詳細解答

STEP I 2006，第 14 題

概率與統計 • 不放回抽樣

> **STEP I 2006，第 14 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/06-s1-q14

關鍵思路

在第 r 次才取到紅色糖果，即是先取到 $r-1$ 顆藍色糖果，然後取到紅色那一顆。若每次抽取後放回（with replacement），每次抽取的情況都相同，所以概率只是幾個冪的簡單乘積；之後把 n 當作連續變量，對其對數求導，便可求出使概率最大的 n。若抽取後不放回（without replacement），最簡潔的看法是想像把所有糖果都抽出來：紅色那一顆落在任何一個位置的可能性都相等。

解答

以下 $n \geq 0$ 均為整數，而 R 表示首次取到紅色糖果時是第幾次抽取。

第 (i) 部分

由於糖果會放回袋中並搖勻，各次抽取互相獨立（independent），而每次抽取時，取到紅色糖果的概率是 $\frac{1}{n+1}$，取到藍色糖果的概率是 $\frac{n}{n+1}$。事件 $R=r$ 表示首 $r-1$ 次抽到藍色，而第 r 次抽到紅色，所以根據獨立性，

$$P(R=r) = \left(\frac{n}{n+1}\right)^{r-1} \frac{1}{n+1} = \frac{n^{r-1}}{(n+1)^r}, \qquad r = 1, 2, 3, \ldots$$

（這就是成功概率為 $\frac{1}{n+1}$ 的幾何分佈（geometric distribution）。）

現在固定 r，把上式看成 n 的函數：

$$f(n) = \frac{n^{r-1}}{(n+1)^r}.$$

若 $r=1$，則 $f(n) = \frac{1}{n+1}$，它在 $n=0$ 時最大（沒有藍色糖果時，第一顆必定是紅色）。

因此假設 $r \geq 2$。這時 $f(0) = 0$，而當 $n > 0$ 時 $f(n) > 0$，所以我們可以在 $n > 0$ 上取對數，並把 n 當作實變量：

$$\ln f(n) = (r-1)\ln n - r\ln(n+1), \qquad \frac{\mathrm{d}}{\mathrm{d}n}\ln f(n) = \frac{r-1}{n} - \frac{r}{n+1} = \frac{(r-1)-n}{n(n+1)}.$$

想一想 n 必須是整數。那麼求導是否合理？

合理，因為我們只是利用導數（derivative）來找出 f 在哪裏遞增、在哪裏遞減。當 $n > 0$ 時，分母 $n(n+1)$ 是正數，所以導數在 $0 < n < r-1$ 時為正，在 $n > r-1$ 時為負。因此 f 在 $[0, r-1]$ 上嚴格遞增（strictly increasing），在 $[r-1, \infty)$ 上嚴格遞減。由於 $r-1$ 本身是整數，它勝過其他所有整數，無須再比較相鄰的整數。

所以在所有情況下（包括 $r = 1$），使概率最大的是

$$n = r - 1,$$

最大值為 $\dfrac{(r-1)^{r-1}}{r^r}$（當 $r = 1$ 時理解為 1）。例如，當 $r = 3$ 時，最理想的袋中有 $n = 2$ 顆藍色糖果，所得概率為 $\frac{4}{27}$。

第 (ii) 部分

現在各次抽取不再獨立，但有兩個簡單的做法。

*直接相乘。*要 $R = r$，首 $r-1$ 次每次都要抽到藍色，然後抽到紅色。這只有在藍色糖果足夠多時才可能發生，即 $r - 1 \leq n$。已知之前抽到的全是藍色，第 k 次抽取時袋中有 $n + 2 - k$ 顆糖果，其中一顆是紅色。所以

$$P(R = r) = \frac{n}{n+1} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{n-r+2}{n-r+3} \cdot \frac{1}{n-r+2}.$$

每個分子都與下一個分母約去，乘積逐項相消（telescope），得

$$P(R = r) = \begin{cases} \dfrac{1}{n+1}, & 1 \leq r \leq n+1, \\ 0, & r > n+1. \end{cases}$$

*對稱性。*抽到紅色糖果便停止，並不影響它所在的位置。所以不妨想像一直抽下去，直至袋中空無一物。這些（可區分的）糖果共有 $(n+1)!$ 種排列次序，每種的可能性都相等，因此紅色糖果落在 $n+1$ 個位置中任何一個的可能性都相等。這樣得出相同的答案。

至於求最大值，先固定 r。若 $n < r-1$，概率為 0。若 $n \geq r-1$，概率為 $\frac{1}{n+1}$，它隨 n 增大而嚴格遞減。所以使概率最大的，是符合條件的最小的袋，

$$n = r - 1,$$

而最大概率為 $\frac{1}{r}$。（當 $n = r-1$ 時，紅色糖果是 r 顆糖果中任何一顆的可能性都相等，而最後才抽到它，正是 r 個等可能結果的其中一個。）

驗算

- 在 (i) 中，$\sum_{r\geq 1} P(R=r) = \frac{1}{n+1}\cdot\dfrac{1}{1-n/(n+1)} = 1$；在 (ii) 中，$n+1$ 個相等的值 $\frac{1}{n+1}$ 之和為 1。
- 在 (i) 中取 $r=3$：$n=1,2,3$ 分別給出 $\frac{1}{8}$、$\frac{4}{27}\approx 0.148$ 和 $\frac{9}{64}\approx 0.141$，所以 $n=2$ 勝出。
- 我們就 $n\leq 6$ 列出糖果的所有排列次序，驗證了 (ii)。就 (i) 取 $n=3$、$r=2$ 進行模擬，得出 0.188，而準確值為 $\frac{3}{16}=0.1875$。

常見錯誤

- 在 (i) 中寫成 $\left(\frac{n}{n+1}\right)^r\frac{1}{n+1}$：在紅色之前只有 $r-1$ 顆藍色。
- 在 (ii) 中，忘記了當 $r>n+1$ 時答案是 0。正是這一點，令「把 n 取得愈小愈好」不致一直滑落到 $n=0$。

STEP 之外

這兩個求最大值的問題，其實都是*最大似然估計*（maximum likelihood estimation）。假如你看到紅色糖果在第 r 次首次出現，卻不知道 n 是多少，那麼在兩個實驗中，令你的觀察結果最有可能出現的 n 值都是 $n=r-1$。這是大學一年級統計學的基石之一。(ii) 中的對稱性論證是*可交換性*（exchangeability）的一個例子，它令許多不放回抽樣的問題變得容易。這兩個概念在 Blitzstein 與 Hwang 所著的 *Introduction to Probability*（概率論入門教科書）中都有詳細闡述。

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。