離散機率回顧

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觀念，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 代數的觀念會自然地出現。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 $\Omega = \{\text{正}, \text{反}\}$ 。
- (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 $\Omega = \{\text{正}, \text{反}\}^2$ ，或是 $\Omega = \{\text{正正}, \text{正反}, \text{反正}, \text{反反}\}$ 。
- (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 $\Omega = \bigcup_{n \geq 0} \{2, 3, 4, 5, 6\}^n \times \{1\}$ 。

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 $\Omega = \{\square, \square\square, \square\square\square, \square\square\square\square, \square\square\square\square\square, \square\square\square\square\square\square\}$ 。

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：

$$A = \text{「丟出來的點數是偶數」} = \{2, 4, 6\},$$

$$B = \text{「丟出來的點數是三的倍數」} = \{3, 6\},$$

$$C = \text{「丟出來的點數是偶數也是三的倍數」} = A \cap B = \{6\}.$$

(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 $\Omega = \bigcup_{n \geq 0} \{2, 3, 4, 5, 6\}^n \times \{1\}$ ，下列是可以考慮的事件：

$$A = \text{「點數 6 永遠不出現」}$$

$$= \{(a_k)_{1 \leq k \leq n} : n \in \mathbb{N}, a_n = 1, a_k \neq 6, \forall k = 1, \dots, n\},$$

$$B = \text{「點數 2 比點數 3 早出現」}$$

$$= \{(a_k)_{1 \leq k \leq n} : n \in \mathbb{N}, a_n = 1, \min\{i \geq 1 : a_i = 2\} < \min\{i \geq 1 : a_i = 3\}\}.$$

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 $A = \emptyset$ ，則我們說事件 A 絕對不會發生。

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 $\emptyset$ 。

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 $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ 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 $(A_n)_{n \geq 1}$ ，我們有

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n), \quad (0.1)$$

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 $\mu(A) = \text{Card}(A)$ 。那麼， $\mathbb{P} = \text{Card}(\Omega)^{-1}\mu$ 是個機率測度，他是在 Ω 上的均勻測度。

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) $\mathbb{P}(\emptyset) = 0$ 。

(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 $\mathbb{P}(A) \leq \mathbb{P}(B)$ 。

(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 $A_1, \dots, A_n$ ，我們有

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mathbb{P}(A_k). \quad (0.2)$$

(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

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mathbb{P}(A_i).$$

(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

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

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， $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ 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 $(A_n)_{n \geq 1}$ ，也就是說，對於所有 $n \geq 1$ ，序列滿足 $A_n \subseteq A_{n+1} \subset \Omega$ ，則我們有

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}(A_n).$$

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

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \downarrow \mathbb{P}(A_n).$$

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。 $\square$

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

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=1}^n A_k\right),$$

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=1}^n A_k\right).$$

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」。

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：

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

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

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

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 $(A_n)_{n \geq 1}$ ，我們有 $\bigcup_{n \geq 1} A_n \in \mathcal{A}$ 。

若函數 $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ 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 $(A_n)_{n \geq 1}$ ，我們有

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n), \quad (0.3)$$

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)。

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 $X : \Omega \rightarrow \mathbb{R}$ 為 離散實數隨機變數 (discrete real random variable)。

範例 0.2.2：令 $\Omega = \{\text{正}, \text{反}\}$ 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

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

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

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

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 $X : \Omega \rightarrow \mathbb{R}$ 。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

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

第零章 離散機率回顧

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。