1 Important Notions in Measure Theory

The modern approach to study Probability Theory is built on Measure Theory. In order to follow this lecture, it is essential to have a minimum understanding and familiarity with Measure Theory. For this reason, we will start by reviewing some notions in Measure Theory: we will recall the important definitions and theorems along with a few proofs. The proofs provided here are classical ones and the techniques involved will also be seen repeatedly in Probability Theory; those which are omitted require more technical details from Measure Theory. Interested readers are invited to take a careful look at Rudin's book "Real and Complex Analysis".

1.1 Measurable Spaces and Measures

Given a set, we desire to define a measure (測度) which is a function that attributes a mass to (potentially) all the subsets. Additionally, we want a measure to satisfy some specific properties, such as additivity (加法性) or even σ-additivity (σ 加法性). Hence, we will first define measurable spaces and introduce the notion of σ-algebra, which is, vaguely speaking, a collection of subsets on which we can make sense of a measure.

1.1.1 Measurable Sets and σ-algebras

First, let us start with the notion of σ*-algebra*.

Definition 1.1.1 : Let E be a set. A subset $\mathcal{A} \subseteq \mathcal{P}(E)$ is called a σ*-algebra* (σ 代數) if the following properties are satisfied.

(a) $E \in \mathcal{A}$.

(b) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.

(c) If for all $n \in \mathbb{N}$, $A_n \in \mathcal{A}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

The elements of $\mathcal{A}$ are called *measurable sets* (可測集合), or $\mathcal{A}$-measurable sets, if one wants to emphasize the underlying σ-algebra. The couple $(E, \mathcal{A})$ is called a *measurable space* (可測空間).

Remark 1.1.2 : In the above definition, σ means that in the condition (c), the union can be any *countable* union. If we replace this condition by any *finite* union, then the corresponding definition defines what we call an *algebra* (代數), in the sense of *algebra of sets*, or *set algebra*.

In Section 1.1.2 below, we will define the notion of measure, which is closely related to the σ-algebra of the considered space. Moreover, later in Probability Theory, we will also see the importance and applications of σ-algebra. For example, in Chapter 5, we will define the notion of *conditional probability* and *conditional expectation*.

Example 1.1.3 : Given a set E. We have two extremal σ-algebras: the finest (最精緻的) is $\mathcal{A} = \mathcal{P}(E)$ and the coarsest (最粗糙) is $\mathcal{A} = \{\varnothing, E\}$. The latter is also called the trivial (平凡) σ-algebra.

In order to construct other examples of σ-algebras, we introduce the notion of *generated σ-algebra* (生成 σ 代數) .

Definition 1.1.4 : Let $\mathcal{C}$ be a subset of $\mathcal{P}(E)$. Then the smallest σ-algebra containing $\mathcal{C}$ exists, denoted by $\sigma(\mathcal{C})$ and is given by,

$$\sigma(\mathcal{C}) = \bigcap_{\substack{\mathcal{A} \text{ is a } \sigma\text{-algebra} \\ \text{s.t. } \mathcal{C} \subseteq \mathcal{A}}} \mathcal{A}.$$

In the definition, the intersection is well defined because $\mathcal{A} = \mathcal{P}(E)$ is a σ-algebra containing $\mathcal{C}$.

Definition 1.1.5 : Let E be a set, and $\mathcal{O}$ be a set of subsets of E satisfying

(a) $E \in \mathcal{O}$ and $\varnothing \in \mathcal{O}$.

(b) For any finitely many open sets $U_1, \dots, U_n \in \mathcal{O}$, we have $U_1 \cap \cdots \cap U_n \in \mathcal{O}$.

(c) For any family of open sets $(U_i)_{i \in I}$ with $U_i \in \mathcal{O}$, we have $\bigcup_{i \in I} U_i \in \mathcal{O}$.

Then, we call $(E, \mathcal{O})$ a *topological space* (拓撲空間) and the elements in $\mathcal{O}$ open sets (開集).

Definition 1.1.6 : Let $(E, \mathcal{O})$ be a topological space.

(1) We denote by $\sigma(\mathcal{O})$ the σ-algebra generated by $\mathcal{O}$, the open sets of the topological space.

(2) It is also called *Borel σ-algebra* (伯雷爾 σ 代數) of E, which can be denoted by $\mathcal{B}(E)$, if there is a canonical choice for $\mathcal{O}$.

(3) We call *Borel sets* (伯雷爾集合) the elements of $\mathcal{B}(E)$.

In other words, Borel σ-algebra is the smallest σ-algebra containing all the open sets of E.

Next, whenever we talk about topological spaces, the σ-algebra we consider will always be its Borel σ-algebra. For example, when we want to equip $\mathbb{R}$ or $\mathbb{R}^d$ with a σ-algebra to turn them into measurable spaces, we take the open sets to be given by any equivalent norm on $\mathbb{R}$ or $\mathbb{R}^d$, and the corresponding Borel σ-algebra. Actually, in the class of Probability Theory, the most of the spaces we will discuss fall into this setting.

Question 1.1.7: Prove that the σ-algebra generated by each of the three following sets of intervals is equal to $\mathcal{B}(\mathbb{R})$,

(1) $\mathcal{I} = \{(a, b) : a < b, a, b \in \mathbb{R}\}$,

(2) $\mathcal{I} = \{(-\infty, a) : a \in \mathbb{R}\}$,

(3) $\mathcal{I} = \{(-\infty, a) : a \in \mathbb{Q}\}$.

The next important notion to introduce is *prodcut σ-algebra.*

Definition 1.1.8 : Given two measurable spaces $(E_1, \mathcal{A}_1)$ and $(E_2, \mathcal{A}_2)$, we can define the *product σ-algebra* (積 σ 代數) on their product space $E_1 \times E_2$,

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2).$$

Question 1.1.9: Show that $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

1.1.2 Measures

Given a measurable space $(E, \mathcal{A})$, we want to define a *measure* on it.

Definition 1.1.10 : A function $\mu : \mathcal{A} \longrightarrow [0, \infty]$ is called a *positive measure* (正測度) or simply *measure* (測度) if the following axioms (公理) are satisfied,

- $\mu(\varnothing) = 0$,
- (σ-additivity) For any disjoint sequence $(A_n)_{n\geqslant 1}$ on $\mathcal{A}$, we have

$$\mu\Big(\bigcup_{n\geqslant 1} A_n\Big) = \sum_{n\geqslant 1} \mu(A_n). \tag{1.1}$$

Moreover, we call $(E, \mathcal{A}, \mu)$ a *measure space* (測度空間) .

Proposition 1.1.11 : *Let $(E, \mathcal{A}, \mu)$ be a measure space. We have the following properties.*

(1) If $A, B \in \mathcal{A}$, $A \subseteq B$, $\mu(A) \leqslant \mu(B)$ and $\mu(A) < \infty$, then

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

(2) If $A, B \in \mathcal{A}$, then

$$\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B).$$

(3) (Lower continuity) Let $(A_n)_{n\geqslant 1}$ be a sequence of increasing elements in $\mathcal{A}$ (i.e., $A_n \subseteq A_{n+1}$ for all n). Then, we have

$$\mu\Big(\bigcup_{n\geqslant 1} A_n\Big) = \lim_{n\to\infty} \uparrow \mu(A_n).$$

(4) (Upper continuity) Let $(B_n)_{n\geqslant 1}$ be a sequence of decreasing elements in $\mathcal{A}$ (i.e., $B_{n+1} \subseteq B_n$ for all n). If there exists n_0 such that $\mu(B_{n_0}) < \infty$, then

$$\mu\Big(\bigcap_{n\geqslant 1} B_n\Big) = \lim_{n\to\infty} \downarrow \mu(B_n).$$

(5) If $(A_n)_{n\geqslant 1}$ is a sequence of elements in $\mathcal{A}$, then

$$\mu\Big(\bigcup_{n\geqslant 1} A_n\Big) \leqslant \sum_{n\geqslant 1} \mu(A_n).$$

Proof : See Exercise 1.5. □

Definition 1.1.12 : Given a measure μ.

- The quantity $\mu(E)$ is called the *total mass* (總質量) of μ.
- If $\mu(E) = 1$, we say that μ is a *probability measure* (機率測度) .
- If $\mu(E) < \infty$, we say that μ is a *finite measure* (有限測度) .
- If there exists an increasing measurable subsets (E_n) such that $E = \cup E_n$ and that for all n, we have $\mu(E_n) < \infty$, then we say that μ is a σ*-finite measure* (σ 有限測度) .
- Let $x \in E$ and suppose that $\{x\} \in \mathcal{A}$. If $\mu(\{x\}) > 0$, then we say that x is an *atom* (原子) of μ and that μ is an *atomic measure* (原子測度) . Otherwise, we say that μ is an *atomless measure* (無原子測度) .

Example 1.1.13 :

(1) If μ is a finite measure on $(E, \mathcal{A})$, then $\frac{\mu}{\mu(E)}$ is a probability measure on the same measurable space.

(2) On the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the Lebesgue measure μ is the unique measure such that $\mu([a, b]) = \mu((a, b)) = b - a$ for any $a < b$. Its uniqueness is a consequence of the monotone class lemma, see Corollary 1.1.19 and Remark 1.1.20 for details. Its existence can be derived from Theorem 1.2.29.

(3) (Dirac measure) Given a measurable space $(E, \mathcal{A})$. Let $x \in E$ and suppose $\{x\} \in \mathcal{A}$. The measure $\mu = \delta_x$ is the *Dirac measure* (狄拉克測度) at x.

Proposition 1.1.14 (Upper and lower continuity) **:** *Let $(E, \mathcal{A})$ be a measurable space. Define $\mu : \mathcal{A} \longrightarrow [0, \infty]$ satisfying*

- $\mu(\varnothing) = 0$,
- *(Finite additivity) For any* finite *disjoint sequence of $(A_n)_{1\leqslant n\leqslant N}$ with values in $\mathcal{A}$, we have*

$$\mu\Big(\bigcup_{n=1}^{N} A_n\Big) = \sum_{n=1}^{N} \mu(A_n).$$

Then, the following properties are equivalent.

(i) (σ-additivity) μ satisfies σ-additivity mentioned in Eq. (1.1), that is, $(E, \mathcal{A}, \mu)$ is a measure space.

(ii) (Lower continuity) For any increasing sequence $(A_n)_{n\geqslant 1}$ with values in $\mathcal{A}$, we have

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \uparrow \mu(A_n).$$

(iii) (Upper continuity) For any decreasing sequence $(A_n)_{n\geqslant 1}$ with values in $\mathcal{A}$, if there exists n_0 such that $\mu(A_{n_0}) < \infty$, then we have

$$\mu\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \downarrow \mu(A_n).$$

Proof : Let us prove (i) $\Longrightarrow$ (ii). Given an increasing sequence $(A_n)_{n\geqslant 1}$, define the sequence $(B_n)_{n\geqslant 1}$ as follows. Set $B_1 := A_1$ and $B_n := A_n \backslash A_{n-1}$ for all $n \geqslant 2$. Under this construction, the elements B_n are pairwise disjoint and for all $n \geqslant 1$, we have

$$\bigcup_{k=1}^{n} B_k = A_n = \bigcup_{k=1}^{n} A_k \qquad \text{and} \qquad \bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k.$$

As a consequence of σ-additivity, we find

$$\begin{aligned}\mu\left(\bigcup_{k=1}^{\infty} A_k\right) &= \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) \\ &= \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

Let us prove (ii) $\Longrightarrow$ (i). Consider a sequence $(A_n)_{n\geqslant 1}$ of disjoint elements and define the sequence $(B_n)_{n\geqslant 1}$ as follows,

$$\forall n \geqslant 1, \qquad B_n := \bigcup_{k=1}^{n} A_k.$$

We note that $(B_n)_{n\geqslant 1}$ is an increasing sequence and using the property (ii), we find

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

Let us prove (ii) $\Longrightarrow$ (iii). Consider a decreasing sequence $(A_n)_{n\geqslant 1}$ with n_0 such that $\mu(A_{n_0}) < \infty$. Set $B_n := A_n^c \cap A_{n_0}$ for all $n \geqslant 1$. Then, the sequence $(B_n)_{n\geqslant 1}$ is increasing. Using the property (ii), we find

$$\begin{aligned}\mu\left(\bigcap_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcap_{n=n_0}^{\infty} (A_n \cap A_{n_0})\right) = \mu\left(A_{n_0} \backslash \Big(\bigcup_{n=n_0}^{\infty} B_n\Big)\right) \\ &= \mu(A_{n_0}) - \mu\left(\bigcup_{n=n_0}^{\infty} B_n\right) \\ &= \mu(A_{n_0}) - \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

To finish the proof, we use a similar method to prove (iii) $\Longrightarrow$ (ii). □

1.1.3 Monotone Classes

In the previous subsections, we defined the notion of σ-algebras and measures. However, σ-algebras are not easy to manipulate directly, since in general, one does not have a way to write down a generic element from them. Apart from this, in the definition of a measure, we talk about countable unions of *disjoint* sequences of subsets; but in the definition of a σ-algebra, this can be countable unions of *any* sequence of subsets. As a consequence, when we need to construct a measure on a measurable space, or to prove some specific conditions such as the uniqueness of the measure, it can be a bit tricky. In this subsection, we will introduce the notion of *monotone classes*, state the monotone class lemma (Theorem 1.1.18) and discuss about the uniqueness of the measure (Corollary 1.1.19).

Definition 1.1.15 : Let $\mathcal{M}$ be a subset of $\mathcal{P}(E)$. We call $\mathcal{M}$ a *monotone class* (單調類) if the following conditions are satisfied.

(a) $E \in \mathcal{M}$.

(b) If $A, B \in \mathcal{M}$ and $A \subseteq B$, then $B \backslash A \in \mathcal{M}$.

(c) If $(A_n)_{n\in\mathbb{N}}$ is an increasing sequence of elements in $\mathcal{M}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$.

Remark 1.1.16 : Any σ-algebra is also a monotone class.

As in the case of σ-algebras, we can also define the notion of a *generated monoton class* (生成單調類) .

Definition 1.1.17 : For any $\mathcal{C} \subseteq \mathcal{P}(E)$, there exists a smallest monotone class such that it contains $\mathcal{C}$, denoted

$$\mathcal{M}(\mathcal{C}) = \bigcap_{\substack{\mathcal{M} \text{ is a monotone class} \\ \text{s.t. } \mathcal{C} \subseteq \mathcal{M}}} \mathcal{M}.$$

The following lemma tells us the conditions under which a generated monotone class and a generated σ-algebra are equal.

Theorem 1.1.18 (Monotone class lemma) **:** *If $\mathcal{C} \subseteq \mathcal{P}(E)$ is closed under finite intersection, then $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.*

This is one of the most important statements in Measure Theory. We only provide the necessary steps for the proof, the complete proof being an exercise.

Proof : From the above remark, we know that $\mathcal{M}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$, so we only need to prove the other inclusion. We need to prove the two following points.

(1) If $\mathcal{M}$ is a monotone class and is closed under finite intersections, then $\mathcal{M}$ is a σ-algebra.

(2) Prove that $\mathcal{M}(\mathcal{C})$ is closed under finite intersections in two steps:

(a) Given $A \in \mathcal{C}$, let $\mathcal{M}_1 = \{B \in \mathcal{M}(C) : A \cap B \in \mathcal{M}(\mathcal{C})\}$ and prove $\mathcal{M}_1 = \mathcal{M}(\mathcal{C})$.

(b) Given $B \in \mathcal{M}(\mathcal{C})$, let $\mathcal{M}_2 = \{A \in \mathcal{M}(\mathcal{C}) : A \cap B \in \mathcal{M}(\mathcal{C})\}$ and prove $\mathcal{M}_2 = \mathcal{M}(\mathcal{C})$. □

The monotone class lemma allows us to show the uniqueness of the measure.

Corollary 1.1.19 : *Let μ and ν be two measures on the measurable space $(E, \mathcal{A})$. Assume that there exists a subset $\mathcal{C} \subseteq \mathcal{A}$ which is closed under finite intersections satisfying $\sigma(\mathcal{C}) = \mathcal{A}$ and such that for all $A \in \mathcal{C}$, we have $\mu(A) = \nu(A)$.*

(1) (Finite measure) If $\mu(E) = \nu(E) < \infty$, then $\mu = \nu$.

(2) (σ-finite measure) If there exists a increasing sequence (E_n) in $\mathcal{C}$ such that $E = \cup E_n$ and that for all n, we have $\mu(E_n) = \nu(E_n)$, then $\mu = \nu$.

Proof :

(1) Let $\mathcal{G} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$. From the hypothesis, we have $\mathcal{C} \subseteq \mathcal{G}$. Using the properties of a measure, we can show that $\mathcal{G}$ is a monotone class. Moreover, since $\mathcal{C}$ is closed under finite intersections, according to the monotone class lemma, we have $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{A}$. This proves that $\mathcal{G} = \mathcal{A}$, which is $\mu = \nu$.

(2) For all n, we can define the restricted measures of μ and ν on E_n,

$$\forall A \in \mathcal{A}, \quad \mu_n(A) = \mu(A \cap E_n), \quad \nu_n(A) = \nu(A \cap E_n).$$

From the first part of the proof, we obtain $\mu_n = \nu_n$, then from the property (3) of a measure, we get

$$\forall A \in \mathcal{A}, \quad \mu(A) = \lim \uparrow \mu_n(A \cap E_n), \quad \nu(A) = \lim \uparrow \nu_n(A \cap E_n).$$

This proves the desired statement. □

Remark 1.1.20 : The above corollary states that if there exists two measures λ and λ' on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for all intervals (a, b), we have $\lambda((a, b)) = \lambda'((a, b)) = b - a$, then they must be the same measure $\lambda = \lambda'$.

Question 1.1.21: Please explain why it is important to assume that μ and ν are finite measures or σ-finite measures in Corollary 1.1.19. In other words, is it possible to find infinite measures μ and ν such that $\mu(A) = \nu(A)$ for all $A \in \mathcal{C}$, but $\mu \neq \nu$?