詳細解答

STEP III 2019，第 12 題

概率與統計 • 隨機子集

> **STEP III 2019，第 12 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2019 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

本題在 *Beyond STEP* 第 9.12 節隨機子集中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch09/random-subsets.html

關鍵思路

從 S 中均勻隨機（uniformly at random）地選取一個子集（subset），等同於逐一考慮整數 $1, 2, \ldots, n$，並為每個整數擲一枚公平硬幣，以決定它是否放入子集。一旦看清這一點，題中每個事件（event）都變成關於每個整數各自的陳述，各整數之間互相獨立（independent），而答案就是某個小概率的 n 次方。

解答

數算子集

對 n 個整數中的每一個決定它「在內」或「在外」，便完全確定了 S 的一個子集。每個整數有 2 種選擇，各自獨立地作出，而不同的選擇模式給出不同的子集。因此，連同 $\varnothing$（全部在外）和 S（全部在內）計算在內，子集的數目為

$$2 \times 2 \times \cdots \times 2 = 2^n.$$

> ? **想一想** 為甚麼可以把各整數視為獨立的擲硬幣？

設 A 在 T 上均勻分佈。2^n 種在內／在外模式各對應一個子集，所以每種模式的概率都是 $2^{-n} = (\frac{1}{2})^n$。這正是擲 n 枚獨立公平硬幣時得到某一特定模式的概率。因此事件「$1 \in A$」、「$2 \in A$」、…、「$n \in A$」互相獨立，各自的概率為 $\frac{1}{2}$。由於 $A_1, \ldots, A_m$ 彼此之

間亦是獨立選取的，所以全部 mn 個事件「$i \in A_k$」都互相獨立，各自的概率為 $\frac{1}{2}$。

第 (i) 部分

2^n 個子集中恰好有一半包含 1（把每個包含 1 的子集與移走 1 後的同一子集配對）。所以

$$\mathrm{P}(1 \in A_1) = \tfrac{1}{2}.$$

第 (ii) 部分

對整數 i，設 E_i 為「i 並非同時屬於 A_1 和 A_2」這一事件。那麼 $A_1 \cap A_2 = \varnothing$ 當且僅當每個 E_i 都發生。由獨立性，

$$\mathrm{P}(i \in A_1 \text{ 且} i \in A_2) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \qquad \text{所以} \qquad \mathrm{P}(E_i) = \tfrac{3}{4}.$$

事件 E_i 只取決於整數 i 的硬幣，所以 $E_1, \ldots, E_n$ 互相獨立，並且

$$\mathrm{P}(A_1 \cap A_2 = \varnothing) = \mathrm{P}(E_1) \cdots \mathrm{P}(E_n) = \left(\tfrac{3}{4}\right)^n,$$

此即所求。

同一論證適用於任意數目的集合。整數 i 屬於 $A_1 \cap A_2 \cap A_3$ 的概率為 $(\frac{1}{2})^3 = \frac{1}{8}$，所以它不在該交集（intersection）中的概率為 $\frac{7}{8}$，從而

$$\mathrm{P}(A_1 \cap A_2 \cap A_3 = \varnothing) = \left(\tfrac{7}{8}\right)^n.$$

一般而言，i 同時屬於全部 m 個集合的概率為 2^{-m}，所以

$$\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_m = \varnothing) = \left(1 - \frac{1}{2^m}\right)^n.$$

第 (iii) 部分

現在 $A_1 \subseteq A_2$ 的意思是沒有任何整數屬於 A_1 而不屬於 A_2。對每個整數 i，（是否在 A_1？是否在 A_2？）共有四種等可能的情況，其中只有「在 A_1、不在 A_2」一種是不容許的。所以每個整數以概率 $\frac{3}{4}$ 符合要求，且互相獨立，因此

$$\mathrm{P}(A_1 \subseteq A_2) = \left(\tfrac{3}{4}\right)^n.$$

（這與第 (ii) 部分吻合：$A_1 \subseteq A_2$ 當且僅當 A_1 與 A_2 的補集（complement）不相交（disjoint），而均勻隨機子集的補集仍然是均勻隨機的。）

至於 $A_1 \subseteq A_2 \subseteq A_3$，考慮某一整數 i 在 (A_1, A_2, A_3) 中的歸屬模式，以 1 表示「在內」，0 表示「在外」。鏈（chain）條件表示一旦 i 屬於某個 A_k，它就屬於其後所有集合，所以模式必須是以下其中之一：

$$000, \quad 001, \quad 011, \quad 111.$$

這是 8 種等可能模式中的 4 種，所以每個整數符合要求的概率為 $\frac{4}{8}=\frac{1}{2}$，從而

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3) = \left(\tfrac{1}{2}\right)^n.$$

對於由 m 個集合組成的鏈，容許的模式是一段 0 之後接一段 1：它由 1 開始出現的位置決定，這位置是 $1, 2, \ldots, m$ 之一，或者「從不出現」。這樣在 2^m 種模式中共有 $m+1$ 種容許模式，所以

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq \cdots \subseteq A_m) = \left(\frac{m+1}{2^m}\right)^n.$$

驗算

- 當 $m=2$ 時，一般公式給出 $(1-\frac{1}{4})^n$ 和 $(\frac{3}{4})^n$；當 $m=3$ 時，給出 $(\frac{7}{8})^n$ 和 $(\frac{4}{8})^n$，與特殊情況吻合。當 $m=1$ 時，給出 $(\frac{1}{2})^n$（只有 $A_1=\varnothing$ 符合）和 1（單一集合必然是一條鏈），正如所料。
- 當 $n=1$ 時，所有情況都可以用手列出。例如在四種選擇中，只有 $A_1=A_2=\{1\}$ 的交集非空，得出 $\frac{3}{4}$。
- 用電腦列出所有集合選擇的驗算，確認了各公式在 n 和 m 較小時均成立。例如當 $n=3$ 時，$\mathrm{P}(A_1 \cap A_2 = \varnothing) = \frac{27}{64}$，而 $\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3) = \frac{1}{8}$。

常見錯誤

- 以「A_1 的每個元素都必須在 A_2 中」為理由，答 $\mathrm{P}(A_1 \subseteq A_2)$ 為 $\left(\frac{1}{2}\right)^n$，卻沒有考慮到 A_1 本身的大小也是隨機的。
- 把鏈的模式數目算成 m 而非 $m+1$，漏掉了整數不屬於任何集合的那種模式。

STEP 之外

把隨機子集表示為一串獨立的擲硬幣結果（即它的指示向量（indicator vector））是組合學（combinatorics）中的標準工具，而這裏對鏈的計數正是偏序集（posets）及布爾格（Boolean lattice）理論的起點。同樣的想法只需很少額外功夫，便可得出鏈的期望（expected）數目，或隨機集合覆蓋 S 的概率。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（《概率導論》）正是以這種精神發展獨立性及指示變量（indicator variables）的理論。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。