详细解答

STEP I 1998，第 13 题

概率与统计 · 放回抽样

STEP I 1998，第 13 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/98-s1-q13

关键思路

这是一道条件概率（conditional probability）的题目：作为条件的事件是“在 n 次抽取中刚好吃掉一粒软糖”，而我们按这粒软糖是在何时吃掉的，把这个事件拆开。每一部分都是几个简单概率的乘积，因为袋中的内容只会在吃掉红色软糖的那一刻改变。把各部分加起来，就是一个等比级数（geometric series），两个 n 次幂之差正是由此而来。

解答

设 $N = r + b$ 为开始时软糖的总数。我们假设 $r \geq 1$ 及 $b \geq 1$（若 $b = 0$ 而 $n \geq 2$，则不可能刚好只吃掉一粒软糖，也就没有可作条件的事件）。

每次抽取时的袋子。 非红色的软糖会放回袋中，所以袋子只会在吃掉红色软糖时改变。吃掉红色软糖之前，袋中有 N 粒软糖，其中 b 粒非红色；之后袋中有 $N-1$ 粒软糖，其中仍有 b 粒非红色。由于每次抽取前都把袋子彻底摇匀，每次抽取都是从当时袋中的软糖里重新作一次均匀的随机选取，所以我们可以把各次抽取的概率逐一相乘。

按吃掉的时间拆分。 对 $k = 1, 2, \ldots, n$，设 E_k 为以下事件：在 n 次抽取中唯一吃掉的那粒软糖，是在第 k 次抽取时吃掉的。那么第 $1, \ldots, k-1$ 次是从原来的袋中抽到非红色，第 k 次是从原来的袋中抽到红色，而第 $k+1, \ldots, n$ 次是从少了一粒的袋中抽到非红色：

$$\mathrm{P}(E_k) = \left(\frac{b}{N}\right)^{k-1} \frac{r}{N} \left(\frac{b}{N-1}\right)^{n-k} = \frac{r\, b^{n-1}}{N^k\, (N-1)^{n-k}}.$$

事件 $E_1, \ldots, E_n$ 互斥（disjoint），而它们的并集（union）就是“刚好吃掉一粒软糖”这个事件 E。所以

$$\mathrm{P}(E) = r\,b^{n-1} \sum_{k=1}^{n} \frac{1}{N^k (N-1)^{n-k}} = \frac{r\,b^{n-1}}{N^n (N-1)^n} \sum_{k=1}^{n} N^{n-k} (N-1)^k,$$

这里我们把每一项的分子和分母同乘以 $N^n (N-1)^n$。

> **想一想** 怎样求 $\sum_{k=1}^{n} N^{n-k}(N-1)^k$ 的和?

这是一个等比级数，首项为 $N^{n-1}(N-1)$，公比（common ratio）为 $\frac{N-1}{N}$。更快捷的做法，是利用大家熟悉的因式分解（factorisation）

$$x^n - y^n = (x-y)(x^{n-1} + x^{n-2}y + \cdots + y^{n-1})$$

取 $x = N$、$y = N-1$，于是 $x - y = 1$：

$$\sum_{k=1}^{n} N^{n-k}(N-1)^k = (N-1) \sum_{j=0}^{n-1} N^{n-1-j}(N-1)^j = (N-1)(N^n - (N-1)^n).$$

因此

$$\mathrm{P}(E) = \frac{r\,b^{n-1}(N^n - (N-1)^n)}{N^n (N-1)^{n-1}}.$$

条件概率。“我在最后一次抽取时吃掉它，而且只吃了一粒”这个事件就是 E_n，其概率为 $\mathrm{P}(E_n) = \dfrac{r\,b^{n-1}}{N^n}$。由于 $E_n \subseteq E$，

$$\begin{aligned}\mathrm{P}(E_n \mid E) = \frac{\mathrm{P}(E_n)}{\mathrm{P}(E)} &= \frac{r\,b^{n-1}}{N^n} \cdot \frac{N^n (N-1)^{n-1}}{r\,b^{n-1}(N^n - (N-1)^n)} \\ &= \frac{(N-1)^{n-1}}{N^n - (N-1)^n} = \frac{(r+b-1)^{n-1}}{(r+b)^n - (r+b-1)^n},\end{aligned}$$

这正是所求的结果。

留意 r 和 b 完全消去，只剩下它们的总和。其实上面的计算还说明了更多：已知 E 发生，软糖在第 k 次抽取时被吃掉的概率与 $\left(\frac{N-1}{N}\right)^k$ 成正比，所以较早的抽取比较后的抽取更有可能。这是合理的：早些吃掉红色软糖，袋中非红色软糖所占的比例便较大，余下各次“再没有抽到红色”的机会也就较高。

验算

▶ $n = 1$：公式给出 $\dfrac{1}{N - (N-1)} = 1$，正确，因为唯一的一次抽取就是最后一次。

- $n = 2$：公式给出 $\dfrac{N-1}{2N-1}$，略小于 $\frac{1}{2}$，与“较早的抽取可能性稍高”一致。
- 取 $r = 3$、$b = 5$、$n = 4$，公式给出 $343/1695 \approx 0.2024$，而若四次抽取的可能性均等，则应为 $\frac{1}{4}$。模拟 4×10^5 次的结果为 0.2028；另外，逐一列举所有抽取序列并追踪袋中内容，所得的精确结果在 $1 \le r, b \le 4$ 及 $1 \le n \le 8$ 的所有情况下都与公式完全吻合。

常见错误

- 在红色软糖被吃掉*之后*，仍然用 b/N 作为抽到非红色的概率。这时袋中只有 $N-1$ 粒软糖，而这正是本题的关键所在：如果由始至终都用 b/N，每个 E_k 的概率便会相同，答案就变成 $1/n$。
- 忘记取条件：单是 $\mathrm{P}(E_n) = rb^{n-1}/N^n$ 并不是答案，因为题目已告诉我们刚好吃掉了一粒软糖。

STEP 之外

这里的袋子是一个简单的马可夫链（Markov chain）：它的状态是余下红色软糖的数目，而状态向下移动的机会只取决于当前的状态。以这类链的最终状态为条件，反过来问它所经过的路径，正是“隐马可夫模型”（hidden Markov models）以及对历史过程作贝叶斯推断（Bayesian inference）背后的想法。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门）正好以这个程度讲解条件概率和马可夫链，而 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）则有许多同类风格的题目。

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。