詳細解答

STEP II 1992，第 16 題

概率與統計 • 條件概率

STEP II 1992，第 16 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/92-s2-q16

本題在 *Beyond STEP* 第 12.17 節巴拿赫的火柴盒中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch12/matchboxes.html

關鍵思路

這是巴拿赫（Banach）火柴盒問題的一個變體。題目告訴了我們何時發現糖果不夠（第 N 次伸手），所以這是一個條件概率（conditional probability）問題。因此我們要計算兩個聯合概率（joint probability）：第一次失敗發生在第 N 次伸手而且拿的是拖肥糖，以及第一次失敗發生在第 N 次伸手而且拿的是薄荷糖。兩者都是「頭 $N-1$ 次伸手中每種糖恰好各有若干次，然後下一次是指定的一種」這類計算，屬於二項（binomial）概率的計算。

解答

假設各次伸手互相獨立（independent），每次拿拖肥糖的概率為 p，拿薄荷糖的概率為 q，其中 $0 < p < 1$。設 T 為以下事件（event）：她在第 N 次伸手時想拿拖肥糖，卻發現拖肥糖那包已經空了，而且之前各次伸手都沒有發現任何一包是空的。設 M 為把拖肥糖換成薄荷糖後的相應事件。我們要求的是

$$\mathrm{P}(\text{拖肥糖} \mid \text{首次發現時的伸手次數為} N) = \frac{\mathrm{P}(T)}{\mathrm{P}(T) + \mathrm{P}(M)}.$$

T 的概率

要 T 發生，拖肥糖那包必須在 $N-1$ 次伸手之後已經空了。到目前為止沒有一次伸手失敗，所以這些伸手每一次都拿到了一顆糖。因此頭 $N-1$ 次伸手中恰好有 n 次是拿拖肥糖（多一次就會失敗），其餘 $N-1-n$ 次是拿薄荷糖。然後第 N 次伸手是拿拖肥糖。

反過來，假設頭 $N-1$ 次伸手中恰好有 n 次拿拖肥糖，以及 $N-1-n \le m$ 次拿薄荷糖，次序不限。那麼之前不可能有任何一次伸手失敗：在任何階段，她伸手拿拖肥糖都不超過 n 次，拿薄荷糖都不超過 m 次。這樣，如果第 N 次伸手是拿拖肥糖，T 就會發生。

> **想一想** 頭 $N-1$ 次伸手的次序難道不要緊嗎？不是要確保她不會提早把糖拿光嗎？

不要緊。她只有在某一包已空而又伸手進去的時候，才會發現那一包是空的。提早把拖肥糖拿光（即拿第 n 顆拖肥糖之時）並無問題，只要她在第 N 次伸手之前不再伸手拿拖肥糖便可。這一點自動成立，因為頭 $N-1$ 次伸手中只有 n 次是拿拖肥糖。所以全部 $\binom{N-1}{n}$ 種次序都可以，於是

$$\mathrm{P}(T) = \binom{N-1}{n} p^n q^{N-1-n} \cdot p = \binom{N-1}{n} p^{n+1} q^{N-1-n}, \qquad n \le N-1 \le n+m,$$

其他情況下 $\mathrm{P}(T) = 0$。如果 $N-1 < n$，她不可能已經把拖肥糖拿光。如果 $N-1 > n+m$，她就需要多於 m 顆薄荷糖。

M 的概率

把兩種糖的角色對調，用完全相同的方法可得

$$\mathrm{P}(M) = \binom{N-1}{m} q^{m+1} p^{N-1-m}, \qquad m \le N-1 \le n+m,$$

其他情況下 $\mathrm{P}(M) = 0$。

答案

這次發現必定發生在第 $\min(n,m)+1$ 次至第 $n+m+1$ 次伸手之間。第 $n+m+2$ 次伸手是不可能的，因為到那時兩包糖都已經空了，早就應該有一次伸手失敗。

- 如果 $N \le m$（即 $\mathrm{P}(M) = 0$ 但 $\mathrm{P}(T) > 0$），她肯定是在拿拖肥糖：概率為 1。
- 如果 $N \le n$（即 $\mathrm{P}(T) = 0$），概率為 0。

- 如果 $\max(n,m)+1 \le N \le n+m+1$，兩種情況都有可能，而

$$\mathrm{P}(\text{拖肥糖}) = \frac{\binom{N-1}{n} p^{n+1} q^{N-1-n}}{\binom{N-1}{n} p^{n+1} q^{N-1-n} + \binom{N-1}{m} q^{m+1} p^{N-1-m}}.$$

在最後一種情況，我們可以把分子和分母同時除以 $\mathrm{P}(T)$ 來化簡。二項式係數（binomial coefficient）中的公因數 $(N-1)!$ 互相約去，而各個冪則合併成 $(q/p)^{n+m+2-N}$。由此得出

$$\mathrm{P}(\text{拖肥糖}) = \left[1 + \frac{n!\,(N-1-n)!}{m!\,(N-1-m)!} \left(\frac{q}{p}\right)^{n+m+2-N}\right]^{-1}.$$

驗算

- 對稱性：如果 $n=m$ 而且 $p=q=\frac{1}{2}$，兩個式子相等，答案是 $\frac{1}{2}$，正如所料。
- 取 $n=2$、$m=3$、$p=\frac{1}{2}$、$N=5$。那麼 $\mathrm{P}(T) = \binom{4}{2}/2^5 = \frac{3}{16}$，而 $\mathrm{P}(M) = \binom{4}{3}/2^5 = \frac{1}{8}$，所以答案是 $\frac{3/16}{5/16} = \frac{3}{5}$。拖肥糖較少，所以先把拖肥糖拿光的可能性較大。
- 對 $N = 1,\ldots,n+m+1$ 求和，有 $\sum_N \big(\mathrm{P}(T)+\mathrm{P}(M)\big) = 1$，因為終究必有一次伸手會失敗。我們已就所有 $n,m \le 6$ 及若干個 p 值，以逐步的精確枚舉核對過這一點以及上述公式。以 $n=4$、$m=3$、$N=7$、$p=0.55$ 進行 2×10^5 次試驗的模擬（simulation），結果是 0.527，而精確值是 0.528。

常見錯誤

- 漏掉第 N 次伸手本身所對應的最後一個因數 p（或 q）。
- 答了 $\mathrm{P}(T)$，而不是條件概率 $\mathrm{P}(T)/(\mathrm{P}(T)+\mathrm{P}(M))$。題目已告訴我們，發現是在第 N 次伸手時發生的。
- 以為所指的事件是「拖肥糖在第 N 次伸手時拿光」。拿走最後一顆拖肥糖並不算發現，下一次再去拿才是發現。

STEP 之外

第一次失敗發生的時間，其分佈是兩個負二項（negative binomial）分佈的混合（mixture）。負二項分佈計算的是取得指定次數的成功所需的試驗次數。巴拿赫原本的火柴盒問題（$n=m$，$p=\frac{1}{2}$）問的是：在發現的一刻，另一個盒子裏還剩下多少根火柴。其答案可以用斯特靈公式（Stirling's formula）來近似，得出平均值（mean）約為 $2\sqrt{n/\pi}$。

關於負二項分佈，可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率入門教材）。費勒（Feller）的經典著作 *An Introduction to Probability Theory and Its Applications* 第 1 卷（概率論及其應用導論）中，有一段關於火柴盒問題的著名討論。

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。