詳細解答

STEP I 1998，第 13 題

概率與統計 • 放回抽樣

STEP I 1998，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

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

關鍵思路

這是一道條件概率（conditional probability）的題目：作為條件的事件是「在 n 次抽取中剛好吃掉一粒軟糖」，而我們按這粒軟糖是在何時吃掉的，把這個事件拆開。每一部分都是幾個簡單概率的乘積，因為袋中的內容只會在吃掉紅色軟糖的那一刻改變。把各部分加起來，就是一個等比級數（geometric series），兩個 n 次冪之差正是由此而來。

解答

設 $N = r + b$ 為開始時軟糖的總數。我們假設 $r \geq 1$ 及 $b \geq 1$（若 $b = 0$ 而 $n \geq 2$，則不可能剛好只吃掉一粒軟糖，也就沒有可作條件的事件）。

每次抽取時的袋子。 非紅色的軟糖會放回袋中，所以袋子只會在吃掉紅色軟糖時改變。吃掉紅色軟糖之前，袋中有 N 粒軟糖，其中 b 粒非紅色；之後袋中有 $N-1$ 粒軟糖，其中仍有 b 粒非紅色。由於每次抽取前都把袋子徹底搖勻，每次抽取都是從當時袋中的軟糖裏重新作一次均勻的隨機選取，所以我們可以把各次抽取的概率逐一相乘。

按吃掉的時間拆分。 對 $k = 1, 2, \ldots, n$，設 E_k 為以下事件：在 n 次抽取中唯一吃掉的那粒軟糖，是在第 k 次抽取時吃掉的。那麼第 $1, \ldots, k-1$ 次是從原來的袋中抽到非紅色，第 k 次是從原來的袋中抽到紅色，而第 $k+1, \ldots, n$ 次是從少了一粒的袋中抽到非紅色：

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

事件 $E_1, \ldots, E_n$ 互斥（disjoint），而它們的併集（union）就是「剛好吃掉一粒軟糖」這個事件 E。所以

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

這裏我們把每一項的分子和分母同乘以 $N^n (N-1)^n$。

想一想 怎樣求 $\sum_{k=1}^{n} N^{n-k}(N-1)^k$ 的和？

這是一個等比級數，首項為 $N^{n-1}(N-1)$，公比（common ratio）為 $\frac{N-1}{N}$。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 $x = N$、$y = N-1$，於是 $x - y = 1$：

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

因此

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

條件概率。 「我在最後一次抽取時吃掉它，而且只吃了一粒」這個事件就是 E_n，其概率為 $\mathrm{P}(E_n) = \dfrac{r\,b^{n-1}}{N^n}$。由於 $E_n \subseteq E$，

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

這正是所求的結果。

留意 r 和 b 完全消去，只剩下它們的總和。其實上面的計算還說明了更多：已知 E 發生，軟糖在第 k 次抽取時被吃掉的概率與 $(\frac{N-1}{N})^k$ 成正比，所以較早的抽取比較後的抽取更有可能。這是合理的：早些吃掉紅色軟糖，袋中非紅色軟糖所佔的比例便較大，餘下各次「再沒有抽到紅色」的機會也就較高。

驗算

▶ $n = 1$：公式給出 $\dfrac{1}{N - (N-1)} = 1$，正確，因為唯一的一次抽取就是最後一次。

- $n = 2$：公式給出 $\dfrac{N-1}{2N-1}$，略小於 $\frac{1}{2}$，與「較早的抽取可能性稍高」一致。
- 取 $r = 3$、$b = 5$、$n = 4$，公式給出 $343/1695 \approx 0.2024$，而若四次抽取的可能性均等，則應為 $\frac{1}{4}$。模擬 4×10^5 次的結果為 0.2028；另外，逐一列舉所有抽取序列並追蹤袋中內容，所得的精確結果在 $1 \le r, b \le 4$ 及 $1 \le n \le 8$ 的所有情況下都與公式完全吻合。

常見錯誤

- 在紅色軟糖被吃掉之後，仍然用 b/N 作為抽到非紅色的概率。這時袋中只有 $N-1$ 粒軟糖，而這正是本題的關鍵所在：如果由始至終都用 b/N，每個 E_k 的概率便會相同，答案就變成 $1/n$。
- 忘記取條件：單是 $\mathrm{P}(E_n) = rb^{n-1}/N^n$ 並不是答案，因為題目已告訴我們剛好吃掉了一粒軟糖。

STEP 之外

這裏的袋子是一個簡單的馬可夫鏈（Markov chain）：它的狀態是餘下紅色軟糖的數目，而狀態向下移動的機會只取決於當前的狀態。以這類鏈的最終狀態為條件，反過來問它所經過的路徑，正是「隱馬可夫模型」（hidden Markov models）以及對歷史過程作貝葉斯推斷（Bayesian inference）背後的想法。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率入門）正好以這個程度講解條件概率和馬可夫鏈，而 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率導論）則有許多同類風格的題目。

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。