詳細解答

# STEP II 2008，第 13 題

概率與統計 • 不放回抽取

> **STEP II 2008，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/08-s2-q13

## 關鍵思路

不必追蹤每一枚籌碼：只追蹤黑色的就夠了。每枚黑色籌碼的去向由兩次抽取決定，而每次抽取時，某一枚指定籌碼被抽中的機會就是（抽出的數目）/（袋中的數目）。按第一次抽取的結果分情況討論，兩部分都會化成簡短的和，而且化簡的程度遠超你的預期。

## 解答

在整個解答中，「隨機抽出」是指袋中每一組 $k$ 枚籌碼被抽出的可能性均等。我們反覆用到的事實是：若從一個裝有 $m$ 枚籌碼的袋中隨機抽出 $k$ 枚，則根據對稱性（symmetry），某一枚指定的籌碼在其中的概率為 $k/m$（這 $m$ 枚籌碼各自被抽中的可能性均等，而平均有 $k$ 枚被抽中）。

第一次轉移之後，袋 $P$ 有 $n-k$ 枚籌碼，袋 $Q$ 有 $n+k$ 枚。第二次轉移時，從這 $n+k$ 枚中抽出 $k$ 枚，移到 $P$。

### 第 (i) 部分

設 $B$ 為那枚黑色籌碼，它起初在 $P$ 中。按 $B$ 在第一次轉移時有沒有移動來分情況。

- $B$ 沒有從 $P$ 中被抽出：概率為 $\dfrac{n-k}{n}$。這時 $B$ 自始至終留在 $P$ 中，因為第二次轉移只會從 $Q$ 取出籌碼。
- $B$ 從 $P$ 中被抽出：概率為 $\dfrac{k}{n}$。這時 $B$ 是 $Q$ 中 $n+k$ 枚籌碼之一，而它回到 $P$ 當且僅當它在被抽回的 $k$ 枚之中，其概率為 $\dfrac{k}{n+k}$。

這兩種情況互斥（disjoint），並且涵蓋了所有可能，所以

$$\begin{aligned}\mathrm{P}(B\text{ 最終在}P) &= \frac{n-k}{n} + \frac{k}{n}\cdot\frac{k}{n+k} = \frac{(n-k)(n+k)+k^2}{n(n+k)}\\ &= \frac{n^2}{n(n+k)} = \frac{n}{n+k}.\end{aligned}$$

這個概率隨 $k$ 嚴格遞減，所以在 $0 \le k \le n$ 的範圍內，它只在 $k = 0$ 時取得最大值，此時其值為 1。（這並不意外：如果甚麼都不移動，黑色籌碼必定留在原處。）

**第 (ii) 部分**

現在設 $B_1$ 為起初在 $P$ 中的黑色籌碼，$B_2$ 為起初在 $Q$ 中的那一枚。設 $S$ 為兩者最終在同一個袋中的事件。同樣按 $B_1$ 在第一次轉移時有沒有移動來分情況。

**情況 1：$B_1$ 留在 $P$ 中**（概率為 $\frac{n-k}{n}$）。這時無論之後發生甚麼，$B_1$ 最終都在 $P$ 中，所以 $S$ 發生當且僅當 $B_2$ 在從 $n+k$ 枚籌碼中抽出的 $k$ 枚之中，這些籌碼都在 $Q$ 內：

$$\mathrm{P}(S \mid \text{情況 1}) = \frac{k}{n+k}.$$

**情況 2：$B_1$ 移到 $Q$**（概率為 $\frac{k}{n}$）。這時兩枚黑色籌碼都在 $Q$ 的 $n+k$ 枚籌碼之中，而 $S$ 發生的條件是第二次抽取把它們兩枚都取走，或者兩枚都不取走。

> **想一想** 從 $n+k$ 枚籌碼中隨機抽出一個 $k$ 元子集，它同時包含兩枚指定籌碼，或者兩枚都不包含的概率是多少？

點算子集的數目。共有 $\binom{n+k}{k}$ 種可能性均等的選法。同時包含兩枚黑色籌碼的選法，是從餘下的 $n+k-2$ 枚中選出另外 $k-2$ 枚；兩枚都不包含的選法，則是從 $n+k-2$ 枚白色籌碼中選出全部 $k$ 枚。相除可得

$$\begin{aligned}\mathrm{P}(\text{兩枚都有}) &= \binom{n+k-2}{k-2}\Big/\binom{n+k}{k} = \frac{k(k-1)}{(n+k)(n+k-1)},\\ \mathrm{P}(\text{兩枚都沒有}) &= \binom{n+k-2}{k}\Big/\binom{n+k}{k} = \frac{n(n-1)}{(n+k)(n+k-1)}.\end{aligned}$$

（這些公式在 $k = 0, 1$ 時同樣成立，此時 $\mathrm{P}(\text{兩枚都有}) = 0$。）所以

$$\mathrm{P}(S \mid \text{情況 2}) = \frac{k(k-1)+n(n-1)}{(n+k)(n+k-1)}.$$

**把各情況合併**。根據全概率定律（law of total probability），

$$\begin{aligned}\mathrm{P}(S) &= \frac{n-k}{n}\cdot\frac{k}{n+k} + \frac{k}{n}\cdot\frac{k(k-1)+n(n-1)}{(n+k)(n+k-1)}\\ &= \frac{k\big[(n-k)(n+k-1)+k^2-k+n^2-n\big]}{n(n+k)(n+k-1)}.\end{aligned}$$

在方括號內，$(n-k)(n+k-1) = n^2 - k^2 - n + k$，所以方括號等於 $2n^2 - 2n = 2n(n-1)$。因此

$$\boxed{\mathrm{P}(S) = \frac{2k(n-1)}{(n+k)(n+k-1)}}$$

**求最大值**。當 $k = 0$ 時概率為 0，所以取 $1 \le k \le n$，並求其倒數的最小值。撇開正常數 $2(n-1)$ 不計，

$$\frac{(n+k)(n+k-1)}{k} = k + (2n-1) + \frac{n(n-1)}{k},$$

所以我們需要在整數範圍內求 $g(k) = k + \dfrac{n(n-1)}{k}$ 的最小值。比較相鄰兩項：

$$g(k+1) - g(k) = 1 - \frac{n(n-1)}{k(k+1)}.$$

當 $k(k+1) < n(n-1)$，即 $k < n-1$ 時，這個差為負數；而當 $k = n-1$ 時，它恰好為零，因為此時 $k(k+1) = (n-1)n$。所以 $g$ 由 $k = 1$ 到 $k = n-1$ 嚴格遞減，並且 $g(n) = g(n-1)$。因此概率取得最大值之處為

$$k = n-1 \quad \text{及} \quad k = n,$$

由於 $n \ge 2$，兩者都在容許的範圍內。當 $k = n-1$ 時，最大值為

$$\frac{2(n-1)^2}{(2n-1)(2n-2)} = \frac{n-1}{2n-1},$$

而 $k = n$ 給出相同的值，$\dfrac{2n(n-1)}{2n(2n-1)} = \dfrac{n-1}{2n-1}$。（把 $g$ 看成連續函數時，其最小值點為 $\sqrt{n(n-1)}$，嚴格位於 $n-1$ 與 $n$ 之間；這兩個整數處的值是完全相等，而非近似相等。）

## 驗算

- 第 (i) 部分取 $k = n$：得 $\frac{1}{2}$。這時全部 $2n$ 枚籌碼都經過 $Q$，其中 $n$ 枚返回，所以黑色籌碼最終落在兩個袋中的可能性均等。吻合。
- 第 (ii) 部分取 $k = 0$ 得 0，這是正確的，因為兩枚黑色籌碼永遠不會相遇。取 $n = 2$、$k = 1$ 得 $\frac{2}{3 \cdot 2} = \frac{1}{3}$，你可以親手驗證。
- 我們對 $2 \le n \le 7$ 及每一個 $k$，以分數精確地列舉了每一對可能的抽取結果；兩條公式在所有情況下都吻合，而 (ii) 中取得最大值之處總是 $\{n-1, n\}$，例如最大值為 $\frac{3}{7} \approx 0.4286$，對應 $n = 4$。

## 常見錯誤

- 在第二次抽取時用了 $k/n$ 而不是 $k/(n+k)$：袋 $Q$ 已增至 $n+k$ 枚籌碼。

- 在第 (ii) 部分的情況 2 中，忘記了「兩枚都沒有被抽出」的可能：如果兩枚黑色籌碼都留在 $Q$ 中，它們仍然在一起。
- 用微積分求最大值，再把 $\sqrt{n(n-1)}$ 捨入為單一個整數。離散的比較顯示，有兩處同樣取得最大值。

## STEP 之外

兩部分背後的工具是超幾何分佈（hypergeometric distribution）：在不放回的隨機抽樣中，樣本內帶記號物件的數目。先以第一階段的結果為條件（conditioning），再把它用於第二階段，這就是全概率定律；同樣的兩階段推理，也是容器之間混合與擴散模型的基礎，例如埃倫費斯特甕模型（Ehrenfest urn）。Blitzstein and Hwang, *Introduction to Probability*（概率論入門教材）正是以這種方式處理超幾何分佈與條件概率。

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。