{X}_n, \boldsymbol{X}$ are random vectors.*

(a) If $\boldsymbol{c} \in \mathbb{R}^d$, then $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{c} \Leftrightarrow \boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{c}$.

(b) If $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$ and $\|\boldsymbol{X}_n\|^r \leq Z$ for all n, some $r > 0$ and some random variable Z with $\mathbb{E}Z < \infty$, then $\boldsymbol{X}_n \xrightarrow{r} \boldsymbol{X}$.

(c) If $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}, \boldsymbol{X}_n \geq 0$ and $\mathbb{E}\boldsymbol{X}_n \to \mathbb{E}\boldsymbol{X} < \infty$, then $\mathbb{E}\|\boldsymbol{X}_n - \boldsymbol{X}\| \to 0$.

(d) $\boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}$ if and only if every subsequence of $\{\boldsymbol{X}_n\}$ has a subsequence almost surely converge to $\boldsymbol{X}$.

Proof.

(a) With part (c) of Theorem 5, we need to show $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{c} \Rightarrow \boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{c}$. First, in one dimension, $\mathbb{P}\{|X_n - c| \leq \varepsilon\} = \mathbb{P}\{X_n \leq c + \varepsilon\} - \mathbb{P}\{X_n \leq c - \varepsilon\} \to 1$ for every $\varepsilon > 0$.

(b) Note that $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$ and $\|\boldsymbol{X}_n\|^r \leq Z$ implies $\|\boldsymbol{X}\|^r \leq Z$ and $\|\boldsymbol{X}_n - \boldsymbol{X}\| \to 0$ a.s., so $\|\boldsymbol{X}_n - \boldsymbol{X}\|^r \leq (\|\boldsymbol{X}_n\| + \|\boldsymbol{X}\|)^r \leq 2^r Z$. Now apply the dominated convergence theorem to $\|\boldsymbol{X}_n - \boldsymbol{X}\|^r$.

(c) In one dimension, note that $\mathbb{E}|X_n - X| = 2\mathbb{E}(X - X_n)_+ - \mathbb{E}(X - X_n)$. The second term converges to zero because $\mathbb{E}X_n \to \mathbb{E}X$. $(X - X_n)_+ \leq X_+$ and $\mathbb{E}X_+ < \infty$ implies that the the first term converges to zero by dominated convergence theorem.

(d) Hint: using Lemma 10.

□

Remark. Part (b) and (c) of Theorem 11 gives a method of deducing convergence n the r-th mean from almost sure convergence. Additionally, it can be strengthened by replacing $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$ with $\boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}$.

Remark. The following statement gives a simple sufficient condition for almost sure convergence.

(a) If $\sum \mathbb{E}\|\boldsymbol{X}_n - \boldsymbol{X}\|^2 < \infty$, then $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$ and $\boldsymbol{X}_n \xrightarrow{qm} \boldsymbol{X}$.

(b) If $\sum \mathbb{E}\|\boldsymbol{X}_n - \boldsymbol{X}\|^r < \infty$, then $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$ and $\boldsymbol{X}_n \xrightarrow{r} \boldsymbol{X}$.

Remark. Part (c) of Theorem 11 is usually stated in terms of densities, and can be strengthened in one dimension as follows:

(a) If $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$ and $\mathbb{E}\|\boldsymbol{X}_n\| \to \mathbb{E}\|\boldsymbol{X}\| < \infty$, then $\mathbb{E}\|\boldsymbol{X}_n - \boldsymbol{X}\| \to 0$.

(b) if $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$ and $\mathbb{E}\|\boldsymbol{X}_n\|^2 \to \mathbb{E}\|\boldsymbol{X}\|^2 < \infty$, then $\mathbb{E}\|\boldsymbol{X}_n - \boldsymbol{X}\|^2 \to 0$.