

# Notes on Probability Theory

Kai-Jyun Wang\*

Fall 2025

The notes are based on the lecture of Prof. David Anderson at University of Wisconsin-Madison in 2025-2026. The course structure mainly follows Durrett. The course assumes a certain amount of knowledges in real analysis. For some classic results in real analysis, one can refer to my notes on real analysis.

---

\*National Taiwan University, Department of Economics.

# Contents

<b>1 Probability Space</b>	<b>3</b>
1.1 Probability Space . . . . .	3
1.2 Random Variable . . . . .	12
1.3 Independence . . . . .	18
1.4 Convergence of Random Variables . . . . .	23
<b>2 Asymptotic Theory</b>	<b>26</b>
2.1 Law of Large Number . . . . .	26
2.2 Convergence in Distribution . . . . .	41
2.3 Characteristic Functions . . . . .	50
2.4 Central Limit Theorems . . . . .	57
<b>3 Conditional Expectation and Martingale</b>	<b>62</b>
3.1 Conditional Expectation . . . . .	62
<b>4 Stochastic Processes</b>	<b>65</b>
4.1 Poisson Process . . . . .	65

# 1. Probability Space

## 1.1. Probability Space

### Definition 1.1

Let  $\Omega$  be a set. A collection of subsets  $\mathcal{F}$  forms a  **$\sigma$ -algebra** if

- (a)  $\emptyset \in \mathcal{F}$ .
- (b)  $A \in \mathcal{F}$  implies  $A^c \in \mathcal{F}$ .
- (c) If  $A_i \in \mathcal{F}$  are countably many sets,  $\cup_i A_i \in \mathcal{F}$ .

The dual  $(\Omega, \mathcal{F})$  is called a **measurable space** and the sets falling in  $\mathcal{F}$  are said to be **measurable**.

### Definition 1.2

Let  $(\Omega, \mathcal{F})$  be a measurable space. A set function  $\mu : \mathcal{F} \rightarrow [0, \infty]$  is a **measure** if

- (a)  $\mu(\emptyset) = 0$ .
- (b) For countably many disjoint  $A_i \in \mathcal{F}$ ,  $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ .

The triple  $(\Omega, \mathcal{F}, \mu)$  is called a **measure space**.

### Definition 1.3

A **probability space** is a measure space  $(\Omega, \mathcal{F}, P)$  such that  $P(\Omega) = 1$ .

### Lemma 1.4

Let  $\mathcal{S}$  be a collection of sets. Then there exists the smallest  $\sigma$ -algebra containing  $\mathcal{S}$ .

*Proof.* Let  $\mathcal{F}$  be the intersection of all  $\sigma$ -algebra containing  $\mathcal{S}$ .  $\mathcal{F}$  is non-empty since the power set is a  $\sigma$ -algebra containing  $\mathcal{S}$ . Now it is clear that  $\emptyset \in \mathcal{F}$  since  $\emptyset \in \mathcal{A}$  for every  $\sigma$ -algebra  $\mathcal{A}$  containing  $\mathcal{S}$ . If  $A \in \mathcal{F}$ ,  $A \in \mathcal{A}$  for all  $\mathcal{A}$  containing  $\mathcal{S}$  and  $A^c \in \mathcal{A}$  for all  $\mathcal{A}$ . Thus  $A^c \in \mathcal{F}$ . Finally, if  $A_i \in \mathcal{F}$  are countably many sets, then each  $A_i$  lies in every  $\mathcal{A}$  containing  $\mathcal{S}$ ; so does  $\cup_i A_i$  and thus  $\cup_i A_i \in \mathcal{F}$ . The minimality follows by the construction of  $\mathcal{F}$ . ■

### Definition 1.5

For any collection of sets  $\mathcal{S}$ , the smallest  $\sigma$ -algebra is denoted as  $\sigma(\mathcal{S})$ .

### Theorem 1.6

Let  $(\Omega, \mathcal{F}, P)$  be a probability space. Then

- (a) If  $A, B \in \mathcal{F}$  and  $A \subset B$ , then  $P(A) \leq P(B)$ .
- (b) For countably many  $A_i \in \mathcal{F}$ ,  $P(\cup_i A_i) \leq \sum_i P(A_i)$ .
- (c) If  $A_i \nearrow A$ ,  $P(A_i) \rightarrow P(A)$ .
- (d) If  $A_i \searrow A$ ,  $P(A_i) \rightarrow P(A)$ .

*Proof.* (a) and (b) are clear. For (c), write  $E_i = A_i - A_{i-1}$  and  $A_0 = \emptyset$ . Then since  $E_i$  are disjoint and  $A_n = \cup_{i=1}^n E_i$ ,

$$P(A_n) = P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) \rightarrow \sum_i P(E_i) = P(\cup_i E_i) = P(A)$$

as  $n \rightarrow \infty$ .

For (d), note that  $A_i^c \nearrow A^c$ . Thus  $1 - P(A_i) = P(A_i^c) \rightarrow P(A^c) = 1 - P(A)$ . Thus  $P(A_i) \rightarrow P(A)$ . ■

### Definition 1.7

The **Borel  $\sigma$ -algebra** is the  $\sigma$ -algebra generated by all open sets.

### Definition 1.8

Let  $P$  be a probability measure on  $(\mathbb{R}, \mathcal{B})$ . The **distribution function**  $F$  is defined as

$$F(x) = P((-\infty, x])$$

for  $x \in \mathbb{R}$ .

### Proposition 1.9

The distribution function in  $(\mathbb{R}, \mathcal{B})$  satisfies that

- (a)  $F(x) \leq F(y)$  for all  $x \leq y$ .
- (b)  $F(x) \rightarrow F(y)$  as  $x \rightarrow y^+$ .
- (c)  $F(-\infty) = 0$  and  $F(\infty) = 1$ .

*Proof.* For (a), note that  $(-\infty, x] \subset (-\infty, y]$  and

$$F(x) = P((-\infty, x]) \leq P((-\infty, y]) = F(y).$$

For (b), notice that for  $x_n \rightarrow y^+$ ,  $(-\infty, x_n] \searrow (-\infty, y]$ . Hence

$$F(x_n) = P((-\infty, x_n]) \rightarrow P((-\infty, y]) = F(y).$$

Similarly, taking  $x_n \rightarrow \pm\infty$  gives (c). ■

### Definition 1.10

A collection  $\mathcal{S}$  of sets is called an **algebra** if

- (a)  $\emptyset \in \mathcal{S}$ .
- (b) If  $A \in \mathcal{S}$ , then  $A^c \in \mathcal{S}$ .
- (c) If  $A, B \in \mathcal{S}$ , then  $A \cup B \in \mathcal{S}$ .

### Remark

An algebra is closed under finite unions. It is also clear that a  $\sigma$ -algebra is an algebra, while the converse is not true. An example is the collection of all finite unions of intervals in  $\mathbb{R}$ .

### Definition 1.11

A collection  $\mathcal{S}$  of sets is called a **semi-algebra** if

- (a) If  $A, B \in \mathcal{S}$ , then  $A \cap B \in \mathcal{S}$ .
- (b) If  $A \in \mathcal{S}$ , then  $A^c$  can be written as a finite disjoint union of sets in  $\mathcal{S}$ .

### Remark

A semi-algebra must contain  $\emptyset$  since for any  $A \in \mathcal{S}$ ,  $A^c = \cup_i A_i$ , where  $A_i \in \mathcal{S}$  are disjoint. Then  $A \cap A_1 = \emptyset \in \mathcal{S}$ .

### Remark

An example of being a semi-algebra but not an algebra is the collection of all intervals of the form  $(a_i, b_i]$  for  $-\infty \leq a_i < b_i \leq \infty$  with the empty set.

### Lemma 1.12

If  $\mathcal{S}$  is a semi-algebra, then  $\bar{\mathcal{S}} = \{\text{finite disjoint unions of sets in } \mathcal{S}\}$  forms an algebra.

*Proof.* It has been shown that  $\emptyset \in \mathcal{S}$ . For  $A, B \in \bar{\mathcal{S}}$ , write  $A = \cup_{i=1}^n A_i$  and  $B = \cup_{j=1}^m B_j$  for disjoint  $A_i, B_j \in \mathcal{S}$ , respectively. Then  $A \cap B = \cup_{i,j} (A_i \cap B_j) \in \bar{\mathcal{S}}$ . Thus  $\bar{\mathcal{S}}$  is closed under intersection. Now if  $A \in \bar{\mathcal{S}}$ ,  $A = \cup_{i=1}^n A_i$  for disjoint  $A_i \in \mathcal{S}$ . Then  $A^c = \cap_{i=1}^n A_i^c$ . By the definition of semi-algebra,  $A_i^c$  can be written as a finite disjoint union of sets in  $\mathcal{S}$  and thus  $A_i^c \in \bar{\mathcal{S}}$ . Since  $\bar{\mathcal{S}}$  is closed under finite intersection,  $A^c = \cap_{i=1}^n A_i^c \in \bar{\mathcal{S}}$ . Finally, for  $A, B \in \bar{\mathcal{S}}$ ,  $A \cup B = (A^c \cap B^c)^c \in \bar{\mathcal{S}}$ . We conclude that  $\bar{\mathcal{S}}$  is indeed an algebra. ■

### Definition 1.13

Suppose  $\mathcal{S}$  is a semi-algebra.  $\bar{\mathcal{S}} = \{\text{finite disjoint unions of sets in } \mathcal{S}\}$  is called the **algebra generated by  $\mathcal{S}$** .

### Definition 1.14

Let  $\mathcal{S}$  be an algebra. A set function  $\mu_0 : \mathcal{S} \rightarrow [0, \infty]$  is called a **premeasure** if

- (a)  $\mu_0(\emptyset) = 0$ .
- (b) For countable disjoint  $A_i \in \mathcal{S}$  such that  $\cup_i A_i \in \mathcal{S}$ ,

$$\mu_0(\cup_i A_i) = \sum_i \mu_0(A_i).$$

### Theorem 1.15

Let  $\nu$  be a set function on a semi-algebra  $\mathcal{S}$  such that  $\nu(\emptyset) = 0$ . Suppose that

- (a) if  $A \in \mathcal{S}$  and  $A = \cup_{i=1}^n A_i$  for disjoint  $A_i \in \mathcal{S}$ , then  $\nu(A) = \sum_{i=1}^n \nu(A_i)$ ;
- (b) if  $A_i \in \mathcal{S}$  are countably many sets and  $A = \cup_i A_i \in \mathcal{S}$ , then  $\nu(A) \leq \sum_i \nu(A_i)$ .

Then  $\nu$  can be extended to a unique premeasure  $\mu_0$  on the algebra generated by  $\mathcal{S}$ .

*Proof.* We first show the existence. From [lemma 1.12](#) we know that  $\mathcal{S}$  generates an algebra  $\mathcal{A} = \{\text{finite disjoint union of sets in } \mathcal{S}\}$ . Define our candidate  $\mu_0$  by  $\mu_0(A) = \sum_i \nu(A_i)$  for

$A = \cup_i A_i$  where  $A_i \in \mathcal{S}$  are disjoint. To see that  $\mu_0$  is well-defined, suppose  $A = \cup_i B_i$  for disjoint  $B_i \in \mathcal{S}$ . Observe that

$$A_i = \cup_j (A_i \cap B_j) \quad \text{and} \quad B_j = \cup_i (A_i \cap B_j)$$

are finite disjoint unions. Then

$$\sum_i \nu(A_i) = \sum_i \sum_j \nu(A_i \cap B_j) = \sum_j \sum_i \nu(A_i \cap B_j) = \sum_j \nu(B_j)$$

by (a). Thus  $\mu_0$  is well-defined.

Now we check that  $\mu_0$  is a premeasure. Clearly  $\mu_0(\emptyset) = 0$ . For finitely many disjoint  $A_i \in \mathcal{A}$  such that  $\cup_i A_i \in \mathcal{A}$ , we can write  $A_i = \cup_j B_{ij}$  for disjoint  $B_{ij} \in \mathcal{S}$ . Then (a) implies that

$$\mu_0(\cup_i A_i) = \mu_0(\cup_{i,j} B_{ij}) = \sum_{i,j} \nu(B_{ij}) = \sum_i \sum_j \mu_0(B_{ij}) = \sum_i \mu_0(A_i).$$

Next, for countably many disjoint  $A_i \in \mathcal{A}$  such that  $A = \cup_i A_i \in \mathcal{A}$ , write  $A_i = \cup_j B_{ij}$ , where  $B_{ij} \in \mathcal{S}$  are finite disjoint for each  $i$ . Then  $\mu_0(A_i) = \sum_j \nu(B_{ij})$  and

$$\sum_i \mu_0(A_i) = \sum_i \sum_j \nu(B_{ij}).$$

Without loss of generality, we may choose  $A_i$  to be those in  $\mathcal{S}$  since otherwise we can replace  $A_i$  by  $B_{ij}$ . We assume that  $A_i \in \mathcal{S}$  from now on. Since  $A \in \mathcal{A}$ ,  $A = \cup_i C_i$  for finite disjoint  $C_i \in \mathcal{S}$ .  $C_i = \cup_j (C_i \cap A_j)$ . Thus (b) gives that

$$\nu(C_i) \leq \sum_j \nu(C_i \cap A_j).$$

Then

$$\mu_0(A) = \sum_i \nu(C_i) \leq \sum_i \sum_j \nu(C_i \cap A_j) = \sum_j \sum_i \nu(C_i \cap A_j) = \sum_j \nu(A_j) = \sum_j \mu_0(A_j).$$

For the opposite inequality, set  $B_n = \cup_{i=1}^n A_i$  and  $C_n = A - B_n$ . Since  $\mathcal{A}$  is an algebra,  $C_n \in \mathcal{A}$  and the finite additivity shows that

$$\mu_0(A) = \sum_{i=1}^n \mu_0(A_i) + \mu_0(C_n) \geq \sum_{i=1}^n \mu_0(A_i).$$

Taking  $n \rightarrow \infty$  gives the desired inequality and thus  $\mu_0$  is  $\sigma$ -additive on  $\mathcal{A}$ .

Finally, if  $\mu_1$  is another premeasure on  $\mathcal{A}$  extending  $\nu$ , then for  $A = \cup_i A_i$  for disjoint  $A_i \in \mathcal{S}$ ,

$$\mu_1(A) = \sum_i \nu(A_i) = \mu_0(A).$$

■

### Definition 1.16

A collection of sets  $\mathcal{P}$  is called a  **$\pi$ -system** if  $A, B \in \mathcal{P}$  implies  $A \cap B \in \mathcal{P}$ .

### Definition 1.17

A collection of sets  $\mathcal{L}$  is called a  **$\lambda$ -system** if

- (a)  $\Omega \in \mathcal{L}$ .
- (b) If  $A, B \in \mathcal{L}$  and  $A \subset B$ , then  $B - A \in \mathcal{L}$ .
- (c) If  $A_i \in \mathcal{L}$  and  $A_i \nearrow A$ , then  $A \in \mathcal{L}$ .

### Theorem 1.18 (Sierpiński-Dynkin $\pi$ - $\lambda$ )

If  $\mathcal{P}$  is a  $\pi$ -system and  $\mathcal{L}$  is a  $\lambda$ -system containing  $\mathcal{P}$ , then  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .

*Proof.* First we show that a collection  $\mathcal{S}$  is a  $\sigma$ -algebra if and only if it is both a  $\pi$ -system and a  $\lambda$ -system. Suppose first that  $\mathcal{S}$  is a  $\pi$ -system and a  $\lambda$ -system.  $\emptyset = \Omega - \Omega \in \mathcal{S}$ . If  $A \in \mathcal{S}$ , then  $A^c = \Omega - A \in \mathcal{S}$ . For  $A, B \in \mathcal{S}$ ,  $A \cup B = (A^c \cap B^c)^c \in \mathcal{S}$  since we have shown that  $\mathcal{S}$  is closed under complement and intersection by being a  $\pi$ -system. Thus  $\mathcal{S}$  is also closed under finite unions. If  $A_i \in \mathcal{S}$  are countably many sets, let  $B_n = \bigcup_{i=1}^n A_i \in \mathcal{S}$ . Then  $B_n \nearrow \bigcup_i A_i$  and thus  $\bigcup_i A_i \in \mathcal{S}$ .

Conversely, if  $\mathcal{S}$  is a  $\sigma$ -algebra, then for  $A, B \in \mathcal{S}$ ,  $A \cap B = (A^c \cup B^c)^c \in \mathcal{S}$ . Thus  $\mathcal{S}$  is a  $\pi$ -system. If  $A, B \in \mathcal{S}$  and  $A \subset B$ , then  $B - A = B \cap A^c \in \mathcal{S}$ . Finally, if  $A_i \in \mathcal{S}$  and  $A_i \nearrow A$ , then  $A = \bigcup_i (A_i - A_{i-1}) \in \mathcal{S}$  with  $A_0 = \emptyset$ . Thus  $\mathcal{S}$  is a  $\lambda$ -system.

Now set  $\mathcal{L}$  to be the smallest  $\lambda$ -system containing  $\mathcal{P}$ . It suffices to show that  $\mathcal{L}$  is also a  $\pi$ -system and thus by the above conclusion,  $\mathcal{L}$  is a  $\sigma$ -algebra containing  $\mathcal{P}$ ; hence  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .

To show that  $\mathcal{L}$  is a  $\pi$ -system, let  $A, B \in \mathcal{L}$ . If  $A, B \in \mathcal{P}$ ,  $A \cap B \in \mathcal{P} \subset \mathcal{L}$ . To extend the result for general  $A, B \in \mathcal{L}$ , we first fix  $B \in \mathcal{P}$  and define

$$\mathcal{L}_B = \{A \mid A \cap B \in \mathcal{L}\}.$$

We claim that  $\mathcal{L}_B$  is a  $\lambda$ -system containing  $\mathcal{P}$ . For  $A \in \mathcal{P}$ ,  $A \cap B \in \mathcal{L}$ . Thus  $\mathcal{P} \subset \mathcal{L}_B$ . Clearly  $\Omega \in \mathcal{L}_B$ . If  $E, F \in \mathcal{L}_B$  and  $E \subset F$ , then

$$(F - E) \cap B = (F \cap B) - (E \cap B) \in \mathcal{L}.$$

Thus  $F - E \in \mathcal{L}_B$ . Finally, if  $E_i \in \mathcal{L}_B$  and  $E_i \nearrow E$ , then

$$E \cap B = \bigcup_i (E_i \cap B) \in \mathcal{L}.$$

Hence  $E \in \mathcal{L}_B$  and we conclude that  $\mathcal{L}_B$  is a  $\lambda$ -system. Since it is a  $\lambda$ -system containing  $\mathcal{P}$ , it also contains the smallest  $\lambda$ -system  $\mathcal{L}$  with the intersection property. Thus  $A \cap B \in \mathcal{L}$  whenever  $A \in \mathcal{L}$  and  $B \in \mathcal{P}$ .

Next, fix  $A \in \mathcal{L}$  and define  $\mathcal{L}_A = \{B \mid A \cap B \in \mathcal{L}\}$ . Clearly  $\mathcal{L}_A$  contains  $\mathcal{L}$  and  $\Omega \in \mathcal{L}_A$ . If  $E, F \in \mathcal{L}_A$  and  $E \subset F$ , then

$$(F - E) \cap A = (F \cap A) - (E \cap A) \in \mathcal{L}.$$

Thus  $F - E \in \mathcal{L}_A$ . Finally, if  $E_i \in \mathcal{L}_A$  and  $E_i \nearrow E$ , then

$$E \cap A = \cup_i(E_i \cap A) \in \mathcal{L}.$$

Hence  $E \in \mathcal{L}_A$  and we conclude that  $\mathcal{L}_A$  is a  $\lambda$ -system. Since it contains  $\mathcal{L}$ ,  $A, B \in \mathcal{L}$  implies  $A \cap B \in \mathcal{L}$ ; in other words,  $\mathcal{L}$  is a  $\pi$ -system and the proof is complete. ■

### Corollary 1.19

*Let  $\mu$  and  $\nu$  be two probability measures agreeing on a  $\pi$ -system  $\mathcal{P}$ , i.e.,  $\mu(A) = \nu(A)$  for all  $A \in \mathcal{P}$ . Then  $\mu(A) = \nu(A)$  for all  $A \in \sigma(\mathcal{P})$ .*

*Proof.* Put

$$\mathcal{L} = \{A \mid \mu(A) = \nu(A)\}.$$

We claim that  $\mathcal{L}$  is a  $\lambda$ -system containing  $\mathcal{P}$ . It is clear that by our assumption,  $\mathcal{P} \subset \mathcal{L}$  and  $\Omega \in \mathcal{L}$ . If  $A, B \in \mathcal{L}$  and  $A \subset B$ , then

$$\mu(B - A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B - A).$$

Thus  $B - A \in \mathcal{L}$ . Finally, if  $A_i \in \mathcal{L}$  and  $A_i \nearrow A$ , then

$$\mu(A) = \lim_{i \rightarrow \infty} \mu(A_i) = \lim_{i \rightarrow \infty} \nu(A_i) = \nu(A).$$

Hence  $A \in \mathcal{L}$  and we conclude that  $\mathcal{L}$  is a  $\lambda$ -system. By the Sierpiński-Dynkin  $\pi$ - $\lambda$  theorem,  $\sigma(\mathcal{P}) \subset \mathcal{L}$ ; in other words,  $\mu$  and  $\nu$  agree on  $\sigma(\mathcal{P})$ . ■

### Definition 1.20

*A measure  $\mu$  on a measurable space  $(\Omega, \mathcal{F})$  is called  **$\sigma$ -finite** if there exists countable  $A_i \in \mathcal{F}$  such that  $\cup_i A_i = \Omega$  and  $\mu(A_i) < \infty$ .*

### Definition 1.21

*A set function  $\mu^* : 2^\Omega \rightarrow [0, \infty]$  is called an **outer measure** if*

- (a)  $\mu^*(\emptyset) = 0$ .
- (b) If  $A \subset B$ , then  $\mu^*(A) \leq \mu^*(B)$ .
- (c) For countably many  $A_i \subset \Omega$ ,  $\mu^*(\cup_i A_i) \leq \sum_i \mu^*(A_i)$ .

### Definition 1.22

*Let  $\mu^*$  be an outer measure. A set  $A \subset \Omega$  is said to be **Carathéodory measurable** or  $\mu^*$ -*

**measurable** if for all  $E \subset \Omega$ ,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

**Lemma 1.23**

Let  $\mu^*$  be an outer measure on  $\Omega$ . Then the collection of all  $\mu^*$ -measurable sets forms a  $\sigma$ -algebra  $\mathcal{F}$  and  $\mu^*|_{\mathcal{F}}$  is a measure.

*Proof.* Put

$$\mathcal{F} = \{A \subset \Omega \mid \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset \Omega\}.$$

We first show that  $\mathcal{F}$  is a  $\sigma$ -algebra. Clearly  $\emptyset \in \mathcal{F}$  and if  $A \in \mathcal{F}$ , then  $A^c \in \mathcal{F}$ . For  $A, B \in \mathcal{F}$ , let  $C = A \cup B$ . The property of outer measure gives that  $\mu^*(E) \leq \mu^*(E \cap C) + \mu^*(E \cap C^c)$ . To see the opposite inequality, note that  $C = A \cup (B \cap A^c)$  and

$$\begin{aligned} \mu^*(E \cap C) + \mu^*(E \cap C^c) &\leq \mu^*(E \cap A) + \mu^*(E \cap B \cap A^c) + \mu^*(E \cap A^c \cap B^c) \\ &= \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E). \end{aligned}$$

Hence  $C \in \mathcal{F}$  and  $\mathcal{F}$  is closed under finite unions. For countable disjoint  $A_i \in \mathcal{F}$  with  $A = \cup_i A_i$ , let  $B_n = \cup_{i=1}^n A_i \in \mathcal{F}$ . Then

$$\mu^*(E \cap A) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B_n) = \sum_{i=1}^n \mu^*(E \cap A_i).$$

Taking  $n \rightarrow \infty$  gives that

$$\mu^*(E \cap A) \geq \sum_i \mu^*(E \cap A_i) \geq \mu^*(E \cap A)$$

by the  $\sigma$ -subadditivity of outer measure. Hence  $\mu^*(E \cap A) = \sum_i \mu^*(E \cap A_i)$ . Note also that  $E \cap A^c \subset E \cap B_n^c$  so  $\mu^*(E \cap A^c) \leq \mu^*(E \cap B_n^c)$ . Thus

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \sum_{i=1}^n \mu^*(E \cap A_i) + \mu^*(E \cap A^c) \rightarrow \mu^*(E \cap A) + \mu^*(E \cap A^c) \geq \mu^*(E)$$

by the  $\sigma$ -subadditivity of outer measure. We conclude that  $\mathcal{F}$  is a  $\sigma$ -algebra.

Finally, denote  $\mu^*|_{\mathcal{F}}$  by  $\mu$ . Clearly  $\mu(\emptyset) = 0$ . For countably many disjoint  $A_i \in \mathcal{F}$  such that  $A = \cup_i A_i \in \mathcal{F}$ , let  $B_n = \cup_{i=1}^n A_i \in \mathcal{F}$ . Then

$$\mu(A) = \mu(B_n) + \mu(A \cap B_n^c) \geq \mu(B_n) = \sum_{i=1}^n \mu(A_i) \rightarrow \sum_i \mu(A_i) \geq \mu(A).$$

Hence  $\mu(A) = \sum_i \mu(A_i)$  and  $\mu$  is a measure on  $\mathcal{F}$ . ■

**Theorem 1.24** (Carathéodory Extension)

Let  $\nu$  be a finitely additive,  $\sigma$ -subadditive set function on a semi-algebra  $\mathcal{S}$  such that  $\nu(\emptyset) = 0$ . Then  $\nu$  can be extended to a measure on  $\sigma(\mathcal{S})$ .

*Proof.* By theorem 1.15,  $\nu$  can be extended to a premeasure  $\mu_0$  on the algebra  $\mathcal{A}$  generated by  $\mathcal{S}$ . Define the outer measure by

$$\mu^*(A) = \inf \left\{ \sum_i \mu_0(E_i) \mid A \subset \cup_i E_i, E_i \in \mathcal{A} \right\}$$

for all  $A \subset \Omega$  with the convention that  $\inf \emptyset = \infty$ . We check that  $\mu^*$  is indeed an outer measure. Clearly  $\mu^*(\emptyset) = 0$ . If  $A \subset B$ , then any cover of  $B$  by sets in  $\mathcal{A}$  is also a cover of  $A$  and hence  $\mu^*(A) \leq \mu^*(B)$ . For countably many  $A_i \subset \Omega$ , we can find  $\{E_{ij}\}_j$  covering  $A_i$  such that

$$\sum_j \mu_0(E_{ij}) \leq \mu^*(A_i) + 2^{-i}\epsilon$$

for some  $\epsilon > 0$ . Then  $\cup_{i,j} E_{ij}$  covers  $\cup_i A_i$  and

$$\mu^*(\cup_i A_i) \leq \sum_i \sum_j \mu_0(E_{ij}) \leq \sum_i \mu^*(A_i) + \epsilon.$$

Since  $\epsilon$  is arbitrary, we conclude that  $\mu^*(\cup_i A_i) \leq \sum_i \mu^*(A_i)$  and  $\mu^*$  is indeed an outer measure.

It follows from lemma 1.23 that the collection of all  $\mu^*$ -measurable sets forms a  $\sigma$ -algebra  $\mathcal{F}$  and  $\mu^*$  restricted on  $\mathcal{F}$  is a measure. It is clear that  $\mathcal{A} \subset \mathcal{F}$  and  $\sigma(\mathcal{S}) \subset \sigma(\mathcal{A}) \subset \mathcal{F}$  and  $\mu = \mu^*|_{\sigma(\mathcal{S})}$  is also a measure. Finally, for  $A, A_i \in \mathcal{S}$  where  $A_i$  covers  $A$ ,

$$\mu(A) = \mu^*(A) \leq \nu(A) \leq \sum_i \nu(A \cap A_i) \leq \sum_i \nu(A_i).$$

Taking the infimum over all such covers, we get  $\nu(A) = \mu^*(A)$  and  $\mu$  is indeed an extension of  $\nu$ . ■

**Remark**

If the measures are probability measures, then we have that the extension is unique by corollary 1.19.

**Theorem 1.25**

If  $F$  is non-decreasing, right-continuous and satisfies that  $F(-\infty) = 0$ ,  $F(\infty) = 1$ , then there is a unique probability measure such that

$$P((-\infty, x]) = F(x).$$

*Proof.* Define

$$\mathcal{S} = \{(a, b] \mid -\infty \leq a < b \leq \infty\} \cup \{\emptyset\}.$$

It is clear that  $\mathcal{S}$  is a semi-algebra. Define the set function  $P : \mathcal{S} \rightarrow [0, 1]$  by

$$P((a, b]) = F(b) - F(a)$$

and  $P(\emptyset) = 0$ . For disjoint, at most countable  $(a_i, b_i] \in \mathcal{S}$ , we define

$$P(\cup_i (a_i, b_i]) = \sum_i P((a_i, b_i]) = \sum_i F(b_i) - F(a_i).$$

It is clear that  $P$  is finitely additive. If  $(a, b] = \cup_i (a_i, b_i]$  for disjoint  $(a_i, b_i] \in \mathcal{S}$ , we may assume without loss of generality that  $a = a_1 < b_1 < b_2 < \dots < b_n = b$  and

$$P((a, b]) = F(b) - F(a) = \sum_i F(b_i) - F(a_i) = \sum_i P((a_i, b_i]).$$

Hence  $P$  is  $\sigma$ -additive. It now follows from the Carathéodory extension theorem that  $P$  can be extended uniquely to a probability measure on  $\sigma(\mathcal{S}) = \mathcal{B}$ . ■

### **Remark**

*This theorem shows that the distribution function completely characterizes the probability measure. In other words, the term “distribution function” can refer to either the CDF or the probability measure.*

## 1.2. Random Variable

### Definition 1.26

Let  $\Omega$  be a probability space. A **random variable**  $X$  is a measurable function  $X : \Omega \rightarrow (S, \mathcal{S})$ , where  $(S, \mathcal{S})$  is a measurable space.

### Remark

The codomain is often taken to be  $(\mathbb{R}, \mathcal{B})$  or  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , but it is also possible to define random functions, i.e.,  $(S, \mathcal{S})$  is a function space.

### Definition 1.27

Let  $X : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$  be a random variable. The **distribution** of  $X$  is the pushforward measure of  $P$  under  $X$ , i.e.,

$$\mu_X(A) = P(X \in A) = P(X^{-1}(A)), \quad A \in \mathcal{S}.$$

### Definition 1.28

Let  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B})$  be a random variable. The **cumulative distribution function** of  $X$  is defined as

$$F_X(x) = P(X \leq x) = P(X_1 \leq x_1, \dots, X_d \leq x_d), \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

### Proposition 1.29

Let  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$  be a random variable and  $F$  be its cumulative distribution function. Then,

- (a)  $F$  is non-decreasing, i.e.,  $x \leq y$  implies  $F(x) \leq F(y)$ ;
- (b)  $F(-\infty) = 0$  and  $F(\infty) = 1$ ;
- (c)  $F$  is right-continuous, i.e.,  $\lim_{y \rightarrow x^+} F(y) = F(x)$ ;
- (d)  $F(x^-) = P(X < x)$ ;
- (e)  $P(X = x) = F(x) - F(x^-)$ .

*Proof.* (a) comes from that  $\{X \leq x\} \subset \{X \leq y\}$  for  $x \leq y$ .

Take  $a_n \rightarrow \infty$ . Then  $\{X \leq a_n\} \nearrow \Omega$  and  $\{X \leq -a_n\} \searrow \emptyset$ . By theorem 1.6, we have that

$$F(a_n) = P(X \leq a_n) \rightarrow P(\Omega) = 1, \quad F(-a_n) = P(X \leq -a_n) \rightarrow P(\emptyset) = 0.$$

(c) is similar to (b). Take  $y_n \rightarrow x^+$ , then  $\{X \leq y_n\} \searrow \{X \leq x\}$ . By theorem 1.6, we have that

$$F(y_n) = P(X \leq y_n) \rightarrow P(X \leq x) = F(x).$$

For (d), take  $x_n \rightarrow x^-$ , then  $\{X \leq x_n\} \nearrow \{X < x\}$ . By theorem 1.6, we have that

$$F(x_n) = P(X \leq x_n) \rightarrow P(X < x).$$

For (e),  $P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x^-)$ . ■

### Theorem 1.30

Let  $F$  be a non-decreasing, right-continuous function satisfying that  $F(-\infty) = 0$  and  $F(\infty) = 1$ . Then there is a random variable  $X$  such that

$$F(x) = \mu_X((-\infty, x]).$$

*Proof.* Put  $\Omega = (0, 1)$ ,  $\mathcal{F} = \mathcal{B}$ ,  $P$  be the Lebesgue measure and  $X(\omega) = \sup \{x \mid F(x) < \omega\}$ .

Notice that

$$\begin{aligned} \{X \leq x\} &= \{\omega \in \Omega \mid \sup \{y \mid F(y) < \omega\} \leq x\} \\ &= \{\omega \in \Omega \mid \text{for all } y > x, F(y) \geq \omega\} \\ &= \{\omega \in \Omega \mid F(x) \geq \omega\}. \end{aligned}$$

Hence  $P(X \leq x) = P(\{\omega \in \Omega \mid \omega \leq F(x)\}) = F(x)$ . ■

### Definition 1.31

If  $X$  and  $Y$  are random variables mapping to some measurable space  $(S, \mathcal{S})$ , then  $X$  and  $Y$  are said to be **equal in distribution** if  $\mu_X = \mu_Y$ , denoted by  $X \stackrel{d}{=} Y$ .

### Definition 1.32

Let  $X : \Omega \rightarrow \mathbb{R}$  be a random variable with distribution  $F$ .  $f : \mathbb{R} \rightarrow \mathbb{R}$  is said to be the **density** of  $X$  if

$$F(x) = \int_{-\infty}^x f(y) dy$$

for all  $x \in \mathbb{R}$ .

### Remark

If  $f$  and  $g$  are both densities of  $X$ , then  $f = g$  a.e.

### Remark

If  $\mu_X \ll \lambda$ , where  $\lambda$  is the Lebesgue measure, then by Radon-Nikodym theorem, there is a density  $f$  such that

$$\mu_X(A) = \int_A f(x) d\lambda(x)$$

for all  $A \in \mathcal{B}$ . Or equivalently,  $F$  is absolutely continuous.

### Example

Not all random variables have densities, even when its CDF is continuous. Consider the

*Cantor function*

$$F(x) = \begin{cases} \sum_n \frac{a_n}{2^n}, & x = \sum_n \frac{2a_n}{3^n} \in C \text{ for some } \{a_n\} \in \{0, 1\}^{\mathbb{N}} \\ \sup_{y \leq x, y \in C} F(y), & x \in [0, 1] - C \\ 0, & x < 0 \\ 1, & x > 1 \end{cases}$$

where  $C$  is the Cantor set. Then  $F$  is a valid CDF, but has no density.

### Definition 1.33

A probability measure  $P$  is said to be **discrete** if there is a countable set  $S$  such that  $P(S^c) = 0$ . A random variable  $X$  is said to be **discrete** if its distribution is.

### Theorem 1.34

Suppose  $X : (\Omega, \mathcal{F}) \rightarrow (S, \sigma(\mathcal{A}))$  and  $\mathcal{A}$  is a collection of subsets in  $S$ . If  $X^{-1}(A) \in \mathcal{F}$  for all  $A \in \mathcal{A}$ , then  $X$  is a random variable.

*Proof.* Set  $\mathcal{G} = \{A \subset S \mid X^{-1}(A) \in \mathcal{F}\}$ . Clearly  $\emptyset \in \mathcal{G}$  and if  $A \in \mathcal{G}$ ,  $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathcal{F}$ , so  $A^c \in \mathcal{G}$ . If  $A_n \in \mathcal{G}$ , then  $X^{-1}(\cup_n A_n) = \cup_n X^{-1}(A_n) \in \mathcal{F}$ , so  $\cup_n A_n \in \mathcal{G}$ . Hence  $\mathcal{G}$  is a  $\sigma$ -algebra containing  $\mathcal{A}$ , so  $\sigma(\mathcal{A}) \subset \mathcal{G}$ . It follows that  $X^{-1}(B) \in \mathcal{F}$  for all  $B \in \sigma(\mathcal{A})$ , so  $X$  is a random variable. ■

### Corollary 1.35

If  $X_i$  are random variables, then

$$\inf_i X_i, \quad \sup_i X_i, \quad \liminf_{i \rightarrow \infty} X_i, \quad \limsup_{i \rightarrow \infty} X_i$$

are all random variables.

*Proof.* Since the sets of the form  $(-\infty, x]$  generate  $\mathcal{B}$ , it suffices to check that the inverse images of these sets are in  $\mathcal{F}$ . For  $\inf_i X_i$ ,

$$\left\{ \inf_i X_i \leq x \right\} = \cup_i \{X_i \leq x\} \in \mathcal{F}.$$

For  $\sup_i X_i$ , since  $\sup_i X_i = -\inf_i (-X_i)$ , it is also a random variable. Finally, write

$$\liminf_i X_i = \sup_n \inf_{i \geq n} X_i, \quad \limsup_i X_i = \inf_n \sup_{i \geq n} X_i.$$

The results follow from the measurability of  $\inf_i X_i$  and  $\sup_i X_i$ . ■

### Definition 1.36

Let  $X$  be a random variable.  $\sigma(X)$  is the smallest  $\sigma$ -algebra such that  $X$  is measurable.

### Remark

If  $X : \Omega \rightarrow (S, \mathcal{S})$ , then  $\sigma(X) = X^{-1}(\mathcal{S})$ .

### Definition 1.37

Let  $X$  be a random variable. The **expectation** of  $X$  is defined as

$$\mathbb{E}[X] = \int X d\mathbb{P}.$$

### Theorem 1.38 (Jensen's Inequality)

Let  $X : \Omega \rightarrow \mathbb{R}^d$  be a random variable such that  $\mathbb{E}[\|X\|_1] < \infty$  and  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  be a convex function. Then

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)].$$

*Proof.* For any given  $y \in \mathbb{R}^d$ , note that  $\{x \in \mathbb{R}^d \mid \phi(x) > \phi(y)\}$  is an open convex set. By the Hahn-Banach separation theorem, there is a hyperplane  $\{f(x) = a + \langle b, x \rangle\}$  separating  $\{(x, \phi(x)) \in \mathbb{R}^{d+1} \mid \phi(x) > \phi(y)\}$  and  $\{(y, \phi(y))\}$ . Note that  $\phi(y) = f(y)$  and  $\phi(x) \geq f(x)$  for all  $x \in \mathbb{R}^d$ . Take  $y = \mathbb{E}[X]$ , then

$$\phi(\mathbb{E}[X]) = f(\mathbb{E}[X]) = \mathbb{E}[f(X)] \leq \mathbb{E}[\phi(X)].$$

■

### Theorem 1.39 (Hölder's Inequality)

Let  $X, Y$  be random variables and  $p, q \in [1, \infty]$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Then

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}.$$

*Proof.* If  $\mathbb{E}[|X|^p]$  and  $\mathbb{E}[|Y|^q]$  are zero or infinite, the result is trivial. We assume that  $\mathbb{E}[|X|^p] = \mathbb{E}[|Y|^q] = 1$ . For fixed  $y \geq 0$ , set  $\phi(x) = x^p/p + y^q/q - xy$  for  $x \geq 0$ .

$$\phi'(x) = x^{p-1} - y, \quad \phi''(x) = (p-1)x^{p-2} \geq 0.$$

Thus  $\phi$  is convex and minimized at  $x = y^{1/(p-1)}$  with minimum  $\phi(y^{1/(p-1)}) = 0$ . Hence  $x^p/p + y^q/q \geq xy$  for all  $x, y \geq 0$ .

$$\mathbb{E}[|XY|] \leq \mathbb{E}\left[\frac{|X|^p}{p} + \frac{|Y|^q}{q}\right] = \frac{1}{p} + \frac{1}{q} = 1 = \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}.$$

■

### Theorem 1.40 (Markov's Inequality)

If  $X \geq 0$  is a random variable, then for any  $c > 0$ ,

$$\mathbb{P}(X \geq c) \leq \frac{1}{c} \mathbb{E}[X].$$

*Proof.*

$$\mathbb{P}(X \geq c) = \int \mathbf{1}_{\{X \geq c\}} d\mathbb{P} \leq \int \frac{X}{c} d\mathbb{P} = \frac{1}{c} \mathbb{E}[X].$$

**Example**

Suppose  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a non-negative function. Put

$$I_A = \inf_{y \in A} \phi(y),$$

where  $A$  is some measurable set. Then for any random variable  $X$ ,

$$I_A \mathbf{1}\{X \in A\} \leq \phi(x) \mathbf{1}\{X \in A\} \leq \phi(x).$$

Thus

$$I_A P(X \in A) \leq E[\phi(X)].$$

**Corollary 1.41** (Chebyshev's Inequality)

Let  $X$  be a random variable. Then for any  $c > 0$  and  $\alpha \in \mathbb{R}$ ,

$$P(|X - \alpha| \geq c) \leq \frac{1}{c^2} E[(X - \alpha)^2].$$

*Proof.* By the Markov's inequality,

$$P(|X - \alpha| \geq c) = P((X - \alpha)^2 \geq c^2) \leq \frac{1}{c^2} E[(X - \alpha)^2].$$

■

**Theorem 1.42**

Suppose  $X$  is a random variable of  $(S, \mathcal{S})$  with distribution  $\mu$  and  $f : (S, \mathcal{S}) \rightarrow (\mathbb{R}, \mathcal{B})$  is measurable. If either

- (a)  $f \geq 0$ , or
- (b)  $E[|f(X)|] < \infty$ ,

then

$$E[f(X)] = \int f(x) d\mu(x).$$

*Proof.* Suppose first that  $f = \mathbf{1}_A$  for some  $A \in \mathcal{S}$ . Then

$$E[f(X)] = P(X \in A) = P(X^{-1}(A)) = \mu(A) = \int \mathbf{1}_A d\mu.$$

By linearity we can extend this result to simple functions. Now suppose first that (a) holds. For such  $f$ , there is a sequence of simple functions  $s_n \nearrow f$  and  $s_n \circ X \nearrow f \circ X$ . By LMCT,

$$E[f(X)] = E\left[\lim_n s_n(X)\right] = \lim_n E[s_n(X)] = \lim_n \int s_n d\mu = \int f d\mu.$$

Suppose that (b) is the case. Write  $f = f^+ - f^-$  and apply the previous result.

$$\mathbb{E}[f(X)] = \mathbb{E}[f^+(X)] - \mathbb{E}[f^-(X)] = \int f^+ d\mu - \int f^- d\mu = \int f d\mu.$$

■

### **Definition 1.43**

The  **$k$ -th moment** of a random variable  $X$  is  $\mathbb{E}[X^k]$ .

### **Definition 1.44**

The **variance** of a random variable  $X$  is  $\text{Var } \mathbb{E}[(X - \mathbb{E}[X])^2]$ .

### **Definition 1.45**

The **covariance** of two integrable random variables  $X, Y$  is

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

### **Definition 1.46**

For  $1 \leq p < \infty$ , the  $\mathcal{L}^p(\Omega, P)$  space is defined as

$$\mathcal{L}^p(\Omega, P) = \{X : \Omega \rightarrow S \mid X \text{ measurable and } \mathbb{E}[|X|^p] < \infty\}.$$

For  $p = \infty$ ,

$$\mathcal{L}^\infty(\Omega, P) = \{X : \Omega \rightarrow S \mid X \text{ measurable and } \text{ess sup}_{\omega \in \Omega} X(\omega) < \infty\}.$$

### **Proposition 1.47**

Let  $1 \leq p < q \leq \infty$ . Then  $\mathcal{L}^q(P) \subset \mathcal{L}^p(P)$ .

*Proof.* Suppose first that  $q < \infty$ . If  $X \in \mathcal{L}^q(P)$ , then

$$\mathbb{E}[|X|^p] \leq \mathbb{E}[|X|^q \mathbf{1}\{|X| \geq 1\}] + \mathbb{E}[|X|^p \mathbf{1}\{|X| < 1\}] \leq \mathbb{E}[|X|^q] + 1 < \infty.$$

Hence  $X \in \mathcal{L}^p(P)$ . If  $q = \infty$ ,  $X$  is essentially bounded, i.e.,  $X \leq M$  for some  $M \in \mathbb{R}$  almost surely. Hence  $X \in \mathcal{L}^p$ . ■

### 1.3. Independence

#### Definition 1.48

Let  $(\Omega, \mathcal{F}, P)$  be a probability space. Suppose  $\mathcal{F}_\beta \subset \mathcal{F}$ ,  $\beta \in B$  are a collection of sub- $\sigma$ -algebras. Then  $\{\mathcal{F}_\beta\}$  are **independent** if for all finite  $\{\mathcal{F}_i\}_{i=1}^n \subset \{\mathcal{F}_\beta\}$ ,

$$P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i)$$

where  $A_i \in \mathcal{F}_i$ .

#### Definition 1.49

A collection of random variables  $\{X_\beta \mid \beta \in B\}$  on  $(\Omega, \mathcal{F}, P)$  is **independent** if the collection of the generating  $\sigma$ -algebras  $\{\sigma(X_\beta) \mid \beta \in B\}$  is.

#### Remark

In other words,

$$P(\cap_i \{X_{\beta_i} \in A_i\}) = \prod_i P(X_{\beta_i} \in A_i).$$

Note that these random variables can map into different measurable space.

#### Definition 1.50

A collection of events  $\mathcal{S}$  is **independent** if  $\{\mathbf{1}_A \mid A \in \mathcal{S}\}$  is.

#### Proposition 1.51

Let  $X_1, \dots, X_n$  be independent random variables and  $g_1, \dots, g_n$  are measurable functions. Then  $g_1(X_1), \dots, g_n(X_n)$  are independent.

*Proof.* Suppose  $g_i : (S_i, \mathcal{S}_i) \rightarrow (T_i, \mathcal{T}_i)$ . For  $A_i \in \mathcal{T}_i$ ,  $g^{-1}(A_i) \in \mathcal{S}_i$  and

$$P(\cap_i \{g_i(X_i) \in A_i\}) = P(\cap_i \{X_i \in g^{-1}(A_i)\}) = \prod_i P(X_i \in g^{-1}(A_i)) = \prod_i P(g_i(X_i) \in A_i).$$

$g_1(X_1), \dots, g_n(X_n)$  are independent. ■

#### Theorem 1.52

Let  $\mathcal{S}_1, \dots, \mathcal{S}_n$  be a collection of  $\pi$ -system. If  $\Omega \in \mathcal{S}_i$  for all  $i = 1, \dots, n$  and for all  $A_i \in \mathcal{S}_i$ ,

$$P(\cap_i A_i) = \prod_i P(A_i),$$

then  $\sigma(\mathcal{S}_1), \dots, \sigma(\mathcal{S}_n)$  are independent.

*Proof.* Fix  $\mathcal{S}_2, \dots, \mathcal{S}_n$ . Put

$$\mathcal{L} = \left\{ A \in \mathcal{F} \mid P(A \cap (\cap_{i=2}^n A_i)) = P(A) \prod_{i=2}^n P(A_i), A_i \in \mathcal{S}_i \text{ for } i = 2, \dots, n \right\}.$$

We claim that  $\mathcal{L}$  forms a  $\lambda$ -system. First, by assumption we can pick  $A_i = \Omega$  for  $i = 2, \dots, n$  to see that  $\Omega \in \mathcal{L}$ . Suppose that  $A \subset B$ ,  $A, B \in \mathcal{L}$ ,

$$\begin{aligned} P((B - A) \cap (\cap_{i=2}^n A_i)) &= P((B \cap (\cap_{i=2}^n A_i)) - (A \cap (\cap_{i=2}^n A_i))) \\ &= P(B) \prod_{i=2}^n P(A_i) - P(A) \prod_{i=2}^n P(A_i) = P(B - A) \prod_{i=2}^n P(A_i). \end{aligned}$$

Hence  $B - A \in \mathcal{L}$ . Let  $S_j \nearrow S$ ,  $S_j \in \mathcal{L}$ . Then

$$P(S \cap (\cap_{i=2}^n A_i)) = \lim_{j \rightarrow \infty} P(S_j \cap (\cap_{i=2}^n A_i)) = \lim_{j \rightarrow \infty} P(S_j) \prod_{i=2}^n P(A_i) = P(S) \prod_{i=2}^n P(A_i).$$

Thus  $S \in \mathcal{L}$  and  $\mathcal{L}$  is a  $\lambda$ -system. By Dynkin's  $\pi$ - $\lambda$ ,  $\sigma(\mathcal{S}_1), \mathcal{S}_2, \dots, \mathcal{S}_n$  satisfies the product property. Repeat the procedure for  $\mathcal{S}_2, \dots, \mathcal{S}_n$ . We have that  $\sigma(\mathcal{S}_1), \dots, \sigma(\mathcal{S}_n)$  satisfies the product property. That is, they are independent. ■

### Corollary 1.53

Let  $X_1, \dots, X_n$  be  $\mathbb{R}$ -valued random variables. Then they are independent if and only if

$$P(X_1 \leq s_1, \dots, X_n \leq s_n) = \prod_{i=1}^n P(X_i \leq s_i)$$

for all  $s_i \in \mathbb{R}$ ,  $1 \leq i \leq n$ .

*Proof.* The sufficient part is trivial. For the converse, put  $\mathcal{S}_i = \{\{X_i \leq t\} \mid t \in \mathbb{R}\} \cup \{\Omega\}$ . Clearly  $\mathcal{S}_i$  are  $\pi$ -system and  $\Omega \in \mathcal{S}_i$  for all  $i$ .  $\sigma(\mathcal{S}_i)$  are independent and  $\mathcal{S}_i$  generates  $\sigma(X_i)$ . Applying theorem 1.52 shows that  $X_i$  are independent. ■

### Corollary 1.54

If  $\mathcal{F}_{ij}$ ,  $1 \leq i \leq n, 1 \leq j \leq m(i)$  are independent  $\sigma$ -algebras, then  $\mathcal{G}_i = \sigma(\cup_j \mathcal{F}_{ij})$  are independent.

*Proof.* Put  $\mathcal{H}_i = \{\cap_j A_j \mid A_j \in \mathcal{F}_{ij}\}$ . We claim that  $\sigma(\mathcal{H}_i) = \mathcal{G}_i$ . Indeed, by choosing sets of the form

$$(\Omega, \dots, \Omega, A_j, \Omega, \dots, \Omega) \in \mathcal{F}_{i1} \times \dots \times \mathcal{F}_{im(i)},$$

it is clear that  $\cup_j \mathcal{F}_{ij} \subset \mathcal{H}_i$ . Also, if  $A \in \mathcal{H}_i$ , then

$$A = \cap_j A_j = (\cup_j (A_j^c))^c \in \sigma(\cup_j \mathcal{F}_{ij}).$$

Thus  $\cup_j \mathcal{F}_{ij} \subset \mathcal{H}_i \subset \sigma(\cup_j \mathcal{F}_{ij})$  and  $\sigma(\mathcal{H}_i) = \sigma(\cup_j \mathcal{F}_{ij}) = \mathcal{G}_i$ . Also notice that  $\mathcal{H}_i$  contain  $\Omega$  and form  $\pi$ -systems. For  $A_i \in \mathcal{H}_i$ , write  $A_i = \cap_j A_{ij}$ . Then

$$P(\cap_i A_i) = P(\cap_{ij} A_{ij}) = \prod_{ij} P(A_{ij}) = \prod_i P(\cap_j A_{ij}) = \prod_i P(A_i).$$

From theorem 1.52 we know that  $\mathcal{G}_i = \sigma(\mathcal{H}_i)$  are independent. ■

### Corollary 1.55

If  $X_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m(i)$  are independent random variables, then  $Y_i = h_i(X_{i1}, \dots, X_{im(i)})$  are independent provided that  $h_i$  are measurable.

*Proof.* Write  $\mathcal{F}_{ij} = \sigma(X_{ij})$ . We claim that  $\sigma(Y_i) \subset \sigma(\cup_j \mathcal{F}_{ij})$ . Indeed, if  $B_i$  is a measurable set,  $h_i^{-1}(B_i)$  is measurable. Write  $h_i^{-1}(B_i) = C_{i1} \times \dots \times C_{im(i)}$  and since each  $X_{ij}^{-1}(C_{ij}) \in \mathcal{F}_{ij}$ , we see that  $\sigma(Y_i) \subset \sigma(\cup_j \mathcal{F}_{ij})$ . It then follows from corollary 1.54 that  $\sigma(Y_i)$  are independent and  $Y_i$  are independent. ■

### Theorem 1.56

If  $X_1, \dots, X_n$  are independent  $\mathbb{R}$ -valued random variables and the distribution of  $X_i$  is  $\mu_i$ . Then the joint distribution of  $(X_1, \dots, X_n)$  is  $\mu_1 \times \dots \times \mu_n$ .

*Proof.* Let  $\mu$  be the distribution of  $(X_1, \dots, X_n)$ . By definition,

$$\begin{aligned}\mu((X_1, \dots) \in A_1 \times \dots \times A_n) &= \mu(X_1 \in A_1, \dots, X_n \in A_n) \\ &= \prod_{i=1}^n \mu_i(X_i \in A_i) = (\mu_1 \times \dots \times \mu_n)(A_1 \times \dots \times A_n).\end{aligned}$$

Now the sets of the forms  $A = A_1 \times \dots \times A_n$  is a  $\pi$ -system generating the product  $\sigma$ -algebra. By corollary 1.19, the joint distribution is exactly  $\mu_1 \times \dots \times \mu_n$ . ■

### Theorem 1.57

Let  $X, Y$  be two independent random variables. If  $h(x, y)$  satisfies either

- (a)  $E[|h(X, Y)|] < \infty$ , or
- (b)  $h$  is non-negative,

then

$$E[h(X, Y)] = \int \int h d\mu_X d\mu_Y,$$

where  $\mu_X, \mu_Y$  are the distributions of  $X$  and  $Y$ , respectively.

*Proof.* The proof follows directly from Fubini-Tonelli theorem. If one of the assumptions is true, then

$$E[h(X, Y)] = \int_{\mathbb{R}^2} h d(\mu_X \times \mu_Y) = \int \int h d\mu_X d\mu_Y.$$

■

### Remark

If  $h(x, y) = h_1(x)h_2(y)$ , then

$$E[h_1(X)h_2(Y)] = E[h(X, Y)] = \int \int h_1 h_2 d\mu_X d\mu_Y = E[h_1(X)] E[h_2(Y)].$$

### Corollary 1.58

If  $X_1, \dots, X_n$  are independent random variables and

- (a)  $\mathbb{E} [|X_1 \cdots X_n|] < \infty$  or
- (b)  $X_i \geq 0$  for all  $i$ ,

then

$$\mathbb{E} [X_1 \cdots X_n] = \prod_{i=1}^n \mathbb{E} [X_i].$$

*Proof.* Let  $h(x, y) = xy$ . By assumptions, we have either  $\mathbb{E} [|h(X_1, X_2)|] < \infty$  or  $h(X_1, X_2) \geq 0$ . By theorem 1.57,  $\mathbb{E} [X_1 X_2] = \mathbb{E} [X_1] \mathbb{E} [X_2]$ . Substitute  $X_1$  by  $X_1 X_2$  and  $X_2$  by  $X_3$ , we see that  $\mathbb{E} [X_1 X_2 X_3] = \mathbb{E} [X_1] \mathbb{E} [X_2] \mathbb{E} [X_3]$ . Repeat the procedure  $n$  times and the result follows. ■

### Definition 1.59

Let  $X, Y$  be independent random variables with CDF  $F$  and  $G$ , respectively. The **convolution** of two CDF is defined as

$$(F * G)(z) = \int F(z - y) dG(y).$$

#### Remark

If  $F$  and  $G$  are absolutely continuous with respect to the Lebesgue measure, then they have Radon-Nikodym derivatives  $f$  and  $g$ . The definition of convolution becomes

$$(F * G)(z) = \int F(z - y) dG(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x) g(y) dx dy.$$

Then

$$(F * G)'(z) = \int f(z - y) g(y) dy = (f * g)(z),$$

which is exactly the definition of convolution of two functions.

### Proposition 1.60

Let  $X$  and  $Y$  be independent random variables. Then

$$\mathbb{P}(X + Y \leq z) = (F * G)(z).$$

*Proof.* By theorem 1.57,

$$\begin{aligned} \mathbb{P}(X + Y \leq z) &= \mathbb{E} [\mathbf{1}_{\{X + Y \leq z\}}] = \int \int \mathbf{1}_{\{x + y \leq z\}} dF(x) dG(y) \\ &= \int F(z - y) dG(y) = (F * G)(z). \end{aligned}$$

■

#### Remark

Note that the convolution is commutative since

$$(F * G)(z) = \mathbb{P}(X + Y \leq z) = \mathbb{P}(Y + X \leq z) = (G * F)(z).$$

### Remark

For discrete  $X$  and  $Y$ , the convolution becomes

$$P(X + Y = z) = \sum_y P(X = z - y) P(Y = y).$$

### Example

Consider  $X \sim \Gamma(\alpha_1, \beta)$  and  $Y \sim \Gamma(\alpha_2, \beta)$ . Then the density for  $X + Y$  is

$$\begin{aligned} f_{X+Y}(z) &= \int f_X(z-y) f_Y(y) dy \\ &= \int_0^z \frac{1}{\Gamma(\alpha_1)} \beta^{\alpha_1} (z-y)^{\alpha_1-1} e^{-\beta(z-y)} \frac{1}{\Gamma(\alpha_2)} \beta^{\alpha_2} y^{\alpha_2-1} e^{-\beta y} dy \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \beta^{\alpha_1+\alpha_2} e^{-\beta z} \int_0^z (z-y)^{\alpha_1-1} y^{\alpha_2-1} dy \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \beta^{\alpha_1+\alpha_2} e^{-\beta z} z^{\alpha_1+\alpha_2-1} \int_0^1 (1-t)^{\alpha_1-1} t^{\alpha_2-1} dt \\ &= \frac{B(\alpha_1, \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \beta^{\alpha_1+\alpha_2} e^{-\beta z} z^{\alpha_1+\alpha_2-1} = \frac{1}{\Gamma(\alpha_1 + \alpha_2)} \beta^{\alpha_1+\alpha_2} e^{-\beta z} z^{\alpha_1+\alpha_2-1}. \end{aligned}$$

Hence  $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \beta)$ .

## 1.4. Convergence of Random Variables

### Definition 1.61

A sequence of probability measures  $P_n$  on  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  are **consistent** if

$$P_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = P_n((a_1, b_1] \times \cdots \times (a_n, b_n]))$$

for every  $n$ .

### Theorem 1.62 (Kolmogorov Extension)

Suppose that a sequence of probability measures  $P_n$  on  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  are consistent. Then there is a unique probability measure  $P$  on  $(\mathbb{R}^{\mathbb{N}}, \mathcal{B})$  satisfying that

$$P(\{\omega \in \mathbb{R}^{\mathbb{N}} \mid \omega_i \in (a_i, b_i], 1 \leq i \leq n\}) = P_n((a_1, b_1] \times \cdots \times (a_n, b_n]),$$

where  $\mathcal{B}$  is generated by the collection

$$\{\omega \in \mathbb{R}^{\mathbb{N}} \mid \omega_i \in (a_i, b_i], 1 \leq i \leq n, n \in \mathbb{N}\}.$$

*Proof.* Let

$$\mathcal{S} = \{(a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \cdots \mid n \in \mathbb{N}\}.$$

Define  $P$  on  $\mathcal{S}$  to be

$$P((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \cdots) = P_n((a_1, b_1] \times \cdots \times (a_n, b_n]))$$

Clearly,  $\mathcal{S}$  forms a semi-algebra. From the Carathéodory extension theorem, it suffices to show that  $P$  is finitely additive,  $\sigma$ -additive on  $\mathcal{S}$  and  $P(\emptyset) = 0$ . Note that  $P(\emptyset) = P(\emptyset \times \mathbb{R} \times \cdots) = P_1(\emptyset) = 0$ . We verify the first two conditions.

First, if  $A, B \in \mathcal{S}$  are disjoint,  $m \leq n$ ,

$$A = \{\omega \in \mathbb{R}^{\mathbb{N}} \mid \omega_i \in (a_i, b_i], 1 \leq i \leq m\} \quad \text{and} \quad B = \{\omega \in \mathbb{R}^{\mathbb{N}} \mid \omega_i \in (c_i, d_i], 1 \leq i \leq n\},$$

then

$$P(A \cup B) = P_n((\pi_n A) \cup (\pi_n B)) = P_n(\pi_n A) + P_n(\pi_n B) = P(A) + P(B),$$

where  $\pi_n : \omega \rightarrow (\omega_1, \dots, \omega_n)$  is the projection onto the first  $n$  components. Hence  $P$  is finitely additive.

Next, suppose  $A_1, \dots \in \mathcal{S}$  are countably many disjoint measurable sets. Put  $A = \cup_i A_i$ . We can consider the algebra  $\bar{\mathcal{S}} = \{\text{finite disjoint union of sets in } \mathcal{S}\}$  generated by  $\mathcal{S}$ .  $B_n = \cup_{i>n} A_i \in \bar{\mathcal{S}}$ . Thus

$$P(A) = P(B_n) + \sum_{i=1}^n P(A_i)$$

by the previous result. It now suffices to show that  $P(B_n) \rightarrow 0$  for any  $B_n \searrow \emptyset$ . Suppose not,

then there is  $\delta > 0$  such that  $P(B_n) \rightarrow \delta$  as  $B_n \rightarrow \emptyset$  by the monotonicity of  $P$ .

For such  $\{B_n\}$ , we claim that there is a sequence of compact set  $K_n$  such that  $K_n \subset B_n$  and  $P(B_n - K_n) < 2^{-(n+1)}\delta$ . Now since  $B_1 \in \bar{\mathcal{S}}$ , there are disjoint  $E_1^1, \dots, E_{m_1}^1$  such that  $B_1 = \bigcup_{i=1}^{m_1} E_i^1$ . Now since each  $E_i^1$  is of the product of  $(\cdot, \cdot]$ . We can find a compact subset  $K_i^1$  of the product of  $[\cdot, \cdot]$  such that  $P(E_i^1 - K_i^1) < m_1^{-1}2^{-2}\delta$ . Hence  $K_1 = \bigcup_i K_i^1 \subset B_1$  satisfies that

$$P(B_1 - K_1) = \sum_{i=1}^{m_1} P(E_i^1 - K_i^1) < 2^{-2}\delta$$

as desired. Repeat the process and find  $K_n$  inductively. The claim follows.

Now,  $\cap_{n=1}^m K_n \searrow K$  as  $m \rightarrow \infty$ . Also,

$$P(B_m - (\cap_{n=1}^m K_n)) \leq \sum_{n=1}^m P(B_n - K_n) \leq \frac{\delta}{2}.$$

Hence  $\delta/2 \leq P(B_m) - \delta/2 \leq P(\cap_{n=1}^m K_n)$ . We see that  $\cap_{n=1}^m K_n$  is non-empty for each  $m$ . But this implies that  $K \subset \cap_n B_n$  is non-empty, a contradiction. Thus  $P(B_n) \rightarrow 0$ .

Finally, the  $\sigma$ -additivity follows from that we can take  $n \rightarrow \infty$  so that

$$P(A) = \lim_{n \rightarrow \infty} P(B_n) + \sum_{i=1}^n P(A_n) = \sum_i P(A_n).$$

Applying Carathéodory extension theorem, such  $P$  can be extended on  $(\mathbb{R}^N, \mathcal{B})$ . ■

### Remark

With Kolmogorov extension theorem, we can consider a sequence of independent variable  $X_i$  on the product probability space with  $\mathcal{F} = \mathcal{B}$ ,  $\tilde{X}_i : \omega \mapsto \omega_i$  and  $P(B_1 \times \dots \times B_n) = \prod_{i=1}^n \mu_i(B_i)$ , where  $\mu_i$  is the distribution of  $X_i$ .

### Definition 1.63

Let  $X_n$  be a sequence of random variable.  $X_n$  **converges almost surely** to  $X$  if

$$P \left\{ \lim_{n \rightarrow \infty} X_n = X \right\} = 1.$$

We denote it as  $X_n \xrightarrow{a.s.} X$  or  $X_n \rightarrow X$  a.s.

### Definition 1.64

Let  $X_n$  be a sequence of random variable.  $X_n$  **converges in probability** to  $X$  if for every  $\epsilon > 0$ ,

$$P \{ |X_n - X| > \epsilon \} \rightarrow 0$$

as  $n \rightarrow \infty$ . We denote it as  $X_n \xrightarrow{P} X$ .

### Definition 1.65

A sequence of random variable  $X_n \in \mathcal{L}^p$  is said to **converge in  $\mathcal{L}^p$**  to  $X$  if

$$\mathbb{E} [|X_n - X|^p]^{1/p} \rightarrow 0$$

as  $n \rightarrow \infty$ . If  $p = \infty$ , the definition becomes

$$\text{ess sup}_{\omega \in \Omega} |X_n(\omega) - X(\omega)| \rightarrow 0.$$

We denote it as  $X_n \rightarrow X$  in  $\mathcal{L}^p$ .

### Proposition 1.66

Let  $X_n$  be a sequence of independent and identically distributed random variables. Then

- (a) If  $X_n \rightarrow X$  almost surely, then  $X_n \xrightarrow{p} X$ .
- (b) If  $X_n \rightarrow X$  in  $\mathcal{L}^p$ , then  $X_n \xrightarrow{p} X$ .

*Proof.* For (a), given  $\epsilon > 0$ , put

$$E_k = \cup_{n \geq k} \{|X_n - X| > \epsilon\}.$$

Note that  $E_k \searrow E = \{|X_n - X| > \epsilon \text{ for infinitely many } n\} = \{\lim_{n \rightarrow \infty} X_n = X\}^c$ . Hence

$$\mathbb{P} \{|X_k - X| > \epsilon\} \leq \mathbb{P}(E_k) \rightarrow \mathbb{P} \left\{ \lim_{n \rightarrow \infty} X_n = X \right\}^c = 0$$

Hence  $X_n \rightarrow X$  in probability.

For (b), suppose first that  $p < \infty$ . By Markov inequality,

$$\mathbb{P} \{|X_n - X| > \epsilon\} = \mathbb{P} \{|X_n - X|^p > \epsilon^p\} \leq \frac{1}{\epsilon^p} \mathbb{E} [|X_n - X|^p] \rightarrow 0.$$

Let  $p = \infty$ . Note that  $\text{ess sup} |X_n - X| = \inf \{c \mid \mathbb{P} \{|X_n - X| > c\} = 0\}$ . Convergence in  $\mathcal{L}^\infty$  implies that for  $\epsilon > 0$ , there is  $N$  such that if  $n \geq N$ ,  $\inf \{c \mid \mathbb{P} \{|X_n - X| > c\} = 0\} < \epsilon$ . That is,  $\mathbb{P} \{|X_n - X| > \epsilon\} = 0$  for  $n \geq N$ . Hence  $X_n \xrightarrow{p} X$ . ■

## 2. Asymptotic Theory

### 2.1. Law of Large Number

#### Definition 2.1

Let  $X_i$  be random variables with  $\mathbb{E}[X_i^2] < \infty$ . They are called **uncorrelated** if

$$\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \mathbb{E}[X_j].$$

#### Theorem 2.2 (Weak Law of Large Number I)

Suppose that  $X_n$  are uncorrelated random variables with  $\text{Var}[X_n] \leq C\infty$  and  $\mathbb{E}[X_n] = \mu$  for all  $n$ . Let  $S_n = \sum_{i=1}^n X_i$ . Then

$$\frac{1}{n}S_n \rightarrow \mu$$

in  $\mathcal{L}^2$  and hence in probability.

*Proof.* Compute that

$$\mathbb{E}\left[\left(\frac{1}{n}S_n - \mu\right)^2\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \leq \frac{C}{n} \rightarrow 0.$$

Hence  $\frac{1}{n}S_n \rightarrow \mu$  in  $\mathcal{L}^2$  and thus in probability. ■

#### Theorem 2.3 (Weak Law of Large Number II, Khinchin)

Suppose that  $X_i$  is a sequence of independent and identically distributed random variables with  $\mathbb{E}[|X_1|] < \infty$ . Let  $S_n = \sum_{i=1}^n X_i$  and  $\mu = \mathbb{E}[X_1]$ . Then

$$\frac{1}{n}S_n \rightarrow \mu$$

in  $\mathcal{L}^1$  and hence in probability.

*Proof.* By replacing  $X_i$  with  $X_i - \mu$ , we may assume without loss of generality that  $\mu = 0$ . Now, for  $C > 0$ ,

$$0 = \mathbb{E}[X_i] = \mathbb{E}[X_i \mathbf{1}\{|X_i| > C\}] + \mathbb{E}[X_i \mathbf{1}\{|X_i| \leq C\}].$$

Also,

$$\begin{aligned} \frac{1}{n}S_n &= \frac{1}{n} \sum_{i=1}^n X_i \mathbf{1}\{|X_i| > C\} + \frac{1}{n} \sum_{i=1}^n X_i \mathbf{1}\{|X_i| \leq C\} \\ &= \frac{1}{n} \sum_{i=1}^n (X_i \mathbf{1}\{|X_i| > C\} - \mathbb{E}[X_i \mathbf{1}\{|X_i| > C\}]) + \frac{1}{n} \sum_{i=1}^n (X_i \mathbf{1}\{|X_i| \leq C\} - \mathbb{E}[X_i \mathbf{1}\{|X_i| \leq C\}]). \end{aligned}$$

Notice that by LDCT,

$$\mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n (X_i \mathbf{1}\{|X_i| > C\} - \mathbb{E}[X_i \mathbf{1}\{|X_i| > C\}]) \right| \right] \leq 2 \mathbb{E}[|X_1| \mathbf{1}\{|X_1| > C\}] \rightarrow 0$$

as  $C \rightarrow \infty$  since  $|X_1| \mathbf{1}\{|X_1| > C\} \leq |X_1|$  and  $\mathbb{E}[|X_1|] < \infty$ . Also, by Hölder inequality and the independence,

$$\mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n (X_i \mathbf{1}\{|X_i| \leq C\} - \mathbb{E}[X_i \mathbf{1}\{|X_i| \leq C\}]) \right| \right] \leq \sqrt{\frac{1}{n} \text{Var}(X_i \mathbf{1}\{|X_i| \leq C\})} \leq \frac{C}{\sqrt{n}}$$

For any given  $\epsilon > 0$ , there is  $C$  such that  $2 \mathbb{E}[|X_1| \mathbf{1}\{|X_1| > C\}] < \epsilon$  and

$$\begin{aligned} \mathbb{E} \left[ \left| \frac{1}{n} S_n \right| \right] &\leq \mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n (X_i \mathbf{1}\{|X_i| > C\} - \mathbb{E}[X_i \mathbf{1}\{|X_i| > C\}]) \right| \right] \\ &\quad + \mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n (X_i \mathbf{1}\{|X_i| \leq C\} - \mathbb{E}[X_i \mathbf{1}\{|X_i| \leq C\}]) \right| \right] \\ &\leq \epsilon + \frac{C}{\sqrt{n}} \rightarrow \epsilon \end{aligned}$$

as  $n \rightarrow \infty$ . Since  $\epsilon$  can be arbitrarily small, we conclude that  $\frac{1}{n} S_n \rightarrow 0$  in  $\mathcal{L}^1$  and hence in probability. ■

#### Definition 2.4

Let  $A_n$  be a sequence of events.

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$$

and

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n.$$

#### Remark

Observe that

$$\limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ for infinitely many } n\}$$

and

$$\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ for all but finitely many } n\}.$$

#### Theorem 2.5 (Borel-Cantelli I)

Let  $A_n$  be a sequence of events. If  $\sum_n P(A_n) < \infty$ , then

$$P \left( \limsup_{n \rightarrow \infty} A_n \right) = 0.$$

*Proof.* Let  $\epsilon > 0$  be given. By assumption, there is  $n_0$  such that  $\sum_{n \geq n_0} P(A_n) < \epsilon$ . Then

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = P(\cap_{m=1}^{\infty} \cup_{n=m}^{\infty} A_n) \leq P(\cup_{n=n_0}^{\infty} A_n) \leq \sum_{n=n_0}^{\infty} P(A_n) < \epsilon.$$

Since  $\epsilon$  can be arbitrarily small,  $P(\limsup_{n \rightarrow \infty} A_n) = 0$ . ■

### Corollary 2.6

Suppose for  $\epsilon > 0$ ,  $\sum_n P(|X_n - X| > \epsilon) < \infty$ . Then  $X_n \rightarrow X$  almost surely.

*Proof.* Let  $E_k = \{|X_n - X| > k^{-1} \text{ for finitely many } n\}$ . Note that  $E_{k+1} \subset E_k$  and  $E_k \searrow E = \{X_n \rightarrow X\}$ . Now we claim that  $P(E_k) = 1$ . Consider  $E_k^n = \{|X_n - X| > k^{-1}\}$ . For fixed  $k$ , by assumption we have  $\sum_n P(E_k^n) < \infty$ . By Borel-Cantelli,  $P(\limsup_{n \rightarrow \infty} E_k^n) = 0$ . Hence

$$P(E_k) = P(\{|X_n - X| > k^{-1} \text{ for infinitely many } n\}^c) = 1 - P(\limsup_{n \rightarrow \infty} E_k^n) = 1.$$

It now follows by the monotone convergence of measures that  $P(E) = 1$ . ■

### Remark

Intuitively, if the convergence is sufficiently fast, the convergence in probability may recover almost sure convergence.

### Theorem 2.7 (Strong Law of Large Number I)

Let  $X_i$  be independent and identically distributed with  $\mu = E[X_1]$  and  $E[X_1^4] < \infty$ . Let  $S_n = \sum_{i=1}^n X_i$ . Then

$$\frac{1}{n} S_n \rightarrow \mu$$

almost surely.

*Proof.* Note that

$$\begin{aligned} E\left[\left(\frac{1}{n} S_n - \mu\right)^4\right] &= \frac{1}{n^4} \left( \sum_i E[(X_i - \mu)^4] + \sum_{i \neq j} E[(X_i - \mu)^2(X_j - \mu)^2] \right) \\ &\leq \frac{1}{n^3} E[(X_1 - \mu)^4] + \frac{1}{n^4} \binom{n}{2} \binom{4}{2} E[(X_1 - \mu)^2]^2 \leq \frac{C}{n^2} \end{aligned}$$

for some constant  $C$ . By Chebyshev's inequality, for  $\epsilon > 0$ ,

$$P\left\{\left|\frac{1}{n} S_n - \mu\right| > \epsilon\right\} \leq \frac{1}{\epsilon^4} E\left[\left(\frac{1}{n} S_n - \mu\right)^4\right] \leq \frac{C}{\epsilon^2 n^2}$$

is absolute summable. Hence by [corollary 2.6](#),

$$\frac{1}{n} S_n \rightarrow \mu$$

almost surely. ■

### Theorem 2.8

$X_n \xrightarrow{p} X$  if and only if every subsequence of  $X_n$  has a further subsequence converging almost surely.

*Proof.* Suppose first that  $X_n \xrightarrow{p} X$ . Given a subsequence  $X_{n(k)}$ , we can choose  $n(k_1) < n(k_2) < \dots$  such that

$$P(|X_{n(k_i)} - X| > 2^{-i}) < 2^{-i}.$$

Since  $2^{-i}$  is summable, by Borel-Cantelli we have

$$P(|X_{n(k_i)} - X| > 2^{-i} \text{ for infinitely many } i) = 0.$$

In other words,

$$P\{X_{n(k_i)} \rightarrow X\} = P\{|X_{n(k_i)} - X| > 2^{-i} \text{ for infinitely many } i\}^c = 1.$$

For the converse, suppose that  $X_n \not\xrightarrow{p} X$  in probability. Then there exist  $\epsilon, \delta > 0$  and

$$P\{|X_{n(k)} - X| > \epsilon\} \geq \delta.$$

By assumption there is a further subsequence converging almost surely and thus in probability, i.e.,

$$P\{|X_{n(k_j)} - X| > \epsilon\} \rightarrow 0.$$

This is a contradiction. Hence  $X_n \rightarrow X$  in probability. ■

### Corollary 2.9

Suppose  $X_n \xrightarrow{p} X$ . Then the followings are true:

- (a) If  $f$  is continuous, then  $f(X_n) \xrightarrow{p} f(X)$ .
- (b) If  $|X_n| \leq Y$  for some  $Y \in \mathcal{L}^1$ , then  $E[X_n] \rightarrow E[X]$ .

*Proof.* For (a), by theorem 2.8, every subsequence has a further subsequence  $X_{n(k_j)} \rightarrow X$  almost surely and hence  $f(X_{n(k_j)}) \rightarrow f(X)$  almost surely. Then by theorem 2.8 again we see that  $f(X_n) \xrightarrow{p} f(X)$ .

For (b), by theorem 2.8, every subsequence has a furhter subsequence  $X_{n(k_j)} \rightarrow X$  almost surely and LDCT gives  $E[X_{n(k_j)}] \rightarrow E[X]$ . This implies that  $E[X_n] \rightarrow E[X]$  as well. ■

### Definition 2.10

Let  $(\Omega, \mathcal{F}, P)$  be a probability space.

$$\mathcal{L}^0(\Omega) = \{X : \Omega \rightarrow \mathbb{R} \mid X \text{ is } \mathcal{F}\text{-measurable}\}.$$

### Remark

In general, the almost convergence notion on  $\mathcal{L}^0$  is not metrizable, i.e., there is no metric  $d$  on

$\mathcal{L}^0$  such that

$$d(X_n, X) \rightarrow 0 \Leftrightarrow X_n \xrightarrow{p} X \text{ a.s.}$$

To see this, suppose that the almost sure convergence is metrizable. If  $X_n \xrightarrow{p} X$ , any subsequence  $X_{n(k)}$  converges to  $X$  in probability as well. By [theorem 2.8](#), we can find a further subsequence converging almost surely and hence in metric  $d$ , but this implies that  $d(X_n, X) \rightarrow 0$ . Then  $X_n \rightarrow X$  almost surely, which is absurd since convergence in probability does not imply almost sure convergence in general.

However, convergence in probability on  $\mathcal{L}^0$  can be metrized. For instance,

$$d(X, Y) = \mathbb{E} [\max \{|X - Y|, 1\}] .$$

**Theorem 2.11** (Borel-Cantelli II)

Let  $A_n$  be independent events and  $\sum_{n=1}^{\infty} P(A_n) = \infty$ . Then

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

*Proof.* By assumption we have that  $\sum_{n \geq m} P(A_n) = \infty$  for every  $m \in \mathbb{N}$ . Notice that  $1 + x \leq e^x$ . Then

$$\begin{aligned} P(\limsup_{n \rightarrow \infty} A_n) &= \lim_{m \rightarrow \infty} P(\cup_{n \geq m} A_n) = 1 - \lim_{m \rightarrow \infty} P(\cap_{n \geq m} A_n^c) \\ &= 1 - \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} P(\cap_{n=m}^N A_n^c) = 1 - \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} \prod_{n=m}^N P(A_n^c) \\ &= 1 - \lim_{m \rightarrow \infty} \prod_{n=m}^{\infty} (1 - P(A_n)) \geq 1 - \lim_{m \rightarrow \infty} \exp\left(-\sum_{n=m}^{\infty} P(A_n)\right) = 1. \end{aligned}$$

Hence  $P(\limsup_{n \rightarrow \infty} A_n) = 1$ . ■

**Lemma 2.12**

Let  $X$  be a non-negative random variable and  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a differentiable function with  $h(0) = 0$  and  $h' \geq 0$ . Then

$$\mathbb{E}[h(X)] = \int_0^{\infty} h'(t) P(X > t) dt.$$

*Proof.* By Fubini-Tonelli theorem,

$$\begin{aligned} \mathbb{E}[h(X)] &= \mathbb{E}\left[\int_0^X h'(t) dt\right] = \mathbb{E}\left[\int_0^{\infty} \mathbf{1}_{\{t < X\}} h'(t) dt\right] \\ &= \int_0^{\infty} h'(t) \mathbb{E}[\mathbf{1}_{\{t < X\}}] dt = \int_0^{\infty} h'(t) P(X > t) dt. \end{aligned}$$
■

### Proposition 2.13

Suppose that  $X_i$  are independent and identically distributed random variables with  $E[|X_i|] = \infty$ . Let  $S_n = \sum_{i=1}^n X_i$ . Then

$$(a) P\{|X_n| > n \text{ for infinitely many } n\} = 1.$$

$$(b) P\left\{\frac{1}{n}S_n \text{ has finite limit}\right\} = 0.$$

*Proof.* For (a), using lemma 2.12 with  $h$  being identity,

$$\begin{aligned} \infty &= E[|X_1|] = \int_0^\infty P(|X_1| > t) dt \leq \sum_{n=0}^\infty \int_n^{n+1} P(|X_1| > t) dt \\ &\leq \sum_{n=0}^\infty \int_n^{n+1} P(|X_1| > n) dt = \sum_{n=0}^\infty P(|X_n| > n). \end{aligned}$$

Now by the second Borel-Cantelli,  $P\{|X_n| > n \text{ for infinitely many } n\} = 1$ .

For (b), consider  $\omega$  with  $\frac{S_n(\omega)}{n} \rightarrow Y(\omega) \in \mathbb{R}$ . Then for such  $\omega$ ,

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \rightarrow 0$$

as  $n \rightarrow \infty$ . Thus

$$\begin{aligned} P\left\{\frac{1}{n}S_n \text{ has finite limit}\right\} &\leq P\{|X_n| > n \text{ for finitely many } n\} \\ &= 1 - P\{|X_n| > n \text{ for infinitely many } n\} = 0. \end{aligned}$$

(b) follows. ■

### Definition 2.14

A collection of  $\sigma$ -algebra  $\{\mathcal{H}_k\}$  is **pairwise independent** if for any  $\mathcal{H}_1, \mathcal{H}_2 \in \{\mathcal{H}_k\}$ ,

$$P(A \cap B) = P(A)P(B)$$

for any  $A \in \mathcal{H}_1$  and  $B \in \mathcal{H}_2$ .

### Remark

As before, a sequence of random variables  $\{X_k\}$  is pairwise independent if  $\{\sigma(X_k)\}$  is.

### Theorem 2.15 (Strong Law of Large Number II, Kolmogorov)

Let  $X_i$  be pairwise independent, identically distributed random variables with  $E[|X_1|] < \infty$  and  $S_n = \sum_{i=1}^n X_i$ . Then

$$\frac{1}{n}S_n \rightarrow E[X_1] = \mu$$

almost surely.

*Proof.* Since we can always decompose  $X_i = X_i^+ - X_i^-$  and  $X_i^+, X_i^-$  satisfy the assumption of the theorem, we may assume without loss of generality that  $X_i \geq 0$ . Let  $Y_i = X_i \mathbf{1}\{X_i \leq i\}$  and

$T_n = \sum_{i=1}^n Y_i$ . Let  $\alpha > 1$  and put  $k_n = \lfloor \alpha^n \rfloor$ . By Chebyshev inequality, for any given  $\epsilon > 0$  we have

$$\begin{aligned}\sum_{n=1}^{\infty} P\left(\left|\frac{T_{k_n} - \mathbb{E}[T_{k_n}]}{k_n}\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{k_n^2} \text{Var}(T_{k_n}) \\ &= \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} \text{Var}(Y_i) = \frac{1}{\epsilon^2} \sum_{i=1}^{\infty} \text{Var}(Y_i) \sum_{n:k_n \geq i} \frac{1}{k_n^2}.\end{aligned}$$

Since  $1/k^2$  is summable and  $k_n$  repeat at most  $m_\alpha$  times, where  $m_\alpha$  is an integer such that  $\alpha^{m_\alpha+1} \geq \alpha^{m_\alpha} + 1$ , we can find a constant  $c_\alpha > 0$  such that

$$\sum_{n:k_n \geq i} \frac{1}{k_n^2} \leq \frac{c_\alpha}{i^2}.$$

Let  $F$  be the distribution of  $X$ . We have

$$\begin{aligned}\sum_{n=1}^{\infty} P\left(\left|\frac{T_{k_n} - \mathbb{E}[T_{k_n}]}{k_n}\right| > \epsilon\right) &\leq \frac{c_\alpha}{\epsilon^2} \sum_{i=1}^{\infty} \frac{\text{Var}(Y_i)}{i^2} \leq \frac{c_\alpha}{\epsilon^2} \sum_{i=1}^{\infty} \frac{\mathbb{E}[Y_i^2]}{i^2} \\ &= \frac{c_\alpha}{\epsilon^2} \sum_{i=1}^{\infty} \frac{1}{i^2} \int_0^i x^2 dF(x) = \frac{c_\alpha}{\epsilon^2} \sum_{i=1}^{\infty} \sum_{k=0}^{i-1} \frac{1}{i^2} \int_k^{k+1} x^2 dF(x) \\ &= \frac{c_\alpha}{\epsilon^2} \sum_{k=0}^{\infty} \left( \sum_{i=k+1}^{\infty} \frac{1}{i^2} \right) \int_k^{k+1} x^2 dF(x)\end{aligned}$$

Also, notice that there is a constant  $C$  such that

$$\sum_{i=k+1}^{\infty} \frac{1}{i^2} \leq \frac{C}{k+1}.$$

Hence,

$$\begin{aligned}\sum_{n=1}^{\infty} P\left(\left|\frac{T_{k_n} - \mathbb{E}[T_{k_n}]}{k_n}\right| > \epsilon\right) &\leq \frac{c_\alpha C}{\epsilon^2} \sum_{k=0}^{\infty} \frac{1}{k+1} \int_k^{k+1} x^2 dF(x) \\ &\leq \frac{c_\alpha C}{\epsilon^2} \sum_{k=0}^{\infty} \int_k^{k+1} x dF(x) = \frac{c_\alpha C}{\epsilon^2} \mathbb{E}[X_1] < \infty.\end{aligned}$$

Note that for  $\delta > 0$  there is an integer  $M$  such that  $\mathbb{E}[X_1 \mathbf{1}\{X_1 > M\}] \leq \delta \leq \mathbb{E}[X_1]$ .

$$\frac{\mathbb{E}[T_{k_n}]}{k_n} = \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{E}[Y_i] \geq \frac{1}{k_n} \sum_{i=1}^M \mathbb{E}[Y_i] + \frac{1}{k_n} \sum_{i=M+1}^{k_n} \mathbb{E}[X_1] - \delta.$$

Also,

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{E}[Y_i] \leq \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{E}[X_1] = \mathbb{E}[X_1].$$

Taking  $n \rightarrow \infty$  and since  $\delta$  is arbitrary, we conclude that

$$\frac{\mathbb{E}[T_{k_n}]}{k_n} \rightarrow \mathbb{E}[X_1].$$

Thus, by the Borel-Cantelli lemma,

$$\mathbb{P}\left\{\frac{T_{k_n}}{k_n} \not\rightarrow \mathbb{E}[X_1]\right\} = \mathbb{P}\left\{\left|\frac{T_{k_n} - \mathbb{E}[T_{k_n}]}{k_n}\right| > \epsilon \text{ for infinitely many } n\right\} = 0.$$

In other words,  $T_{k_n}/k_n \rightarrow \mathbb{E}[X_1]$  almost surely. Also,

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}\{X_k \neq Y_k\} &= \sum_{k=1}^{\infty} \mathbb{P}\{X_k > k\} \leq \sum_{k=1}^{\infty} \int_{k-1}^k \mathbb{P}(X_1 > t) dt \\ &= \int_0^{\infty} \mathbb{P}(X_1 > t) dt = \mathbb{E}[X_1] < \infty \end{aligned}$$

by [lemma 2.12](#). Hence by Borel-Cantelli lemma,  $X_k \neq Y_k$  for finitely many  $k$  almost surely. This implies that

$$\lim_{n \rightarrow \infty} \frac{1}{k_n} S_{k_n} = \lim_{n \rightarrow \infty} \frac{T_{k_n}}{k_n} = \mathbb{E}[X_1]$$

almost surely. Note that  $S_m$  is monotone and for each  $m$ , we may find  $k(n_m) \leq m \leq k(n_{m+1})$  so that

$$\frac{S_{k(n_m)}}{k(n_{m+1})} \leq \frac{S_m}{m} \leq \frac{S_{k(n_{m+1})}}{k(n_m)}.$$

Take  $m \rightarrow \infty$ , we conclude that

$$\frac{1}{\alpha} \mu \leq \liminf_{m \rightarrow \infty} \frac{S_m}{m} \leq \limsup_{m \rightarrow \infty} \frac{S_m}{m} \leq \alpha \mu$$

almost surely. Taking  $\alpha \rightarrow 1^+$  gives the desired result. ■

### Theorem 2.16

Let  $X_i$  be independent and identically distributed with  $\mathbb{E}[X_1^+] = \infty$  and  $\mathbb{E}[X_1^-] < \infty$ . Let  $S_n = \sum_{i=1}^n X_i$ . Then

$$\frac{1}{n} S_n \rightarrow \infty$$

almost surely.

*Proof.* Write  $X_i = X_i^+ - X_i^-$ . For  $X_i^+$ , consider  $Y_i^M = \min\{X_i^+, M\}$  for some  $M > 0$ . Note that  $Y_i^M$  is independent and identically distributed with finite mean. By the strong law of large

number,

$$\frac{1}{n} \sum_{i=1}^n Y_i^M \rightarrow \mathbb{E}[Y_1^M]$$

almost surely. Hence,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^+ \geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i^M = \mathbb{E}[Y_1^M].$$

Notice that  $Y_1^M \nearrow X_1^+$  as  $M \rightarrow \infty$ . By LMCT,

$$\lim_{M \rightarrow \infty} \mathbb{E}[Y_1^M] = \mathbb{E}[X_1^+] = \infty.$$

We conclude that  $\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^+ = \infty$ . On the other hand, by the strong law of large number,

$$\frac{1}{n} \sum_{i=1}^n X_i^- \rightarrow \mathbb{E}[X_1^-]$$

almost surely. We end up with

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^+ - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^- = \infty$$

almost surely. ■

### Example

Let  $Y_i$  be independent and identically distributed with density

$$f(y) = \mathbf{1}\{y \geq 1\} \frac{1}{c} \frac{1}{y^2},$$

where  $c$  is some normalizing constant. Let  $H_i \sim \text{Ber}(2^{-i})$ . Put  $X_i = Y_i H_i$ . Then  $\mathbb{E}[X_i] = \infty$  for all  $i$ , but since

$$\sum_i \mathbb{P}(X_i > 0) = \sum_i 2^{-i} < \infty,$$

by the Borel-Cantelli lemma,  $X_i \rightarrow 0$  almost surely and

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow 0$$

almost surely.

### Example

$Y \geq 0$  is a random variable with  $\mathbb{E}[Y] = \infty$ . Put  $X_i = Y$  for all  $i$ . Then  $X_i$  are identically

distributed with  $\mathbb{E}[X_i] = \infty$ . But

$$\frac{1}{n} \sum_{i=1}^n X_i = Y \not\rightarrow \infty$$

almost surely.

**Example (Event Streaks)**

$X_i \stackrel{iid}{\sim} \text{Ber}(2^{-1})$ . Let  $L_n$  be the longest streaks of 1 in the first  $n$  trials. We have the following:

$$\lim_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} = 1.$$

To see this, let  $\ell_n$  be the length of the current streaks. For instance, the following sequence

$$1, 0, 1, 1, 1, 0, \dots$$

generates  $\ell_1 = 1, \ell_2 = 0, \ell_6 = 4$ . Observe that  $L_n = \max_{m \leq n} \ell_m$ . Now,

$$\mathbb{P}(\ell_n \geq k) = \sum_{m=k}^n \mathbb{P}(\ell_n = k) = \sum_{m=k}^n 2^{-k-1} \leq 2^{-k}$$

as  $n \rightarrow \infty$ . For  $\epsilon > 0$ ,

$$\mathbb{P}(\ell_n \geq (1 + \epsilon) \log_2(n)) = \mathbb{P}(\ell_n \geq \lceil (1 + \epsilon) \log_2(n) \rceil) \leq 2^{-\lceil (1 + \epsilon) \log_2(n) \rceil} \leq 2^{-(1 + \epsilon) \log_2(n)} = \frac{1}{n^{1+\epsilon}}$$

is summable. By the Borel-Cantelli lemma,

$$\mathbb{P}\{\ell_n \geq (1 + \epsilon) \log_2(n) \text{ for infinitely many } n\} = 0.$$

Hence

$$\mathbb{P}\{\ell_n < (1 + \epsilon) \log_2(n) \text{ for all but finitely many } n\} = 1.$$

That is, for almost every  $\omega$ , there is  $N(\omega)$  such that  $\ell_n < (1 + \epsilon) \log_2(n)$  for  $n \geq N(\omega)$ . For such  $\omega$ , we have

$$L_n(\omega) = \max_{m \leq n} \ell_n(\omega) \leq \max_{m \leq n} (1 + \epsilon) \log_2(n) = (1 + \epsilon) \log_2(n)$$

as  $n > N(\omega)$ . Thus

$$\limsup_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \leq 1 + \epsilon$$

almost surely. Note that

$$\left\{ \limsup_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \leq 1 + \epsilon \right\} \setminus \left\{ \limsup_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \leq 1 \right\}$$

as  $\epsilon \rightarrow 0^+$  and by the monotone convergence of the measures,

$$\limsup_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \leq 1$$

almost surely.

For the other side, note that for large  $n$ , we may split the sequence into blocks of size  $\lceil(1 - \epsilon) \log_2(n)\rceil$  and

$$\frac{n}{\lceil(1 - \epsilon) \log_2(n)\rceil} \geq \frac{n}{\log_2(n)}$$

for large  $n$ .

$$\begin{aligned} \mathbb{P}(L_n \leq (1 - \epsilon) \log_2(n)) &\leq \mathbb{P}(\text{each block did not have all } 1\text{s}) \\ &\leq (1 - 2^{-\lceil(1-\epsilon) \log_2(n)/2\rceil})^{n/\lceil(1-\epsilon) \log_2(n)/2\rceil} \\ &\leq \left(1 - \frac{1}{n^{1-\epsilon}}\right)^{n^{1-\epsilon} \frac{n^\epsilon}{\log_2(n)}} \leq \exp\left(-\frac{n^\epsilon}{\log_2(n)}\right), \end{aligned}$$

which is summable, so by the Borel Cantelli lemma,

$$\mathbb{P}\{L_n \leq (1 - \epsilon) \log_2(n) \text{ for infinitely many } n\} = 0.$$

By a similar argument as above,

$$\liminf_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \geq 1 - \epsilon$$

almost surely and by the monotone convergence of the measures

$$\liminf_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \geq 1$$

almost surely. We conclude that

$$1 \leq \liminf_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \leq \limsup_{n \rightarrow \infty} \frac{L_n}{\log_2(n)} \leq 1$$

and the claim follows.

### Example (Counting Process)

Let  $X_i \in (0, \infty)$  be independent and identically distributed random variable. Put  $\mu = \mathbb{E}[X_1]$ ,  $T_n = \sum_{i=1}^n X_i$  and  $N_t = \sup\{n \mid T_n \leq t\}$ . Then we have the following claim:

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{\mu}$$

almost surely. To see this, note that since  $X_i < \infty$  for all  $i$ ,

$$\lim_{t \rightarrow \infty} N_t = \limsup_{t \rightarrow \infty} \{n \mid T_n \leq t\} = \infty.$$

Now, observe that  $T_{N_t} \leq t \leq T_{N_t+1}$  and hence

$$\frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{T_{N_t+1}}{N_t + 1} \frac{N_t + 1}{N_t}.$$

By the strong law of large number,  $T_{N_t}/N_t \rightarrow \mu$  almost surely. Thus

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{\mu}.$$

**Theorem 2.17** (Glivenko-Cantelli)

Suppose that  $X_i \stackrel{iid}{\sim} F$  with  $X_i \in (-\infty, \infty)$  and

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}$$

is the empirical CDF. Then

$$\|F_n - F\|_\infty \rightarrow 0$$

almost surely when  $n \rightarrow \infty$ .

*Proof.* We first claim that for  $\epsilon > 0$ , we may find a finite partition  $\{t_j\}$  such that  $-\infty = t_0 < \dots < t_j = \infty$  and

$$F(t_{j+1}^-) - F(t_j) \leq \epsilon$$

for all  $j$ . To see the existence of such partition, put  $t_0 = -\infty$  and let

$$t_{j+1} = \sup \{t \in \mathbb{R} \mid F(t) \leq F(t_j) + \epsilon\}.$$

Observe that  $F(t_{j+1}) \geq F(t_j) + \epsilon$ . If not, then  $F(t_{j+1}) < F(t_j) + \epsilon$ . By the right-continuity of  $F$ , there is  $\delta > 0$  such that  $F(t_{j+1} + \delta) \leq F(t_j) + \epsilon$ , contradicting to the definition of  $t_{j+1}$ . It now also follows from the definition that

$$F(t_{j+1}^-) \leq F(t_j) + \epsilon.$$

Finally, since  $F$  is of finite total variation, the jumps of sizes greater than  $\epsilon$  can occur only finitely many times and we conclude the existence of such partition.

Next, by the strong law of large number, for almost every  $\omega$  there is  $N(\omega)$  uniform in  $j$  such that

$$|F_n(t_j) - F(t_j)| \leq \epsilon$$

for all  $n > N(\omega)$ . For any  $t \in [t_j, t_{j+1})$ , we have

$$F(t) - F(t_j) \leq F(t_{j+1}^-) - F(t_j) \leq \epsilon.$$

Again, by the strong law of the large number,

$$F_n(t_{j+1}^-) - F_n(t_j) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{t_j < X_i < t_{j+1}\}} \rightarrow \mathbb{E} [\mathbf{1}_{\{t_j < X_i < t_{j+1}\}}] = F(t_{j+1}^-) - F(t_j)$$

almost surely. That is, for almost every  $\omega$ , there is  $N'(\omega) > N(\omega)$  such that for all  $j$ ,

$$F_n(t_{j+1}^-) - F_n(t_j) \leq F(t_{j+1}^-) - F(t_j) + \epsilon$$

if  $n \geq N'(\omega)$ . Combining the above estimates, if  $n \geq N'(\omega)$ ,

$$\begin{aligned} |F_n(t) - F(t)| &\leq |F_n(t) - F_n(t_j)| + |F_n(t_j) - F(t_j)| + |F(t_j) - F(t)| \\ &\leq |F_n(t_{j+1}^-) - F_n(t_j)| + 2\epsilon \\ &\leq F(t_{j+1}^-) - F(t_j) + 3\epsilon \leq 4\epsilon. \end{aligned}$$

Since  $\epsilon$  is arbitrary, we conclude that  $F_n \rightarrow F$  uniformly for almost every  $\omega$  and the proof is complete.  $\blacksquare$

### Theorem 2.18 (Kolmogorov Maximal Inequality)

Suppose that  $X_i$  are independent with  $\mathbb{E}[X_i] = 0$  and  $\text{Var}[X_i] < \infty$ . Let  $S_n = \sum_{i=1}^n X_i$ . Then

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq x\right) \leq \frac{1}{x^2} \text{Var}[S_n].$$

*Proof.* Let  $A_k = \{|S_k| \geq x \text{ and } |S_j| < x \text{ for } 1 \leq j \leq k-1\}$ . Note that

$$\sum_{k=1}^n \mathbf{1}_{A_k} = \mathbf{1}_{\left\{\max_{1 \leq k \leq n} |S_k| \geq x\right\}}.$$

$$\begin{aligned} \mathbb{E}[S_n^2] &\geq \mathbb{E}\left[S_n^2 \sum_{k=1}^n \mathbf{1}_{A_k}\right] = \sum_{k=1}^n \mathbb{E}[S_n^2 \mathbf{1}_{A_k}] = \sum_{k=1}^n \mathbb{E}[(S_n - S_k + S_k)^2 \mathbf{1}_{A_k}] \\ &= \sum_{k=1}^n \mathbb{E}[(S_n - S_k)^2 \mathbf{1}_{A_k}] + 2 \mathbb{E}[(S_n - S_k)S_k \mathbf{1}_{A_k}] + \mathbb{E}[S_k^2 \mathbf{1}_{A_k}] \\ &\geq \sum_{k=1}^n \mathbb{E}[S_k^2 \mathbf{1}_{A_k}] + 2 \mathbb{E}[(S_n - S_k)S_k \mathbf{1}_{A_k}]. \end{aligned}$$

Notice that  $S_n - S_k \in \sigma(X_{k+1}, \dots, X_n)$  and  $S_k \mathbf{1}_{A_k} \in \sigma(X_1, \dots, X_k)$  are independent. Thus

$$\begin{aligned}\mathbb{E}[S_n^2] &\geq \sum_{k=1}^n \mathbb{E}[S_k^2 \mathbf{1}_{A_k}] + 2 \mathbb{E}[(S_n - S_k) S_k \mathbf{1}_{A_k}] \\ &= \sum_{k=1}^n \mathbb{E}[S_k^2 \mathbf{1}_{A_k}] \geq x^2 \sum_{k=1}^n \mathbb{E}[\mathbf{1}_{A_k}] = x^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq x\right).\end{aligned}$$

Hence

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq x\right) \leq \frac{1}{x^2} \mathbb{E}[S_n^2].$$

■

### Definition 2.19

Let  $(\Omega, \mathcal{F}, P)$  be a probability space. A sub- $\sigma$ -algebra  $\mathcal{G}$  is **P-trivial** if for all  $A \in \mathcal{G}$ ,  $P(A) \in \{0, 1\}$ .

### Theorem 2.20 (Kolmogorov Zero-One Law)

Let  $\mathcal{F}_i$  be independent  $\sigma$ -algebras,  $\mathcal{G}_n = \sigma(\mathcal{F}_1, \dots)$  and  $\mathcal{G}_\infty = \cap_{n=1}^\infty \mathcal{G}_n$ . Then  $\mathcal{G}_\infty$  is P-trivial.

*Proof.* Observe that a  $\sigma$ -algebra  $\mathcal{G}$  satisfies that for  $A \in \mathcal{G}$ ,  $P(A) \in \{0, 1\}$  if  $\mathcal{G}$  is independent of itself. Indeed, if  $\mathcal{G}$  is independent of itself, then for any  $A \in \mathcal{G}$ ,

$$P(A) = P(A \cap A) = P(A)^2$$

implies that  $P(A) = 0$  or  $1$ . Now, for any given  $n$ ,  $\sigma(\mathcal{F}_1, \dots, \mathcal{F}_{n-1})$  is independent with  $\mathcal{G}_n$  and  $\mathcal{G}_\infty \subset \mathcal{G}_n$ . Hence  $\mathcal{G}_\infty$  is independent of  $\sigma(\mathcal{F}_1, \dots, \mathcal{F}_{n-1})$  for all  $n$ .

In particular,  $\mathcal{G}_\infty$  is independent of  $\sigma(\cup_n \mathcal{F}_n)$ . To see this, note that  $\cup_n \sigma(\cup_{k=1}^n \mathcal{F}_k)$  is a  $\pi$  system that generates  $\sigma(\cup_n \mathcal{F}_n)$  and for  $A \in \cup_n \sigma(\cup_{k=1}^n \mathcal{F}_k)$ ,  $A \in \sigma(\cup_{k=1}^n \mathcal{F}_k)$  for some  $n$ . For  $B \in \mathcal{G}_\infty$ ,

$$P(A \cap B) = P(A) P(B)$$

since  $\sigma(\cup_{k=1}^n \mathcal{F}_k)$  and  $\mathcal{G}_\infty$  is independent. Now it follows from theorem 1.52 that  $\mathcal{G}_\infty$  and  $\sigma(\cup_n \mathcal{F}_n)$  are independent. Notice that  $\mathcal{G}_\infty \subset \sigma(\cup_n \mathcal{F}_n)$  and hence  $\mathcal{G}_\infty$  is independent of itself. The proof is complete. ■

### Corollary 2.21

Let  $X_i$  be independent and identically distributed and put  $S_n = \sum_{i=1}^n X_i$ . Then

- (a)  $S_n$  is either almost surely convergent or almost surely divergent.
- (b) If  $\frac{1}{n} S_n$  converges almost surely, its limit is almost surely a constant.

*Proof.* Define  $\mathcal{F}_i = \sigma(X_i)$  and note that  $\{S_n \text{ converges}\} \in \mathcal{G}_\infty = \cap_n \sigma(\cup_{i \geq n} \mathcal{F}_i)$  since

$$\{S_n \text{ converges}\} = \left\{ \lim_{n \rightarrow \infty} \sum_{i \geq n} X_i = 0 \right\}$$

is  $\mathcal{G}_\infty$ -measurable. By the Kolmogorov zero-one law,

$$P\{S_n \text{ converges}\} \in \{0, 1\}.$$

This proves (a).

For (b), note that by a similar argument, we have

$$\left\{ \frac{1}{n} S_n \text{ converges} \right\} \in \mathcal{G}_\infty,$$

where  $\mathcal{G}_\infty$  is  $P$ -trivial. Also,  $\lim_{n \rightarrow \infty} \frac{1}{n} S_n$  is  $\mathcal{G}_\infty$ -measurable. Since  $\mathcal{G}_\infty$  is  $P$ -trivial,

$$F(t) := P\left\{ \lim_{n \rightarrow \infty} \frac{1}{n} S_n \leq t \right\} \in \{0, 1\}.$$

Thus

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n = \sup \{t \mid F(t) = 0\}$$

almost surely, proving that the limit is almost surely a constant. ■

## 2.2. Convergence in Distribution

### Definition 2.22

Let  $\mu_n$  and  $\mu$  be probability measures on  $(S, d)$ . We say that  $\mu_n \rightarrow \mu$  **in distribution** or **weakly** if for all  $f \in C_b(S, \mathbb{R})$ ,

$$\int f d\mu_n \rightarrow \int f d\mu.$$

We denote the convergence as  $\mu_n \xrightarrow{d} \mu$ .

### Remark

The notion of convergence is the smallest topology such that the linear functional of the form  $\ell_f : \mu \rightarrow \int f d\mu$  where  $f \in C_b(S)$  is continuous.

### Definition 2.23

Let  $X_n$  and  $X$  be random variables.  $X_n \xrightarrow{d} X$  if the corresponding distributions  $\mu_n \xrightarrow{d} \mu$ .

### Remark

The convergence in distribution of the random variables does not depend on the space where the random variables are defined; in fact, they can be defined on different spaces. The definition can also be written as

$$E[f(X_n)] \rightarrow E[f(X)]$$

for all  $f \in C_b(S)$ .

### Theorem 2.24 (Scheffé)

If  $f_n : S \rightarrow \mathbb{R}$  are density functions such that  $f_n \rightarrow f$  almost everywhere, where  $f$  is a density function, then

$$\sup_{B \in \mathcal{B}} \left| \int_B f_n dx - \int_B f dx \right| \rightarrow 0$$

as  $n \rightarrow 0$ . In particular, when  $S = \mathbb{R}$ , taking  $B = [-\infty, x]$  gives the uniform convergence of the CDFs.

*Proof.* Since

$$\sup_{B \in \mathcal{B}} \left| \int_B f_n dx - \int_B f dx \right| \leq \sup_{B \in \mathcal{B}} \int_B |f_n - f| dx \leq \int |f_n - f| dx,$$

the theorem follows once we prove that  $f_n \rightarrow f$  in  $L^1$ . Now, since  $|f_n - f| \rightarrow 0$  almost everywhere and

$$|f_n - f| \leq |f_n| + |f| \quad \Rightarrow \quad 0 \leq |f_n| + |f| - |f_n - f|.$$

By the assumptions that  $f_n$  and  $f$  are density functions,

$$\int f_n dx = 1 = \int f dx.$$

By the Fatou's lemma,

$$\begin{aligned} 2 \int |f| dx &= \int \liminf_{n \rightarrow \infty} |f_n| + |f| - |f_n - f| \leq \liminf_{n \rightarrow \infty} \int f_n dx + \int f dx - \int |f_n - f| dx \\ &= 2 \int f dx - \limsup_{n \rightarrow \infty} \int |f_n - f| dx. \end{aligned}$$

Hence

$$\limsup_{n \rightarrow \infty} \int |f_n - f| dx \leq 0 \quad \Rightarrow \quad \int |f_n - f| dx \rightarrow 0.$$

Hence  $f_n \rightarrow f$  in  $\mathcal{L}^1$  and the proof is complete.  $\blacksquare$

### Proposition 2.25

Let  $X_n, X$  be random variables on a separable metric space  $(S, d)$ . Then  $X_n \xrightarrow{P} X$  implies  $X_n \xrightarrow{d} X$ .

*Proof.* Let  $X_n \xrightarrow{P} X$ . To show that  $X_n \xrightarrow{d} X$ , we need to show that  $E[f(X_n)] \rightarrow E[f(X)]$ . Suppose that this does not hold. There is a subsequence  $X_{n_k}$  such that  $X_{n_k} \rightarrow X$  almost surely by theorem 2.8. By continuity we have  $f(X_{n_k}) \rightarrow f(X)$  almost surely. Since  $f$  is bounded, the bounded convergence theorem implies that  $E[f(X_{n_k})] \rightarrow E[f(X)]$ , contradicting to our hypothesis. Hence  $E[f(X_n)] \rightarrow E[f(X)]$  and  $X_n \xrightarrow{d} X$ .  $\blacksquare$

### Theorem 2.26 (Skorokhod Representation)

Suppose  $\mu_n \xrightarrow{d} \mu$  on  $\mathbb{R}$ . Then there are corresponding random variables  $X_n, X$  for  $\mu_n$  and  $\mu$  such that  $X_n \sim \mu_n$ ,  $X \sim \mu$  and  $X_n \rightarrow X$  almost surely.

*Proof.* Let  $F_n$  and  $F$  be the CDFs for  $\mu_n$  and  $\mu$ , respectively. Take  $\Omega = [0, 1]$ ,  $\mathcal{F} = \mathcal{B}$  and  $P$  be the Lebesgue measure on  $[0, 1]$ . Put

$$X_n(\omega) = \sup \{x \in \mathbb{R} \mid F_n(x) < \omega\} \quad \text{and} \quad X(\omega) = \sup \{x \in \mathbb{R} \mid F(x) < \omega\}$$

with the convention that  $\sup \emptyset = -\infty$ . Then

$$P\{X_n \leq x\} = P\{\omega \mid \omega \leq F_n(x)\} = F_n(x) \quad \text{and} \quad P\{X \leq x\} = P\{\omega \mid \omega \leq F(x)\} = F(x).$$

It now suffices to show that  $X_n \rightarrow X$  almost surely. Indeed, since  $F_n, F$  are CDFs, there are only at most countable discontinuities. Let  $\omega$  be a point of continuity of  $X$ . We may find another continuity point such that  $F(y) < \omega$ . The convergence in distribution implies that  $F_n(y) \rightarrow F(y)$ . Hence for  $n$  large enough, we have  $F_n(y) < \omega$  and hence  $X_n(\omega) > y$ . Thus

$$\liminf_{n \rightarrow \infty} X_n(\omega) \geq y$$

for all  $y \leq X(\omega)$ . Thus

$$\liminf_{n \rightarrow \infty} X_n(\omega) \geq X(\omega).$$

Similarly, pick a continuity point  $y$  such that  $F(y) \geq \omega$  would give

$$\limsup_{n \rightarrow \infty} X_n(\omega) \leq y$$

for all  $y \geq X(\omega)$  and thus

$$\limsup_{n \rightarrow \infty} X_n(\omega) \leq X(\omega).$$

Combining the above results gives that  $X_n \rightarrow X$  almost surely, since  $X$  is continuous almost surely. ■

### Corollary 2.27

Let  $g \geq 0$  be a continuous measurable function on  $\mathbb{R}$  and  $X_n \xrightarrow{d} X$ . Then

$$E[g(X)] \leq \liminf_{n \rightarrow \infty} E[g(X_n)].$$

*Proof.* Let  $Y_n$  and  $Y$  be the Skorokhod representations for  $X_n$  and  $X$ , respectively. Since now  $g(Y_n) \rightarrow g(Y)$  almost surely, the Fatou's lemma shows that

$$E[g(X)] = E[g(Y)] \leq \liminf_{n \rightarrow \infty} E[g(Y_n)] = \liminf_{n \rightarrow \infty} E[g(X_n)].$$

■

### Theorem 2.28 (Helly-Bray)

Suppose that  $X_n$  and  $X$  are  $\mathbb{R}$ -valued random variables. Then  $X_n \xrightarrow{d} X$  if and only if

$$E[g(X_n)] \rightarrow E[g(X)]$$

for all  $g \in C_b(\mathbb{R})$ .

*Proof.* Assume first that  $X_n \xrightarrow{d} X$ . By the Skorokhod representation theorem, we may assume that  $X_n$  and  $X$  are defined on the same probability space and  $X_n \rightarrow X$  almost surely. Now, since for all  $g \in C_b(\mathbb{R})$ ,  $g(X_n) \rightarrow g(X)$  almost surely and are uniformly bounded, the bounded convergence theorem implies that

$$E[g(X_n)] \rightarrow E[g(X)].$$

Conversely, suppose that  $E[g(X_n)] \rightarrow E[g(X)]$  for all  $g \in C_b(\mathbb{R})$ . Let  $F_n$  and  $F$  be the distribution functions for  $X_n$  and  $X$  respectively and  $x$  be a continuity point of  $F$ . For  $\epsilon > 0$ , consider

$$g_\epsilon(y) = \begin{cases} 1 & y \leq x \\ 1 - \frac{y-x}{\epsilon} & x < y \leq x + \epsilon \\ 0 & y \geq x + \epsilon. \end{cases}$$

Clearly  $g_\epsilon \in C_b(\mathbb{R})$ . Let  $g(y) = \mathbf{1}\{y \leq x\}$ .

$$\limsup_{n \rightarrow \infty} F_n(x) = \limsup_{n \rightarrow \infty} \mathbb{E}[g(X_n)] \leq \limsup_{n \rightarrow \infty} \mathbb{E}[g_\epsilon(X_n)] = \mathbb{E}[g_\epsilon(X)] \leq F(x + \epsilon).$$

Since  $\epsilon$  is arbitrary and  $F$  is continuous at  $x$ , we have

$$\limsup_{n \rightarrow \infty} F_n(x) \leq F(x).$$

On the other hand,

$$\liminf_{n \rightarrow \infty} F_n(x) = \liminf_{n \rightarrow \infty} \mathbb{E}[g(X_n)] \geq \liminf_{n \rightarrow \infty} \mathbb{E}[g_\epsilon(X_n + \epsilon)] = \mathbb{E}[g_\epsilon(X + \epsilon)] \geq F(x + \epsilon) \geq F(x).$$

Hence  $F_n(x) \rightarrow F(x)$  and the proof is complete.  $\blacksquare$

### Remark

The theorem gives an equivalent definition for the convergence in distribution in  $\mathbb{R}$ -valued case.

### Theorem 2.29 (Continuous Mapping Theorem)

Let  $X_n \xrightarrow{d} X$  in  $\mathbb{R}$  and  $g$  be a measurable function continuous  $\mu_X$ -almost surely. Then  $g(X_n) \xrightarrow{d} g(X)$ .

*Proof.* By the Skorokhod representation theorem, we may assume that  $X_n$  and  $X$  are on the same space and  $X_n \rightarrow X$  almost surely. By the continuity of  $g$ , we have  $g(X_n) \rightarrow g(X)$  almost surely. For all  $f \in C_b(\mathbb{R})$ ,  $f(g(X_n)) \rightarrow f(g(X))$  almost surely as well. Since  $f \circ g$  is bounded, the bounded convergence theorem gives  $\mathbb{E}[f(g(X_n))] \rightarrow \mathbb{E}[f(g(X))]$ . By the Helly-Bray theorem, this implies that  $g(X_n) \xrightarrow{d} g(X)$ .  $\blacksquare$

### Remark

If  $g$  is bounded, then  $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(X)]$  directly by applying bounded convergence theorem on  $g(X_n)$  and  $g(X)$ .

### Example

$X_n \sim U[-\frac{1}{n}, \frac{1}{n}] \xrightarrow{d} \delta_0$ . Let  $g(x) = \mathbf{1}\{x \geq 0\}$ . Then  $g(X_n) \sim Ber(\frac{1}{2}) \not\xrightarrow{d} g(X) \sim \delta_1$ .

### Definition 2.30

A metric space  $(S, d)$  is **Polish** if it is complete and separable.

### Theorem 2.31 (Portmanteau)

Suppose that  $(S, d)$  is a Polish space. Let  $X_n, X : \Omega \rightarrow S$  be random variables. Then the followings are equivalent:

$$(a) \quad X_n \xrightarrow{d} X.$$

$$(b) \quad \text{If } G \subset S \text{ is an open set, then } \liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G).$$

- (c) If  $F \subset S$  is a closed set, then  $\limsup_{n \rightarrow \infty} P(X_n \in F) \leq P(X \in F)$ .
- (d) If  $A \subset S$  satisfies  $P(X \in \partial A) = 0$ , then  $\lim_{n \rightarrow \infty} P(X_n \in A) = P(X \in A)$ .

*Proof.* We start from proving (a) implies (b). Let  $G \subset S$  be an open set. For any continuous function with  $0 \leq f \leq \mathbf{1}_G$ , we have  $E[f(X_n)] \rightarrow E[f(X)]$  and  $E[f(X_n)] \leq P(X_n \in G)$ . Hence

$$E[f(X)] \leq \liminf_{n \rightarrow \infty} P(X_n \in G).$$

Take  $f \nearrow \mathbf{1}_G$  and by LMCT, since  $f(X) \nearrow \mathbf{1}_G(X)$ ,  $E[\mathbf{1}_G(X)] \leq \liminf_{n \rightarrow \infty} P(X_n \in G)$ .

To see that (b) and (c) are equivalent, note that if  $\liminf_{n \rightarrow \infty} P(X_n \in G) \geq P(X \in G)$  for all open  $G$ , for every closed  $F$ ,

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(X_n \in F) &= -\liminf_{n \rightarrow \infty} -P(X_n \in F) = 1 - \liminf_{n \rightarrow \infty} P(X_n \in F^c) \\ &\leq 1 - P(X \in F^c) = P(X \in F). \end{aligned}$$

Conversely, if  $\limsup_{n \rightarrow \infty} P(X_n \in F) \leq P(X \in F)$  for every closed  $F$ , then

$$\liminf_{n \rightarrow \infty} P(X_n \in G) = 1 - \limsup_{n \rightarrow \infty} P(X_n \in G^c) \geq 1 - P(X \in G^c) = P(X \in G).$$

Now, assume that (b) and (c) holds. If  $A \subset S$  satisfies that  $P(X \in \partial A) = 0$ , then we have  $P(X \in \bar{A}) = P(X \in A^\circ)$ . Now,

$$\begin{aligned} P(X \in A^\circ) &\leq \liminf_{n \rightarrow \infty} P(X_n \in A^\circ) \leq \liminf_{n \rightarrow \infty} P(X_n \in A) \\ &\leq \limsup_{n \rightarrow \infty} P(X_n \in A) \leq \limsup_{n \rightarrow \infty} P(X_n \in \bar{A}) \leq P(X \in \bar{A}) = P(X \in A^\circ). \end{aligned}$$

Hence  $\lim_{n \rightarrow \infty} P(X_n \in A) = P(X \in A)$ .

Finally, assume (d) holds. We prove that (d) implies (c) and (b) implies (a). For any closed  $F$ , consider the closed  $\epsilon$ -neighborhood

$$F_\epsilon = \{s \in S \mid d(F, s) \leq \epsilon\}.$$

Now  $\partial F_\epsilon = \{s \in S \mid d(F, s) = \epsilon\}$  are disjoint for distinct  $\epsilon > 0$ . Since  $X$  can accumulate mass on only countably many such sets,  $P(X \in \partial F_\epsilon) = 0$  for almost every  $\epsilon$ . Picking  $\epsilon_k \rightarrow 0$  such that  $P(X \in \partial F_{\epsilon_k}) = 0$ , we have  $P(X_n \in F) \leq P(X_n \in F_\epsilon)$ . Taking  $n \rightarrow \infty$ ,

$$\limsup_{n \rightarrow \infty} P(X_n \in F) \leq \limsup_{n \rightarrow \infty} P(X_n \in F_{\epsilon_k}) = P(X \in F_{\epsilon_k}).$$

Now taking  $\epsilon_k \rightarrow 0$ ,

$$\limsup_{n \rightarrow \infty} P(X_n \in F) \leq P(X \in F).$$

To see that (b) implies (a), let  $f \in C_b(S)$  with  $f \geq 0$ . By [lemma 2.12](#), (b), and Fatou's lemma,

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_0^\infty \mathbb{P}(f(X) > t) dt \leq \int_0^\infty \liminf_{n \rightarrow \infty} \mathbb{P}(f(X_n) > t) dt \\ &\leq \liminf_{n \rightarrow \infty} \int_0^\infty \mathbb{P}(f(X_n) > t) dt = \liminf_{n \rightarrow \infty} \mathbb{E}[f(X_n)].\end{aligned}$$

Suppose  $f \leq C$ , replacing  $f$  with  $C - f$  gives

$$C - \mathbb{E}[f(X)] \leq \liminf_{n \rightarrow \infty} C - \mathbb{E}[f(X_n)] = C - \limsup_{n \rightarrow \infty} \mathbb{E}[f(X_n)].$$

Hence  $\mathbb{E}[f(X)] \geq \limsup_{n \rightarrow \infty} \mathbb{E}[f(X_n)]$ . We see that  $\mathbb{E}[f(X)] = \lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)]$  for  $f \geq 0$ ,  $f \in C_b(S)$ . For general  $f$ , write  $f = f^+ - f^-$  and the conclusion follows by applying above arguments to  $f^+$  and  $f^-$ .  $\blacksquare$

### Example

For a random variable  $X$  on  $\mathbb{R}^d$ , consider the perturbated version  $X_n = X + \sigma_n Z$  with  $\sigma_n \searrow 0$  and  $Z$  being independent of  $X$ ,  $\mu_Z \ll \lambda$  where  $\lambda$  is the Lebesgue measure. We claim that  $X_n \xrightarrow{d} X$ . For any open  $G$  and  $\epsilon > 0$ , define

$$G_{-\epsilon} = \left\{ x \in \mathbb{R}^d \mid \overline{B_\epsilon(x)} \subset G \right\}.$$

Then

$$\mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G_{-\epsilon}, \|\sigma_n Z\| \leq \epsilon) = \mathbb{P}(X \in G_{-\epsilon}) \mathbb{P}(\|\sigma_n Z\| \leq \epsilon).$$

Taking  $n \rightarrow \infty$ ,  $\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G_{-\epsilon})$ . Then taking  $\epsilon \rightarrow 0$  gives

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G).$$

By the portmanteau theorem,  $X_n \xrightarrow{d} X$ .

### Theorem 2.32 (Helly's Selection)

For every sequence of distribution  $F_n$  on  $\mathbb{R}$ , there is a subsequence  $F_{n_k} \rightarrow F$  pointwise at continuities where  $F$  is a non-decreasing, right-continuous function.

*Proof.* Since  $\mathbb{R}$  is separable, there is a countable dense subset  $D$ . For  $x_1 \in D$ , pick a subsequence  $F_{n_1}$  such that  $F_{n_1}$  converges at  $x_1$ . This is possible due to the Bolzano-Weierstrass theorem. Proceed with  $x_2 \in D$  and extract subsequence from  $F_{n_1}$ . Continue this process and take the diagonal. We obtain a subsequence  $F_{n_k}$  such that for every  $x_m \in D$ ,  $F_{n_k}(x_m)$  converges. For general  $x \in \mathbb{R}$ , define

$$F(x) = \inf_{y \in D, y > x} \lim_{k \rightarrow \infty} F_{n_k}(y)$$

where  $x_m \in D$  is such that  $x_m \nearrow x$ . It is clear that  $F$  is non-decreasing.

Next, for each  $x \in \mathbb{R}$  and  $\epsilon > 0$ , there is  $y \in D$  with  $y > x$  such that

$$F(y) = \lim_{k \rightarrow \infty} F_{n_k}(y) \leq F(x) + \epsilon.$$

Hence  $F$  is right-continuous.

If  $x$  is a continuity of  $F$ , then for  $\epsilon > 0$  we can also choose  $z \in D$  with  $z < x$  such that  $F(z) \geq F(x) - \epsilon$ . Now

$$F_{n_k}(z) \leq F_{n_k}(x) \leq F_{n_k}(y).$$

Taking  $k \rightarrow \infty$ ,

$$F(x) - \epsilon \leq F(z) \leq \liminf_{k \rightarrow \infty} F_{n_k}(x) \leq \limsup_{k \rightarrow \infty} F_{n_k}(x) \leq F(y) \leq F(x) + \epsilon.$$

Since  $\epsilon$  is arbitrary, we conclude that  $F_{n_k}(x) \rightarrow F(x)$ . Hence  $F$  is the desired function. ■

### Remark

A sequence of distribution function  $F_n$  converges pointwise at continuities to  $F$  does not imply that  $F$  is a distribution function. One may consider  $F_n(x) = \mathbf{1}_{\{x \geq n\}}$ . As  $n \rightarrow \infty$ ,  $F_n$  converges pointwise at continuities to  $F = 0$ .

### Definition 2.33

A collection of probability measures  $\{\mu_\alpha\}$  on  $S$  is **uniformly tight** or simply **tight** if for every  $\epsilon > 0$ , there is a compact set  $K \subset S$  such that  $\mu_\alpha(K^c) < \epsilon$  for every  $\alpha$ .

### Remark

An alternative definition is that

$$\liminf_{n \rightarrow \infty} \mu_n(B_r) \rightarrow 1$$

as  $r \rightarrow \infty$  for  $\{\mu_n\}$  with  $\mu_n$  being defined on normed space.

### Definition 2.34

A sequence of probability measures  $\{\mu_n\}$  on  $\mathbb{R}$  is said to converge **vaguely** if

$$\int f d\mu_n \rightarrow \int f d\mu$$

for all  $f \in C_c(\mathbb{R})$ .

### Remark

Since  $C_c(\mathbb{R}) \subset C_b(\mathbb{R})$ , converging weakly implies converging vaguely. Also, the  $\mu$  in the definition of the vague convergence is not necessarily a probability distribution. This is exactly what we see in the remark of Helly's selection theorem.

### Theorem 2.35

Let  $\mu_n$  be probability measures on  $\mathbb{R}^d$  such that  $\mu_n \xrightarrow{v} \mu$ . Then the followings are equivalent:

- (a)  $\{\mu_n\}$  is tight.
- (b)  $\mu$  is a probability measure, i.e.,  $\mu(\mathbb{R}^d) = 1$ .
- (c)  $\mu_n \xrightarrow{d} \mu$ .

*Proof.* Assuming (c), then

$$\int f d\mu_n \rightarrow \int f d\mu$$

for all  $f \in C_b(\mathbb{R}^d)$ . Take  $f = 1$  shows (b).

Suppose that (b) holds. For any open ball  $B_r \subset \mathbb{R}^d$ , consider  $f \in C_c(\mathbb{R}^d)$  with  $0 \leq f \leq \mathbf{1}_{B_r}$ . For every  $n$ , we have

$$\int f d\mu_n \leq \mu_n(B_r).$$

Also,

$$\int f d\mu_n \rightarrow \int f d\mu \leq \mu(B_r)$$

by the vague convergence. Thus

$$\int f d\mu = \liminf_{n \rightarrow \infty} \int f d\mu \leq \liminf_{n \rightarrow \infty} \mu_n(B_r).$$

Taking  $f \nearrow \mathbf{1}_{B_r}$  gives  $\mu(B_r) \leq \liminf_{n \rightarrow \infty} \mu_n(B_r)$ . Now taking  $r \rightarrow \infty$  gives  $\liminf_{n \rightarrow \infty} \mu_n(B_r) \rightarrow 1$  as  $r \rightarrow \infty$  and hence  $\{\mu_n\}$  is tight.

Suppose (a) is true. Fix any  $f \in C_b(\mathbb{R}^d)$ . For every open ball  $B_r$ , consider  $g_r \in C_c(\mathbb{R}^d)$  with  $f|_{B_r} \leq g_r \leq f$ . Then

$$\left| \int f d\mu_n - \int f d\mu \right| \leq \left| \int g_r d\mu_n - \int g_r d\mu \right| + \int |f - g_r| d\mu_n + \int |f - g_r| d\mu.$$

Since  $\{\mu_n\}$  is tight, for every  $\epsilon > 0$  there is some  $M$  such that  $r > M$  implies  $\mu_n(B_r^c) < \epsilon$  and

$$\left| \int f d\mu_n - \int f d\mu \right| \leq \left| \int g_r d\mu_n - \int g_r d\mu \right| + \|f\|_\infty \epsilon + \int |f - g_r| d\mu.$$

Taking  $n \rightarrow \infty$  yields

$$\limsup_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \leq \|f\|_\infty \epsilon + \int |f - g_r| d\mu.$$

Taking  $r \rightarrow \infty$  gives  $g_r \rightarrow f$  and

$$\limsup_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \leq \|f\|_\infty \epsilon.$$

Since  $\epsilon$  is arbitrary, we conclude that  $\int f d\mu_n \rightarrow \int f d\mu$  and  $\mu_n \xrightarrow{d} \mu$ . ■

**Theorem 2.36** (Prokhorov)

Suppose  $\mu_n$  is a sequence of probability measures on  $\mathbb{R}$ .  $\{\mu_n\}$  is tight if and only if every subsequence of  $\mu_n$  has a further subsequence converging weakly.

*Proof.* Suppose that  $\{\mu_n\}$  is tight. By Helly's selection theorem, for every subsequence of  $\mu_n$  we may extract a further subsequence converging vaguely to a measure  $\mu$ . By the [theorem 2.35](#),  $\mu$  is a probability measure.

Conversely, suppose that  $\{\mu_n\}$  is not tight. For every  $\epsilon > 0$  and  $k \in \mathbb{N}$ , we can pick  $\mu_{n_k}$  such that  $\mu_{n_k}(B_k^c) \geq \epsilon$ . For this sequence, we may extract a further subsequence  $\mu_{n(k_j)}$  such that it converges weakly to  $\mu$  and it follows from [theorem 2.35](#) that  $\{\mu_{n(k_j)}\}$  is tight. However  $\mu_{n(k_j)}(B_{k_j}^c) \geq \epsilon$  for every  $k_j$ , posing a contradiction. Hence  $\{\mu_n\}$  is tight. ■

### Proposition 2.37

Let  $X_n$  be a sequence of random variables. If there is  $\phi(x) \geq 0$  such that  $\phi(x) \rightarrow \infty$  as  $\|x\| \rightarrow \infty$  and  $E[\phi(X_n)] \leq C$  for all  $n$ , then  $\{X_n\}$  is tight.

*Proof.* Notice that

$$C \geq E[\phi(X_n)] \geq E[\phi(X_n) \mathbf{1}\{\|X_n\| \geq r\}] \geq \left( \inf_{\|x\| \geq r} \phi(x) \right) P(\|X_n\| \geq r).$$

Hence

$$P(\|X_n\| \geq r) \leq \frac{C}{\inf_{\|x\| \geq r} \phi(x)} \rightarrow 0$$

as  $r \rightarrow \infty$ . Thus  $\{X_n\}$  tight. ■

## 2.3. Characteristic Functions

### Definition 2.38

Let  $X$  be an  $\mathbb{R}$ -valued random variable. The characteristic function of  $X$  is defined as

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)].$$

### Remark

The characteristic function always exists for every  $t \in \mathbb{R}$  since  $x \mapsto \cos(tx)$  and  $x \mapsto \sin(tx)$  are bounded functions.

### Remark

If  $X$  has distribution  $\mu$ , we also write  $\hat{\mu} = \varphi_X$ . It is sometimes called the **Fourier transform** of the probability measure  $\mu$ .

### Example

If  $X \sim U[a, b]$ , then

$$\varphi_X(t) = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)] = \int_a^b \frac{\cos(tx)}{b-a} dx + i \int_a^b \frac{\sin(tx)}{b-a} dx = \frac{e^{ibt} - e^{iat}}{(b-a)it}.$$

### Example

If

$$X = \begin{cases} n & \text{with prob. } \frac{1}{2} \\ -n & \text{with prob. } \frac{1}{2}, \end{cases}$$

then

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \frac{1}{2}e^{int} + \frac{1}{2}e^{-int} = \cos(nt).$$

### Example

If  $X \sim \text{Poisson}(\lambda)$ , then

$$\varphi_X(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = \exp(\lambda(e^{it} - 1)).$$

### Proposition 2.39

Let  $X$  be an  $\mathbb{R}$ -valued random variable. Then the followings are true:

- (a)  $\varphi_X(0) = 1$ .
- (b)  $\varphi_X(-t) = \overline{\varphi_X(t)}$ .
- (c)  $|\varphi_X(t)| \leq 1$ .
- (d)  $\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$ .
- (e) If  $X$  and  $Y$  are independent, then  $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$ .

*Proof.* (a) is trivial. For (b),  $\varphi_X(-t) = \mathbb{E}[e^{-itX}] = \mathbb{E}[\cos(tX)] - i\mathbb{E}[\sin(tX)] = \overline{\varphi_X(t)}$ .

For (c),  $|\varphi_X(t)| \leq \mathbb{E}[|e^{itX}|] \leq \mathbb{E}[1] = 1$ .

For (d),  $\varphi_{aX+b}(t) = \mathbb{E}[e^{iatX+ibt}] = e^{ibt} \mathbb{E}[e^{iatX}] = e^{ibt} \varphi_X(at)$ .

For (e),  $\varphi_{X+Y}(t) = \mathbb{E}[e^{itX+itY}] = \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}] = \varphi_X(t)\varphi_Y(t)$ . ■

### Remark

For (e), inductively we have that for independent variables  $X_n$ ,

$$\varphi_{X_1+\dots+X_n}(t) = \prod_{i=1}^n \varphi_{X_i}(t).$$

### Example

$Z \sim N(0, 1)$ . Then

$$\begin{aligned}\varphi_Z(t) &= \mathbb{E}[\cos(tZ)] + i\mathbb{E}[\sin(tZ)] \\ &= \int \cos(tx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx + i \int \sin(tx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int \cos(tx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.\end{aligned}$$

Now

$$\frac{\varphi_Z(t) - \varphi_Z(s)}{t-s} = \int \frac{\cos(tx) - \cos(sx)}{t-s} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

By the mean value theorem,  $|\cos(tx) - \cos(sx)| \leq |\sin(c)| |tx - sx| \leq |x| |t - s|$  for some constant lying between  $tx$  and  $sx$ . Hence

$$\left| \frac{\cos(tx) - \cos(sx)}{t-s} \right| \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \leq |x| \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

which is integrable. It follows by LDCT that taking  $t \rightarrow s$

$$\begin{aligned}\varphi'_Z(s) &= - \int \sin(sx)x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \int \sin(sx) d\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\right) \\ &= -s \int \cos(sx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = -s\varphi_Z(s).\end{aligned}$$

Solving the ODE with initial condition  $\varphi_Z(0) = 1$  gives

$$\varphi_Z(t) = \exp\left(-\frac{t^2}{2}\right).$$

In general, if  $X \sim N(\mu, \sigma^2)$ ,  $X = \mu + \sigma Z$  and

$$\varphi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right).$$

**Theorem 2.40** (Inversion Formula)

Let  $\mu$  be a probability measure on  $\mathbb{R}$  with characteristic function  $\varphi$ . Then

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \mu(a, b) + \frac{1}{2}\mu\{a, b\}.$$

*Proof.* Put

$$g(T, \lambda) = \int_{-T}^T \frac{\sin(\lambda t)}{t} dt = \int_0^T \frac{\sin(\lambda t)}{t} dt.$$

Note that  $g(T, \lambda) \rightarrow 2 \operatorname{sgn}(\lambda) \frac{\pi}{2} = \pi \operatorname{sgn}(\lambda)$ . Now

$$\begin{aligned} f(T) &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt \\ &= \frac{1}{2\pi} \int_{-T}^T \int_a^b e^{-ity} \varphi(t) dy dt \\ &= \frac{1}{2\pi} \int_{-T}^T \int_a^b \int e^{-it(y-x)} d\mu(x) dy dt \\ &= \frac{1}{2\pi} \int \int_{-T}^T \int_a^b e^{-it(y-x)} dy dt d\mu(x) \\ &= \frac{1}{2\pi} \int \int_{-T}^T \frac{\sin(t(b-x)) - \sin(t(a-x))}{t} dt d\mu(x) \\ &\quad + \frac{1}{2\pi} \int \int_{-T}^T \frac{[\cos(t(b-x)) - \cos(t(a-x))]}{t} dt d\mu(x), \end{aligned}$$

where the fourth equality uses Fubini's theorem, which is valid since  $[-T, T] \times [a, b] \times \mathbb{R}$  is of finite measure ( $\mu(\mathbb{R}) = 1$ ) and  $|e^{-it(y-x)}| \leq 1$ . Notice that

$$t \mapsto \frac{[\cos(t(b-x)) - \cos(t(a-x))]}{t}$$

is a bounded odd function. Hence

$$\begin{aligned} f(T) &= \frac{1}{2\pi} \int \int_{-T}^T \frac{\sin(t(b-x)) - \sin(t(a-x))}{t} dt d\mu(x) \\ &= \frac{1}{2\pi} \int g(T, b-x) - g(T, a-x) d\mu(x). \end{aligned}$$

Since  $g(T, \lambda) \rightarrow \pi \operatorname{sgn}(\lambda)$ ,

$$g(T, b-x) - g(T, a-x) \rightarrow \begin{cases} 2\pi & \text{if } x \in (a, b) \\ \pi & \text{if } x \in \{a, b\} \\ 0 & \text{otherwise} \end{cases}$$

as  $T \rightarrow \infty$ . By the bounded convergence theorem,

$$f(T) \rightarrow \frac{1}{2\pi} (2\pi\mu(a, b) + \pi\mu\{a, b\}) = \mu(a, b) + \frac{1}{2}\mu\{a, b\}$$

as  $T \rightarrow \infty$ . ■

### Remark

The theorem establishes that the characteristic function is unique, i.e., if  $\varphi_X = \varphi_Y$ , then  $X \stackrel{d}{=} Y$ .

### Corollary 2.41

Let  $X$  be a random variable on  $\mathbb{R}$  with characteristic function  $\varphi_X$ . Then  $X$  is symmetric if and only if  $\varphi_X$  is real.

*Proof.* Suppose that  $X$  is symmetric. Then  $\varphi_X(t) = \varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$ . Hence  $\varphi_X(t)$  is real. Conversely, if  $\varphi_X$  is real,  $\varphi_X(t) = \overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$ . By the uniqueness of characteristic functions,  $X \stackrel{d}{=} -X$  and  $X$  is symmetric. ■

### Lemma 2.42

Let  $\mu$  be a probability measure on  $\mathbb{R}$  with characteristic function  $\varphi$ . Then for  $r > 0$ , we have

$$\mu\{x \mid |x| \geq r\} \leq \frac{r}{2} \int_{-2/r}^{2/r} 1 - \varphi(t) dt$$

*Proof.* By Fubini's theorem, since  $(t, x) \mapsto |1 - e^{itx}|$  is integrable on  $[-2/r, 2/r] \times \mathbb{R}$ ,

$$\begin{aligned} \int_{-2/r}^{2/r} 1 - \varphi(t) dt &= \int_{-2/r}^{2/r} \int 1 - e^{itx} d\mu(x) dt = \int \int_{-2/r}^{2/r} 1 - e^{itx} dt d\mu(x) \\ &= \frac{4}{r} \int 1 - \frac{\sin(2x/r)}{2x/r} d\mu(x) \geq \frac{2}{r} \mu \left\{ x \left| \left| \frac{2x}{r} \right| \geq 2 \right. \right\} \end{aligned}$$

since  $\sin(y) \leq y/2$  for all  $y \geq 2$ . Rearrange the inequality

$$\mu\{x \mid |x| \geq r\} \leq \frac{r}{2} \int_{-2/r}^{2/r} 1 - \varphi(t) dt.$$
■

### Proposition 2.43

Let  $\{\mu_\alpha\}$  be a set of probability measures on  $\mathbb{R}$  and  $\{\varphi_\alpha\}$  be their characteristic function. Then  $\{\mu_\alpha\}$  is tight if and only if  $\{\varphi_\alpha\}$  is equicontinuous at zero.

*Proof.* Suppose that  $\{\varphi_\alpha\}$  is equicontinuous at zero. Note that  $\varphi_\alpha(0) = 1$  for all  $\alpha$ . For every  $\epsilon > 0$ , there is  $r > 0$  such that  $1 - \varphi_\alpha(t) < \epsilon$  for all  $t \in [-2/r, 2/r]$ . Then [lemma 2.42](#) gives

$$\mu_\alpha\{x \mid |x| \geq r\} \leq \frac{r}{2} \int_{-2/r}^{2/r} 1 - \varphi_\alpha(t) dt \leq \frac{r}{2} \cdot 2 \cdot \frac{2}{r} \epsilon = 2\epsilon.$$

Since  $\epsilon$  is arbitrary,  $\{\mu_\alpha\}$  is tight.

Conversely, assume that  $\{\mu_\alpha\}$  is tight. Let  $X_\alpha \sim \mu_\alpha$  be the corresponding random variables. For any  $t \in \mathbb{R}$ ,

$$|\mu_\alpha(t) - \mu_\alpha(0)| = |\mathbb{E}[e^{itX_\alpha} - 1]| \leq \mathbb{E}[|1 - e^{itX_\alpha}|] \leq \mathbb{E}[\min\{2, |tX_\alpha|\}],$$

where the last inequality is due to  $|1 - e^{itX_\alpha}| \leq 2$  and

$$|1 - e^{itX_\alpha}| = \left| -i \int_0^{tX_\alpha} e^{is} ds \right| \leq \int_0^{|tX_\alpha|} |e^{is}| ds = |tX_\alpha|.$$

Since  $\{X_\alpha\}$  is tight, for every  $\epsilon > 0$  there is  $M > 0$  such that  $P(|X_\alpha| \geq M) < \epsilon$ . Hence

$$|\mu_\alpha(t) - \mu_\alpha(0)| \leq \mathbb{E}[\min\{2, |tX_\alpha|\}] \leq 2P(|X_\alpha| \geq M) + 2M|t| \leq 2(\epsilon + M|t|).$$

Take  $|t| \rightarrow 0$  and since  $\epsilon$  is arbitrary, the equicontinuity follows.  $\blacksquare$

#### Theorem 2.44 (Levy's Continuity Theorem)

Suppose  $\mu_n$  are random variables on  $\mathbb{R}$  with characteristic function  $\varphi_n(t)$ . Then

- (a) If  $\mu_n \xrightarrow{d} \mu$ , then  $\varphi_n(t) \rightarrow \varphi(t)$  for all  $t \in \mathbb{R}$  where  $\varphi$  is the characteristic function of  $\mu$ .
- (b) If  $\varphi_n(t) \rightarrow \varphi(t)$  and  $\varphi(t)$  is continuous at 0, then  $\{\mu_n\}$  is tight and  $\mu_n \xrightarrow{d} \mu$  where  $\mu$  has characteristic function  $\varphi$ .

*Proof.* For (a), suppose that  $\mu_n \xrightarrow{d} \mu$ . Then for every  $f \in C_b(\mathbb{R})$ ,

$$\int f d\mu_n \rightarrow \int f d\mu.$$

In particular, taking  $f : x \mapsto \cos(tx)$  and  $f : x \mapsto \sin(tx)$  shows that

$$\varphi_n(t) = \int \cos(tx) d\mu_n + i \int \sin(tx) d\mu_n \rightarrow \int \cos(tx) d\mu + i \int \sin(tx) d\mu = \varphi(t).$$

For (b), by lemma 2.42 and bounded convergence theorem,

$$\limsup_{n \rightarrow \infty} \mu_n \{x \mid |x| \geq r\} \leq \lim_{n \rightarrow \infty} \frac{r}{2} \int_{-2/r}^{2/r} 1 - \varphi_n(t) dt = \frac{r}{2} \int_{-2/r}^{2/r} 1 - \varphi(t) dt.$$

Since  $\varphi(t)$  is continuous at 0,

$$\frac{r}{2} \int_{-2/r}^{2/r} 1 - \varphi(t) dt \rightarrow 0$$

as  $r \rightarrow \infty$ . This implies that  $\{\mu_n\}$  is tight.

Now, suppose that  $\mu_n$  does not converge weakly. Then we can find  $\epsilon \geq 0$ ,  $f \in C_b$  and a subsequence  $\mu_{n_k}$  such that

$$\left| \int f d\mu_{n_k} - \int f d\mu \right| \geq \epsilon$$

for every  $k$ . By Prokhorov's theorem, we can extract a further subsequence converging weakly, which is a contradiction. Hence  $\mu_n \xrightarrow{d} \mu$ . ■

### Theorem 2.45

Suppose that  $X$  is a random variable on  $\mathbb{R}$  with  $\mathbb{E}[|X|^n] < \infty$  and has characteristic function  $\varphi(t)$ . Then  $\varphi \in C^n(\mathbb{R})$  and

$$\varphi^{(n)}(t) = \mathbb{E}[(iX)^n e^{itX}].$$

*Proof.* We prove the result for  $n = 1$ . The other orders follow inductively. For  $t \in \mathbb{R}$ ,

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \mathbb{E}\left[e^{itX} \frac{e^{ihX} - 1}{h}\right].$$

Note that

$$\left|e^{itx} \frac{e^{ihx} - 1}{h}\right| \leq \left|\frac{\cos(hx) - 1}{h}\right| + \left|\frac{\sin(hx)}{h}\right| \leq 2|x|.$$

Since  $\mathbb{E}[2|X|] < \infty$ , LDCT gives

$$\varphi'(t) = \lim_{h \rightarrow 0} \mathbb{E}\left[e^{itX} \frac{e^{ihX} - 1}{h}\right] = \mathbb{E}[e^{itX}(iX)].$$
■

### Remark

The converse does not hold in general. We provide an example where  $\varphi(x)$  is differentiable at 0 but  $\mathbb{E}[|X|] = \infty$ . Consider the random variable defined by  $P(X = \pm k) = \frac{c}{k^2 \log(k)}$  for  $k \geq 2$  where  $c$  is some normalizing constant. Then

$$\mathbb{E}[|X|] = 2c \sum_{k=2}^{\infty} \frac{1}{k \log(k)}.$$

Since the integral

$$\int_2^{\infty} \frac{1}{x \log(x)} dx = \log(\log(x))|_2^{\infty} = \infty,$$

the first moment is not finite. However,

$$\varphi(t) = 2c \sum_{k=2}^{\infty} \frac{\cos(kt)}{k^2 \log(k)}, \quad \text{and} \quad \varphi'(0) = 2c \sum_{k=2}^{\infty} \frac{0}{k \log(k)} = 0.$$

### Corollary 2.46 (Taylor Expansion)

Let  $\varphi$  be a characteristic function of a random variable  $X$  with  $\mathbb{E}[|X|^n] < \infty$ . Then as  $t \rightarrow 0$ ,

$$\varphi(t) = \sum_{k=0}^n \frac{(it)^k \mathbb{E}[X^k]}{k!} + o(t^n).$$

*Proof.* By theorem 2.45, we can consider the Taylor expansion of  $\varphi$  up to order  $n$  at  $t = 0$ .

$$\varphi(t) = \sum_{k=0}^n \frac{\varphi^{(k)}(0)t^k}{k!} + o(t^n) = \sum_{k=0}^n \frac{\mathbb{E}[(iX)^k] t^k}{k!} + o(t^n) = \sum_{k=0}^n \frac{(it)^k \mathbb{E}[X^k]}{k!} + o(t^n)$$

as  $t \rightarrow 0$ . ■

## 2.4. Central Limit Theorems

**Theorem 2.47** (Central Limit Theorem I, Lindeberg-Levy)

Let  $X_i$  be independent and identically distributed random variables with  $E[X_i] = 0$  and  $E[X_i^2] = 1$ . Let  $X \sim N(0, 1)$  and  $S_n = \sum_{i \leq n} X_i$ . Then

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{d} X$$

as  $n \rightarrow \infty$ .

*Proof.* Notice that  $\varphi_X(t) = \exp(-t^2/2)$  and by [corollary 2.46](#),

$$\varphi_{\frac{1}{\sqrt{n}} S_n}(t) = \prod_{i=1}^n \varphi_{X_i}\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + o(1/n)\right)^n \rightarrow \exp(-t^2/2)$$

as  $n \rightarrow \infty$  for fixed  $t$ . Hence  $\frac{1}{\sqrt{n}} S_n \xrightarrow{d} X$  by the Levy's continuity theorem.  $\blacksquare$

**Theorem 2.48** (Central Limit Theorem II, Lindeberg-Feller)

For each  $n$ , let  $X_{n,i}$ ,  $1 \leq i \leq n$  be random variables with mean 0 and  $S_n = \sum_{i \leq n} X_{n,i}$ . Suppose

- (a) For given  $n$ ,  $X_{n,i}$  are independent.
- (b)  $\text{Var}[\sum_{i=1}^n X_{n,i}] = \sum_{i=1}^n E[X_{n,i}^2] \rightarrow \sigma^2$ . as  $n \rightarrow \infty$ .
- (c) For all  $\epsilon > 0$ ,  $\lim_{n \rightarrow \infty} \sum_{i=1}^n E[X_{n,i}^2 \mathbf{1}\{|X_{n,i}| > \epsilon\}] = 0$ .

Then  $S_n \xrightarrow{d} N(0, \sigma^2)$ .

*Proof.* We begin the proof of the theorem by the following claim. For  $t \in \mathbb{R}$ , we have

$$\left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| \leq \min \left\{ \frac{2|t|^n}{n!}, \frac{|t|^{n+1}}{(n+1)!} \right\}.$$

To see this, let  $g_n(t)$  be the difference on the left hand side. For  $n = 0$ , the inequality holds trivially. Now suppose that the inequality holds for  $n$ .

$$g_{n+1}(t) = e^{it} - \sum_{k=0}^{n+1} \frac{(it)^k}{k!} = i \int_0^t e^{is} - \sum_{k=0}^n \frac{(is)^k}{k!} ds = i \int_0^t g_n(s) ds.$$

Now

$$|g_{n+1}(t)| \leq \int_0^t \frac{2|s|^n}{n!} ds = \frac{2|t|^{n+1}}{(n+1)!} \quad \text{and} \quad |g_{n+1}(t)| \leq \int_0^t \frac{|s|^{n+1}}{(n+1)!} ds = \frac{|t|^{n+2}}{(n+2)!}.$$

The claim follows by induction.

Now notice that for given  $n$ , if  $Z_{n,i} \sim N(0, \sigma_{n,i}^2)$  are independent with  $\sigma_{n,i}^2 = E[X_{n,i}^2]$ , then  $\sum_{i \leq n} Z_{n,i} \xrightarrow{d} Z \sim N(0, \sigma^2)$ . Let the characteristic functions for  $Z_{n,i}$  be  $\psi_{n,i}$ . Note that for

complex numbers  $z_i$  and  $w_i$  with  $|z_i|, |w_i| \leq 1$ , we have

$$\left| \prod_{i \leq n} z_i - \prod_{i \leq n} w_i \right| \leq \sum_{i \leq n} |z_i - w_i|.$$

If  $n = 2$ ,  $|z_1 z_2 - w_1 w_2| \leq |z_1 z_2 - z_1 w_2| + |z_1 w_2 - w_1 w_2| \leq |z_2 - w_2| + |z_1 - w_1|$ . The general case follows by induction. It follows that if  $\varphi_{n,i}$  are characteristic functions for  $X_{n,i}$ ,

$$\begin{aligned} \left| \prod_{i \leq n} \varphi_{n,i}(t) - \prod_{i \leq n} \psi_{n,i}(t) \right| &\leq \sum_{i \leq n} |\varphi_{n,i}(t) - \psi_{n,i}(t)| \\ &\leq \sum_{i \leq n} \left| \varphi_{n,i}(t) - 1 + \frac{\sigma_{n,i}^2 t^2}{2} \right| + \sum_{i \leq n} \left| \psi_{n,i}(t) - 1 + \frac{\sigma_{n,i}^2 t^2}{2} \right| \\ &\leq \sum_{i \leq n} \mathbb{E} \left[ \left| e^{itX_{n,i}} - 1 + \frac{X_{n,i}^2 t^2}{2} \right| \right] + \sum_{i \leq n} \mathbb{E} \left[ \left| e^{itZ_{n,i}} - 1 + \frac{Z_{n,i}^2 t^2}{2} \right| \right]. \end{aligned}$$

Now by the claim we have for all  $\epsilon > 0$ ,

$$\begin{aligned} \mathbb{E} \left[ \left| e^{itX_{n,i}} - 1 + \frac{X_{n,i}^2 t^2}{2} \right| \right] &\leq \mathbb{E} \left[ \min \left\{ \frac{2|tX_{n,i}|^2}{2!}, \frac{|tX_{n,i}|^3}{3!} \right\} \right] \\ &\leq |t|^2 \mathbb{E} \left[ |X_{n,i}|^2 \mathbf{1} \{ |X_{n,i}| > \epsilon \} \right] + \frac{|t|^3}{6} \mathbb{E} \left[ |X_{n,i}|^3 \mathbf{1} \{ |X_{n,i}| \leq \epsilon \} \right] \\ &\leq |t|^2 \mathbb{E} \left[ |X_{n,i}|^2 \mathbf{1} \{ |X_{n,i}| > \epsilon \} \right] + \frac{\epsilon |t|^3}{6} \mathbb{E} \left[ |X_{n,i}|^2 \mathbf{1} \{ |X_{n,i}| \leq \epsilon \} \right] \\ &\leq |t|^2 \mathbb{E} \left[ |X_{n,i}|^2 \mathbf{1} \{ |X_{n,i}| > \epsilon \} \right] + \frac{\epsilon |t|^3 \sigma_{n,i}^2}{6} \end{aligned}$$

Let  $\sigma_n^2 = \sum_{i \leq n} \sigma_{n,i}^2$ . The claim applies to  $Z_{n,i}$  gives

$$\begin{aligned} \sum_{i \leq n} \mathbb{E} \left[ \left| e^{itZ_{n,i}} - 1 + \frac{Z_{n,i}^2 t^2}{2} \right| \right] &\leq \frac{|t|^3}{6} \sum_{i \leq n} \mathbb{E} \left[ |Z_{n,i}|^3 \right] = \frac{|t|^3}{6} \sum_{i \leq n} \sigma_{n,i}^3 \mathbb{E} [|Y|^3] \\ &\leq \frac{|t|^3}{6} \mathbb{E} [|Y|^3] \sigma_n^2 \sup_{i \leq n} \sigma_{n,i} \end{aligned}$$

Hence

$$\begin{aligned} \left| \prod_{i \leq n} \varphi_{n,i}(t) - \prod_{i \leq n} \psi_{n,i}(t) \right| &\leq \sum_{i \leq n} |t|^2 \mathbb{E} \left[ |X_{n,i}|^2 \mathbf{1} \{ |X_{n,i}| > \epsilon \} \right] + \frac{\epsilon |t|^3 \sigma_{n,i}^2}{6} + \frac{|t|^3}{6} \mathbb{E} [|Y|^3] \sigma_n^2 \sup_{i \leq n} \sigma_{n,i} \\ &\rightarrow \frac{\epsilon |t|^3}{6} \sigma^2 + \frac{|t|^3}{6} \mathbb{E} [|Y|^3] \sigma^2 \epsilon \end{aligned}$$

as  $n \rightarrow \infty$  since

$$\begin{aligned}\sup_{i \leq n} \sigma_{n,i} &= \left( \sup_{i \leq n} \sigma_{n,i}^2 \right)^{1/2} = \left( \sup_{i \leq n} \mathbb{E} [X_{n,i}^2 \mathbf{1} \{|X_{n,i}| > \epsilon\}] + \mathbb{E} [X_{n,i}^2 \mathbf{1} \{|X_{n,i}| \leq \epsilon\}] \right)^{1/2} \\ &\leq \left( \epsilon^2 + \sum_{i \leq n} \mathbb{E} [X_{n,i}^2 \mathbf{1} \{|X_{n,i}| > \epsilon\}] \right)^{1/2} \rightarrow \epsilon\end{aligned}$$

as  $n \rightarrow \infty$ . Since  $\epsilon$  is arbitrary, we conclude that  $|\prod_{i \leq n} \varphi_{n,i}(t) - \prod_{i \leq n} \psi_{n,i}(t)| \rightarrow 0$  and  $S_n \xrightarrow{d} N(0, \sigma^2)$ .  $\blacksquare$

**Corollary 2.49** (Central Limit Theorem III, Lyapunov)

Let  $X_j$  be independent with  $\mathbb{E}[X_j] = 0$  and  $\alpha_n^2 = \sum_{j \leq n} \text{Var}(X_j)$ . Suppose that there is  $\delta > 0$  such that  $\mathbb{E}[|X_j|^{2+\delta}] < \infty$  and

$$\lim_{n \rightarrow \infty} \frac{1}{\alpha_n^{2+\delta}} \sum_{j \leq n} \mathbb{E}[|X_j|^{2+\delta}] = 0.$$

Then

$$\frac{1}{\alpha_n} \sum_{j \leq n} X_j \xrightarrow{d} N(0, 1).$$

*Proof.* Put  $Y_{n,j} = \alpha_n^{-1} X_j$ . Note that

$$\text{Var} \left[ \sum_{j \leq n} Y_{n,j} \right] = \frac{1}{\alpha_n^2} \cdot \alpha_n^2 = 1$$

and for  $\epsilon > 0$ ,

$$\sum_{j \leq n} \mathbb{E}[Y_{n,j}^2 \mathbf{1} \{|Y_{n,j}| > \epsilon\}] = \frac{1}{\alpha_n^2} \sum_{j \leq n} \mathbb{E}[X_{n,j}^2 \mathbf{1} \{ |X_{n,j}|^\delta > (\alpha_n \epsilon)^\delta \}] \leq \frac{1}{\alpha_n^{2+\delta} \epsilon^\delta} \sum_{j \leq n} \mathbb{E}[|X_{n,j}|^{2+\delta}] \rightarrow 0$$

as  $n \rightarrow \infty$ . Hence by the Lindeberg-Feller theorem,  $\alpha_n^{-1} \sum_{j \leq n} X_j \xrightarrow{d} N(0, 1)$ .  $\blacksquare$

### Example

Let  $X_j \sim \text{Ber}(p_j)$  be independent random variables and  $S_n = \sum_{j \leq n} X_j$ . Then

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} \xrightarrow{d} N(0, 1)$$

if  $\text{Var}(S_n) \rightarrow \infty$ . To see this, write  $Y_j = X_j - p_j$  and  $\sigma_j^2 = \text{Var}(Y_j) = \mathbb{E}[Y_j^2] = p_j(1 - p_j)$ . Let  $\delta > 0$  be given. Since  $|Y_j| \leq 1$ ,  $\mathbb{E}[|Y_j|^{2+\delta}] \leq \mathbb{E}[Y_j^2] = \sigma_j^2$  and

$$\frac{1}{\text{Var}(S_n)^{2+\delta}} \sum_{j \leq n} \mathbb{E}[|Y_j|^{2+\delta}] \leq \frac{1}{\text{Var}(S_n)^{2+\delta}} \sum_{j \leq n} \sigma_j^2 = \text{Var}(S_n)^{-\delta} \rightarrow 0.$$

By the Lyapunov theorem, we conclude that

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} \xrightarrow{d} N(0, 1).$$

### Theorem 2.50

For given  $n$ ,  $X_{n,j} \sim \text{Ber}(p_{n,j})$  with  $1 \leq j \leq n$  are independent. Let  $S_n = \sum_{j \leq n} X_{n,j}$ . If

$$(a) \sum_{j \leq n} p_{n,j} \rightarrow \lambda \in (0, \infty)$$

$$(b) \max_{j \leq n} p_{n,j} \rightarrow 0,$$

then  $S_n \xrightarrow{d} \text{Poisson}(\lambda)$ .

*Proof.* Let  $\varphi_{n,j}(t) = 1 + p_{n,j}(e^{it} - 1)$  be the characteristic functions for  $X_{n,j}$ . Then

$$\begin{aligned} \log(\varphi_{S_n}(t)) &= \sum_{j=1}^n \log(\varphi_{n,j}(t)) = \sum_{j=1}^n \log(1 + p_{n,j}(e^{it} - 1)) \\ &= \sum_{j=1}^n p_{n,j}(e^{it} - 1) + o(p_{n,j}) \rightarrow \lambda(e^{it} - 1), \end{aligned}$$

where we use the fact that  $\log(1 + x) = x + o(x)$  as  $x \rightarrow 0$  and condition (b) implies that the Taylor expansion is valid. Hence  $\varphi_{S_n}(t) \rightarrow \exp(\lambda(e^{it} - 1))$  and  $S_n \xrightarrow{d} \text{Poisson}(\lambda)$ . ■

### Example

The independence assumption of the random variables can sometimes be relaxed. Consider the random permutations of  $\{1, \dots, n\}$ . We are interested in finding the distribution of the number of fixed points for the random permutations. Define  $X_{n,j} = \mathbf{1}\{\pi(j) = j\}$  where  $\pi$  is a random permutation and  $S_n = \sum_{j \leq n} X_{n,j}$ . Put  $A_{n,j} = \{X_{n,j} = 1\}$ . Then

$$\mathbb{P}(S_n > 0) = \mathbb{P}(\bigcup_{j=1}^n A_{n,j}) = \sum_j \mathbb{P}(A_{n,j}) - \sum_{k \leq j} \mathbb{P}(A_{n,j} \cap \mathbb{P}(A_{n,k})) + \dots + (-1)^{n+1} \mathbb{P}(A_{n,1} \cap \dots \cap A_{n,n}).$$

Note that

$$\mathbb{P}(A_{n,i_1} \cap \dots \cap A_{n,i_r}) = \frac{(n-r)!}{n!}.$$

Hence

$$\mathbb{P}(S_n > 0) = n \cdot \frac{(n-1)!}{n!} - \binom{n}{2} \frac{(n-2)!}{n!} + \dots + (-1)^{n+1} \frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1} \frac{1}{n!}.$$

Thus  $\mathbb{P}(S_n = 0) = 1 - 1 + \frac{1}{2!} - \dots + (-1)^n \frac{1}{n!} \rightarrow e^{-1}$  as  $n \rightarrow \infty$ . Fix  $k$  entries. The remaining  $n-k$  entries with no fixed points has combinatorics  $a_{n-k}$ . Thus

$$\mathbb{P}(S_n = k) = \binom{n}{k} \frac{a_{n-k}}{n!}, \quad \text{and} \quad \mathbb{P}(S_{n-k} = 0) = \frac{a_{n-k}}{(n-k)!}.$$

So

$$P(S_n = k) = \binom{n}{k} \frac{(n-k)! P(S_{n-k} = 0)}{n!} \rightarrow \frac{1}{k!} e^{-1}$$

for given  $k$ . We see that  $S_n \xrightarrow{d} Poisson(1)$ .

### 3. Conditional Expectation and Martingale

#### 3.1. Conditional Expectation

##### Definition 3.1

Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $X \in \mathcal{L}^1(\Omega)$ . The **conditional expectation** of  $X$  with respect to a  $\sigma$ -algebra  $\mathcal{G} \subset \mathcal{F}$  is a  $\mathcal{G}$ -measurable random variable  $Z$  such that

$$\int_A Z dP = \int_A X dP$$

for every  $A \in \mathcal{G}$ .

##### Remark

Define

$$v(A) = \int_A X dP.$$

It is clear that  $v \ll P$  — in particular on  $\mathcal{G}$  — and hence the Radon-Nikodym theorem implies that there is a almost surely unique random variable  $Z \in \mathcal{L}^1(\Omega, \mathcal{G}, P)$  such that  $v(A) = \int_A Z dP$ . Hence the conditional expectation is well-defined. We may denote the conditional expectation of  $X$  with respect to  $\mathcal{G}$  as  $E[X|\mathcal{G}]$ . For random variable  $Y$ , we write  $E[X|Y] = E[X|\sigma(Y)]$ .

##### Proposition 3.2

Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{G} \subset \mathcal{F}$  is a  $\sigma$ -algebra. The followings are true:

- (a)  $E[X|\mathcal{F}] = X$ .
- (b) If  $\sigma(X)$  is independent of  $\mathcal{G} \subset \mathcal{F}$ , then  $E[X|\mathcal{G}] = E[X]$ .
- (c) For  $c \in \mathbb{R}$ ,  $E[cX + Y|\mathcal{G}] = cE[X|\mathcal{G}] + E[Y|\mathcal{G}]$ .
- (d) If  $X \leq Y$  almost surely, then  $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$  almost surely.
- (e)  $E[|E[X|\mathcal{G}]|] \leq E[|X|]$ .
- (f) If  $f$  is convex,  $f(E[X|\mathcal{G}]) \leq E[f(X)|\mathcal{G}]$ .

*Proof.* For (a), note that  $X$  is  $\mathcal{F}$ -measurable and for every  $A \in \mathcal{F}$ ,  $E[X\mathbf{1}_A] = E[X\mathbf{1}_A]$ . Hence  $X$  satisfies the required condition for it to be the conditional expectation.  $E[X|\mathcal{F}] = X$  by the uniqueness of the conditional expectation.

For (b), since the constant is  $\mathcal{G}$ -measurable and  $E[X\mathbf{1}_A] = E[X]E[\mathbf{1}_A] = E[E[X]\mathbf{1}_A]$  for every  $A \in \mathcal{G}$ , we have  $E[X|\mathcal{G}] = E[X]$ .

For (c), it is clear that  $cE[X|\mathcal{G}] + E[Y|\mathcal{G}]$  is measurable. Also, for every  $A \in \mathcal{G}$ ,

$$\begin{aligned} E[(cX + Y)\mathbf{1}_A] &= cE[X\mathbf{1}_A] + E[Y\mathbf{1}_A] = cE[E[X|\mathcal{G}]\mathbf{1}_A] + E[E[Y|\mathcal{G}]\mathbf{1}_A] \\ &= E[(cE[X|\mathcal{G}] + E[Y|\mathcal{G}])\mathbf{1}_A]. \end{aligned}$$

Hence  $E[cX + Y|\mathcal{G}] = cE[X|\mathcal{G}] + E[Y|\mathcal{G}]$ .

For (d), let  $A = \{E[Y|\mathcal{G}] - E[X|\mathcal{G}] > 0\} = \{E[Y - X|\mathcal{G}] > 0\}$ . By the definition of conditional expectation, since  $A \in \mathcal{G}$ ,

$$E[(E[Y|\mathcal{G}] - E[X|\mathcal{G}])\mathbf{1}_A] = E[E[Y - X|\mathcal{G}]\mathbf{1}_A] = E[(Y - X)\mathbf{1}_A] \geq 0$$

due to the assumption that  $X \leq Y$  almost surely. If  $P(A) > 0$ , we also have

$$E[(E[Y|\mathcal{G}] - E[X|\mathcal{G}])\mathbf{1}_A] < 0,$$

which is a contradiction. Thus  $P(A) = 0$  and we conclude that  $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$  almost surely.

For (e), we may take  $A = \{E[X|\mathcal{G}] \geq 0\}$ . Then

$$E[|E[X|\mathcal{G}]|] = E[E[X|\mathcal{G}]\mathbf{1}_A] - E[E[X|\mathcal{G}]\mathbf{1}_{A^c}] = E[X\mathbf{1}_A] - E[X\mathbf{1}_{A^c}] \leq E[|X|],$$

proving (e).

For (f), note that by the convexity, there are  $a, b$  such that  $f(x) \geq ax + b$  and  $f(x_0) = x_0$  for some  $x_0 = E[X|\mathcal{G}]$ ,

$$f(E[X|\mathcal{G}]) = aE[X|\mathcal{G}] + b = E[aX + b|\mathcal{G}] \leq E[f(X)|\mathcal{G}]$$

by (c) and (d). ■

### Proposition 3.3

Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $X_n, X \in \mathcal{L}^1(\Omega)$  and  $\mathcal{G} \subset \mathcal{F}$  be a  $\sigma$ -algebra. Then the followings are true:

- (a) If  $X_n \nearrow X$  and  $X_n \geq 0$ , then  $E[X_n|\mathcal{G}] \nearrow E[X|\mathcal{G}]$  almost surely.
- (b) If  $X_n \rightarrow X$  almost surely and  $|X_n| \leq Y \in \mathcal{L}^1(\Omega)$ , then  $E[X_n|\mathcal{G}] \rightarrow E[X|\mathcal{G}]$  almost surely.
- (c)  $E[\liminf_{n \rightarrow \infty} X_n|\mathcal{G}] \leq \liminf_{n \rightarrow \infty} E[X_n|\mathcal{G}]$  almost surely.

*Proof.* For (b), by assumption and LDCT we have  $X_n \rightarrow X$  in  $\mathcal{L}^1$ . The proposition 3.2 (e) implies that  $E[|E[X_n - X|\mathcal{G}]|] \leq E[|X_n - X|] \rightarrow 0$ . Hence (b) follows.

Now for (a), proposition 3.2 (d) implies that  $E[X_n|\mathcal{G}]$  is increasing. Since  $X_n \leq |X| \in \mathcal{L}^1$ , (b) shows that  $E[X_n|\mathcal{G}] \rightarrow E[X|\mathcal{G}]$ .

Again, by proposition 3.2 (d),  $E[\inf_{k \geq n} X_k|\mathcal{G}] \leq E[X_m|\mathcal{G}]$  for every  $m \geq n$ . Thus

$$E\left[\inf_{k \geq n} X_k|\mathcal{G}\right] \leq \inf_{m \geq n} E[X_m|\mathcal{G}].$$

Since  $\inf_{k \geq n} X_k \nearrow \liminf_{n \rightarrow \infty} X$ , we may take  $n \rightarrow \infty$  and the result from (a) gives

$$E \left[ \liminf_{n \rightarrow \infty} X_n | \mathcal{G} \right] \leq \liminf_{n \rightarrow \infty} E [X_n | \mathcal{G}]$$

as desired.  $\blacksquare$

### Proposition 3.4

If  $\sigma(X) \subset \mathcal{G}$  and  $E [|Y|], E [|XY|] < \infty$ , then  $E [XY | \mathcal{G}] = X E [Y | \mathcal{G}]$ .

*Proof.* Suppose first that  $X = \mathbf{1}_G$  where  $G \in \mathcal{G}$ . Then for any  $A \in \mathcal{G}$ ,

$$E [X E [Y | \mathcal{G}] \mathbf{1}_A] = E [\mathbf{1}_G \mathbf{1}_A E [Y | \mathcal{G}]] = E [\mathbf{1}_G Y \mathbf{1}_A] = E [E [\mathbf{1}_G Y | \mathcal{G}] \mathbf{1}_A] = E [E [XY | \mathcal{G}] \mathbf{1}_A]$$

since  $G \cap A \in \mathcal{G}$ . By the linearity this extends to the case where  $X$  is simple. For  $X \geq 0$ , there are simple functions  $X_n \nearrow X$  almost surely and hence  $X_n Y \rightarrow XY$  with  $|X_n Y| \leq |XY| \in \mathcal{L}^1$ . By proposition 3.3 (b),  $E [X_n Y | \mathcal{G}] \rightarrow E [XY | \mathcal{G}]$  almost surely. Also,  $|E [X_n Y | \mathcal{G}] \mathbf{1}_A| \leq E [|X_n Y| | \mathcal{G}] \mathbf{1}_A \leq E [|XY|] \mathbf{1}_A \in \mathcal{L}^1$ .

$$E [X E [Y | \mathcal{G}] \mathbf{1}_A] = \lim_{n \rightarrow \infty} E [X_n E [Y | \mathcal{G}] \mathbf{1}_A] = \lim_{n \rightarrow \infty} E [E [X_n Y | \mathcal{G}] \mathbf{1}_A] = E [E [XY | \mathcal{G}] \mathbf{1}_A]$$

where the first equality follows by  $|X_n E [Y | \mathcal{G}] \mathbf{1}_A| \leq |XY \mathbf{1}_A| \in \mathcal{L}^1$  and then applying proposition 3.2 (f) with  $f(x) = |x|$ ,  $X_n E [Y | \mathcal{G}] \mathbf{1}_A \rightarrow X E [Y | \mathcal{G}] \mathbf{1}_A$  and LDCT. The general case follows from the decomposition  $X = X^+ - X^-$  and  $X_n = X_n^+ - X_n^-$ .  $\blacksquare$

### Proposition 3.5 (Law of Iterated Expectation, Tower Property)

Let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$  are  $\sigma$ -algebras. Then

- (a)  $E [E [X | \mathcal{H}] | \mathcal{G}] = E [X | \mathcal{H}]$ .
- (b)  $E [E [X | \mathcal{G}] | \mathcal{H}] = E [X | \mathcal{H}]$ .

*Proof.* For (a), since  $E [X | \mathcal{H}]$  is  $\mathcal{G}$ -measurable, proposition 3.4 implies that  $E [E [X | \mathcal{H}] | \mathcal{G}] = E [X | \mathcal{H}] E [1 | \mathcal{G}] = E [X | \mathcal{H}]$ .

For (b), first note that both sides are  $\mathcal{H}$ -measurable. For  $A \in \mathcal{H} \subset \mathcal{G}$ ,

$$E [E [E [X | \mathcal{G}] | \mathcal{H}] \mathbf{1}_A] = E [E [X | \mathcal{G}] \mathbf{1}_A] = E [X \mathbf{1}_A] = E [E [X | \mathcal{H}] \mathbf{1}_A].$$

The conclusion follows.  $\blacksquare$

## **4. Stochastic Processes**

### **4.1. Poisson Process**