Topic 1: Probabilistic Models

Lecture Outline

Basic probabilistic model: an experiment

- Review of Classical Probability (古典機率)
- Sample space (樣本空間)
- Events (事件)
- Probability measure

  →Three axioms of probability (機率三大公理)
- Total Probability Theorem (全機率定理)

Reading: 高中數學課本、Bertsekas & Tsitsiklis 1.1-1.2

Review of Classical Probability

考慮有 N 個元素的樣本空間(sample space) S，且假設**每個樣本點出現的機會均等**。S 中的事件 A 有 k 個元素，則事件 A 發生的機率定義為

$$\mathrm{P}(A) \triangleq \frac{n(A)}{n(S)} = \frac{k}{N}$$

其中 $n(A)$ 表示事件 A 中樣本點個數。此定義是由 Laplace (法國人，1749~1827)所提出，稱為古典機率(classical probability)定義法。

Example:

一副撲克牌有52張，均勻洗牌後任取一張，若每張被取出機會相等，試求取出的牌是黑桃的機率為多少?

以上是高中數學。

Limitations of Classical Probability

- It cannot handle events with an *infinite* number of possible outcomes

  Ex: Possible value of received electro-magnetic wave in your smartphone

- It also cannot handle events where each outcome is *not equally likely*

  Ex: 丟擲一顆不公正的骰子，出現偶數的機率?

These **limitations** make classical probability inapplicable for complicated tasks.

We need something more!

Components of a Probabilistic Model

A probabilistic model is a mathematical description of a process (which is called *experiment*) with an outcome that is not fully predictable, e.g. a coin flip.

Components:

- Outcome: each experiment produces exactly one *outcome* (Ex: 銅板的正面或反面). An outcome is also called a sample point (樣本點)
- Sample space: A list of all outcomes of an experiment
  - →A sample space is a set
  - →Any *subset* of the sample space is called an *event* (事件)
- Algebra of events (set theory): language for manipulating collections of elementary events

Experiment

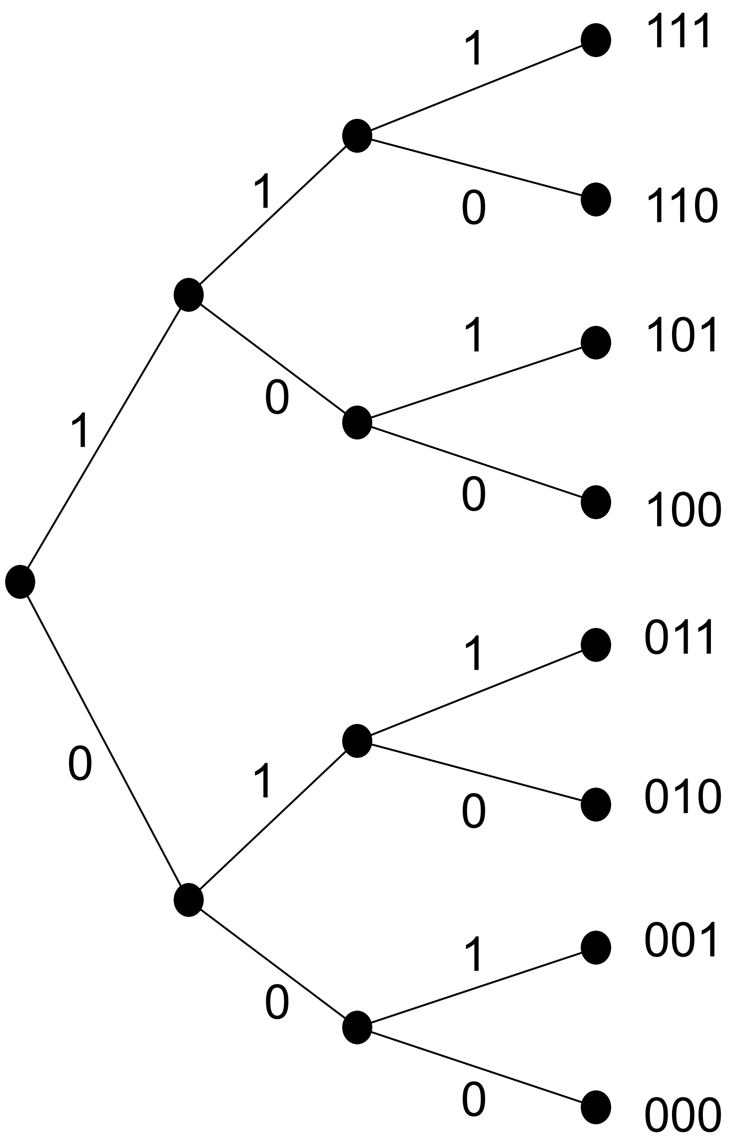
***Example*:**
Three flips of a fair coin labeled 1 (or *heads*) on one side and 0 (or *tails*) on the other.

Many ways to represent outcomes:

- List or table: each column corresponds to the number of the flip

  000
  001
  010
  011
  100
  101
  110
  111

- Sequential description: a *tree*

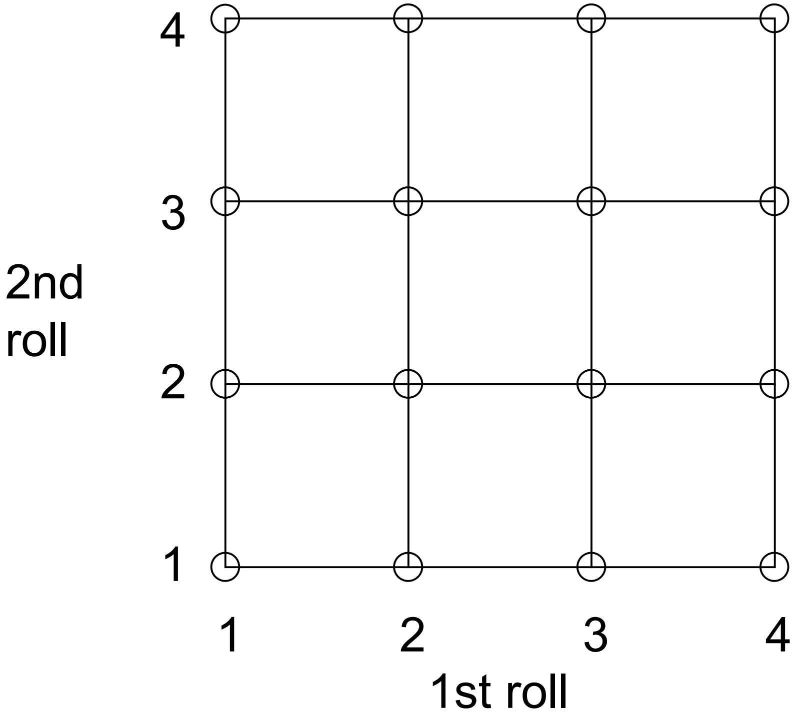

1
111
1
0
110
1
101
1
0
0
100
1
011
0
1
0
010
1
001
0
0
000

Example:

Roll a fair four-sided die twice

Outcome of the experiment

We can use a table or tree as above to represent all possible outcomes, or a graphical representation as below:



Sample Space (樣本空間)

> A *sample space* is a collection of all of the possible elementary outcomes (*sample points*)

- A sample space is an example of a *set*
  - →Dealing with a sample space requires using basic *set theory*

We define

- Sample points are all *disjoint* or *mutually exclusive* (互斥), *i.e.*, they are separate and distinct outcomes

- Sample points are *collectively exhaustive*, *i.e.*, together they make up the entire sample space, they constitute all possible elementary outcomes

We often use **Ω** to denote the sample space. A sample point ω in the sample space **Ω** can be described by $\omega \in \mathbf{\Omega}$

Event (事件)

Definition

Events are just *subsets* of the sample space, *i.e.*, sets of elements which belong to **Ω**

Examples of events:

- Event A: Flip three coins and get 010 (1 means Head, 0 means Tail)
  - →This is a simple event (elementary event, which is just an outcome)
  - →A={010}
- Event B: Flip three coins and get exactly one 1
  - →More complicated event, consisting of three elementary events 001, 010, and 100
  - →$B=\{001, 010, 100\}$
- Event C: Flip three coins and get a result such that the sum of values is 2
  - → $C=\{011,101, 110\}$

Remarks:

1. 符號上使用左右大括號來表示集合、事件
2. 「機率」的計算必和「事件」有關。我們對特殊事件會發生的機率感興趣

Events

We usually use capital letter to denote an event F, *a subset of* $\mathbf{\Omega}$, and write $F \subset \mathbf{\Omega}$

- A set F might have only one point (i.e. outcome) in it, e.g., $F = \{\omega\}$ for a specific $\omega \in \mathbf{\Omega}$

  →An outcome is itself an event (*elementary event*)

- All sets are subsets of themselves

  → the entire sample space $\mathbf{\Omega}$ is an event

- A set might have *no points* in it, *i.e.*, the *empty set* ϕ

  → This is the event that "nothing happens"

Events

- Use *algebra of events, i.e., set theory,* to manipulate events
  - Combine simple events into complicated events
  - Decompose complicated events into simpler events
  - Perform operations, *e.g. complement*, *intersection*, and *union* on sets

- Three basic operations on sets
  - Complement: $F^C=\{\omega: \omega\notin F\}$ *補集* (或稱 *餘集*) → All the sample points of Ω that are not in *F*

  - Intersection: $F\cap G=\{\omega: \omega\in F$ and $\omega\in G\}$ 交集

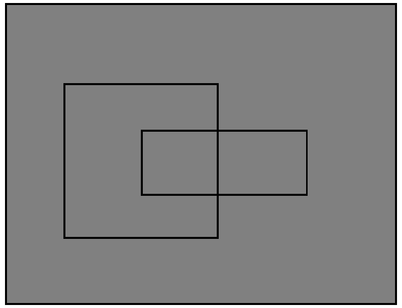
    →the points that are in both sets

    →If F and G have no points in common, then they are said to be *disjoint* or *mutually exclusive*, i.e. $F\cap G=\Phi$, the null set (空集合)

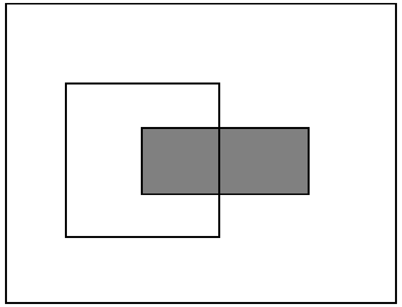
  - Union: $F\cup G=\{\omega: \omega\in F$ or $\omega\in G\}$ 聯集

    →the points that are either in one set or the other, or both

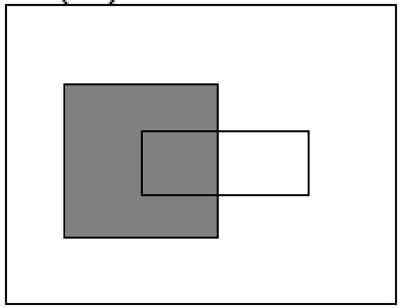
- The definitions above can be illustrated by *Venn diagrams*



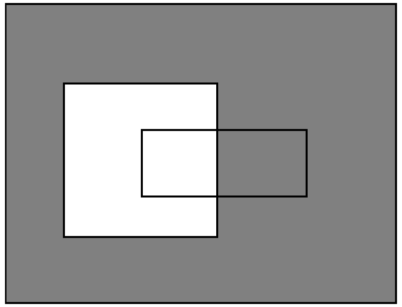
(a) Ω



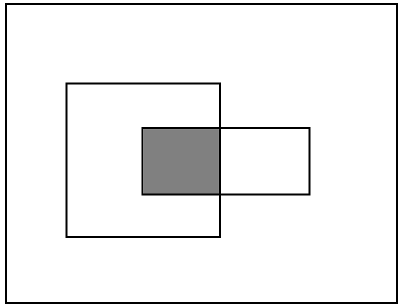
(b) G



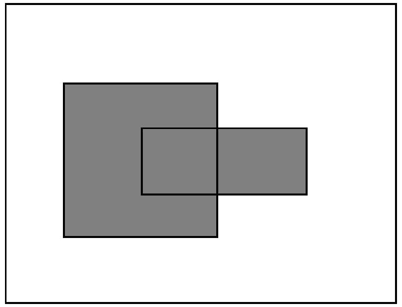
(c) F



(d) F^c



(e) $F \cap G$



(f) $F \cup G$

Set Theory Basics

- Fundamental set relations:
  - $F \cup G = G \cup F$ commutative law
  - $F \cup (G \cup H) = (F \cup G) \cup H$ associative law
  - $F \cap (G \cup H) = (F \cap G) \cup (F \cap G)$ distributive law
  - $(F^C)^C = F$
  - $F^C \cap F = \phi$ (空集合)
  - $(F \cap G)^C = F^C \cup G^C$ DeMorgan's law
  - $F \cap \Omega = F$

- Other relations:
  - $(F \cup G)^C = F^C \cap G^C$ DeMorgan's other law

Set Theory

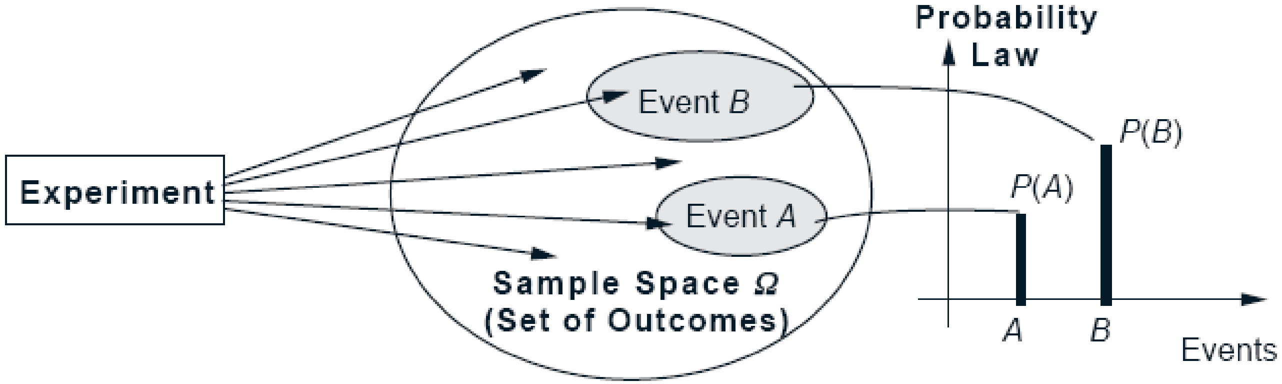
- Other relations:
  - $F \cap G = G \cap F$ commutative law
  - $F \cap (G \cap H) = (F \cap G) \cap H$ associative law
  - $F \cup (G \cap H) = (F \cup G) \cap (F \cup G)$ distributive law
  - $F^C \cup F = \Omega$
  - $F \cup \phi = F$ (ϕ 空集合)
  - $F \cup (F \cap G) = F = F \cap (F \cup G)$
  - $F \cup \Omega = \Omega$
  - $F \cap \phi = \phi$
  - $F \cup G = F \cup (F^C \cap G) = F \cup (G-F)$ → *之後會用到此特別的集合分割*
  - $F \cup F = F$
  - $F \cap F = F$

Probability Law

- We will want to assign probabilities to particular events occurring as a result of an experiment

  →Probability law

- The **probability law**, which assigns a nonnegative number $P(A)$ to event A that encodes our knowledge or belief about the collective "likelihood" of the elements of A, *formally* called the **probability** of A



Probability

Probability Law

An assignment of a real number $P(F)$ to every event F such that the following three simple axioms are satisfied:

- *Non-negativity*: $P(F) \geqq 0$ for all event F
- *Normalization*: $P(\mathbf{\Omega}) = 1$
- *Additivity*: If F and G are *disjoint* ($F \cap G = \phi$), then

$$P(F \cup G) = P(F) + P(G)$$

The function $P(.)$ defined for all subsets of Ω is called a probability measure if and only if it satisfies the three axioms of probability

Intuition for Probability Axioms

Intuitions behind the probability axioms:

Mimic relative frequencies (相對發生次數), *i.e.*, how frequent an event occurs.
For example, perform a sequence of N experiments (such as roll a dice N times):

The relative frequency of an event F is

$$\frac{\text{number of times F occurs}}{N}$$

- Relative frequency is nonnegative (Non-negativity)
- The relative frequency of something happening is 1 (Normalization)
- Relative frequency of the union of two *disjoint* events add (Additivity)

Elementary Properties

Assume that $P(.)$ is a probability measure defined on a sample space Ω. Then, we can use the three axioms to prove the following properties:

1. $P(F^C) = 1 - P(F)$

2. $P(F) \le 1$

3. $P(\phi) = 0$
4. If an event F is the union of a finite collection of disjoint events $\{F_i;\, i=1,\ldots,n\}$, i.e., if $F_i \cap F_k = \phi$ (空集合) for $i \neq k$, and $F = \cup_{i=1}^{n} F_i$, then

$$P(F) = \sum_{i=1}^{n} P(F_i)$$

Elementary Properties

5. Total Probability Theorem: (全機率定理、加法法則)

If $\{F_i; i=1,2,...,K\}$ is a finite partition of **Ω**, *i.e.*, if $F_i \cap F_k = \phi$ ，when $i \neq k$, and $F_1 \cup F_2 ... \cup F_K = \Omega$, then for any event G, we have

$$P(G) = \sum_{i=1}^{K} P(G \cap F_i)$$

這是一個非常重要的結果!

我們可透過將G拆解，來計算複雜事件G的機率值。

Example

A fair coin is flipped twice. Let G be the event that the first flip results in Heads. What is the probability of event G?

Total Probability Example

Example:

一棟有40住戶的大樓只有30個車位，每年必須透過抽籤方式決定哪一用戶於該年度有車位可停。抽籤方式如下:

將1~40號的號碼球放置不透明箱中，每戶輪流抽取一球，取後不放回。抽出1~30號的住戶可得車位使用權。

請問，先抽有利還是後抽?

Elementary Properties

6. If $F \subset G$, then $P(F) \leq P(G)$

7. $P(F \cup G) = P(F) + P(G) - P(F \cap G)$

8. $P(F \cup G) \leq P(F) + P(G)$ (*Union bound* or Bonferroni inequality)

Discrete Probability Law

The fourth property provides explanations to classical probability, which you've learned in high school, using counting arguments (排列組合)

Consider a sample space with a *finite number of sample points* $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$

Since sample points are by definition disjoint ($\{\omega_i\} \cap \{\omega_j\} = \phi$ for $i \neq j$), it follows from the fourth property that for any event F

$$P(F) = \sum_{k:\omega_k \in F} P(\{\omega_k\})$$

- We can find the *probability of an event with finite sample points* by adding up the probabilities of all the sample points in the event
  - →Need to know how many such k's (# of sample points)
  - →Need to know $P(\{\omega_k\})$ for each k
- This is great for computing probabilities for *discrete experiments* (coins, dice, etc.), but not for continuous experiments! (Why not?)

Example

Discrete *Uniform* Probability Law

If all sample points are *equally probable*, i.e., $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ and

$$P(\{\omega_k\}) = \frac{1}{N}; k = 1, 2, \ldots, N$$

This implies

$$P(F) = \frac{\text{number of sample points in F}}{\text{total number of sample points}}$$

- We can compute probabilities by ***counting*** for experiments with finite sample spaces. This is called discrete uniform probability law, equivalent to those given by 古典機率論
- Examples:
  $\rightarrow$ Probability of getting exactly one 1 in three coin flips: $\frac{3}{8}$

  $\rightarrow$ Probability get an odd number of 1's in three coin flips: $\frac{4}{8} = \frac{1}{2}$

What if discrete, but not finite?

Consider *countably infinite* sample space, such as the set of all integers

- Uniform law does not make sense, since the sum will go to infinity
- Nonuniform law is more reasonable with the additivity property

We need a stronger form of Axiom 3: *countable additivity*, limiting form of Axiom 3

If $F_1, F_2, F_3 \ldots$ are disjoint, then $P\left(\bigcup_{k=1}^{\infty} F_k\right) = \sum_{k=1}^{\infty} P(F_k)$

Example:

Suppose the sample space is {1,2,3,…}, which is countable but infinite, and $P(n) = 2^{-n}$.

$$\begin{aligned} \Pr(\text{outcome is even}) &= \mathrm{P}(\{2,4,6,8,\ldots\}) \\ &= \mathrm{P}(2)+\mathrm{P}(4)+\mathrm{P}(6)+\ldots = \sum_{k=1}^{\infty} P(2k) \\ &= \sum_{k=1}^{\infty} P(2^{-2k}) = \frac{1}{3} \end{aligned}$$

Continuous Probability Law

What if continuous sample space, i.e., uncountably infinite outcomes?

- For continuous sample space, since the outcomes are not countable, the idea of adding up probabilities of elementary points does *NOT* work!

  → Calculus comes into play (sum → integral)

Examples:

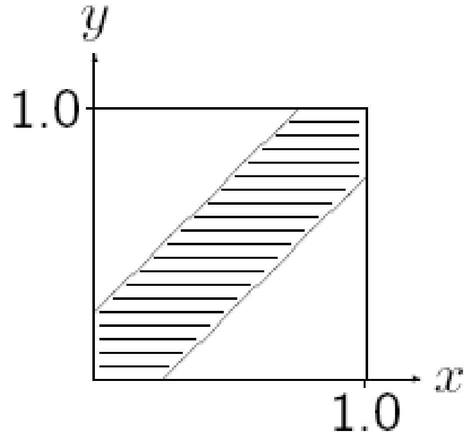
1. A wheel of fortune is continuously calibrated from 0 to 1, so the possible outcomes of an experiment consisting of a single spin are the numbers in the interval $\Omega = [0, 1]$.

2. What is the probability of the event consisting of a single element?

Example (Ex. 1.5)

Romeo and Juliet will have a date. Each arrives late with a random delay of up to 1 hour, where the pair of delays is equivalent to that achievable by spinning two identical fair wheels. Each will wait only 1/4 of an hour before leaving.

What is the probability that Romeo and Juliet will meet?



Crosshatched region $= \{(x, y) : |x - y| \le \frac{1}{4}\}$

Answer:

$$\frac{\text{Area of crosshatched region}}{\text{Area of sample space}} = 1 - 2 \times \frac{1}{2}(.75)^2 = .4375$$