香港中文大學
數學系
概率練習

1 排列與組合

兩種核心計數技術是排列和組合。**排列**計數的是選擇順序重要的distinct 安排數目。

1.1 說明示例：從5 張牌中有序選擇3 張

我們有五張標記為$1, 2, 3, 4, 5$ 的distinct 牌。我們選擇三張牌，且選擇的順序/次序很重要。

$$5 \times 4 \times 3 = 60$$

$\boxed{①}$	$\boxed{②}$	$\boxed{③}$
（5 張牌中選1 張）	（剩餘4 張中選1 張）	（剩餘3 張中選1 張）

1.2 逐步邏輯推理：

我們將三張選中的牌放入有序位置①、②、③：

1. 位置①：有5 個可用數字：$1, 2, 3, 4, 5$（共5 個選擇）。
2. 位置②：位置① 已用一張牌，剩餘4 張未用牌（共4 個選擇）。
3. 位置③：位置① 和② 已用兩張牌，僅剩3 張未用牌（共3 個選擇）。

由計數乘法原理，將每個位置的選擇數相乘：

$$\text{總有序結果數} = 5 \times 4 \times 3 = \mathbf{60}$$

1.3 一般排列定義與公式

從n 個distinct 物品中選擇k 個distinct 物品且保留順序時，此計數稱為排列，記作$\boldsymbol{{}_nP_k}$。

基本乘積形式（有限降序乘積）：

$${}_nP_k = n\,(n-1)\,(n-2)\cdots(n-k+1), \quad \text{對於} k \leq n$$

1.4 階乘特殊情況（$k = n$）：

若我們選擇所有n 個物品（$k = n$）：

$${}_nP_n = n\,(n-1)\,(n-2)\cdots 2 \cdot 1$$

此乘積定義為n 的**階乘**，寫作$n!$（讀作「n 階乘」）。

1.5 階乘等價推導（逐步代數）：

將分子和分母乘以尾隨乘積$(n-k)(n-k-1)\cdots 2\cdot 1=(n-k)!$：

$$\begin{aligned}{}_nP_k &= n\,(n-1)\cdots(n-k+1)\\ &= \frac{\bigl[n\,(n-1)\cdots(n-k+1)\bigr]\cdot\bigl[(n-k)\cdots 2\cdot 1\bigr]}{(n-k)\cdots 2\cdot 1}\\ &= \frac{n!}{(n-k)!}\end{aligned}$$

1.6 牌示例的數值驗證（$n=5,\ k=3$）

$$\begin{aligned}{}_5P_3 &= \frac{5!}{(5-3)!}=\frac{5!}{2!}\\ 5! &= 5\times 4\times 3\times 2\times 1=120\\ 2! &= 2\times 1=2\\ {}_5P_3 &= \frac{120}{2}=\mathbf{60}\end{aligned}$$

與直接乘積計算結果匹配，確認正確。

2 組合的定義與公式

組合計數的是選擇順序不重要的物品選擇方式數目。

2.1 示例1：從標記為$1,2,3,4,5$ 的5 張牌中選擇3 張無序牌

直接計算：

$$\frac{5\times 4\times 3}{3!}=10$$

1: 解釋順序等價 **假**設選中的值為$1,2,3$。

- 若順序重要（排列），有$3!=3\times 2\times 1=6$ 個distinct 有序安排：

$$(1,2,3),\ (1,3,2),\ (2,1,3),\ (2,3,1),\ (3,1,2),\ (3,2,1)$$

- 對於組合，所有6 個有序分組計為**一個相同的無序組**。為消除k 個選中物品的冗餘排列，我們將總排列數除以$k!$。

2: 使用階乘重寫分子 分子$5\times 4\times 3$ 是排列${}_5P_3=\frac{5!}{(5-3)!}=\frac{5!}{2!}$。展開階乘定義：

$$5!=5\times 4\times 3\times 2\times 1,\quad 2!=2\times 1$$

從分子和分母消去$2!$ 後剩餘$5\times 4\times 3$。將其代回組合表達式：

$$\frac{5\times 4\times 3}{3!}=\frac{\frac{5!}{2!}}{3!}=\frac{5!}{3!\cdot 2!}$$

3: 逐步驗證數值

$$3! = 3 \times 2 \times 1 = 6$$
$$5 \times 4 \times 3 = 60$$
$$\frac{60}{6} = \mathbf{10}$$

一般組合公式　從n 個distinct 物品中選擇k 個物品且不考慮順序時，組合數記作${}_nC_k$（也寫作二項式係數$\binom{n}{k}$）。

$${}_nC_k = \frac{{}_nP_k}{k!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{k}, \quad \text{對於} k \le n$$

2.2 示例2：從500 條領帶中選擇5 條

問題：求從500 條領帶中無序挑選5 條的方式數。由組合公式，$n = 500,\ k = 5$：

$$\binom{500}{5} = \frac{500!}{5! \cdot (500-5)!} = \frac{500!}{5! \cdot 495!}$$

1: 為手動計算簡化分數（消去495!）：

$$\begin{aligned}\binom{500}{5} &= \frac{500 \times 499 \times 498 \times 497 \times 496 \times 495!}{5! \times 495!} \\ &= \frac{500 \times 499 \times 498 \times 497 \times 496}{5 \times 4 \times 3 \times 2 \times 1}\end{aligned}$$

2: 算術計算：　分子乘積：

$$500 \times 499 = 249500,$$
$$249500 \times 498 = 124251000,$$
$$124251000 \times 497 = 61752747000,$$
$$61752747000 \times 496 = 30629362512000$$

分母$5! = 120$

$$\binom{500}{5} = \frac{30629362512000}{120} = \mathbf{255244687600}$$

3 帶重複的排列與組合

到目前為止，我們已涵蓋無重複選擇的計數問題。當允許物品重複時，我們使用帶重複的排列和組合的單獨公式。

1) 帶重複的排列（重複排列）：符號${}_n\Pi_k$

${}_n\Pi_k$ 表示從n 個distinct 類型中有序選擇k 個物品，允許物品重複使用（允許重複）。

$$\boldsymbol{{}_n\Pi_k = n^k}$$

2) 帶重複的組合（重複組合）：符號$_n\boldsymbol{H_k}$

$_nH_k$ 表示從n 個distinct 類型中無序選擇k 個物品，允許物品重複使用（允許重複）。公式轉換為標準二項式係數：

$$\boldsymbol{{}_nH_k = \binom{n+k-1}{k}}$$

3.1 示例3

從$\{1,2,3,4,5\}$ 中選擇三個數字，依序排列以形成三位自然數（允許重複）。求有多少個此類三位數是5 的倍數。

逐步解題推理： 一個自然數是5 的倍數，若其個位數等於0 或5。我們可用的數字僅為$1,2,3,4,5$，因此個位數**必須固定為**5。

- 位置3（個位）：僅1 個有效選擇（5）

- 位置1（百位）和2（十位）：我們從$\{1,2,3,4,5\}$ 中選擇2 個數字，允許重複（$n=5,\ k=2$）這是前兩個數字的帶重複排列問題：$_n\Pi_k = n^k$

$$\begin{aligned} n = 5&, \quad k = 2 \\ {}_5\Pi_2 &= 5^2 \\ 5^2 &= 5\times 5 = \mathbf{25} \end{aligned}$$

總有效三位5 的倍數：**25**

3.2 示例4

四人為三位候選人A、B、C 進行匿名投票。求distinct 投票結果分布的總數。

逐步解題推理： **投**票是匿名的，因此只關心每位候選人得到的票數（投票者的順序不重要）。這是帶重複的組合情景：

- $n=3$ 個distinct 候選人類型（A、B、C）

- $k=4$ 個總投票要分配公式：$_nH_k = \binom{n+k-1}{k}$

$${}_3H_4 = \binom{3+4-1}{4} = \binom{6}{4}$$

二項式係數簡化：$\binom{6}{4} = \binom{6}{2}$

$$\begin{aligned} \binom{6}{2} &= \frac{6!}{2!\,(6-2)!} = \frac{6!}{2!\,4!} \\ 6! &= 6\times 5\times 4! \\ \binom{6}{2} &= \frac{6\times 5\times 4!}{(2\times 1)\times 4!} = \frac{30}{2} = \mathbf{15} \end{aligned}$$

總唯一匿名投票分布：**15**

4 概率

特定事件發生的**概率**取值於區間$[0, 1]$。例如，擲公平硬幣正面朝上的概率為$\frac{1}{2}$。

當擲硬幣時，僅存在兩種結果：正面或反面。總共有2 個可能結果，其中正面有恰好1 個有利結果。

- 概率值為0 表示事件不可能發生（永遠不會發生）。

- 概率值為1 表示事件必然發生（一定會發生）。

要數學上計算概率，我們首先定義所有可能結果的完整集合。

- 硬幣擲擲樣本空間：$S = \{正面, 反面\}$

- 標準六面骰子樣本空間：$S = \{1, 2, 3, 4, 5, 6\}$

此所有可能結果的完整集合稱為**樣本空間**，記作S。

4.1 形式古典概率定義

設：

- $n(S) =$ 樣本空間S 中distinct 可能結果的總數

- $n(A) =$ 特定事件A 的有利結果數

事件A 的古典概率，寫作$P(A)$，定義為：

$$P(A) = \frac{\text{事件}A\text{ 的有利結果數}}{\text{可能結果總數}} = \boldsymbol{\frac{n(A)}{n(S)}}$$

4.2 幾何概率定義

對於幾何問題，其中S 代表總區域/面積，且$A \subset S$（事件區域完全位於樣本空間區域內）：

$$P(A) = \frac{\text{區域}\boldsymbol{A}\text{ 的面積}}{\text{區域}\boldsymbol{S}\text{ 的面積}}$$

4.3 統計概率與大數法則

在實際實驗試驗中：假設相同試驗重複n 次總計，事件A 在其中k 次被觀測到發生。事件A 的**相對頻率**（統計概率）為：

$$\frac{k}{n}$$

當試驗次數n 增長至無限大時，相對頻率收斂至固定常數值$\mathcal{P}$，稱為**數學概率** $P(A)$。此收斂性質是**大數法則**，以極限形式形式表達：

$$\boldsymbol{\lim_{n\to\infty} \frac{k}{n} = P(A)}$$

4.4 基本概率公理與性質

設A 為樣本空間S 內的任意事件：

1. 有界範圍：$\boldsymbol{0 \leq P(A) \leq 1}$ 對任何有效事件$A \subseteq S$
2. 完整樣本空間的必然性：$\boldsymbol{P(S) = 1}$
3. 不可能空事件：$\boldsymbol{P(\emptyset) = 0}$（$\emptyset$ = 空集，無結果）
4. 互斥事件的可加性：若A 和B 不能同時發生（不交，$A \cap B = \emptyset$）：

$$\boldsymbol{P(A \cup B) = P(A) + P(B)}$$

5. 補集規則：設A^c（也寫作$\overline{A}$）表示補集事件（A未發生的所有結果）：

$$\boldsymbol{P(A^c) = 1 - P(A)}$$

逐步示例計算（硬幣擲擲正面）

$$\begin{aligned} S &= \{\text{正面}, \text{反面}\} \implies n(S) = 2 \\ A &= \{\text{正面}\} \implies n(A) = 1 \\ P(\text{正面}) &= \frac{n(A)}{n(S)} = \frac{1}{2} = \mathbf{0.5} \end{aligned}$$

逐步示例計算（骰子擲擲，事件= 擲出3）

$$\begin{aligned} S &= \{1, 2, 3, 4, 5, 6\} \implies n(S) = 6 \\ A &= \{3\} \implies n(A) = 1 \\ P(\text{擲出}3) &= \frac{1}{6} \approx \mathbf{0.1667} \end{aligned}$$

4.5 示例5：抽取兩顆同色球的概率

問題陳述：口袋包含3 個黑球、2 個白球和1 個紅球（總計$3 + 2 + 1 = 6$ 個球）。同時隨機抽取兩個球。計算兩個選中球顏色相同的概率。

1: 計算從6 個球中無序挑選2 個球的總方式數 同時抽取兩個球是無序選擇，因此我們使用組合$\binom{n}{k} = {}_nC_k$。總樣本空間結果：

$$n(S) = \binom{6}{2}$$

展開並逐步計算：

$$\begin{aligned} \binom{6}{2} &= \frac{6!}{2!\,(6-2)!} = \frac{6!}{2! \cdot 4!} \\ 6! &= 6 \times 5 \times 4! \\ \binom{6}{2} &= \frac{6 \times 5 \times 4!}{(2 \times 1) \times 4!} = \frac{30}{2} = \mathbf{15} \end{aligned}$$

2: 計數有利結果（兩顆球顏色相同） **僅**黑色和白色顏色至少有兩個球；僅有1 個紅球，因此

1. 從3 個黑球中挑選2 個黑球的方式：$\binom{3}{2}$

$$\binom{3}{2} = \frac{3!}{2! \cdot 1!} = \frac{3 \times 2!}{2! \times 1} = \mathbf{3}$$

2. 從2 個白球中挑選2 個白球的方式：$\binom{2}{2}$

$$\binom{2}{2} = \frac{2!}{2! \cdot 0!} = \frac{2!}{2! \times 1} = \mathbf{1} \quad (\text{由定義}0! = 1)$$

總有利結果$n(A)$ = 黑球對+ 白球對：

$$n(A) = \binom{3}{2} + \binom{2}{2} = 3 + 1 = \mathbf{4}$$

3: 古典概率計算

$$P(\text{顏色相同}) = \frac{n(A)}{n(S)} = \frac{4}{15}$$

示例6：次品超幾何概率

問題設定：

總產品：1000

次品：3

正常（非次品）產品：$1000 - 3 = 997$ 我們無放回隨機選擇10 個產品。

計算兩個概率：

1. 10 個選中物品中無次品的概率。
2. 至少有一個選中物品為次品的概率。

1: 從1000 個物品中挑選10 個的方式總數 這是無序組合選擇：

$$n(S) = \binom{1000}{10}$$

部分(1): 10 個選中物品中零次品 **要**零次品，我們從997 個正常產品中選擇所有10 個單位，從3 個次品中選擇0 個次品。事件A_0（0 個次品）的有利結果數：

$$n(A_0) = \binom{997}{10} \times \binom{3}{0}$$

由古典概率公式：

$$P(A_0) = \frac{\binom{997}{10}\binom{3}{0}}{\binom{1000}{10}}$$

回想$\binom{m}{0} = 1$ 對任何整數$m \geq 0$，因此$\binom{3}{0} = 1$，簡化為：

$$P(A_0) = \frac{\binom{997}{10}}{\binom{1000}{10}}$$

2. $P(A_0)$ **的數值簡化　展**開二項式比率以避免巨大階乘值：

$$\begin{aligned}\frac{\binom{997}{10}}{\binom{1000}{10}} &= \frac{\frac{997!}{10!\cdot 987!}}{\frac{1000!}{10!\cdot 990!}} \\ &= \frac{997!\cdot 990!}{987!\cdot 1000!} \\ &= \frac{990\times 989\times 988}{1000\times 999\times 998}\end{aligned}$$

逐步算術計算：

$$990\times 989\times 988 = 967360680, \quad 1000\times 999\times 998 = 997002000$$
$$P(0\text{ 個次品}) = \frac{967360680}{997002000} \approx \mathbf{0.9702695}$$

部分(2)：至少一個選中次品　設事件B = 至少一個次品單位。B 的補集是恰好事件A_0（零次品單位）。由補集概率規則：

$$P(B) = 1 - P(A_0)$$

用二項式係數書寫的公式：

$$P(\text{至少1 個次品}) = 1 - \frac{\binom{997}{10}\binom{3}{0}}{\binom{1000}{10}}$$

部分(2) 的數值計算

$$P(\text{至少1 個次品}) = 1 - 0.9702695 = \mathbf{0.0297305}$$

5　條件概率

條件概率是數據分析中的基礎概念。它表示在事件A 已經發生的先前條件下，事件B 發生的概率。此條件概率寫作$\boldsymbol{P(B \mid A)}$。

形式定義

$$\boldsymbol{P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B \cap A)}{P(A)}}, \quad \text{僅當} P(A) > 0 \text{ 時有效}$$

圖中的記號$P_A(B)$ 是$P(B \mid A)$ 的等價簡寫。

從定義推導的乘積規則

重新排列條件概率公式得到聯合概率的乘積規則：

$$\boldsymbol{P(A \cap B) = P(A)\, P(B \mid A) = P(B)\, P(A \mid B) = P(B \cap A)}$$

對n 個事件$A_1, A_2, \ldots, A_n$ 的一般乘積規則

$$\begin{aligned} & P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \\ = {} & P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n)}{P(A_1 \cap A_2 \cap \cdots \cap A_{n-1})} \\ = {} & \boldsymbol{P(A_1 \cap A_2 \cap \cdots \cap A_n)} \end{aligned}$$

事件A 的補集分割恆等式

對於任何事件A，不交分割成立：

$$A = (A \cap B) \cup (A \cap B^c), \quad (A \cap B) \cap (A \cap B^c) = \emptyset$$

由不交集的概率可加性：

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

重新排列：

$$\boldsymbol{P(A \cap B) = P(A) - P(A \cap B^c)}$$

示例7

已知值：

$$P(A) = \frac{21}{25}, \quad P(A \cap B^c) = \frac{1}{5}$$

目標：計算$P(B \mid A)$

1: 將分割恆等式代入條件公式

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A) - P(A \cap B^c)}{P(A)}$$

2: 逐步代入數值分數 首先將$\frac{1}{5}$ 轉換為分母25 以進行統一減法：

$$\frac{1}{5} = \frac{1 \times 5}{5 \times 5} = \frac{5}{25}$$

將值代入分子：

$$P(A) - P(A \cap B^c) = \frac{21}{25} - \frac{5}{25} = \frac{21 - 5}{25} = \frac{16}{25}$$

3: 將分子除以$P(A)$

$$P(B \mid A) = \frac{\frac{16}{25}}{\frac{21}{25}}$$

共同分母25 消去：

$$P(B \mid A) = \mathbf{\frac{16}{21}}$$

6 貝葉斯定理

貝葉斯定理透過納入相關條件概率的先驗知識來計算事件的概率。此工具對於不確定性下的數學決策至關重要，並廣泛用於評估信息等無形資產的價值。

關鍵術語定義

- **先驗概率**：$P(A)$ — 在觀測到新數據或證據之前計算的事件A 的概率。
- **後驗概率**：$P(A \mid B)$ — 在觀測到新證據/事件B 發生後，事件A 的修訂、更新概率。在$P(A \mid B)$ 中，事件B 是觀測條件，$P(A \mid B)$ 在看到B 後更新我們對A 的信念。

貝葉斯定理透過結合已知的先驗概率和從觀測事件得出的條件似然來推導後驗概率。

樣本空間的分割與全概率法則

設樣本空間S 分割為不交事件$A_1, A_2, \ldots, A_n$：

$$A_i \cap A_j = \emptyset \quad (i \neq j), \qquad S = A_1 \cup A_2 \cup \cdots \cup A_n$$

對於任何任意事件$B \subseteq S$：

$$B = S \cap B = \big(A_1 \cup A_2 \cup \cdots \cup A_n\big) \cap B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$$

所有集合$A_i \cap B$ 相互排斥，因此由有限可加性：

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

應用概率乘積規則$P(A_i \cap B) = P(A_i)\, P(B \mid A_i)$ 得到**全概率法則：**

$$\boldsymbol{P(B) = \sum_{i=1}^{n} P(A_i)\, P(B \mid A_i)} = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n)$$

貝葉斯定理公式

從條件概率定義：

$$P(A_j \mid B) = \frac{P(A_j \cap B)}{P(B)}$$

代入乘積規則$P(A_j \cap B) = P(A_j)P(B \mid A_j)$ 和$P(B)$ 的全概率：

$$\boldsymbol{P(A_j \mid B) = \frac{P(A_j)\, P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i)\, P(B \mid A_i)}}$$

- $P(A_j)$：分割事件A_j 的先驗概率
- $P(A_j \mid B)$：給定觀測事件B 後A_j 的後驗概率

6.1 示例8

問題設定：

三台工廠機器A, B, C 分別生產總產出的50%, 30%, 20%。次品率：$P(\text{次品} \mid A) = 0.04$, $P(\text{次品} \mid B) = 0.03$, $P(\text{次品} \mid C) = 0.02$。

1. 求隨機選中產品為次品的總概率（$P(X)$，其中$X =$ 次品）。
2. 求次品產品由機器C 生產的後驗概率（$P(C \mid X)$）。

1: 定先驗概率

$$P(A) = 0.5, \quad P(B) = 0.3, \quad P(C) = 0.2$$

條件次品似然：

$$P(X \mid A) = 0.04, \quad P(X \mid B) = 0.03, \quad P(X \mid C) = 0.02$$

2: $P(X)$ 的全概率法則

$$\begin{aligned} P(X) &= P(A)P(X \mid A) + P(B)P(X \mid B) + P(C)P(X \mid C) \\ &= (0.5 \times 0.04) + (0.3 \times 0.03) + (0.2 \times 0.02) \\ &= 0.02 + 0.009 + 0.004 \\ &= \mathbf{0.033} \end{aligned}$$

3: $P(C \mid X)$ 的貝葉斯定理

$$\begin{aligned} P(C \mid X) &= \frac{P(C)\,P(X \mid C)}{P(A)P(X \mid A) + P(B)P(X \mid B) + P(C)P(X \mid C)} \\ &= \frac{0.2 \times 0.02}{0.033} \\ &= \frac{0.004}{0.033} = \frac{\mathbf{4}}{\mathbf{33}} \approx 0.1212 \end{aligned}$$

7 隨機變量

考慮擲兩枚公平硬幣的實驗。所有distinct 結果的樣本空間為：

$$S = \{\,(\text{正面}, \text{正面}),\ (\text{正面}, \text{反面}),\ (\text{反面}, \text{正面}),\ (\text{反面}, \text{反面})\,\}$$

對於公平硬幣擲擲，樣本空間中每個結果同等可能，因此每個單一結果的概率等於$\mathbf{\frac{1}{4}}$。

定義X 為**兩次硬幣擲擲中觀測到的反面數**。隨機變量X 將每個樣本結果映射到實數值：

- 結果(正面, 正面) $\mapsto X = 0$ 個反面

- 結果(正面, 反面), (反面, 正面) $\mapsto X = 1$ 個反面
- 結果(反面, 反面) $\mapsto X = 2$ 個反面

此映射函數X 定義為**隨機變量**。

隨機變量的定義

隨機變量的行為類似於計算機編程中的變量，只是其實現值由概率機會而非固定賦值決定。形式上，隨機變量是將樣本空間S 的每個結果（樣本點）映射到實數線$\mathbb{R}$ 上實數的函數。

使用隨機變量將定性概率事件轉換為定量數值，這簡化了隨機實驗的計算、比較和統計分析。記號慣例：

- 大寫字母（例如X, Y, Z）：表示隨機變量函數本身
- 小寫字母（例如x, y, z）：表示隨機變量可能取的特定數值

對X（2 次硬幣擲擲的反面數）的逐步概率質量計算

總同等可能樣本結果：$n(S) = 4$，每個概率$\frac{1}{4}$。

1. 值$x = 0$（零反面）：僅1 個有利結果(H,H)

$$P(X = 0) = \frac{\text{結果數，}X = 0}{n(S)} = \frac{1}{4} = 0.25$$

2. 值$x = 1$（一個反面）：兩個有利結果(H,T), (T,H)

$$P(X = 1) = \frac{2}{4} = \frac{1}{2} = 0.50$$

3. 值$x = 2$（兩個反面）：僅1 個有利結果(T,T)

$$P(X = 2) = \frac{1}{4} = 0.25$$

總概率和為1 的驗證

X 的所有互斥可能值覆蓋完整樣本空間：

$$P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = \frac{1 + 2 + 1}{4} = \frac{4}{4} = \mathbf{1}$$

樣本空間$\to X \to$ 實數映射

8 離散概率分布

離散隨機變量 X 是僅能取可數集合的distinct 實值$x_1, x_2, \ldots, x_n$ 的隨機變量。**離散概率分布**完全指定離散隨機變量X 的每個可能值x_i 的概率$P(X = x_i)$。

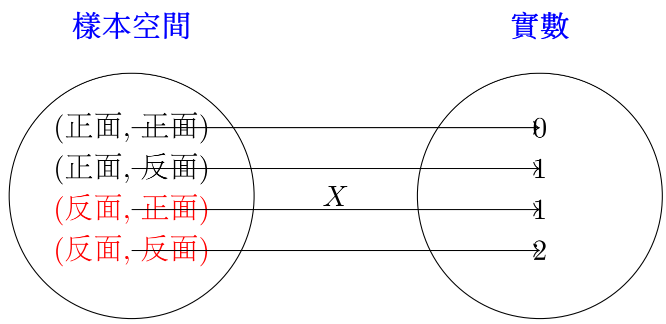

Figure 1: 透過隨機變量X（反面數）將樣本空間結果映射到實數

離散分布的表格格式

X	x_1	x_2	$\ldots$	x_n	總和
概率	$P(X = x_1)$	$P(X = x_2)$	$\ldots$	$P(X = x_n)$	1

8.1 示例9：兩次硬幣擲擲（計反面數X）

實驗：同時擲兩枚公平硬幣。X = 觀測到的反面數。

樣本空間：$S = \{(H, H), (H, T), (T, H), (T, T)\}$，每個結果同等可能，概率$\frac{1}{4}$。

逐步映射和概率計算：

1. $X = 0$：僅結果(H, H)
$$P(X = 0) = \frac{1}{4}$$
2. $X = 1$：結果$(H, T), (T, H)$（2 種情況）
$$P(X = 1) = \frac{2}{4} = \frac{1}{2}$$
3. $X = 2$：僅結果(T, T)
$$P(X = 2) = \frac{1}{4}$$

樣本空間→ X → 概率映射

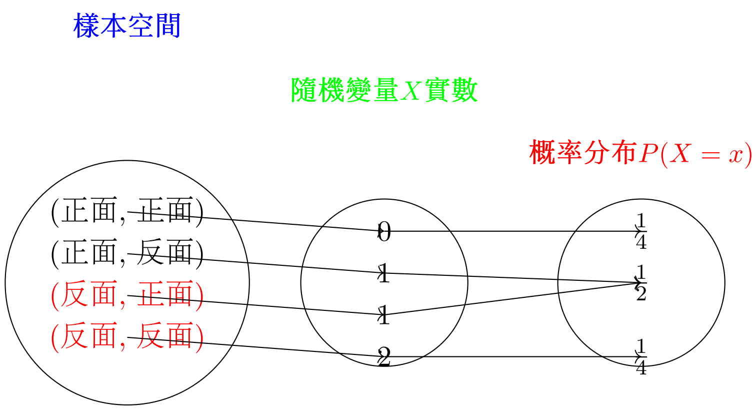

Figure 2: 將樣本結果映射到X 值，然後映射到概率質量

概率質量函數（PMF）定義

將概率值$P(X = x_i)$ 賦予X 的每個可能離散值x_i 的函數$f(x)$ 稱為X 的**概率質量函數（pmf）**：

$$f(x) = \begin{cases} P(X = x_i) & \text{對於}x = x_1,\, x_2,\, \ldots,\, x_n \\ 0 & \text{對於所有其他實數}x \end{cases}$$

三個基本PMF 性質

1. 有界非負性：
$$\boldsymbol{0 \le f(x_i) \le 1 \quad (i = 1, 2, \ldots, n)}$$
2. 總質量和為1（完整樣本空間覆蓋）：
$$\boldsymbol{f(x_1) + f(x_2) + \cdots + f(x_n) = \sum_{i=1}^{n} f(x_i) = 1}$$
3. 區間概率：對於界限$a \le b$，概率$P(a \le X \le b)$ 等於區間$[a, b]$ 內所有x 的pmf 值之和：
$$\boldsymbol{P(a \le X \le b) = \sum_{a \le x \le b} f(x)}$$

8.2 示例10：兩次硬幣示例PMF 的驗證

PMF 值：$f(0) = \frac{1}{4},\ f(1) = \frac{1}{2},\ f(2) = \frac{1}{4}$

1. 界限檢查：$0 < \frac{1}{4} < 1,\ 0 < \frac{1}{2} < 1,\ 0 < \frac{1}{4} < 1$（滿足性質1）

2. 總和：
$$f(0) + f(1) + f(2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = \frac{1 + 2 + 1}{4} = \frac{4}{4} = \mathbf{1}$$

3. 示例區間計算：$P(0 \le X \le 1)$
$$P(0 \le X \le 1) = f(0) + f(1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

9 連續概率分布

連續隨機變量 X 是可取不可數無限實值集合的隨機變量。對於任何連續隨機變量，X 等於單一精確點值x 的概率始終為零：

$$\boldsymbol{P(X = x) = 0}$$

因為點概率為零，離散概率質量函數（pmf）無法描述連續分布。相反，我們引入**概率密度函數（pdf）**，記作$f(x)$，以建模連續概率行為。

pdf $f(x)$ 是離散pmf 的連續類比。函數$f(x)$ 若滿足三個核心性質，則qualifies 為連續隨機變量X 的有效概率密度函數：

1. 對所有實輸入非負：
$$\boldsymbol{f(x) \ge 0 \quad \forall x \in \mathbb{R}}$$

2. 整個實線上的總積分等於1（總概率質量）：

$$\int_{-\infty}^{\infty} \boldsymbol{f(x)\,dx = 1}$$

3. 區間概率等於$f(x)$ 在區間界限$[a, b]$ 上的定積分。對於連續變量，端點的包含或排除不改變概率（因為$P(X = a) = P(X = b) = 0$）：

$$\begin{aligned}\boldsymbol{P(a \leq X \leq b)} &= P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) \\ &= \int_{\boldsymbol{a}}^{\boldsymbol{b}} \boldsymbol{f(x)\,dx}\end{aligned}$$

區間概率的圖形解釋

概率$P(a \leq X \leq b)$ 在幾何上對應pdf 曲線$y = f(x)$ 在垂直線$x = a$ 和$x = b$ 之間的面積。

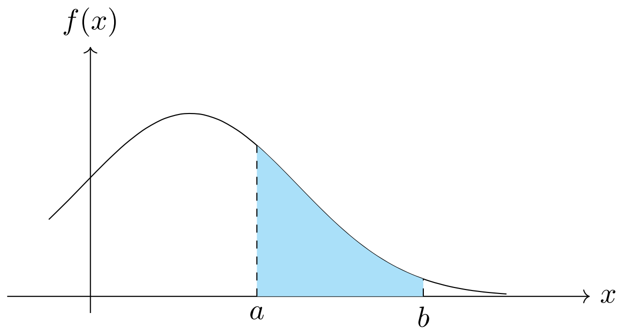

Figure 3: 陰影面積$= P(a \leq X \leq b) = \int_a^b f(x)\,dx$

註：PDF 中「密度」的詞源

我們將概率解釋為質量，將區間長度解釋為一維體積：

1. 比率：$\dfrac{\text{概率}}{\text{區間長度}} = \dfrac{\text{質量}}{\text{體積}}$
2. 比率$\frac{\text{質量}}{\text{體積}}$ 是物理密度的標準定義。
3. 重新排列的關係：

$$\left(\frac{\text{概率}}{\text{區間長度}}\right) \times (\text{區間長度}) = \text{概率}$$

將密度（每單位長度的概率）乘以區間長度，得到該區段的總概率。此關係賦予名稱概率密度函數。

9.1 示例12：（均勻連續分布檢驗）

取$[0, 2]$ 上的均勻pdf：

$$f(x) = \begin{cases} \frac{1}{2} & 0 \leq x \leq 2 \\ 0 & \text{其他情況} \end{cases}$$

1. 性質1 檢查：$f(x) \geq 0$ 對所有實x（支撐上$\frac{1}{2} > 0$，之外為0）
2. 性質2 總積分：

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^2 \frac{1}{2}dx = \frac{1}{2}x\Big|_0^2 = \frac{1}{2}(2-0) = \mathbf{1}$$

3. 區間概率示例：$P(0.5 \leq X \leq 1.5)$

$$P(0.5 \leq X \leq 1.5) = \int_{0.5}^{1.5} \frac{1}{2}dx = \frac{1}{2}(1.5-0.5) = \mathbf{0.5}$$

也確認$P(X = 1) = 0$（單點積分$\int_1^1 \frac{1}{2}dx = 0$）