4 Convergence of Random Variables

In topology, we know that the notion of convergence can differ depending on the topology that the space is equipped with. Some notions of convergence can be compared but some other do not. In this chapter, we will discuss different notions of convergence of random variables and see which notions are stronger or weaker than the others.

4.1 Convergence in Probability

Given a sequence of random variables $(X_n)_{n\geqslant 1}$ with values in $\mathbb{R}^d$ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In Measure Thoery, we have already discussed the following two notions of convergence,

- Almost sure convergence: if $\mathbb{P}(\{\omega \in \Omega \mid X(\omega) = \lim_{n\to\infty} X_n(\omega)\}) = 1$, then we write $X_n \xrightarrow{\text{a.s.}} X$.
- For $p \in [1, \infty)$, convergence in L^p: if $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$, then we write $X_n \xrightarrow{L^p} X$.

Definition 4.1.1 : If the following holds for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0,$$

then we say that $(X_n)_{n\geqslant 1}$ *converges in probability* (機率收斂) to X, denoted,

$$X_n \xrightarrow{(\mathbb{P})} X.$$

Then, we show that the notion of convergence in probability is metrizable.

Let $\mathcal{L}^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$ be the space consisting of random variables with values in $\mathbb{R}^d$. We define the following equivalence relation on this space: $X \sim Y$ if and only if X and Y are equal almost surely. We write $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$ for its quotient space, on which we can define the following distance (距離),

$$d(X, Y) = \mathbb{E}\left[|X - Y| \wedge 1\right]. \tag{4.1}$$

Proposition 4.1.2 : *The definition in Eq. (4.1) is indeed a distance characterizing the convergence in probability, i.e., $(X_n)_{n\geqslant 1}$ converges in probability to X if and only if $d(X_n, X)$ tends to 0. Moreover, the metric space $(L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P}), d)$ is complete.*

Proof : It is not hard to show that Eq. (4.1) defines a distance. One only needs to use the definition of the quotient space and the definition of a distance. Next, we show that this distance is equivalent to the convergence in probability. First, we assume that (X_n) converges in probability to X. Then, for any given $\varepsilon \in (0,1)$, we have,

$$\begin{aligned}\mathbb{E}\left[|X_n - X| \wedge 1\right] &\leqslant \mathbb{E}\left[|X_n - X|\mathbb{1}_{|X_n - X|\leqslant\varepsilon}\right] + \mathbb{E}\left[(|X_n - X| \wedge 1)\mathbb{1}_{|X_n - X|>\varepsilon}\right] \\ &\leqslant \varepsilon + \mathbb{P}\left(|X_n - X| > \varepsilon\right).\end{aligned}$$

This implies that $\limsup d(X_n, X) \leqslant \varepsilon$. Since $\varepsilon > 0$ can be arbitrarily small, we obtain $\lim d(X_n, X) = 0$. Conversely, if $\lim d(X_n, X) = 0$, then for any $\varepsilon \in (0,1)$, we have,

$$\mathbb{P}\left(|X_n - X| > \varepsilon\right) \leqslant \varepsilon^{-1}\,\mathbb{E}\left[|X_n - X| \wedge 1\right] = \varepsilon^{-1} d(X_n, X) \longrightarrow 0.$$

Now, we show the completeness of the metric space $(L^0_{\mathbb{R}^d}, d)$. Given a Cauchy sequence (X_n) for the distance d, we can extract a subsequence $Y_k = X_{n_k}$ such that for all $k \geqslant 1$,

$$d(Y_k, Y_{k+1}) \leqslant 2^{-k}.$$

Then, we have,

$$\mathbb{E}\left[\sum_{k=0}^{\infty}(|Y_{k+1} - Y_k| \wedge 1)\right] = \sum_{k=0}^{\infty} d(Y_k, Y_{k+1}) < \infty.$$

Hence, $\sum(|Y_{k+1} - Y_k| \wedge 1) < \infty$, a.s., meaning that $\sum |Y_{k+1} - Y_k| < \infty$, a.s. and the following definition makes sense,

$$X = Y_0 + \sum_{k=0}^{\infty}(Y_{k+1} - Y_k).$$

From the above construction, we know that (Y_k) converges a.s. to X and using the dominated convergence theorem, we reach at,

$$d(Y_k, X) = \mathbb{E}\left[|Y_k - X| \wedge 1\right] \longrightarrow 0.$$

So (Y_k) converges in probability to X. Finally, since (X_k) is a Cauchy sequence, (X_k) also converges in probability to X. $\square$

Proposition 4.1.3 : *Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables. Then the following properties are true.*

(1) If (X_n) converges a.s. or in L^p for $p \geqslant 1$ to X, then it also converges in probability to X.

(2) If (X_n) converges in probability to X, then there exists a subsequence (X_{n_k}) converging a.s. to X.

Proof : We have already proven (2) in Proposition 4.1.2, so we focus on the proof of (1). If (X_n) converges almost surely to X, then from the dominated convergence theorem, we have,

$$d(X_n, X) = \mathbb{E}\left[|X_n - X| \wedge 1\right] \longrightarrow 0.$$

If (X_n) converges in L^p to X, then

$$d(X_n, X) \leqslant \|X_n - X\|_1 \leqslant \|X_n - X\|_p \longrightarrow 0.$$

$\square$

Proposition 4.1.4 : *Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables converging to X in probability. Suppose that there exists $r \in (1,\infty)$ such that $(X_n)_{n\geqslant 1}$ is bounded in L^r. Then, for all $p \in [1,r)$, the sequence $(X_n)_{n\geqslant 1}$ converges to X in L^p.*

Proof : From the assumption, consider $C > 0$ such that for all $n \geqslant 1$, $\mathbb{E}\left[|X_n|^r\right] \leqslant C$. From (2) of Proposition 4.1.3, we may find a subsequence $(n_k)_{k\geqslant 1}$ such that $X_{n_k} \xrightarrow{\text{a.s.}} X$. Then, the Fatou's lemma implies

$$\mathbb{E}\left[|X|^r\right] = \mathbb{E}\left[\lim_{k\to\infty} |X_{n_k}|^r\right] \leqslant \liminf_{k\to\infty} \mathbb{E}\left[|X_{n_k}|^r\right] \leqslant C.$$

For all $p \in [1,r)$ and $\varepsilon > 0$, we have,

$$\begin{aligned}\mathbb{E}\left[|X_n - X|^p\right] &= \mathbb{E}\left[|X_n - X|^p \mathbb{1}_{|X_n - X|\leqslant\varepsilon}\right] + \mathbb{E}\left[|X_n - X|^p \mathbb{1}_{|X_n - X|>\varepsilon}\right] \\ &\leqslant \varepsilon^p + \mathbb{E}\left[|X_n - X|^r\right]^{p/r} \mathbb{P}(|X_n - X| > \varepsilon)^{1-p/r} \\ &\leqslant \varepsilon^p + 2^p C^{p/r}\, \mathbb{P}\left(|X_n - X| > \varepsilon\right)^{1-p/r}.\end{aligned}$$

where in the second line we apply the Hölder's inequality and in the third line we apply the Minkowski's inequality,

$$\mathbb{E}\left[|X_n - X|^r\right]^{p/r} = \|X_n - X\|_r^p \leqslant (\|X_n\|_r + \|X\|_r)^p \leqslant (2C^{1/r})^p = 2^p C^{p/r}.$$

Since (X_n) converges in probability to X, we find,

$$\limsup_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] \leqslant \varepsilon^p.$$

Since ε can be arbitrarily small, (X_n) converges to X in L^p. □

4.2 All-or-none Law

The *all-or-none law* is also known as *Kolmogorov's 0-1 law* (Kolmogorov 零一律) , which states that the probability that a *tail event* (尾端事件) occurs is either 0 or 1.

Definition 4.2.1 : Let $(X_n)_{n\geqslant 1}$ be a sequence of independent random variables with values in any metric space. For all $n \geqslant 1$, define $\mathcal{B}_n$ to be the following σ-algebra,

$$\mathcal{B}_k = \sigma(X_k : k \geqslant n).$$

We also define the *asymptotic σ-algebra* (漸進 σ 代數) $\mathcal{B}_\infty$ to be,

$$\mathcal{B}_\infty = \bigcap_{n=1}^{\infty} \mathcal{B}_n.$$

A measurable event in $\mathcal{B}_\infty$ is called a *tail event* (尾端事件) .

Theorem 4.2.2 : *Using the above notations, $\mathcal{B}_\infty$ is a trivial σ-algebra, i.e., for all $B \in \mathcal{B}_\infty$, we have $\mathbb{P}(B) = 0$ or 1.*

Proof : For all $n \geqslant 1$, let

$$\mathcal{D}_n = \sigma(X_k : k \leqslant n).$$

From Question 3.1.19, we know that for all $n \geqslant 1$, the σ-algebras $\mathcal{D}_n$ and $\mathcal{B}_{n+1}$ are independent, so $\mathcal{D}_n$ is also independent of $\mathcal{B}_\infty$. This implies,

$$\forall A \in \bigcup_{n=1}^{\infty} \mathcal{D}_n, \qquad \forall B \in \mathcal{B}_\infty, \qquad \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Since $\cup \mathcal{D}_n$ is closed under finite intersections, from Proposition 3.1.18, we know that $\mathcal{B}_\infty$ is independent of the following σ-algebra,

$$\sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{D}_n\Big) = \sigma(X_n : n \geqslant 1).$$

Since $\mathcal{B}_\infty$ is also included in the above σ-algebra, so $\mathcal{B}_\infty$ is independent of itself, which means that for all $B \in \mathcal{B}_\infty$, we have,

$$\mathbb{P}(B) = \mathbb{P}(B \cap B) = \mathbb{P}(B)^2,$$

implying $\mathbb{P}(B) = 0$ or 1. ☐

Question 4.2.3: Given an i.i.d. sequence of random variables $(X_n)_{n\geqslant 1}$ and define σ-algebras $\mathcal{B}_k$ and $\mathcal{B}_\infty$ as in Definition 4.2.1. Show the following properties.

(1) If Y is a real random variable that is $\mathcal{B}_\infty$-measurable, then it is almost surely a constant.

(2) If $\frac{1}{n}(X_1 + \ldots X_n)$ converges almost surely, deduce from the previous question that its limit must almost surely be a constant.

The following proposition is an important application of the 0-1 law.

Proposition 4.2.4 : *Let $(X_n)_{n\geqslant 1}$ be an i.i.d. sequence of random variables with distribution $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$. For all $n \geqslant 1$, let $S_n = X_1 + \ldots, +X_n$. Then,*

$$\text{a.s.,} \qquad \sup_n S_n = +\infty \qquad \text{and} \qquad \inf_n S_n = -\infty.$$

In particular, it means that almost surely, there exists an infinity of n such that $S_n = 0$.

Remark 4.2.5 : We toss a fair coin. Assume that we earn one dollar when we get a head and loose one dollar when we get a tail, then during the whole game, the net asset can be as positive as possible and also as negative as possible.

Proof : For all $p \geqslant 1$, we define the following event,

$$A_p = \{-p \leqslant \inf_{n\geqslant 1} S_n \leqslant \sup_{n\geqslant 1} S_n \leqslant p\}.$$

We note that the limit of (A_p) is,

$$A_\infty := \lim_{p\to\infty} \uparrow A_p = \{-\infty < \inf_n S_n \leqslant \sup S_n < \infty\}.$$

First, we want to prove that for all $p \geqslant 1$, $\mathbb{P}(A_p) = 0$. For all $k > 2p$ and $j \geqslant 0$, let

$$B_{j,k} = \{X_{jk+1} = \cdots = X_{jk+k} = 1\}.$$

Then, we have,

$$\bigcup_{j=0}^{\infty} B_{j,k} \subseteq A_p^c. \tag{4.2}$$

Since $(B_{j,k})_{j\geqslant 0}$ is an independent sequence of events with $\sum \mathbb{P}(B_{j,k}) = \infty$, from Borel–Cantelli lemma, the event on the left side of Eq. (4.2) has probability 1, meaning that $\mathbb{P}(A_p) = 0$. By the continuity of probability measures, we obtain $\mathbb{P}(A) = \lim \uparrow \mathbb{P}(A_p) = 0$, which means,

$$\mathbb{P}(\{\inf_n S_n = -\infty\} \cup \{\sup_n S_n = \infty\}) = 1.$$

By symmetry, we have,

$$\mathbb{P}(\inf_n S_n = -\infty) = \mathbb{P}(\sup_n S_n = \infty),$$

So both probabilities are non-zero.

Finally, we want to use the 0-1 law (Theorem 4.2.2) to show that the probabilities in the above formula are both 1. We first note that, for all $k \geqslant 1$,

$$\{\sup_n S_n = \infty\} = \{\sup_{n\geqslant k}(X_k + \cdots + X_n) = \infty\} \in \mathcal{B}_k,$$

meaning that the event $\{\sup_n S_n = \infty\}$ is $\mathcal{B}_k$-measurable. So it is also measurable with respect to the intersection of all the $\mathcal{B}_k$'s, which is $\mathcal{B}_\infty$. □

4.3 Strong Law of Large Numbers

The goal of this section is to show that if an i.i.d. sequence of random variables (X_n) is in L^1, then its arithmetic average $\frac{1}{n}(X_1 + \ldots + X_n)$ converges almost surely to $\mathbb{E}[X_1]$.

In Proposition 3.3.6, we proved that under a stronger assumption that $\mathbb{E}[|X_1|^4] < \infty$, the statement holds. But here we want to look for the minimal condition so that the theorem holds.

Theorem 4.3.1 : *Let $(X_n)_{n\geqslant 1}$ be an i.i.d. sequence of random variables with distribution in L^1. Then,*

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{a.s.} \mathbb{E}[X_1].$$

Remark 4.3.2 : The integrability is the optimal assumption since in the above statement, the limit needs to exist. If the random variables are non-negative and $\mathbb{E}[X_1] = \infty$, we may apply this theorem to $(X_n \wedge k)_{n \geqslant 1}$ to get

$$\Big[\frac{1}{n}(X_1 + \cdots + X_n)\Big] \wedge k = \frac{1}{n}(X_1 \wedge k + \cdots + X_n \wedge k) \xrightarrow{\text{a.s.}} \mathbb{E}[X_1 \wedge k],$$

then by using the monotone convergence theorem while taking the limit $k \longrightarrow \infty$ to derive,

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{\text{a.s.}} +\infty.$$

Remark 4.3.3 : Later in Section 6.6, we will also show that this convergence also holds in L^1.

Proof : Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n \geqslant 1$. Set $a > \mathbb{E}[X_1]$ and the random variable,

$$M = \sup_{n \geqslant 0}(S_n - na) \in [0, \infty].$$

We will show this theorem by showing (i) and the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) :

(i) $\mathbb{P}(M < \infty) > 0$, or equivalently, $\mathbb{P}(M = \infty) < 1$;

(ii) $M < \infty$, a.s.;

(iii) the strong law of large numbers.

We explain first why (ii) implies (iii) the strong law of large numbers. From the definition of M, for all n, we have $S_n \leqslant na + M$ and from the property in (ii), we obtain,

$$\limsup_{n \to \infty} \frac{1}{n} S_n \leqslant a, \qquad \text{a.s.}$$

Since a can be any number larger than $\mathbb{E}[X_1]$ and arbitrarily close to $\mathbb{E}[X_1]$, we have,

$$\limsup_{n \to \infty} \frac{1}{n} S_n \leqslant \mathbb{E}[X_1], \qquad \text{a.s.}$$

If we replace X_n with $-X_n$, then we get,

$$\liminf_{n \to \infty} \frac{1}{n} S_n \geqslant \mathbb{E}[X_1], \qquad \text{a.s.}$$

The two formulas above together imply the strong law of large numbers.

Then we show that (i) implies (ii). We first note that for any $k \geqslant 0$, the event $\{M < \infty\}$ can be rewritten as,

$$\begin{aligned}\{M < \infty\} &= \{\sup_{n \geqslant 0}(S_n - na) < \infty\} \\ &= \{\sup_{n \geqslant k}(S_n - S_k - (n-k)a) < \infty\} \in \sigma(X_{k+1}, X_{k+2}, \dots).\end{aligned}$$

Hence, we know that the event $\{M < \infty\}$ is measurable with respect to the asymptotic σ-algebra $\mathcal{B}_\infty$, which means that $\mathbb{P}(M < \infty) = 0$ or 1. So, (i) implies (ii).

Finally, let us show (i). For all $k \geqslant 0$, define the following random variable,

$$M_k = \sup_{0 \leqslant n \leqslant k} (S_n - na),$$
$$M'_k = \sup_{0 \leqslant n \leqslant k} (S_{n+1} - S_1 - na).$$

Since the vectors $(X_1, \ldots, X_k)$ and $(X_2, \ldots, X_{k+1})$ have the same distribution and we have $M_k = F_k(X_1, \ldots, X_k)$ and $M'_k = F_k(X_2, \ldots, X_{k+1})$ for some deterministic (確定性) function $F_k : \mathbb{R}^k \longrightarrow \mathbb{R}$, we deduce that the following two random variables also have the same distribution,

$$M = \lim_{k \to \infty} \uparrow M_k \quad \text{and} \quad M' = \lim_{k \to \infty} \uparrow M'_k.$$

Moreover, from the definition, we know that for all $k \geqslant 1$,

$$M_{k+1} = \sup\Big(0, \sup_{1 \leqslant n \leqslant k+1} (S_n - na)\Big) = \sup(0, M'_k + X_1 - a) = M'_k - \inf(a - X_1, M'_k).$$

M_k and M'_k being both in L^1 and have the same distribution, we get,

$$\mathbb{E}[\inf(a - X_1, M'_k)] = \mathbb{E}[M'_k] - \mathbb{E}[M_{k+1}] = \mathbb{E}[M_k] - \mathbb{E}[M_{k+1}] \leqslant 0$$

Next, since $M'_k \geqslant 0$, we have $|\inf(a - X_1, M'_k)| \leqslant |a - X_1|$, then it follows from the dominated convergence theorem that

$$\mathbb{E}[\inf(a - X_1, M')] = \lim_{k \to \infty} \mathbb{E}[\inf(a - X_1, M'_k)] \leqslant 0.$$

Finally, if $\mathbb{P}(M = \infty) = 1$, then we also have $\mathbb{P}(M' = \infty) = 1$, meaning that $\inf(a - X_1, M') = a - X_1$ a.s. But we chose $a > \mathbb{E}[X_1]$, which contradicts the fact that $\mathbb{E}[a - X_1] \leqslant 0$. □

4.4 Convergence in Distribution

4.4.1 Definition and Examples

The functional space $C_b(\mathbb{R}^d)$ consists of continuous bounded functions from $\mathbb{R}^d$ to $\mathbb{R}$. We can define a norm on this space to be the supremum of the function.

$$\|\varphi\|_\infty = \sup_{x \in \mathbb{R}^d} |\varphi(x)|.$$

Definition 4.4.1 : Given a sequence of probability measures $(\mu_n)_{n \geqslant 1}$ on $\mathbb{R}^d$. We say that $(\mu_n)_{n \geqslant 1}$ *converges weakly* (弱收斂) to μ, denoted,

$$\mu_n \Longrightarrow \mu,$$

if the following condition holds,

$$\forall \varphi \in C_b(\mathbb{R}^d), \qquad \int \varphi \, \mathrm{d}\mu_n \longrightarrow \int \varphi \, \mathrm{d}\mu. \tag{4.3}$$

Given a sequence of random variables $(X_n)_{n\geqslant 1}$ with values in $\mathbb{R}^d$. If the sequence of distributions $(\mathbb{P}_{X_n})_{n\geqslant 1}$ converges weakly to $\mathbb{P}_X$, then we say that $(X_n)_{n\geqslant 1}$ *converges weakly* (弱收斂) or *converges in distribution* (分佈收斂) to X, denoted,

$$X_n \xrightarrow{\mathcal{L}} X \qquad \text{or} \qquad X_n \xrightarrow{(d)} X.$$

The weak convergence of random variables is also equivalent to the following condition,

$$\forall \varphi \in C_b(\mathbb{R}^d), \qquad \mathbb{E}[\varphi(X_n)] \longrightarrow \mathbb{E}[\varphi(X)].$$

Remark 4.4.2 : We add some comments to the above definition.

(1) The space of measures on $\mathbb{R}^d$ can be seen as a subspace of the dual space $C_b(\mathbb{R}^d)^*$. As such, Eq. (4.3) can be understood as the convergence in the weak-* topology. However, in Probability Theory, it is called "weak convergence".[1]

(2) In the definition, there is an abuse of notations: when we say that the sequence of random variables $(X_n)_{n\geqslant 1}$ converges weakly to X, there is no uniqueness for X; the only mathematical object with uniqueness is the distribution $\mathbb{P}_X$. Therefore, in a more precise language, we say that the sequence of random variables $(X_n)_{n\geqslant 1}$ converges in distribution to $\mathbb{P}_X$.

(3) When we talk about other notions of convergence of random variables, we need them to live on the same probability space but not for the convergence in distribution. The random variables can *be defined on different spaces* and that is why this notion is important in Probability Theory.

Example 4.4.3 : Below we give two examples of the convergence in distribution.

(1) If X_n has the uniform distribution on $\{\frac{k}{2^n} : 1 \leqslant k \leqslant 2^n\}$, then X_n converges in distribution to the uniform distribution on $[0,1]$ because the Riemann summation of a continuous function approximates its Riemann integral.

(2) If X_n has the Gaussian distribution $\mathcal{N}(0,\sigma_n^2)$ with $\sigma_n^2 \longrightarrow 0$, then X_n converges in distribution to a random variable which is almost surely zero.

Question 4.4.4: If $\mu_n \Longrightarrow \mu$, look for a continuous but unbounded function and a discontinuous bounded function such that Eq. (4.3) does not hold.

Question 4.4.5: Let $(X_n)_{n\geqslant 1}$ and X be random variables with values in $\mathbb{Z}^d$. Prove that X_n converges in distribution to X if and only if,

$$\forall x \in \mathbb{Z}^d, \qquad \mathbb{P}(X_n = x) \longrightarrow \mathbb{P}(X = x).$$

Question 4.4.6: Assume that for all $n \geqslant 1$, the random variable X_n has a density, denoted $\mathbb{P}_{X_n}(\mathrm{d}x) = p_n(x)\,\mathrm{d}x$. Suppose

(1) $p_n(x) \longrightarrow p(x), \qquad \mathrm{d}x\text{-a.s.}$,

[1] In French mathematics literature, there is another term to distinguish this notion, called "convergence étroite".

(2) there exists a non-negative function q such that $\int_{\mathbb{R}^d} q(x)\,\mathrm{d}x < \infty$ and

$$\forall n, \qquad p_n(x) \leqslant q(x), \qquad \mathrm{d}x\text{-a.s.}.$$

Prove that p is the density function of a probability measure on $\mathbb{R}^d$ and that X_n converges in distribution to $p(x)\,\mathrm{d}x$.

Question 4.4.7: If $(X_n)_{n\geqslant 1}$ converges in distribution to X, is it true that for any $B \in \mathcal{B}(\mathbb{R}^d)$, we also have the following convergence?

$$\mathbb{P}(X_n \in B) \longrightarrow \mathbb{P}(X \in B).$$

4.4.2 Equivalent Conditions for Convergence in Distribution

The following theorem gives some properties which are equivalent to the weak convergence of probability measures.

Theorem 4.4.8 (The Portmanteau Theorem [2]) **:** *Let $(\mu_n)_{n\geqslant 1}$ and μ be probability measures on $\mathbb{R}^d$. Then, the four following conditions are equivalent.*

(1) The sequence $(\mu_n)_{n\geqslant 1}$ converges weakly to μ.

(2) For any open set $G \subseteq \mathbb{R}^d$,

$$\liminf_{n\to\infty} \mu_n(G) \geqslant \mu(G).$$

(3) For any closed set $F \subseteq \mathbb{R}^d$,

$$\limsup_{n\to\infty} \mu_n(F) \leqslant \mu(F).$$

(4) For any Borel set $B \subseteq \mathbb{R}^d$, if $\mu(\partial B) = 0$, then,

$$\lim_{n\to\infty} \mu_n(B) = \mu(B).$$

Proof : We first prove that (1) $\Rightarrow$ (2). If G is an open set in $\mathbb{R}^d$, we can construct a sequence of continuous bounded functions $(\varphi_p)_{p\geqslant 1}$ such that $0 \leqslant \varphi_p \leqslant 1$ and $\varphi_p \uparrow \mathbb{1}_G$. For example, take $\varphi_p(x) = pd(x, G^c) \wedge 1$. We have,

$$\begin{aligned}\liminf_{n\to\infty} \mu_n(G) &= \liminf_{n\to\infty} \Big(\lim_{p\to\infty} \uparrow \int \varphi_p \,\mathrm{d}\mu_n \Big) \\ &\geqslant \sup_{p\geqslant 1} \Big(\liminf_{n\to\infty} \int \varphi_p \,\mathrm{d}\mu_n \Big) \\ &= \sup_{p\geqslant 1} \Big(\int \varphi_p \,\mathrm{d}\mu \Big) = \mu(G).\end{aligned}$$

The equivalent relation (2) $\Leftrightarrow$ (3) is not hard to prove. Taking the complement interchanges the role

[2] In the first edition of the book "Convergence of Probability Measures" from Patrick Billingsley (1968), he mentioned that this result can be tracked back to an article of Aleksandrov in 1940. Later, in the second edition (1999) of the same book, he dedicated this theorem to Jean-Pierre Portmanteau, and cited the article "Hope for the empty set?" (Espoir pour l'ensemble vide?) from the journal "Annales de l'Université de Felletin". However, this person, the university and the article do not exist.

of an open set and a closed set and also changes the direction of the inequality.

Prove that (2)+(3) $\Rightarrow$ (4). Let $B \in \mathcal{B}(\mathbb{R}^d)$. Then, we have,

$$\limsup \mu_n(B) \leqslant \limsup \mu_n(\overline{B}) \leqslant \mu(\overline{B}),$$
$$\liminf \mu_n(B) \geqslant \liminf \mu_n(\mathring{B}) \geqslant \mu(\mathring{B}),$$

Due to the assumption that $\mu(\partial B) = 0$, we have $\mu(\overline{B}) = \mu(\mathring{B})$, implying,

$$\limsup \mu_n(B) = \liminf \mu_n(B) = \lim \mu_n(B).$$

Finally, we prove that (4) $\Rightarrow$ (1). Let $\varphi \in C_b(\mathbb{R}^d)$. We can separate the positive and the negative part of φ into $\varphi = \varphi^+ - \varphi^-$, so we can assume that φ is a non-negative function. Since φ is bounded, we can take $K > 0$ such that $0 \leqslant \varphi \leqslant K$. From the Fubini's theorem, we obtain,

$$\int \varphi(x)\mu(\mathrm{d}x) = \int \Big(\int_0^K \mathbb{1}_{\{t\leqslant\varphi(x)\}}\,\mathrm{d}t\Big)\mu(\mathrm{d}x) = \int_0^K \Big(\int \mathbb{1}_{\{t\leqslant\varphi(x)\}}\mu(\mathrm{d}x)\Big)\,\mathrm{d}t.$$

Let $E_t^\varphi = \{x \in \mathbb{R}^d : \varphi(x) \geqslant t\}$. Then the above formula rewrites,

$$\int \varphi(x)\mu(\mathrm{d}x) = \int_0^K \mu(E_t^\varphi)\,\mathrm{d}t.$$

Similarly, for all n, we have,

$$\int \varphi(x)\mu_n(\mathrm{d}x) = \int_0^K \mu_n(E_t^\varphi)\,\mathrm{d}t.$$

We can notice that $\partial E_t^\varphi \subseteq \{x \in \mathbb{R}^d : \varphi(x) = t\}$ and it follows from Exercise 1.13 that there exists at most countably many t such that

$$\mu(\{x \in \mathbb{R}^d : \varphi(x) = t\}) > 0.$$

Hence, from the assumption (4), we have,

$$\mu_n(E_t^\varphi) \longrightarrow \mu(E_t^\varphi), \qquad \mathrm{d}t\text{-a.s.},$$

and the dominated convergence theorem implies,

$$\int \varphi(x)\mu_n(\mathrm{d}x) = \int_0^K \mu_n(E_t^\varphi)\,\mathrm{d}t \xrightarrow[n\to\infty]{} \int_0^K \mu(E_t^\varphi)\,\mathrm{d}t = \int \varphi(x)\mu(\mathrm{d}x).$$

$\square$

Question 4.4.9: Consider real-valued random variables $(X_n)_{n\geqslant 1}$ and X and we write $(F_{X_n})_{n\geqslant 1}$ and F_X for their distributions. Then, the sequence of random variables $(X_n)_{n\geqslant 1}$ converges in distribution to X if and only if for all the points of continuity x of F_X, the distribution function $F_{X_n}(x)$ converges to $F(x)$.

Proposition 4.4.10 : *If $(X_n)_{n\geqslant 1}$ converges in probability to X, then $(X_n)_{n\geqslant 1}$ also converges in distribution to X.*

Proof : First, assume that $(X_n)_{n\geqslant 1}$ converges almost surely to X. In this case, for any $\varphi \in C_b(\mathbb{R}^d)$, $\varphi(X_n)$ converges almost surely to $\varphi(X)$ and the dominated convergence theorem implies $\mathbb{E}[\varphi(X_n)] \longrightarrow \mathbb{E}[\varphi(X)]$. This shows that $(X_n)_{n\geqslant 1}$ converges in distribution to X.

In a more general setting, we show by contradiction. Suppose that $(X_n)_{n\geqslant 1}$ does not converge in distribution to X, then we can find $\varphi \in C_b(\mathbb{R}^d)$ such that $\mathbb{E}[\varphi(X_n)]$ does not converge to $\mathbb{E}[\varphi(X)]$. Let $\varepsilon > 0$ and a subsequence $(n_k)_{k\geqslant 1}$ such that for all k, we have,

$$\forall k \geqslant 1, \qquad |\mathbb{E}[\varphi(X_{n_k})] - \mathbb{E}[\varphi(X)]| \geqslant \varepsilon.$$

From Proposition 4.1.3, we can find a subsequence $(X_{n_{k_l}})_{l\geqslant 1}$ that converges almost surely to X, but the proof from the first part gives a contradiction. □

Proposition 4.4.11 : *Show that if $(X_n)_{n\geqslant 1}$ converges in distribution to X which is almost surely a constant, then (X_n) also converges in probability to X.*

Proof : See Exercise 4.15. □

We define $C_c(\mathbb{R}^d)$ to be the set of continuous and *compactly supported* (緊緻支撐) functions.

Proposition 4.4.12 : *Let (μ_n) and μ be probability measures on $\mathbb{R}^d$. Let H be a subset of the normed space $(C_b(\mathbb{R}^d), \|\cdot\|_\infty)$ and assume that its closure (閉包) contains $C_c(\mathbb{R}^d)$. Then, the following properties are equivalent.*

(i) The sequence of probability distributions (μ_n) converges weakly to μ.

(ii) We have,

$$\forall \varphi \in C_c(\mathbb{R}^d), \qquad \int \varphi \,\mathrm{d}\mu_n \longrightarrow \int \varphi \,\mathrm{d}\mu.$$

(iii) We have,

$$\forall \varphi \in H, \qquad \int \varphi \,\mathrm{d}\mu_n \longrightarrow \int \varphi \,\mathrm{d}\mu.$$

Proof : Since $C_c(\mathbb{R}^d) \subseteq C_b(\mathbb{R}^d)$ and $H \subseteq C_b(\mathbb{R}^d)$, there is nothing to prove for (i) ⇒ (ii) and (i) ⇒ (iii). Now, we prove (ii) ⇒ (i). Consider a continuous bounded function $\varphi \in C_b(\mathbb{R}^d)$ and a seuqnece (f_k) of functions in $C_c(\mathbb{R}^d)$ such that $0 \leqslant f_k \leqslant 1$ and $\lim \uparrow f_k = 1$, then for any k, we have $\varphi f_k \in C_c(\mathbb{R}^d)$ so,

$$\int \varphi f_k \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int \varphi f_k \,\mathrm{d}\mu.$$

Moreover, we have,

$$\Big|\int \varphi \,\mathrm{d}\mu_n - \int \varphi f_k \,\mathrm{d}\mu_n\Big| \leqslant \Big(\sup_x |\varphi(x)|\Big)\Big(1 - \int f_k \,\mathrm{d}\mu_n\Big),$$
$$\Big|\int \varphi \,\mathrm{d}\mu - \int \varphi f_k \,\mathrm{d}\mu\Big| \leqslant \Big(\sup_x |\varphi(x)|\Big)\Big(1 - \int f_k \,\mathrm{d}\mu\Big).$$

Hence, for all k, we have,

$$\begin{aligned}\limsup_{n\to\infty}\left|\int\varphi\,\mathrm{d}\mu_n-\int\varphi\,\mathrm{d}\mu\right| &\leqslant \left(\sup_x|\varphi(x)|\right)\left(\limsup_{n\to\infty}\left(1-\int f_k\,\mathrm{d}\mu_n\right)+\left(1-\int f_k\,\mathrm{d}\mu\right)\right),\\ &= 2\left(\sup_x|\varphi(x)|\right)\left(1-\int f_k\,\mathrm{d}\mu\right).\end{aligned}$$

The above formula being true for all k, we can take $k\to\infty$ to obtain,

$$\int\varphi\,\mathrm{d}\mu_n\longrightarrow\int\varphi\,\mathrm{d}\mu.$$

Next, we prove (iii) $\Rightarrow$ (ii). Let $\varphi\in C_c(\mathbb{R}^d)$. Using the density of H, for all $k\geqslant 1$, we can find $\varphi_k\in H$ such that $\|\varphi-\varphi_k\|\leqslant 1/k$, so for all k, we have,

$$\begin{aligned}&\limsup_{n\to\infty}\left|\int\varphi\,\mathrm{d}\mu_n-\int\varphi\,\mathrm{d}\mu\right|\\ &\leqslant\limsup_{n\to\infty}\left(\left|\int\varphi\,\mathrm{d}\mu_n-\int\varphi_k\,\mathrm{d}\mu_n\right|+\left|\int\varphi_k\,\mathrm{d}\mu_n-\int\varphi_k\,\mathrm{d}\mu\right|+\left|\int\varphi_k\,\mathrm{d}\mu-\int\varphi\,\mathrm{d}\mu\right|\right)\\ &\leqslant\frac{2}{k}.\end{aligned}$$

Since k can be arbitrarily large, we obtain $\int\varphi\,\mathrm{d}\mu_n\longrightarrow\int\varphi\,\mathrm{d}\mu$. □

Remark 4.4.13 : For a sequence $(\mu_n)_{n\geqslant 1}$ of probability measures on $\mathbb{R}^d$ and a measure μ, we say that μ_n converges vaguely (淡收斂) to μ if

$$\forall f\in C_c(\mathbb{R}^d),\qquad \int f\,\mathrm{d}\mu_n\xrightarrow[n\to\infty]{}\int f\,\mathrm{d}\mu.$$

According to Proposition 4.4.12, when we know that μ is also a *probability measure*, the weak convergence and the vague convergence are equivalent; but in general, without the assumption that μ has a total mass 1, the Fatou's lemma can only give us $\mu(\mathbb{R}^d)\leqslant 1$.

We may consider the following example,

$$\forall f\in C_c(\mathbb{R}^d),\qquad f(n)=\int_{\mathbb{R}^d}f(x)\delta_n(\mathrm{d}x)\xrightarrow[n\to\infty]{}0,$$

implying that μ_n converges vaguely to 0, but 0 is clearly not a probability measure because its total mass is 0. The main reason is that, when the test functions at are disposition are functions from $C_c(\mathbb{R}^d)$, we might have some probability masses that "escape to infinity", which cannot be captured by functions in $C_c(\mathbb{R}^d)$; however, the test functions in $C_b(\mathbb{R}^d)$ are able to capture this phenomenon. This intuition also provides another explanation for the equivalence between the weak convergence and the vague convergence in the case that μ is a probability measure.

Question 4.4.14: Given a sequence $(\mu_n)_{n\geqslant 1}$ of probability measures on $\mathbb{R}^d$. We say that $(\mu_n)_{n\geqslant 1}$ is a tight (緊密) sequence if for all $\varepsilon>0$, there exists a compact set $K_\varepsilon\subseteq\mathbb{R}^d$ such that

$$\mu_n(K_\varepsilon)\geqslant 1-\varepsilon.$$

Given a measure μ on $\mathbb{R}^d$ and prove that the following properties are equivalent.

(1) μ_n converges weakly to μ.

(2) μ_n converges vaguely to μ and $(\mu_n)_{n\geqslant 1}$ is tight.

Theorem 4.4.15 (Lévy's continuity theorem) **:** *Let $(\mu_n)_{n\geqslant 1}$ be a sequence of probability measures on $\mathbb{R}^d$. Then, $(\mu_n)_{n\geqslant 1}$ converges weakly to μ if and only if,*

$$\forall \xi \in \mathbb{R}^d, \qquad \widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi).$$

Similarly, the sequence of random variables $(X_n)_{n\geqslant 1}$ converges in distribution to X if and only if,

$$\forall \xi \in \mathbb{R}^d, \qquad \Phi_{X_n}(\xi) \longrightarrow \Phi_X(\xi).$$

Proof : We only need to prove the first part of the statement. If $(\mu_n)_{n\geqslant 1}$ converges weakly to μ, then from the definition of the weak convergence, since for any fixed $\xi \in \mathbb{R}$, both $x \mapsto \cos(\xi x)$ and $x \mapsto \sin(\xi x)$ are bounded, we obtain

$$\forall \xi \in \mathbb{R}^d, \qquad \widehat{\mu}_n(\xi) = \int e^{\mathrm{i}\,\xi\cdot x}\mu_n(\mathrm{d}x) \longrightarrow \int e^{\mathrm{i}\,\xi\cdot x}\mu(\mathrm{d}x) = \widehat{\mu}(\xi).$$

Then we show its converse. Assume that for all $\xi \in \mathbb{R}^d$, we have $\widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi)$ and we want to show that the sequence of probability measures $(\mu_n)_{n\geqslant 1}$ converges weakly. We want to use (3) from Proposition 4.4.12. To simplify the proof, we also assume that $d = 1$.

Let $f \in C_c(\mathbb{R})$. For all $\sigma > 0$, let

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{x^2}{2\sigma^2}\Big).$$

In the proof of Theorem 2.4.15, from Eq. (2.10), we know that $g_\sigma * f$ converges simply to f. Then using Question 2.4.16, since f is compactly supported, we have the uniform convergence of $g_\sigma * f$ to f. If we let

$$H = \{\varphi = g_\sigma * f : f \in C_c(\mathbb{R}) \text{ and } \sigma > 0\},$$

then $C_c(\mathbb{R}) \subseteq \overline{H}$, so it is enough to show that,

$$\forall f \in C_c(\mathbb{R}), \qquad \int g_\sigma * f \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int g_\sigma * f \,\mathrm{d}\mu.$$

From Eq. (2.8) and Eq. (2.9) in the proof of Theorem 2.4.15, for any probability measure ν on $\mathbb{R}$, we have,

$$\int g_\sigma * f \,\mathrm{d}\nu = \frac{1}{\sqrt{2\pi\sigma^2}} \int f(x)\Big(\int e^{\mathrm{i}\,\xi x} g_{1/\sigma}(\xi)\widehat{\nu}(-\xi)\,\mathrm{d}\xi\Big)\,\mathrm{d}x$$

From the assumption, for all $\xi \in \mathbb{R}$, $\widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi)$ and $|\widehat{\mu}_n(\xi)| \leqslant 1$, so using the dominated convergence theorem, we obtain,

$$\int e^{\mathrm{i}\,\xi x} g_{1/\sigma}(\xi)\widehat{\mu}_n(-\xi)\,\mathrm{d}\xi \xrightarrow[n\to\infty]{} \int e^{\mathrm{i}\,\xi x} g_{1/\sigma}(\xi)\widehat{\mu}(-\xi)\,\mathrm{d}\xi.$$

Since the left side of the above formula also satisfies $|\cdot| \leqslant 1$, we apply again the dominated convergence

theorem to conclude,
$$\int g_\sigma * f \,\mathrm{d}\mu_n \xrightarrow[n\to\infty]{} \int g_\sigma * f \,\mathrm{d}\mu.$$
□

4.5 Applications of Convergence in Distribution

4.5.1 Convergence of Emperical Measures

Let $(X_n)_{n\geqslant 1}$ be an i.i.d. sequence of random variables with values in $\mathbb{R}^d$. We can think of these random variables as values observed in a series of independent and identical random experiments. In statistics, we wish to derive the distribution of X_1 from the observations $X_1(\omega), \ldots, X_n(\omega)$ (ω being a point in the probability space).

Taking a national poll as example. Let N be the Taiwanese population. The Taiwanese number i has its own vector $a(i) \in \mathbb{R}^d$ representing its data, such as age, income, health condition, political tendency, etc. When we are given a measurable set $A \in \mathcal{B}(\mathbb{R}^d)$ (e.g., fans of Han Kuo-Yu with annual income above 1 million and above 50 years old), we want to know the proportion of Taiwanese popultion whose vector $a(i)$ belongs to this set. Alternatively speaking, we want to estimate,

$$\mu(A) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_A(a(i)).$$

When N is large, it is impossible to compute precisely this value, and the ultimate goal of a national poll is to find a representative set of n people from the Taiwanese population so that we can have a reasonable estimate of $\mu(A)$. If we consider uniform random variables $Y_1, \ldots, Y_n$ with values in $\{1, \ldots, N\}$ (i.e., choose n Taiwanese uniformly at random) and write $X_j = a(Y_j)$, $X_1, \ldots, X_n$ be i.i.d. random variables with the following distribution,

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \qquad \mathbb{P}_{X_1}(A) = \mathbb{P}(a(Y_1) \in A) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_A(a(i)) = \mu(A).$$

From this sample, we obtain the following estimation,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_A(X_i(\omega)) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}(A).$$

Coming back to the original problem, we want to know whether the above estimate is close to the theoretical value $\mu(A)$, which means that we want to know whether the *emperical measure* defined below converges to $\mathbb{P}_{X_1}$ when n tends to infinity,

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}.$$

The following theorem gives a positive answer.

Theorem 4.5.1 : *Let $(X_n)_{n\geqslant 1}$ be an i.i.d. sequence of random variables with values in $\mathbb{R}^d$. For all $\omega \in \Omega$*

and $n \geqslant 1$*, define* $\mu_{n,\omega}$ *be the following* emperical measure (經驗測度) *on* $\mathbb{R}^d$*,*

$$\mu_{n,\omega} = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}.$$

Then, when $n \to \infty$*, we have the following convergence result,*

$$\mu_{n,\omega} \Longrightarrow \mathbb{P}_{X_1}.$$

Remark 4.5.2 : This theorem does not provide any convergence speed, so we do not know at which rate $\mu_{n,\omega}$ converges to $\mathbb{P}_{X_1}$.

Proof : Let H be a countable dense subset of $C_c(\mathbb{R}^d)$. If $\varphi \in H$, from the strong law of large numbers, we have,

$$\frac{1}{n}\sum_{i=1}^{n} \varphi(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[\varphi(X_1)].$$

The above formula rewrites,

$$\int \varphi \, \mathrm{d}\mu_{n,\omega} \xrightarrow{\text{a.s.}} \int \varphi \, \mathrm{d}\mathbb{P}_{X_1}.$$

Due to the countability of H, we can obtain,

$$\text{a.s.} \quad \forall \varphi \in H, \quad \int \varphi \, \mathrm{d}\mu_{n,\omega} \xrightarrow[n\to\infty]{} \int \varphi \, \mathrm{d}\mathbb{P}_{X_1}.$$

We conclude using Proposition 4.4.12. □

4.5.2 Central Limit Theorem

In Theorem 4.3.1, we obtained the strong law of large numbers: if $(X_n)_{n\geqslant 1}$ is an i.i.d. sequence of random variables where each term is integrable, then the following result holds almost surely,

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{\text{a.s.}} \mathbb{E}[X_1].$$

After this, we can study the *speed* of the above convergence. To be more precise, we want to understand the behavior of the following quantity when n is arbitrarily large,

$$\frac{1}{n}(X_1 + \cdots + X_n) - \mathbb{E}[X_1]. \tag{4.4}$$

Let us start with a simple computation: assume that X_i is square-integrable, then we note that,

$$\mathbb{E}[(X_1 + \cdots + X_n - n\,\mathbb{E}[X_1])^2] = \mathrm{Var}(X_1 + \cdots + X_n) = n\,\mathrm{Var}(X_1).$$

This means that $(X_1 + \cdots + X_n - n\,\mathbb{E}[X_1])^2$ is linear in n, i.e., Eq. (4.4) and $\frac{1}{\sqrt{n}}$ are of the same order.

Below we first state the one-dimensional version of the central limit theorem (中央極限定理), the higher-dimensional version being discussed later in Section 4.5.3.

Theorem 4.5.3 (*Central limit theorem* (中央極限定理)) **:** *Let $(X_n)_{n\geqslant 1}$ be an i.i.d. sequence of real-valued random variables where each term is square-integrable. Let $\sigma^2 = \mathrm{Var}(X_1)$. Then, we have,*

$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n - n\,\mathbb{E}[X_1]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

In other words, for all $a, b \in \overline{\mathbb{R}}$ with $a < b$, we have the following convergence,

$$\lim_{n\to\infty} \mathbb{P}\left(n\,\mathbb{E}[X_1] + a\sqrt{n} \leqslant X_1 + \cdots + X_n \leqslant n\,\mathbb{E}[X_1] + b\sqrt{n}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathrm{d}x.$$

Proof : The second part of the theorem being a direct consequence of the first part (Question 4.4.9 and Exercise 4.13), we only need to prove the first part. Additionally, we can replace X_n with $X_n - \mathbb{E}[X_n]$, so that we can assume $\mathbb{E}[X_n] = 0$. Let

$$Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

We want to use Theorem 4.4.15 to show this theorem. The characteristic function of the random variable Z_n writes,

$$\Phi_{Z_n}(\xi) = \mathbb{E}\left[\exp\left(\mathrm{i}\,\xi\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\right)\right] = \mathbb{E}\left[\exp\left(\mathrm{i}\,\frac{\xi}{\sqrt{n}} X_1\right)\right]^n = \Phi_{X_1}\Big(\frac{\xi}{\sqrt{n}}\Big)^n.$$

The series expansion in Proposition 2.4.17 gives, when $\xi \longrightarrow 0$,

$$\Phi_{X_1}(\xi) = 1 + \mathrm{i}\,\xi\,\mathbb{E}[X_1] - \frac{1}{2}\xi^2\,\mathbb{E}[X_1^2] + o(\xi^2) = 1 - \frac{\sigma^2\xi^2}{2} + o(\xi^2).$$

Hence, for any given $\xi \in \mathbb{R}$, when $n \longrightarrow \infty$, we have,

$$\Phi_{X_1}\Big(\frac{\xi}{\sqrt{n}}\Big) = 1 - \frac{\sigma^2\xi^2}{n} + o\Big(\frac{1}{n}\Big).$$

So we get,

$$\lim_{n\to\infty} \Phi_{Z_n}(\xi) = \lim_{n\to\infty} \Big(1 - \frac{\sigma^2\xi^2}{2n} + o\Big(\frac{1}{n}\Big)\Big)^n = \exp\left(-\frac{\sigma^2\xi^2}{2}\right) = \Phi_U(\xi),$$

where U has the distribution $\mathcal{N}(0, \sigma^2)$. To conclude, we have shown the central limit theorem using Theorem 4.4.15. $\square$

4.5.3 Central Limit Theorem in Higher Dimensions

Suppose that we have an i.i.d. sequence of integrable random variables $(X_n := (X_n^{(1)}, \dots, X_n^{(d)}))_{n \geqslant 1}$ with values in $\mathbb{R}^d$. We can apply the strong law of large numbers to each of its component $X_n^{(i)}$ to obtain,

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{\text{a.s.}} \mathbb{E}[X_1],$$

where $\mathbb{E}[X_1]$ is a vector consisting of the expectations of each component. If $(X_n)_{n \geqslant 1}$ is square-integrable, we can apply the same approach to deduce the central limit theorem for each of the component. However, this approach is not enough to get the central limit theorem for the d-dimensional vector for the simple reason that the marginal distributions are not sufficient to describe the distribution of the whole vector. In fact, the higher-dimensional version of the central limit theorem involves the multivariate normal distribution that was discussed in Section 3.4.1.

Theorem 4.5.4 (High-dimensional central limit theorem) **:** *Let $(X_n)_{n \geqslant 1}$ be an i.i.d. sequence of random variables with values in $\mathbb{R}^d$. Assume that they are all square-integrable, then we have,*

$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n - n\,\mathbb{E}[X_1]) \xrightarrow{(d)} \mathcal{N}(0, K_{X_1}).$$

Proof : The proof is exactly the same as in the one-dimensional case. Without loss of generality, suppose that $\mathbb{E}[X_1] = 0$. For all $\xi \in \mathbb{R}^d$, we have,

$$\mathbb{E}\left[\exp\left(\mathrm{i}\,\xi \cdot \left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right)\right] = \mathbb{E}\left[\exp\left(\mathrm{i}\,\frac{\xi}{\sqrt{n}} \cdot X_1\right)\right] = \Phi_{X_1}\left(\frac{\xi}{\sqrt{n}}\right)^n.$$

At the same time, we also have,

$$\Phi_{X_1}\left(\frac{\xi}{\sqrt{n}}\right) = 1 - \frac{1}{2n}\xi^T K_{X_1} \xi + o(n^{-1}).$$

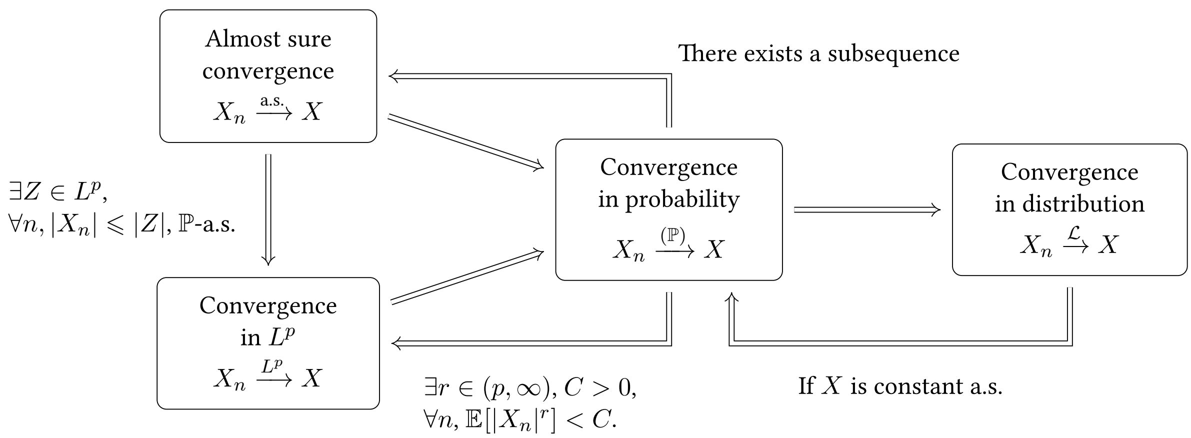
Hence,

$$\lim_{n \to \infty} \mathbb{E}\left[\exp\left(\mathrm{i}\,\xi \cdot \left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right)\right] = \exp\left(-\frac{1}{2}\xi^T K_{X_1} \xi\right).$$

Finally, we conclude with Lévy's continuity theorem (Theorem 4.4.15). □

4.6 Conclusion

We use the following diagram to conclude this chapter, showing different notions of convergence and their relations with each other.


Almost sure convergence
$X_n \xrightarrow{\text{a.s.}} X$
There exists a subsequence
Convergence in probability
$X_n \xrightarrow{(\mathbb{P})} X$
Convergence in distribution
$X_n \xrightarrow{\mathcal{L}} X$
$\exists Z \in L^p$,
$\forall n, |X_n| \leqslant |Z|, \mathbb{P}$-a.s.
Convergence in L^p
$X_n \xrightarrow{L^p} X$
$\exists r \in (p, \infty), C > 0$,
$\forall n, \mathbb{E}[|X_n|^r] < C$.
If X is constant a.s.