0 Reminders on Discrete Probability

離散機率回顧

第零章

The goal of this chapter is to motivate a more general probability theory based on Measure Theory. In particular, we are going to start by recalling a few notions in discrete probability in Section 0.1. These notions had already been studied in the class of Introduction to Probability Theory, and we shall not focus on proving their properties. In Section 0.1.3, we are going to explain the difficulties towards a general theory, where the notion of σ-*algebras* come in naturally. In Section 0.2, we motivate the theory of Lebesgue integration from the notion of expectation, which is a weighted sum in the discrete setting, or a Lebesgue integral with Dirac masses.

0.1 Discrete probability model

In order to describe random phenomena using a mathematical language, we define the following notions.

- Sample space (finite or countable for now): describes all the possible outcomes.
- Events: subsets consisting of some possible outcomes.
- Probability: tells us how "often" an event happens.

0.1.1 Sample spaces and events

Let us start with some basic definitions and examples.

Definition 0.1.1 : In a random experiment, we denote by Ω the set consisting of all the possible outcomes (結果), called *sample space* (樣本空間) . In a discrete probability model, we require Ω to be finite or countable.

Example 0.1.2 : Below are some examples of (finite or countable) sample spaces.

(1) Toss a coin with two sides. We may take the sample space to be $\Omega = \{\text{Head}, \text{Tail}\}$.

(2) Toss a coin with two sides twice. We may take the sample space to be $\Omega = \{\text{Head}, \text{Tail}\}^2$, or $\Omega = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$.

(3) Throw a dice with six faces. We may take the sample space to be $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(4) Throw a dice with six faces until 1 appears. We may take the sample space to be $\Omega = \bigcup_{n\geqslant 0}\{2, 3, 4, 5, 6\}^n \times \{1\}$.

這個章節的目的是給出為什麼在討論更一般的機率論時，我們會需要測度論。在第 0.1 節中，我們會先回顧一些離散機率的概念，這些概念我們已經在機率導論的課程中討論過，所以我們不會著重在證明他們的性質。在第 0.1.3 小節中，我們會解釋在推廣到一般理論時會遇到的困難，這時候可以看到，σ 代數的概念會自然地出現。在第 0.2 節中，透過期望值的概念，我們帶出使用勒貝格積分的動機，因為在離散設定中，期望值是個加權和，或也可以看作是測度為狄拉克測度的勒貝格積分。

第一節 離散機率模型

為了能夠以數學的語言來描述隨機現象，我們需要定義下列概念：

- （有限或可數的）樣本空間：描述哪些是可能出現的結果；
- 事件：某些可能出現結果構成的子集合；
- 機率：告訴我們事件會「多常」發生。

第一小節 樣本空間及事件

讓我們從幾個基本定義及例子出發。

定義 0.1.1 ： 在隨機試驗中，我們將所有可能出現的結果 (outcome) 所構成的集合記為 Ω，稱之為樣本空間 (sample space) 。在離散機率模型當中，我們要求 Ω 是有限或是可數的。

範例 0.1.2 ： 下列是幾個（有限或可數）樣本空間的例子。

(1) 丟一顆只有正反兩面的銅板，我們可以把樣本空間選為 $\Omega = \{正, 反\}$。

(2) 丟一顆只有正反兩面的銅板兩次，我們可以把樣本空間選為 $\Omega = \{正, 反\}^2$，或是 $\Omega = \{正正, 正反, 反正, 反反\}$。

(3) 丟一顆有六個面的骰子，我們可以把樣本空間選為 $\Omega = \{1, 2, 3, 4, 5, 6\}$。

(4) 連續投擲一顆有六個面的骰子直到 1 出現，我們可以把樣本空間選為 $\Omega = \bigcup_{n\geqslant 0}\{2, 3, 4, 5, 6\}^n \times \{1\}$。

Remark 0.1.3 : We note that the choice of the sample space is not unique. For instance, in Example 0.1.2 (3), we may take the sample space to be $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$.

Then comes the notion of *events*. Whether a given event occurs or not depends on the outcome of the random experiment. In mathematical terms, an "event" can be seen as the "subset consisting of all the outcomes (from a random experiment) such that the event occurs".

Example 0.1.4 : Below are some examples of events.

(1) Throw a dice with six faces. Take the sample space to be $\Omega = \{1,2,3,4,5,6\}$, and below are some possible events that we may consider,

$$\begin{aligned} A &= \text{“the outcome is even”} = \{2,4,6\}, \\ B &= \text{“the outcome is a multiple of three”} = \{3,6\}, \\ C &= \text{“the outcome is even and a multiple of three”} = A \cap B = \{6\}. \end{aligned}$$

(2) Throw a dice with six faces until 1 appears. Take the sample space to be $\Omega = \bigcup_{n\geqslant 0}\{2,3,4,5,6\}^n \times \{1\}$, and below are some possible events that we may consider,

$$\begin{aligned} A &= \text{“the outcome 6 never appears”} \\ &= \{(a_k)_{1\leqslant k\leqslant n} : n \in \mathbb{N}, a_n = 1, a_k \neq 6, \forall k = 1,\dots,n\}, \\ B &= \text{“the outcome 2 appears earlier than 3”} \\ &= \{(a_k)_{1\leqslant k\leqslant n} : n \in \mathbb{N}, a_n = 1, \min\{i \geqslant 1 : a_i = 2\} < \min\{i \geqslant 1 : a_i = 3\}\}. \end{aligned}$$

Since an event can be understood as a subset of the sample space, we can apply the usual operators in Set Theory to events, defining the following terminologies.

Definition 0.1.5 : Given two events A and B in the sample space Ω, we have the following notations and interpretations.

- The event A and the event B occur at the same time, denoted $A \cap B$.
- The event A or the event B occurs, denoted $A \cup B$.
- The event A does not occur, denoted A^c.
- If $A = \Omega$, then we say the event A will definitely occur.
- If $A = \varnothing$, then we say the event A will never occur.

註解 0.1.3 ： 我們注意到，樣本空間的選擇並不是唯一的。例如說，在範例 0.1.2 (3) 中，我們也可以把樣本空間取做 $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$。

接著我們要討論的是事件的概念。給定一個事件，事件的成立與否是由隨機試驗的結果所決定的；以數學的語言來理解，「事件」可以看成是由「所有成立的（隨機試驗）結果構成的子集合」。

範例 0.1.4 ： 下列是幾個事件的例子。

(1) 丟一顆有六面的骰子，樣本空間是 $\Omega = \{1,2,3,4,5,6\}$，下列是可以考慮的事件：

$$\begin{aligned} A &= \text{「丟出來的點數是偶數」} = \{2,4,6\}, \\ B &= \text{「丟出來的點數是三的倍數」} = \{3,6\}, \\ C &= \text{「丟出來的點數是偶數也是三的倍數」} = A \cap B = \{6\}. \end{aligned}$$

(2) 連續投擲一顆有六個面的骰子直到 1 出現，我們把樣本空間取做 $\Omega = \bigcup_{n\geqslant 0}\{2,3,4,5,6\}^n \times \{1\}$，下列是可以考慮的事件：

$$\begin{aligned} A &= \text{「點數 6 永遠不出現」} \\ &= \{(a_k)_{1\leqslant k\leqslant n} : n \in \mathbb{N}, a_n = 1, a_k \neq 6, \forall k = 1,\dots,n\}, \\ B &= \text{「點數 2 比點數 3 早出現」} \\ &= \{(a_k)_{1\leqslant k\leqslant n} : n \in \mathbb{N}, a_n = 1, \min\{i \geqslant 1 : a_i = 2\} < \min\{i \geqslant 1 : a_i = 3\}\}. \end{aligned}$$

由於事件可以被看成是樣本空間的子集合，我們可以將集合論中常用到的子集合運算用在事件上面，並且定義下列名詞。

定義 0.1.5 ： 給定兩個樣本空間 Ω 中的事件 A 與 B，我們有下列描述。

- 事件 A 與事件 B 同時發生，記作 $A \cap B$。
- 事件 A 或事件 B 會發生，記作 $A \cup B$。
- 事件 A 不會發生，記作 A^c。
- 若 $A = \Omega$，則我們說事件 A 絕對會發生。
- 若 $A = \varnothing$，則我們說事件 A 絕對不會發生。

0.1.2 Discrete probability spaces

Now, we need to say what the *probability* of an event is. In other words, a probability is a function (called *probability measure*) with good properties (see the axioms in Definition 0.1.6) that assigns a number, or probability of occurrence, in $[0,1]$ to as many events as possible.

Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a subset of the power set of the sample space that contains the events to which we can assign a probability. Looking at Definition 0.1.5, it is natural to require $\mathcal{A}$

- to be closed under finite intersections,
- to be closed under finite unions,
- to be closed by taking the complement,
- to contain Ω,
- and to contain $\varnothing$.

Clearly, $\mathcal{A} = \mathcal{P}(\Omega)$ satisfies all the above properties, and ideally, we want to take $\mathcal{A} = \mathcal{P}(\Omega)$, that is to say, a probability measure is defined on *all* the subsets of the sample space Ω. In practice, however, this can only be achieved for *finite or countable* sample spaces, but not for any sample space. We will see a counterexample in Example 0.1.13.

Below, we introduce the axioms of discrete probability. Note that we may take $\mathcal{A} = \mathcal{P}(\Omega)$, and the property (2) can be strengthened to "to be stable under countable unions". The reason of this will be discussed later in Proposition 0.1.11.

Definition 0.1.6 (Axioms of discrete probability) **:** Given a finite or countable sample space Ω, and $\mathcal{A} = \mathcal{P}(\Omega)$ to be the collection of all the subsets of Ω. A function $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ is said to be a *discrete probability* (離散機率) on $(\Omega, \mathcal{A})$ if

- $\mathbb{P}(\Omega) = 1$;
- (σ-additivity) For any pairwise disjoint sequence $(A_n)_{n\geqslant 1}$ of elements in $\mathcal{A}$, we have

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \tag{0.1}$$

In addition, the elements in $\mathcal{A} = \mathcal{P}(\Omega)$ are called *measurable events* (可測事件) and the triplet $(\Omega, \mathcal{A}, \mathbb{P})$ is called *discrete probability space* (離散機率空間) .

Example 0.1.7 : If Ω is finite, μ is the counting measure, that is $\mu(A) = \mathrm{Card}(A)$ for $A \subseteq \mathcal{P}(\Omega)$. Then $\mathbb{P} = \mathrm{Card}(\Omega)^{-1}\mu$ is a probability measure, which is the *uniform measure* on Ω.

第二小節 離散機率空間

現在，我們要來討論什麼是事件的機率。換句話說，機率是個有一些好性質的函數（稱作機率測度），如同在定義 0.1.6 中的公理所描述的。機率測度會對盡可能多的事件給予一個在 $[0,1]$ 中的數字，代表事件發生的機率。

令 $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ 為樣本空間冪集合的子集合，當中的元素代表我們可以給予機率的事件。透過觀察定義 0.1.5 ，很自然地，我們會希望 $\mathcal{A}$

- 在有限交集下是封閉的；
- 在有限聯集下是封閉的；
- 取補集時是封閉的；
- 包含 Ω；
- 以及包含 $\varnothing$。

顯然地，$\mathcal{A} = \mathcal{P}(\Omega)$ 滿足上面所有的性質，且理想上來說，我們想取 $\mathcal{A} = \mathcal{P}(\Omega)$，也就是機率測度可以在樣本空間 Ω 的所有子集合上定義。在實際情況中，這對於有限或可數的樣本空間來說是可行的，但對任意樣本空間來說是不可能的。稍後在範例 0.1.13 中我們會看到反例。

定義 0.1.6 【離散機率公理】 ： 給定有限或可數的樣本空間 Ω，我們記 $\mathcal{A} := \mathcal{P}(\Omega)$ 為由所有 Ω 子集合構成的集合。若函數 $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ 滿足

- $\mathbb{P}(\Omega) = 1$；
- 【σ 可加性】對任意由 $\mathcal{A}$ 中元素構成的兩兩互不相交序列 $(A_n)_{n\geqslant 1}$，我們有

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n), \tag{0.1}$$

則我們稱之為在 $(\Omega, \mathcal{A})$ 上的離散機率 (discrete probability) 。

此外，我們將 $\mathcal{A} = \mathcal{P}(\Omega)$ 中的元素稱作可測事件 (measurable events) ，將三元組 $(\Omega, \mathcal{A}, \mathbb{P})$ 稱之為離散機率空間 (discrete probability space) 。

範例 0.1.7 ： 如果 Ω 是有限的，μ 是計數測度，也就是說對於 $A \subseteq \mathcal{P}(\Omega)$，我們有 $\mu(A) = \mathrm{Card}(A)$。那麼，$\mathbb{P} = \mathrm{Card}(\Omega)^{-1}\mu$ 是個機率測度，他是在 Ω 上的均勻測度。

You may easily check the following properties for a discrete probability space.

Proposition 0.1.8 : *Given a discrete probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathcal{A} = \mathcal{P}(\Omega)$, then the below properties hold.*

(1) $\mathbb{P}(\varnothing) = 0$.

(2) *We have* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *for any event* $A \in \mathcal{A}$.

(3) *(Monotonicity) For any events* $A, B \in \mathcal{A}$ *with* $A \subseteq B$, *we have* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.

(4) *(Finite-additivity) For any finite sequence* $A_1, \ldots, A_n$ *of pairwise disjoint events from* $\mathcal{A}$, *we have*

$$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathbb{P}(A_k). \tag{0.2}$$

(5) *(sub-σ-additivity) Let* I *be finite or countable and a family* (組) $(A_i)_{i\in I}$ *of events from* $\mathcal{A}$ *indexed by* I. *Then,*

$$\mathbb{P}\Big(\bigcup_{i\in I} A_i\Big) \leqslant \sum_{i\in I} \mathbb{P}(A_i).$$

(6) *For any events* $A, B \in \mathcal{A}$, *we have*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Remark 0.1.9 : In this proposition, the point (4) states finite additivity, which is weaker than σ-additivity required in point (2) of Definition 0.1.6 in general. If the sample space Ω is finite, since there are finitely many events in $\mathcal{P}(\Omega)$, it is not hard to see that finite additivity in the point (4) and σ-additivity in point (2) of Definition 0.1.6 are equivalent. If the sample space Ω is infinite but countable, it is possible to construct an example that is finitely additive but not σ-additive, see Example 0.1.10.

Example 0.1.10 : Consider the triplet $(\mathbb{N}, \mathcal{A}, \mathbb{P})$, where

$$\mathcal{A} = \{A \subseteq \mathbb{N} : A \text{ or } A^c \text{ is finite}\}$$

is an algebra (i.e. satisfying the five conditions in the beginning of Section 0.1.2) and $\mathbb{P}$ is defined as below,

$$\forall A \in \mathcal{A}, \quad \mathbb{P}(A) = \begin{cases} 0 & \text{if } A \text{ is finite,} \\ 1 & \text{if } A^c \text{ is finite.} \end{cases}$$

We can check that $\mathbb{P}$ is finitely additive but not σ-additive. In fact, we have

$$\mathbb{P}(\mathbb{N}) = 1 \neq 0 = \sum_{n\geqslant 1} \mathbb{P}(\{n\})$$

你可以輕易檢查下列離散機率空間的性質。

命題 0.1.8 ： 給定離散機率空間 $(\Omega, \mathcal{A}, \mathbb{P})$，其中 $\mathcal{A} = \mathcal{P}(\Omega)$，則下列性質成立。

(1) $\mathbb{P}(\varnothing) = 0$。

(2) 對於所有事件 $A \in \mathcal{A}$，我們有 $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$。

(3) 【單調性】對於任意事件 $A, B \in \mathcal{A}$ 滿足 $A \subseteq B$，我們有 $\mathbb{P}(A) \leqslant \mathbb{P}(B)$。

(4) 【有限可加性】對於任意由兩兩互不相交 $\mathcal{A}$ 中的事件構成的有限序列 $A_1, \ldots, A_n$，我們有

$$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathbb{P}(A_k). \tag{0.2}$$

(5) 【子 σ 可加性】令 I 為有限或可數集合及由 I 所編號，從 $\mathcal{A}$ 中取出的事件組 (family) $(A_i)_{i\in I}$，則

$$\mathbb{P}\Big(\bigcup_{i\in I} A_i\Big) \leqslant \sum_{i\in I} \mathbb{P}(A_i).$$

(6) 對於任意事件 $A, B \in \mathcal{A}$，我們有

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

註解 0.1.9 ： 在此命題中，第 (4) 點描述的是有限可加性，一般來講，此性質比在定義 0.1.6 (2) 中要求的 σ 可加性還要來得弱。如果樣本空間 Ω 是有限的，則由於 $\mathcal{P}(\Omega)$ 中最多只有有限多種不同的事件，因此不難看出來，此性質 (4) 中的有限可加性與定義 0.1.6 (2) 中要求的 σ 可加性等價。若樣本空間 Ω 是無窮但可數的，是有辦法構造滿足有限可加性但不滿足 σ 可加性的範例，見範例 0.1.10 。

範例 0.1.10 ： 考慮三元組 $(\mathbb{N}, \mathcal{A}, \mathbb{P})$，其中

$$\mathcal{A} = \{A \subseteq \mathbb{N} : A \text{ 或 } A^c \text{ 是有限的}\}$$

是個代數（也就是滿足第 0.1.2 小節中開頭的前五個條件），且 $\mathbb{P}$ 定義如下：

$$\forall A \in \mathcal{A}, \quad \mathbb{P}(A) = \begin{cases} 0 & \text{若 } A \text{ 是有限的,} \\ 1 & \text{若 } A^c \text{ 是有限的.} \end{cases}$$

我們可以檢查 $\mathbb{P}$ 滿足有限可加性，但並不滿足 σ 可加性。事實上，我們有

$$\mathbb{P}(\mathbb{N}) = 1 \neq 0 = \sum_{n\geqslant 1} \mathbb{P}(\{n\})$$

The following proposition shows us the importance of σ-additivity for the probability $\mathbb{P}$. To be more precise, if we wish $\mathbb{P}$ to be *continuous*, then σ-additivity is necessary and sufficient.

Proposition 0.1.11 (Upper and lower continuity) : *Let Ω be a finite or countable sample space, $\mathbb{P}$: $\mathcal{P}(\Omega) \longrightarrow [0,1]$ be a function that satisfies $\mathbb{P}(\Omega) = 1$ and finite additivity mentioned in Eq. (0.2). The conditions below are equivalent.*

(i) *(σ-additivity) $\mathbb{P}$ satisfies σ-additivity mentioned in Eq. (0.3), that is, $(\Omega, \mathbb{P})$ is a* discrete probability space.

(ii) *(Lower continuity) For any increasing sequence $(A_n)_{n\geqslant 1}$ of events, that is, $A_n \subseteq A_{n+1} \subset \Omega$ for all $n \geqslant 1$, then we have*

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

(iii) *(Upper continuity) For any decreasing sequence $(A_n)_{n\geqslant 1}$ of events, that is $A_{n+1} \subseteq A_n \subset \Omega$ for all $n \geqslant 1$, then we have*

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

Proof : It is a direct consequence of the axioms from Definition 0.1.6 and the properties from Proposition 0.1.8. You may try to prove the equivalence between different statements by yourself. We will also discuss this again in Proposition 0.1.11. □

Corollary 0.1.12 (Countable intersection and union) : *For any sequence $(A_n)_{n\geqslant 1}$ of events, we have*

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big),$$
$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcap_{k=1}^{n} A_k\Big).$$

0.1.3 Towards a general probability space

So far we have defined the notion of discrete probability space. In order to make a more general definition, in the case where the sample space is uncountable, we need to see what may go wrong and adapt the definitions accordingly. From the discussion in Section 0.1.2, we see that the tricky part lies in the definition of a probability measure. Example 0.1.13 below illustrates that on $\Omega = [-1, 2]$, which is uncountable, we cannot define a "uniform probability measure" on $\mathcal{A} = \mathcal{P}(\Omega)$.

下面的命題 0.1.11 告訴我們 σ 可加性對機率 $\mathbb{P}$ 的重要性：當我們希望 $\mathbb{P}$ 有連續性時，σ 可加性會是充分且必要的。

命題 0.1.11 【上連續與下連續】 ： 令 Ω 為有限或可數的樣本空間，$\mathbb{P} : \mathcal{P}(\Omega) \longrightarrow [0,1]$ 為滿足 $\mathbb{P}(\Omega) = 1$ 及式 (0.2) 中有限可加性的函數。下列為等價條件：

(i) 【σ 可加性】$\mathbb{P}$ 滿足式 (0.3) 中的 σ 可加性，也就是說 $(\Omega, \mathbb{P})$ 是個離散機率空間。

(ii) 【下連續】對於任意遞增的事件序列 $(A_n)_{n\geqslant 1}$，也就是說，對於所有 $n \geqslant 1$，序列滿足 $A_n \subseteq A_{n+1} \subset \Omega$，則我們有

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

(iii) 【上連續】對於任意遞減的事件序列 $(A_n)_{n\geqslant 1}$，也就是說，對於所有 $n \geqslant 1$，序列滿足 $A_{n+1} \subseteq A_n \subset \Omega$，則我們有

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

證明：這是個能夠從定義 0.1.6 中引理，以及命題 0.1.8 中性質所推得的直接結果。你可以嘗試自己證明不同敘述之間的等價關係。我們也會在命題 0.1.11 當中重新做討論。 □

系理 0.1.12 【可數交集與聯集】 ： 對於任意的事件序列 $(A_n)_{n\geqslant 1}$，我們有

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big),$$
$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcap_{k=1}^{n} A_k\Big).$$

第三小節 推廣到一般機率空間

到目前為止，我們定義了離散機率空間的概念。為了能夠在更廣泛的框架下，像是當樣本空間為不可數時，來定義這些概念，我們需要去檢查哪些定義可能會出錯，並且去做出相對應的修改。從第 0.1.2 小節中的討論，我們看出來定義機率測度是比較麻煩的。下面的範例 0.1.13 會告訴我們，在不可數的 $\Omega = [-1, 2]$ 上，我們無法在 $\mathcal{A} = \mathcal{P}(\Omega)$ 上定義「均勻機率測度」。

Example 0.1.13 (Vitali set) **:** Let us define the following equivalence relation $x\mathcal{R}y \Leftrightarrow x - y \in \mathbb{Q}$ on the set of real numbers $\mathbb{R}$. This provides us with the set $\mathbb{R}/\mathbb{Q}$ of equivalence classes and for each equivalence class $\bar{x} \in \mathbb{R}/\mathbb{Q}$, we may pick up a representant v_x on $[0,1]$. We consider the Vitali set defined as below,

$$V = \{v_x : x \in \mathbb{R}/\mathbb{Q}\}.$$

The set $[0,1] \cap \mathbb{Q}$ is countable and we can enumerate its elements as $\{q_k : k \in \mathbb{N}\}$ then define $V_k = V + q_k$. It is not hard to prove the following inclusion relation,

$$[0,1] \subseteq \bigcup_{k \geqslant 1} V_k \subseteq [-1,2].$$

We want to consider the uniform probability $\mathbb{P}$ on $\Omega = [-1,2]$, that is such that $\mathbb{P}(I) = \frac{1}{3}(b-a)$ for any interval $I = [a,b] \subseteq \Omega$. Moreover, we may also notice that it is translation invariant, i.e., $\mathbb{P}(I) = \mathbb{P}(I+q)$ for any interval I such that I and $I+q$ are both included in Ω. If the probability $\mathbb{P}$ can be defined on all the subsets of $\mathcal{P}(\Omega)$, then the σ-additivity implies that

$$\frac{1}{3} \leqslant \sum_{k \geqslant 1} \mathbb{P}(V_k) = \sum_{k \geqslant 1} \mathbb{P}(V) \leqslant 1,$$

which is impossible. In fact, if $\mathbb{P}(V) = 0$, then the inequality on the left is not satisfied; if $\mathbb{P}(V) > 0$, the inequality on the right is not satisfied.

As a consequence, we should define a probability measure on a *proper subset* of $\mathcal{P}(\Omega)$ instead of the whole power set. To be more precise, we need to make a choice of a subset $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ with nice properties, so that a probability measure can be defined on it.

Definition 0.1.14 (Axioms of general probability) **:** Given a sample space Ω, and a subset $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ satisfying

(a) $\Omega \in \mathcal{A}$,

(b) for $A \in \mathcal{A}$, we also have $A^c \in \mathcal{A}$,

(c) for any sequence $(A_n)_{n\geqslant 1}$ with elements in $\mathcal{A}$, we have $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$.

A function $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ is said to be a *probability measure* (機率測度) on $(\Omega, \mathcal{A})$ if

- $\mathbb{P}(\Omega) = 1$;
- (σ-additivity) For any pairwise disjoint sequence $(A_n)_{n\geqslant 1}$ of elements in $\mathcal{A}$, we have

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \tag{0.3}$$

In addition, the elements in $\mathcal{A}$ are called *measurable events* (可測事件) and the triplet $(\Omega, \mathcal{A}, \mathbb{P})$ is called *general probability space* (一般機率空間).

範例 0.1.13 【Vitali 集合】**：** 在實數集 $\mathbb{R}$ 上，我們定義下列等價關係：$x\mathcal{R}y \Leftrightarrow x - y \in \mathbb{Q}$。透過此等價關係，我們可以得到由等價類構成的集合 $\mathbb{R}/\mathbb{Q}$，且對每個等價類 $\bar{x} \in \mathbb{R}/\mathbb{Q}$，我們可以取一個在 $[0,1]$ 之間的代表 v_x。我們考慮 Vitali 集合，定義如下：

$$V = \{v_x : x \in \mathbb{R}/\mathbb{Q}\}.$$

集合 $[0,1] \cap \mathbb{Q}$ 為可數，我們可以將裡面的元素編號 $\{q_k : k \in \mathbb{N}\}$，並且定義 $V_k = V + q_k$，則我們不難證明下列包含關係：

$$[0,1] \subseteq \bigcup_{k \geqslant 1} V_k \subseteq [-1,2].$$

若在 $\Omega = [-1,2]$ 上，我們想考慮均勻機率 $\mathbb{P}$，則他會滿足對於任意區間 $I = [a,b] \subseteq \Omega$，我們會有 $\mathbb{P}(I) = \frac{1}{3}(b-a)$。此外，我們還可以注意到，這樣的機率會是平移不變的，也就是說，如果 I 與 $I+q$ 皆包含在 Ω 中，我們有 $\mathbb{P}(I) = \mathbb{P}(I+q)$。假設機率 $\mathbb{P}$ 可以定義在所有子集合 $\mathcal{P}(\Omega)$ 上，則根據 σ 可加性，我們會有

$$\frac{1}{3} \leqslant \sum_{k \geqslant 1} \mathbb{P}(V_k) = \sum_{k \geqslant 1} \mathbb{P}(V) \leqslant 1,$$

而這是不可能的。這原因是，我們可以看出來，如果 $\mathbb{P}(V) = 0$，那麼左方的不等式不滿足；如果 $\mathbb{P}(V) > 0$，右方的不等式不滿足。

因此，我們需要把機率測度定義在 $\mathcal{P}(\Omega)$ 的嚴格子集合上，而不是整個冪集合上。更確切來說，我們需要選擇一個滿足好性質的子集合 $\mathcal{A} \subseteq \mathcal{P}(\Omega)$，使得我們能夠在上面定義機率測度。

定義 0.1.14 【一般機率公理】**：** 給定樣本空間 Ω，以及子集合 $\mathcal{A} \subseteq \mathcal{P}(\Omega)$，滿足：

(a) $\Omega \in \mathcal{A}$；

(b) 對於 $A \in \mathcal{A}$，我們也有 $A^c \in \mathcal{A}$；

(c) 對於任意元素在 $\mathcal{A}$ 中的序列 $(A_n)_{n\geqslant 1}$，我們有 $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$。

若函數 $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ 滿足

- $\mathbb{P}(\Omega) = 1$；
- 【σ 可加性】對任意由 $\mathcal{A}$ 中元素構成的兩兩互不相交序列 $(A_n)_{n\geqslant 1}$，我們有

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n), \tag{0.3}$$

則我們稱之為在 $(\Omega, \mathcal{A})$ 上的機率測度 (probability measure) 。此外，我們將 $\mathcal{A}$ 中的元素稱作可測事件 (measurable events) ，將三元組 $(\Omega, \mathcal{A}, \mathbb{P})$ 稱之為一般機率空間 (general probability space) 。

Remark 0.1.15 : A subset $\mathcal{A}$ satisfying the conditions (a), (b), and (c) in Definition 0.1.14 is called a *σ-algebra*, which is one of the most fundamental notions in Measure Theory and Probability Theory.

0.2 Random variables and expectation

When we carry out a random experiment, we may want to assign a real value to each outcome. For example, when we throw a coin, we may want to assign the value 1 to Head and the value 0 to Tail; or when we throw a dice, we may want to assign the number (or its square) we see on the dice.

Definition 0.2.1 : Consider a discrete probability space $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Any function $X : \Omega \longrightarrow \mathbb{R}$ is called a *discrete real random variable* (離散實數隨機變數) .

Example 0.2.2 : Let $\Omega = \{\text{Head}, \text{Tail}\}$ be a sample space corresponding to throwing a coin once. Fix $p \in [0, 1]$. Take $\mathcal{A} = \mathcal{P}(\Omega)$ and $\mathbb{P}$ to be defined by

$$\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\{\text{Head}\}) = p, \quad \mathbb{P}(\{\text{Tail}\}) = 1 - p, \quad \mathbb{P}(\{\text{Head}, \text{Tail}\}) = 1.$$

We may consider the random variable $X : \Omega \to \{0, 1\}$ defined by

$$X(\text{Head}) = 1, \quad \text{and} \quad X(\text{Tail}) = 0.$$

Expectation is the average value taken by a random variable.

Definition 0.2.3 : Given a discrete real random variable $X : \Omega \to \mathbb{R}$. If

$$\sum_x |x| \, \mathbb{P}(\{X = x\}) < \infty, \tag{0.4}$$

that is, the above series converges absolutely (絕對收斂), then we say that the *expectation* (期望值) of X is well defined, denoted by

$$\mathbb{E}[X] := \sum_{x \in X(\Omega)} x \, \mathbb{P}(\{X = x\}) \in (-\infty, +\infty).$$

Example 0.2.4 : For the random variable X defined in Example 0.2.2, its expectation writes

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p.$$

註解 0.1.15 ： 滿足定義 0.1.14 中條件 (a)、(b) 及 (c) 的子集合 $\mathcal{A}$ 稱作 σ 代數，這是測度論和機率論裡面，最重要的基本概念之一。

第二節 隨機變數及期望值

當我們在進行隨機試驗時，我們會想要對於每個出現的結果，對應一個實數值。舉例來說，當我們丟一顆銅板，我們可能會想要出現正面時，對應 1 的值；出現反面時，對應 0 的值；或是說當我們丟一顆骰子時，我們想要對應骰子上看到的點數（或是他的平方）。

定義 0.2.1 ： 考慮離散機率空間 $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$。我們稱任意函數 $X : \Omega \longrightarrow \mathbb{R}$ 為離散實數隨機變數 (discrete real random variable) 。

範例 0.2.2 ： 令 $\Omega = \{正, 反\}$ 為對應到投擲一顆銅板的樣本空間。固定 $p \in [0, 1]$。取 $\mathcal{A} = \mathcal{P}(\Omega)$ 並將 $\mathbb{P}$ 定義做

$$\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\{正\}) = p, \quad \mathbb{P}(\{反\}) = 1 - p, \quad \mathbb{P}(\{正, 反\}) = 1.$$

我們可以考慮隨機變數 $X : \Omega \to \{0, 1\}$ 如下：

$$X(正) = 1, \quad 以及 \quad X(反) = 0.$$

期望值是隨機變數取值的平均。

定義 0.2.3 ： 給定實隨機變數 $X : \Omega \to \mathbb{R}$。若下列的級數會絕對收斂 (converges absolutely)

$$\sum_x |x| \, \mathbb{P}(\{X = x\}) < \infty, \tag{0.4}$$

則我們說 X 的期望值 (expectation) 是定義良好的，記作

$$\mathbb{E}[X] := \sum_{x \in X(\Omega)} x \, \mathbb{P}(\{X = x\}) \in (-\infty, +\infty).$$

範例 0.2.4 ： 對於範例 0.2.2 中定義的隨機變數 X 來說，他的期望值為

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Remark 0.2.5 : The expectation $\mathbb{E}[X]$ of a random variable X is taken to be a weighted sum over all the possible values taken by X. In the weighted sum, we note that the weight at $x \in X(\Omega)$ is given by the probability $\mathbb{P}(\{X = x\})$. It is not a Riemann summation, but rather a Lebesgue integral where the measure is given by Dirac masses. As a consequence, the Lebesgue integral is a good tool to use when it comes to the expectation in a general probability space.

註解 0.2.5 ： 隨機變數 X 的期望值 $\mathbb{E}[X]$ 為對於 X 所有可能的值，加權之後所得到的和。在加權和中，值 $x \in X(\Omega)$ 的權重會由機率 $\mathbb{P}(\{X = x\})$ 所給定。這不是個黎曼和，而是當我們把測度取做狄拉克測度時的勒貝格積分。因此，在一般一般的機率空間中，在定義期望值時，勒貝格積分會是個好工具。