详细解答

STEP III 2019，第 12 题

概率与统计 • 随机子集

STEP III 2019，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

本题在 *Beyond STEP* 第 9.12 节随机子集中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/random-subsets.html

关键思路

从 S 中均匀随机（uniformly at random）地选取一个子集（subset），等同于逐一考虑整数 $1, 2, \ldots, n$，并为每个整数掷一枚公平硬币，以决定它是否放入子集。一旦看清这一点，题中每个事件（event）都变成关于每个整数各自的陈述，各整数之间互相独立（independent），而答案就是某个小概率的 n 次方。

解答

数算子集

对 n 个整数中的每一个决定它“在内”或“在外”，便完全确定了 S 的一个子集。每个整数有 2 种选择，各自独立地作出，而不同的选择模式给出不同的子集。因此，连同 $\varnothing$（全部在外）和 S（全部在内）计算在内，子集的数目为

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

想一想 为甚么可以把各整数视为独立的掷硬币？

设 A 在 T 上均匀分布。2^n 种在内／在外模式各对应一个子集，所以每种模式的概率都是 $2^{-n} = (\frac{1}{2})^n$。这正是掷 n 枚独立公平硬币时得到某一特定模式的概率。因此事件“$1 \in A$”、“$2 \in A$”、...、“$n \in A$”互相独立，各自的概率为 $\frac{1}{2}$。由于 $A_1, \ldots, A_m$ 彼此之

间亦是独立选取的，所以全部 mn 个事件“$i \in A_k$”都互相独立，各自的概率为 $\frac{1}{2}$。

第 (i) 部分

2^n 个子集中恰好有一半包含 1（把每个包含 1 的子集与移走 1 后的同一子集配对）。所以

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

第 (ii) 部分

对整数 i，设 E_i 为“i 并非同时属于 A_1 和 A_2”这一事件。那么 $A_1 \cap A_2 = \varnothing$ 当且仅当每个 E_i 都发生。由独立性，

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

事件 E_i 只取决于整数 i 的硬币，所以 $E_1, \ldots, E_n$ 互相独立，并且

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

此即所求。

同一论证适用于任意数目的集合。整数 i 属于 $A_1 \cap A_2 \cap A_3$ 的概率为 $(\frac{1}{2})^3 = \frac{1}{8}$，所以它不在该交集（intersection）中的概率为 $\frac{7}{8}$，从而

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

一般而言，i 同时属于全部 m 个集合的概率为 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

现在 $A_1 \subseteq A_2$ 的意思是没有任何整数属于 A_1 而不属于 A_2。对每个整数 i，（是否在 A_1? 是否在 A_2?）共有四种等可能的情况，其中只有“在 A_1、不在 A_2”一种是不容许的。所以每个整数以概率 $\frac{3}{4}$ 符合要求，且互相独立，因此

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

（这与第 (ii) 部分吻合：$A_1 \subseteq A_2$ 当且仅当 A_1 与 A_2 的补集（complement）不相交（disjoint），而均匀随机子集的补集仍然是均匀随机的。）

至于 $A_1 \subseteq A_2 \subseteq A_3$，考虑某一整数 i 在 (A_1, A_2, A_3) 中的归属模式，以 1 表示“在内”，0 表示“在外”。链（chain）条件表示一旦 i 属于某个 A_k，它就属于其后所有集合，所以模式必须是以下其中之一：

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

这是 8 种等可能模式中的 4 种，所以每个整数符合要求的概率为 $\frac{4}{8}=\frac{1}{2}$，从而

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

对于由 m 个集合组成的链，容许的模式是一段 0 之后接一段 1：它由 1 开始出现的位置决定，这位置是 $1,2,\ldots,m$ 之一，或者“从不出现”。这样在 2^m 种模式中共有 $m+1$ 种容许模式，所以

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

验算

- 当 $m=2$ 时，一般公式给出 $(1-\frac{1}{4})^n$ 和 $(\frac{3}{4})^n$；当 $m=3$ 时，给出 $(\frac{7}{8})^n$ 和 $(\frac{4}{8})^n$，与特殊情况吻合。当 $m=1$ 时，给出 $(\frac{1}{2})^n$（只有 $A_1=\varnothing$ 符合）和 1（单一集合必然是一条链），正如所料。
- 当 $n=1$ 时，所有情况都可以用手列出。例如在四种选择中，只有 $A_1=A_2=\{1\}$ 的交集非空，得出 $\frac{3}{4}$。
- 用电脑列出所有集合选择的验算，确认了各公式在 n 和 m 较小时均成立。例如当 $n=3$ 时，$\mathrm{P}(A_1\cap A_2=\varnothing)=\frac{27}{64}$，而 $\mathrm{P}(A_1\subseteq A_2\subseteq A_3)=\frac{1}{8}$。

常见错误

- 以“A_1 的每个元素都必须在 A_2 中”为理由，答 $\mathrm{P}(A_1\subseteq A_2)$ 为 $\left(\frac{1}{2}\right)^n$，却没有考虑到 A_1 本身的大小也是随机的。
- 把链的模式数目算成 m 而非 $m+1$，漏掉了整数不属于任何集合的那种模式。

STEP 之外

把随机子集表示为一串独立的掷硬币结果（即它的指示向量（indicator vector））是组合学（combinatorics）中的标准工具，而这里对链的计数正是偏序集（posets）及布尔格（Boolean lattice）理论的起点。同样的想法只需很少额外功夫，便可得出链的期望（expected）数目，或随机集合覆盖 S 的概率。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（《概率导论》）正是以这种精神发展独立性及指示变量（indicator variables）的理论。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。