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.*

第四章 隨機變數的收斂

在拓撲學中，在同樣的空間，但賦有不同拓撲的情況下，我們有不同的收斂概念，不同的收斂概念之間，有些可以互相比較，而有些則是互不相關。在這章中，我們將討論隨機變數的不同收斂概念，並探討不同概念之間的強弱比較。

第一節 機率收斂

給定一機率空間 $(\Omega, \mathcal{A}, \mathbb{P})$ 並且令 $(X_n)_{n\geqslant 1}$ 為定義在此機率空間之上，值域為 $\mathbb{R}^d$ 的隨機變數序列。在測度論中，我們已經探討過下列兩個收斂概念：

- 殆必收斂：若 $\mathbb{P}(\{\omega \in \Omega \mid X(\omega) = \lim_{n\to\infty} X_n(\omega)\}) = 1$ ，則寫作 $X_n \xrightarrow{\text{a.s.}} X$。
- 對於 $p \in [1, \infty)$，L^p 收斂：若 $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$ ，則寫作 $X_n \xrightarrow{L^p} X$。

定義 4.1.1 ： 若對於所有 $\varepsilon > 0$，我們有

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

則我們稱 $(X_n)_{n\geqslant 1}$ 機率收斂 (converges in probability) 至 X，並記作

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

接著我們要證明機率收斂的概念是可以賦距的。

令 $\mathcal{L}^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$ 為所以值域為 $\mathbb{R}^d$ 的隨機變數構成的空間，在此空間上，我們定義等價關係若且唯若 $X \sim Y$ 則 X 與 Y 殆必相等，並將商空間記作 $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$。在商空間 $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$ 上，我們可以賦予下列距離 (distance)：

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

命題 4.1.2 ： 式 (4.1) 中的定義是個距離，而且可以刻劃機率收斂的概念；換句話說：若且唯若 $(X_n)_{n\geqslant 1}$ 機率收斂至 X，則 $d(X_n, X)$ 趨近 0。此外，賦距空間 $(L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P}), d)$ 是完備的。

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. □

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,

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

證明：驗證式 (4.1) 為距離不難，只需要用到商空間和距離的定義即可。接著，我們證明此距離和機率收斂等價。首先，假設 (X_n) 機率收斂至 X，則對於任意給定的 $\varepsilon \in (0,1)$，我們有

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

則告訴我們 $\limsup d(X_n, X) \leqslant \varepsilon$，但由於 $\varepsilon > 0$ 可以任意小，所以我們得到 $\lim d(X_n, X) = 0$。反之，若 $\lim d(X_n, X) = 0$，則對於任意 $\varepsilon \in (0,1)$，我們有

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

再來我們證明賦距空間 $(L^0_{\mathbb{R}^d}, d)$ 的完備性。給定在距離 d 之下的柯西序列 (X_n)，我們可以找到一個子序列 $Y_k = X_{n_k}$ 使得對於所有 $k \geqslant 1$，

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

則我們有

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

所以 $\sum (|Y_{k+1} - Y_k| \wedge 1) < \infty$, a.s.，也就是 $\sum |Y_{k+1} - Y_k| < \infty$, a.s.，因此我們可以定義

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

而且根據上面的構造，我們知道 (Y_k) 殆必收斂至 X，而且根據勒貝格收斂定理，我們有

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

所以 (Y_k) 機率收斂至 X；由於 (X_k) 是柯西序列，(X_k) 也機率收斂至 X。 □

命題 4.1.3 ： 令 $(X_n)_{n \geqslant 1}$ 為隨機變數序列，則下列性質成立：

(1) 若 (X_n) 殆必收斂或對於 $p \geqslant 1$ 會 L^p 收斂至 X，則它也機率收斂至 X。

(2) 若 (X_n) 機率收斂至 X，則存在子序列 (X_{n_k}) 殆必收斂至 X。

證明：我們在命題 4.1.2 中的證明，已經證明過 (2) 了，因此我們只需要證明 (1)。若 (X_n) 殆必收斂至 X，則根據勒貝格收斂定理，我們有

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

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

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

□

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 $n \geqslant 1$, $\mathbb{E}[|X_n|^r] \leqslant C$. 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

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

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

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

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

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

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

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

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,

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

若 (X_n) 在 L^p 中收斂至 X，則

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

□

命題 4.1.4 ： 令隨機變數序列 $(X_n)_{n\geqslant 1}$ 機率收斂至 X。假設存在 $r \in (1,\infty)$ 使得 $(X_n)_{n\geqslant 1}$ 在 L^r 中是有界的，則對於所有 $p \in [1, r)$，序列 $(X_n)_{n\geqslant 1}$ 在 L^p 中收斂至 X。

證明：根據假設，設 $C > 0$ 使得對於所有 $n \geqslant 1$，我們有 $\mathbb{E}[|X_n|^r] \leqslant C$。透過命題 4.1.3 中的 (2)，我們可以得到子序列 $(n_k)_{k\geqslant 1}$ 使得 $X_{n_k} \xrightarrow{\text{a.s.}} X$。接著，我們使用 Fatou 引理，得到

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

對於所有 $p \in [1, r)$ 以及 $\varepsilon > 0$，我們有

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

其中在第二行中，我們使用了赫爾德定理；在第三行中，我們使用閔可夫斯基不等式：

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

由於 (X_n) 機率收斂至 X，我們得到

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

由於 ε 可以無限小，我們得到 (X_n) 在 L^p 中收斂至 X。 □

第二節 全有全無律

全有全無律又稱作*Kolmogorov* 零一律 (Kolmogorov's 0-1 law) ，敘述尾端事件 (tail event) 的發生機率必定為零或一。

定義 4.2.1 ： 令 $(X_n)_{n\geqslant 1}$ 為獨立隨機變數序列，且值域是任意賦距空間。對於所有 $n \geqslant 1$，定義 $\mathcal{B}_n$ 為下列 σ 代數

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

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 $\frac{1}{n}(X_1 + \dots X_n)$ 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 $S_n = X_1 + \dots, +X_n$. Then,*

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

定義漸進 σ 代數 (asymptotic σ-algebra) $\mathcal{B}_\infty$ 為

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

在 $\mathcal{B}_\infty$ 中的事件也稱作尾端事件 (tail event)。

定理 4.2.2 ： 沿用上面的記號，$\mathcal{B}_\infty$ 是平凡的 σ 代數，也就是說，對於所有的 $B \in \mathcal{B}_\infty$，我們有 $\mathbb{P}(B) = 0$ 或 1。

證明：對於所有 $n \geqslant 1$，令

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

根據問題 3.1.19 ，我們知道對於所有 $n \geqslant 1$，$\mathcal{D}_n$ 與 $\mathcal{B}_{n+1}$ 是獨立的，因此 $\mathcal{D}_n$ 也會與 $\mathcal{B}_\infty$ 獨立。這告訴我們

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

由於 $\cup \mathcal{D}_n$ 在有限交集下是封閉的，根據命題 3.1.18 ，我們知道 $\mathcal{B}_\infty$ 會與下列 σ 代數獨立：

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

由於 $\mathcal{B}_\infty$ 包含在上述 σ 代數之中，所以 $\mathcal{B}_\infty$ 與自己獨立，也就是說，對於所有 $B \in \mathcal{B}_\infty$，我們有

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

也就是說 $\mathbb{P}(B) = 0$ 或 1。 □

問題 4.2.3：給定 i.i.d. 實隨機變數序列 $(X_n)_{n\geqslant 1}$，並且以定義 4.2.1 的方式定義 σ 代數 $\mathcal{B}_k$ 及 $\mathcal{B}_\infty$。試證明下列性質：

(1) 若 Y 是個對於 $\mathcal{B}_\infty$ 可測的實隨機變數，則它殆必為常數。

(2) 若 $\frac{1}{n}(X_1 + \dots X_n)$ 殆必收斂，從上題推得則其極限必須殆必為常數。

下列命題是零一律的重要應用。

命題 4.2.4 ： 令 $(X_n)_{n\geqslant 1}$ 為 i.i.d. 隨機變數序列，並假設他們的分佈是 $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$。對於所有 $n \geqslant 1$，我們令 $S_n = X_1 + \dots, +X_n$，則

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

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 $\mathcal{B}_\infty$. □

此外，這也代表殆必存在無限個 n 使得 $S_n = 0$。

註解 4.2.5 ： 若我們投擲一公正銅板，每次出現正面即得到一元，反之則損失一元，此命題告訴我們，在遊戲過程中，我們淨賺可以是很大的正值，也可以是很大的負值。

證明：對於所有 $p \geqslant 1$，我們令下列事件

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

我們注意到，事件 (A_p) 的極限會是

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

首先，我們想要證明對於所有 $p \geqslant 1$，$\mathbb{P}(A_p) = 0$。對於所有 $k > 2p$ 以及 $j \geqslant 0$，令

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

則我們有

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

由於 $(B_{j,k})_{j\geqslant 0}$ 是個獨立序列，而且 $\sum \mathbb{P}(B_{j,k}) = \infty$，根據 Borel–Cantelli 引理，式 (4.2) 左側事件的機率為 1，也就是 $\mathbb{P}(A_p) = 0$。根據機率測度的連續性，我們得到 $\mathbb{P}(A) = \lim \uparrow \mathbb{P}(A_p) = 0$，也就是說

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

根據對稱性，我們知道

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

所以這兩個機率皆非零。

最後，我們想要使用全有全無律（定理 4.2.2 ）來證明上式的兩個機率皆為 1。我們首先注意到，對於所有 $k \geqslant 1$，

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

所以說，事件 $\{\sup_n S_n = \infty\}$ 對於所有 σ 代數 $\mathcal{B}_k$ 皆為可測的，因此也會對所有 $\mathcal{B}_k$ 的交集 $\mathcal{B}_\infty$ 可測。□

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) ⇒ (ii) ⇒ (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,

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

第三節 強大數法則

此章節的目的是證明當 (X_n) 為在 L^1 中的 i.i.d. 隨機變數序列，則他的算術平均 $\frac{1}{n}(X_1 + \ldots + X_n)$ 殆必收斂至 $\mathbb{E}[X_1]$。

在命題 3.3.6 中，我們證明過在比較強的假設下，也就是 $\mathbb{E}[|X_1|^4] < \infty$，此命題成立，但我們現在希望能夠在最少條件的情況下，證明這樣的定理。

定理 4.3.1 ： 令 $(X_n)_{n\geqslant 1}$ 為分佈在 L^1 中的 i.i.d. 隨機變數序列，則

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

註解 4.3.2 ： 可積性的假設是最佳假設，因為上述收斂的極限必須要存在；當隨機變數序列皆為非負，而且 $\mathbb{E}[X_1] = \infty$，我們可以將此定理應用在 $(X_n \wedge k)_{n\geqslant 1}$ 之上以得到

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

接著取極限 $k \longrightarrow \infty$ 並使用單調收斂定理，進而推得

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

註解 4.3.3 ： 稍後在第 6.6 節中，我們會證明此收斂也會在 L^1 中成立。

證明：對於所有 $n \geqslant 1$，令 $S_n = X_1 + \cdots + X_n$，以及 $S_0 = 0$。設 $a > \mathbb{E}[X_1]$ 及隨機變數

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

為了證明此定理，我們將證明 (i) 及 (i) ⇒ (ii) ⇒ (iii)：

(i) $\mathbb{P}(M < \infty) > 0$ 或是等價條件 $\mathbb{P}(M = \infty) < 1$；

(ii) $M < \infty$, a.s.；

(iii) 強大數定理。

首先解釋為什麼 (ii) 蘊含 (iii) 強大數定理。根據 M 的定義，對於所有 n，我們有 $S_n \leqslant na + M$，接著使用 (ii) 的性質，我們可以得到

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

由於 a 可以是任意比 $\mathbb{E}[X_1]$ 大，且無窮接近於 $\mathbb{E}[X_1]$ 的數，所以我們得到

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

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,

$$\begin{aligned}M_k &= \sup_{0\leqslant n\leqslant k}(S_n - na), \\ M_k' &= \sup_{0\leqslant n\leqslant k}(S_{n+1} - S_1 - na).\end{aligned}$$

Since the vectors $(X_1, \dots, X_k)$ and $(X_2, \dots, X_{k+1})$ have the same distribution and we have $M_k = F_k(X_1, \dots, X_k)$ and $M_k' = F_k(X_2, \dots, X_{k+1})$ 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

若將 X_n 用 $-X_n$ 取代，則我們有

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

上兩式合併便可以得到強大數定理。

接著我們證明 (i) 蘊含 (ii)。首先，我們注意到，對於任意 $k \geqslant 0$，事件 $\{M < \infty\}$ 可以寫為

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

因此我們知道，事件 $\{M < \infty\}$ 對於漸進 σ 代數 $\mathcal{B}_\infty$ 是可測的，也就是說 $\mathbb{P}(M < \infty) = 0$ 或 1。所以 (i) 蘊含 (ii)。

最後我們證明 (i)。對於所有 $k \geqslant 0$，定義下列隨機變數

$$\begin{aligned}M_k &= \sup_{0\leqslant n\leqslant k}(S_n - na), \\ M_k' &= \sup_{0\leqslant n\leqslant k}(S_{n+1} - S_1 - na).\end{aligned}$$

由於向量 $(X_1, \dots, X_k)$ 以及 $(X_2, \dots, X_{k+1})$ 有著相同的分佈，且存在一個確定性 (deterministic) 函數 $F_k : \mathbb{R}^k \longrightarrow \mathbb{R}$ 使得我們有 $M_k = F_k(X_1, \dots, X_k)$ 以及 $M_k' = F_k(X_2, \dots, X_{k+1})$，因此，下面兩個隨機變數也有相同的分佈：

$$M = \lim_{k\to\infty} \uparrow M_k \quad 且 \quad M' = \lim_{k\to\infty} \uparrow M_k'.$$

此外，根據定義，我們知道對於所有 $k \geqslant 1$，

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

由於 M_k 以及 M_k' 皆在 L^1 中且有相同的分佈，我們可以得到

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

接著，由於 $M_k' \geqslant 0$，我們有 $|\inf(a - X_1, M_k')| \leqslant |a - X_1|$，再透過控制收斂定理，我們得到：

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

最後，如果 $\mathbb{P}(M = \infty) = 1$，則我們也會有 $\mathbb{P}(M' = \infty) = 1$，因此 $\inf(a - X_1, M') = a - X_1$ a.s.。但由於我們選擇 $a > \mathbb{E}[X_1]$，這和 $\mathbb{E}[a - X_1] \leqslant 0$ 相互矛盾。□

第四節 分佈收斂

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.

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

第一小節 定義及範例

函數空間 $C_b(\mathbb{R}^d)$ 為連續有界由 $\mathbb{R}^d$ 至 $\mathbb{R}$ 的函數所構成，此外，我們可以賦予此空間由函數的最小上界所定義的範數：

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

定義 4.4.1 ： 給定在 $\mathbb{R}^d$ 上的機率測度序列 $(\mu_n)_{n\geqslant 1}$，若

$$\mu_n \Longrightarrow \mu,$$

則我們說 $(\mu_n)_{n\geqslant 1}$ 弱收斂 (converges weakly) 至 μ，記作

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

給定值域為 $\mathbb{R}^d$ 的隨機變數序列 $(X_n)_{n\geqslant 1}$，若分佈 $(\mathbb{P}_{X_n})_{n\geqslant 1}$ 弱收斂至 $\mathbb{P}_X$，則我們說 $(X_n)_{n\geqslant 1}$ 弱收斂 (converges weakly) 或分佈收斂 (converges in distribution) 至 X，記作

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

隨機變數弱收斂的概念也與下列條件等價：

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

註解 4.4.2 ： 這裡我們對上述的定義做些補充說明：

(1) $\mathbb{R}^d$ 上測度構成的空間可以看作是對偶空間 $C_b(\mathbb{R}^d)^*$ 的子空間，因此式 (4.3) 應該被視為在弱∗拓撲上的收斂才對；但在機率論中，我們將他稱作弱收斂。[1]

(2) 定義中其實有個名稱濫用的問題：當我們說隨機變數序列 $(X_n)_{n\geqslant 1}$ 弱收斂至 X 時，這裡的 X 並沒有唯一性，有唯一性的是極限隨機變數的分佈 $\mathbb{P}_X$；因此，當我們想要使用精準一點的語言時，會說隨機變數序列 $(X_n)_{n\geqslant 1}$ 分佈收斂至 $\mathbb{P}_X$。

(3) 一般討論隨機變數收斂性質的時候，我們必須要假設他們的定義域（機率空間）相同，但當我們探討分佈收斂時，隨機變數的定義域未必要相同，這也是此收斂概念在機率論中重要的原因。

範例 4.4.3 ： 我們給兩個分佈收斂的例子。

(1) 若 X_n 是在 $\{\frac{k}{2^n} : 1\leqslant k\leqslant 2^n\}$ 上的均勻分佈，則 X_n 弱收斂至 $[0,1]$ 上的均勻分佈，這是因為連續函數的黎曼和會趨近其黎曼積分。

[1] 在法文數學界的用法，有另外的詞專門稱呼此收斂概念，稱之為 convergence étroite。

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

(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,

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

(2) 若 X_n 是 $\mathcal{N}(0,\sigma_n^2)$ 的高斯分佈，且 $\sigma_n^2 \longrightarrow 0$，則 X_n 分佈收斂至處處皆零的隨機變數。

問題 4.4.4：若 $\mu_n \Longrightarrow \mu$，找出連續但無界函數，以及不連續但有界函數，使得式 (4.3) 不成立。

問題 4.4.5：令 $(X_n)_{n\geqslant 1}$ 及 X 為值域為 $\mathbb{Z}^d$ 的隨機變數。試證：若且唯若 X_n 分佈收斂至 X，則

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

問題 4.4.6：假設對於所有 $n \geqslant 1$，(X_n) 是有密度函數的隨機變數，記作 $\mathbb{P}_{X_n}(\mathrm{d}x) = p_n(x)\,\mathrm{d}x$。假設

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

(2) 存在非負函數 q 使得 $\int_{\mathbb{R}^d} q(x)\,\mathrm{d}x < \infty$ 且

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

證明 p 是個在 $\mathbb{R}^d$ 上機率測度的密度函數，而且 X_n 分佈收斂至 $p(x)\,\mathrm{d}x$。

問題 4.4.7：若 $(X_n)_{n\geqslant 1}$ 分佈收斂至 X，是否對於任意 $B \in \mathcal{B}(\mathbb{R}^d)$，我們有下列收斂？

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

第二小節 分佈收斂的等價條件

下列定理給我們幾個與機率測度弱收斂的等價性質。

定理 4.4.8 【Portmanteau 定理 [2]】：令 $(\mu_n)_{n\geqslant 1}$ 及 μ 為在 $\mathbb{R}^d$ 上的機率測度，則下列四個條件等價。

(1) 序列 $(\mu_n)_{n\geqslant 1}$ 弱收斂至 μ。

(2) 對任意開集 $G \subseteq \mathbb{R}^d$，

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

(3) 對任意閉集 $F \subseteq \mathbb{R}^d$，

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

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 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,

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

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

[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.

(4) 對任意伯雷爾集合 $B \subseteq \mathbb{R}^d$，若 $\mu(\partial B) = 0$，則

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

證明：我們先證明 (1) $\Rightarrow$ (2)。若 G 為 $\mathbb{R}^d$ 中的開集，我們可以構造連續有界函數序列 $(\varphi_p)_{p\geqslant 1}$ 使得 $0 \leqslant \varphi_p \leqslant 1$ 而且 $\varphi_p \uparrow \mathbb{1}_G$，例如：$\varphi_p(x) = pd(x, G^c) \wedge 1$。我們有

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

等價關係 (2) $\Leftrightarrow$ (3) 不難證明，補集會交換開閉集的角色，以及不等式的方向。

證明 (2)+(3) $\Rightarrow$ (4)。設 $B \in \mathcal{B}(\mathbb{R}^d)$，則我們有

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

但由於 $\mu(\partial B) = 0$ 的假設，我們得到 $\mu(\overline{B}) = \mu(\mathring{B})$，因此

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

最後我們要證明 (4) $\Rightarrow$ (1)。令 $\varphi \in C_b(\mathbb{R}^d)$，我們可以將 φ 正負部份拆解 $\varphi = \varphi^+ - \varphi^-$，因此我們可以假設 φ 是個恆正的函數。由於 φ 有界，令 $K > 0$ 使得 $0 \leqslant \varphi \leqslant K$。根據富比尼定理，我們可以得到

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

設 $E_t^\varphi = \{x \in \mathbb{R}^d : \varphi(x) \geqslant t\}$，則上式可以重新寫作

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

同樣的，對於所有 n，我們有

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

我們可以注意到，$\partial E_t^\varphi \subseteq \{x \in \mathbb{R}^d : \varphi(x) = t\}$，而且根據習題 1.13 ，我們知道最多只有可數多

[2]在 Patrick Billingsley 書《機率測度的收斂》（Convergence of Probability Measures）第一版（1968）中，他說此結果可以追溯到 1940 年 Aleksandrov 的文章；但在第二版（1999）中，卻替此定理掛上了 Jean-Pierre Portmanteau 的名，說出自 1915 年出版在期刊 Annales de l'Université de Felletin 中的文章〈空集合的期望？〉（Espoir pour l'ensemble vide ?），但此人物、大學及文章皆不存在。

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,

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

□

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.

個 t 使得

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

因此根據 (4) 的假設，我們有

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

因此勒貝格收斂定理告訴我們

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

□

問題 4.4.9：考慮實隨機變數 $(X_n)_{n\geqslant 1}$ 及 X，將他們的分佈函數寫作 $(F_{X_n})_{n\geqslant 1}$ 以及 F_X。若且唯若隨機數列序列 $(X_n)_{n\geqslant 1}$ 分佈收斂至 X，則對於所有 F_X 的連續點 x，分佈函數 $F_{X_n}(x)$ 收斂至 $F(x)$。

命題 4.4.10 ： 若 $(X_n)_{n\geqslant 1}$ 機率收斂至 X，則 $(X_n)_{n\geqslant 1}$ 也分佈收列至 X。

證明：首先，假設 $(X_n)_{n\geqslant 1}$ 殆必收斂至 X，此情況下，對於任意 $\varphi \in C_b(\mathbb{R}^d)$，$\varphi(X_n)$ 殆必收斂至 $\varphi(X)$，因此根據勒貝格收斂定理，$\mathbb{E}[\varphi(X_n)] \longrightarrow \mathbb{E}[\varphi(X)]$，所以 $(X_n)_{n\geqslant 1}$ 會分佈收斂至 X。

在一般情況下，我們用反證法。假設 $(X_n)_{n\geqslant 1}$ 不會分佈收斂至 X，那麼可以找到 $\varphi \in C_b(\mathbb{R}^d)$ 使得 $\mathbb{E}[\varphi(X_n)]$ 不會收斂到 $\mathbb{E}[\varphi(X)]$。令 $\varepsilon > 0$ 以及子序列 $(n_k)_{k\geqslant 1}$ 使得對於所有 k，我們有

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

根據命題 4.1.3 ，我們可以找到子序列 $(X_{n_{k_l}})_{l\geqslant 1}$ 殆必收斂至 X，但我們前半段的論證告訴我們，這是矛盾的。□

命題 4.4.11 ： 證明若 $(X_n)_{n\geqslant 1}$ 分佈收斂至 X，且 X 殆必為常數，則 (X_n) 也機率收斂至 X。

證明：參照習題 4.15 。□

我們定義 $C_c(\mathbb{R}^d)$ 為由連續且緊緻支撐 (compactly supported) 的函數構成的集合。

命題 4.4.12 ： 令 (μ_n) 及 μ 為 $\mathbb{R}^d$ 上的機率測度。令 H 為賦範空間 $(C_b(\mathbb{R}^d), \|\cdot\|_\infty)$ 的子集合，並假設其閉包 (closure) 包含 $C_c(\mathbb{R}^d)$。則下列性質等價：

(i) 機率分佈序列 (μ_n) 弱收斂至 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) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iii). Now, we prove (ii) $\Rightarrow$ (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,

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

Hence, for all k, we have,

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

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,

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

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

(ii) 我們有

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

(iii) 我們有

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

證明：由於 $C_c(\mathbb{R}^d) \subseteq C_b(\mathbb{R}^d)$ 以及 $H \subseteq C_b(\mathbb{R}^d)$，(i) $\Rightarrow$ (ii) 以及 (i) $\Rightarrow$ (iii) 是不用證明的。現在我們證明 (ii) $\Rightarrow$ (i)。考慮有界連續函數 $\varphi \in C_b(\mathbb{R}^d)$ 以及 $C_c(\mathbb{R}^d)$ 中的函數序列 (f_k) 使得 $0 \leqslant f_k \leqslant 1$ 且 $\lim \uparrow f_k = 1$，則對於任意 k，我們有 $\varphi f_k \in C_c(\mathbb{R}^d)$，所以

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

此外，我們有

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

因此對於所有 k，我們可以得到

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

由於上式對任意 k 皆成立，我們可以取 $k \to \infty$ 以得到

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

接著我們證明 (iii) $\Rightarrow$ (ii)。令 $\varphi \in C_c(\mathbb{R}^d)$，利用 H 的稠密性，對於所有 $k \geqslant 1$，我們可以找到 $\varphi_k \in H$ 使得 $\|\varphi - \varphi_k\| \leqslant 1/k$，因此，對於所有 k，我們有

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

由於 k 可以任意大，因此我們得到 $\int \varphi \,\mathrm{d}\mu_n \longrightarrow \int \varphi \,\mathrm{d}\mu$。 □

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)

註解 4.4.13 ： 對於 $\mathbb{R}^d$ 上的機率測度序列 $(\mu_n)_{n\geqslant 1}$ 及測度 μ，若

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

則我們說 μ_n 會淡收斂 (converges vaguely) 至 μ。從命題 4.4.12 我們可以得知，在已知 μ 也是個機率測度時，弱收斂與淡收斂的概念是等價的；但如果少了 μ 的總質量為 1 的假設，透過 Fatou 引理，其實我們只會得到 $\mu(\mathbb{R}^d) \leqslant 1$。

我們可以考慮下面的例子：

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

這代表著 μ_n 會淡收斂至 0，但顯然 0 不會是個機率測度，因為總質量為 0。主要的原因在於，當我們的測試函數只能夠選擇 $C_c(\mathbb{R}^d)$ 中的函數時，可能會發生機率質量「跑到無窮遠」的情況，這是無法被 $C_c(\mathbb{R}^d)$ 中的函數所捕捉到的；但在 $C_b(\mathbb{R}^d)$ 中的測試函數是有辦法捕捉到此現象的。此直覺也告訴我們，為什麼在已知 μ 是個機率測度時，若收斂與但收斂的概念會等價的原因。

問題 4.4.14： 給定在 $\mathbb{R}^d$ 上的機率測度序列 $(\mu_n)_{n\geqslant 1}$。若對於所有 $\varepsilon > 0$，存在緊緻集合 $K_\varepsilon \subseteq \mathbb{R}^d$ 使得

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

則我們說 $(\mu_n)_{n\geqslant 1}$ 是個緊密 (tight) 的序列。給定在 $\mathbb{R}^d$ 上的測度 μ，證明下列性質等價：

(1) μ_n 會弱收斂至 μ。

(2) μ_n 會淡收斂至 μ 且 $(\mu_n)_{n\geqslant 1}$ 是緊密的。

定理 4.4.15 【Lévy 連續定理】 ： 令 $(\mu_n)_{n\geqslant 1}$ 為 $\mathbb{R}^d$ 上的機率測度序列，若且唯若 $(\mu_n)_{n\geqslant 1}$ 弱收斂至 μ，則

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

相同的，若且唯若隨機序列 $(X_n)_{n\geqslant 1}$ 分佈收斂至 X，則

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

證明： 我們只需要證明此定理的第一部份。若 $(\mu_n)_{n\geqslant 1}$ 弱收斂至 μ，根據弱收斂本身的定義，由於對於任意固定的 $\xi \in \mathbb{R}$，函數 $x \mapsto \cos(\xi x)$ 和 $x \mapsto \sin(\xi x)$ 都是有界的，我們得到

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

接著證明其逆定理：假設對於所有 $\xi \in \mathbb{R}^d$，我們有 $\widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi)$，我們要證明機率測度序列 $(\mu_n)_{n\geqslant 1}$ 的弱收斂。我們想要使用命題 4.4.12 中的 (3)。為了簡化證明，我們假設 $d = 1$ 的情況。

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 $X_1(\omega), \dots, X_n(\omega)$ (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,

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

令 $f \in C_c(\mathbb{R})$，對於所有 $\sigma > 0$，設

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

在定理 2.4.15 的證明中，根據式 (2.10)，我們知道 $g_\sigma * f$ 簡單收斂至 f；再來根據問題 2.4.16 ，由於 f 是緊緻支撐的，$g_\sigma * f$ 會均勻收斂至 f。若我們設

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

則 $C_c(\mathbb{R}) \subseteq \overline{H}$，所以只需要證明

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

根據定理 2.4.15 證明中的式 (2.8) 及式 (2.9)，對於任意 $\mathbb{R}$ 上的機率測度 ν，我們有

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

根據假設，對於所有 $\xi \in \mathbb{R}$，$\widehat{\mu}_n(\xi) \longrightarrow \widehat{\mu}(\xi)$，且 $|\widehat{\mu}_n(\xi)| \leqslant 1$，因此勒貝格收斂定理給出

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

由於上式左方也滿足 $|\cdot| \leqslant 1$，我們再次使用勒貝格收斂定理，且得證：

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

□

第五節 分佈收斂的應用

第一小節 經驗測度的收斂

令 $(X_n)_{n\geqslant 1}$ 為值域為 $\mathbb{R}^d$ 的 i.i.d. 隨機變數序列。我們可以把這些隨機變數當作在一連串相同且獨立的隨機試驗中，所觀察到的值，而在統計上，我們希望能夠透過所觀察到的 $X_1(\omega), \dots, X_n(\omega)$（$\omega$ 是機率空間中選定的一個點），求 X_1 的分佈。

拿民調來當例子，設 N 為台灣的人口數，編號 i 的台灣人有自己的向量 $a(i) \in \mathbb{R}^d$，代表的是這個人的資料，例如年紀、收入、健康狀況、政治傾向等等，當我們給定一個可測集合 $A \in \mathcal{B}(\mathbb{R}^d)$（例如：年收入百萬以上五十歲以上的男性韓粉），我們想要知道有多少比例的台灣人，參數向量 $a(i)$ 會落在這個可測集合中。換句話說，我們想要估計

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

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 $Y_1, \dots, Y_n$ with values in $\{1, \dots, N\}$ (i.e., choose n Taiwanese uniformly at random) and write $X_j = a(Y_j)$，$X_1, \dots, X_n$ 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,

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

當 N 很大的時候，要準確的計算這個值是不可能的，而民調的目的也在於如何選出有代表性的 n 個台灣人，使得我們能夠正確的估計 $\mu(A)$。若我們取獨立且值域為 $\{1, \dots, N\}$ 的均勻變數 $Y_1, \dots, Y_n$（也就是均勻隨機取 n 個台灣人），並寫 $X_j = a(Y_j)$，$X_1, \dots, X_n$ 為 i.i.d. 隨機變數，此分佈是由下列性質所決定的：

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

根據這樣篩選出來的樣本，我們得到的估計值會是

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

回到原本的問題，我們想要知道這樣的估計值和理論上的值 $\mu(A)$ 是否相近，也就是說，我們想要知道定義如下的經驗測度是否在當 n 趨近無窮大時會逼近 $\mathbb{P}_{X_1}$：

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

下列定理給了我們肯定的答案。

定理 4.5.1 ： 令 $(X_n)_{n\geqslant 1}$ 為值域為 $\mathbb{R}^d$ 的 i.i.d. 隨機變數序列。對於所有 $\omega \in \Omega$ 以及 $n \geqslant 1$，定義 $\mu_{n,\omega}$ 為下列在 $\mathbb{R}^d$ 上的經驗測度 (emperical measure)：

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

則當 $n \to \infty$ 時，我們有下列收斂結果：

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

註解 4.5.2 ： 此定理並沒有給出任何收斂速度，因此我們並不知道 $\mu_{n,\omega}$ 會以多快的速度收斂至 $\mathbb{P}_{X_1}$。

證明：令 H 為可數且在 $C_c(\mathbb{R}^d)$ 中稠密的子集合。若 $\varphi \in H$，根據強大數法則，我們有

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

上式可以重新寫作

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

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

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

由於 H 可數，我們可以得到

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

根據命題 4.4.12 ，得證。□

第二小節 中央極限定理

在定理 4.3.1 中，我們得到了強大數法則：若 $(X_n)_{n\geqslant 1}$ 為 i.i.d. 可積的隨機變數序列，則我們有下列殆必收斂結果

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

下一個我們可以探討的是，上述式子收斂的速度，也就是探討下列項式在 n 趨近無窮大時，與 n 的關係：

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

我們可以先做個簡單的計算：若假設 X_i 皆是平方可積的，那麼我們可以注意到

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

也就是說，$(X_1+\cdots+X_n - n\,\mathbb{E}[X_1])^2$ 與 n 是線性的關係，換句話說，式 (4.4) 和 $\frac{1}{\sqrt{n}}$ 是同等級大小的。

下面，首先我們先討論一維的中央極限定理 (central limit theorem)，高維的版本會在第 4.5.3 小節中有簡單的敘述討論。

> **定理 4.5.3** 【中央極限定理 (Central limit theorem)】 **：** 令 $(X_n)_{n\geqslant 1}$ 為 i.i.d. 實隨機變數序列，且假設其各項皆為平方可積。令 $\sigma^2 = \mathrm{Var}(X_1)$，則我們有
>
> $$\frac{1}{\sqrt{n}}(X_1+\cdots+X_n - n\,\mathbb{E}[X_1]) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2).$$
>
> 換句話說，對於所有 $a,b \in \overline{\mathbb{R}}$ 滿足 $a<b$，我們有下列收斂
>
> $$\lim_{n\to\infty} \mathbb{P}\left(n\,\mathbb{E}[X_1] + a\sqrt{n} \leqslant X_1+\cdots+X_n \leqslant n\,\mathbb{E}[X_1] + b\sqrt{n}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) \mathrm{d}x.$$

證明：定理後半部的敘述是前半部的直接結果（問題 4.4.9 、習題 4.13 ），因此我們只需要證明第一部份。此外，我們可以將 X_n 用 $X_n - \mathbb{E}[X_n]$ 取代，因此可以假設 $\mathbb{E}[X_n]=0$。令

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

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,

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

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

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,

我們想用定理 4.4.15 來證明此定理，隨機變數 Z_n 的特徵函數寫作

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

命題 2.4.17 中的展開式給出，當 $\xi \longrightarrow 0$，

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

因此對任意給定的 $\xi \in \mathbb{R}$，當 $n \longrightarrow \infty$，我們有

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

因此我們得到，

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

其中 U 有著 $\mathcal{N}(0,\sigma^2)$ 的分佈。因此透過定理 4.4.15 ，我們完成中央極限定理的證明。 □

第三小節 高維度的中央極限定理

假設我們有 i.i.d. 可積且值域為 $\mathbb{R}^d$ 的隨機變數序列 $(X_n := (X_n^{(1)},\dots,X_n^{(d)}))_{n\geqslant 1}$，針對每個分量 $X_n^{(i)}$，我們都可以得到大數法則並且推得

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

其中 $\mathbb{E}[X_1]$ 為各個分量的期望值構成的向量。若 $(X_n)_{n\geqslant 1}$ 為平方可積，我們同樣可以針對每個分量去得到相對應的中央極限定理，但是這樣的方法並不足以得到 d 維向量的中央極限定理，因為邊緣分佈不足以決定整個隨機向量的分佈。事實上，要描述高維度的中央極限定理，我們需要我們在第 3.4.1 小節中定義的多元常態分佈。

定理 4.5.4 【高維度的中央極限定理】 ： 令 $(X_n)_{n\geqslant 1}$ 為值域為 $\mathbb{R}^d$ 的 i.i.d. 隨機變數序列，且假設他們皆為平方可積。則我們有

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

證明： 此證明與在一維中相同。不失一般性，假設 $\mathbb{E}[X_1]=0$。對於所有 $\xi \in \mathbb{R}^d$，我們有

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

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,

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

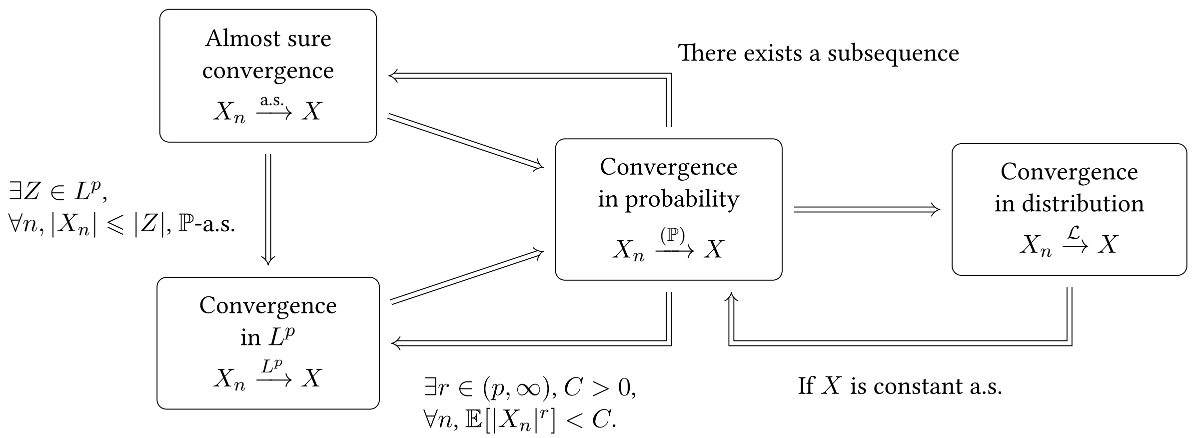
Hence,

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

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.



同時，我們也有

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

因此，

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

最後，我們使用 Lévy 連續定理（定理 4.4.15）總結。□

第六節 總結

最後，我們用下列一張圖來總結這章探討過的各種收斂概念，以及不同收斂概念之間的關係。