# 0 Reminders on Discrete Probability

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 $\sigma$-*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 $\Omega$ the set consisting of all the possible outcomes (結果), called *sample space* (樣本空間) . In a discrete probability model, we require $\Omega$ 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 $\Omega = \bigcup_{n \geqslant 0} \{2, 3, 4, 5, 6\}^n \times \{1\}$.

**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 $\Omega$, 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.

### 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 $\Omega$,
- 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 $\Omega$. 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 $\Omega$, and $\mathcal{A} = \mathcal{P}(\Omega)$ to be the collection of all the subsets of $\Omega$. 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

- $\mathbb{P}(\Omega) = 1$;
- ($\sigma$-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 $\Omega$ is finite, $\mu$ 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 $\Omega$.

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-$\sigma$-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 $\sigma$-additivity required in point (2) of Definition 0.1.6 in general. If the sample space $\Omega$ 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 $\sigma$-additivity in point (2) of Definition 0.1.6 are equivalent. If the sample space $\Omega$ is infinite but countable, it is possible to construct an example that is finitely additive but not $\sigma$-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 $\sigma$-additive. In fact, we have

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

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

**Proposition 0.1.11** (Upper and lower continuity) **:** *Let $\Omega$ 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) *($\sigma$-additivity) $\mathbb{P}$ satisfies $\sigma$-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 $\mathcal{A} = \mathcal{P}(\Omega)$.

**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 $\Omega$. If the probability $\mathbb{P}$ can be defined on all the subsets of $\mathcal{P}(\Omega)$, then the $\sigma$-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 $\Omega$, 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

- $\mathbb{P}(\Omega) = 1$;
- ($\sigma$-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) .

**Remark 0.1.15 :** A subset $\mathcal{A}$ satisfying the conditions (a), (b), and (c) in Definition 0.1.14 is called a $\sigma$*-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

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

**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.