详细解答

STEP II 2008，第 13 题

概率与统计 • 不放回抽取

STEP II 2008，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/08-s2-q13

关键思路

不必追踪每一枚筹码：只追踪黑色的就够了。每枚黑色筹码的去向由两次抽取决定，而每次抽取时，某一枚指定筹码被抽中的机会就是（抽出的数目）/（袋中的数目）。按第一次抽取的结果分情况讨论，两部分都会化成简短的和，而且化简的程度远超你的预期。

解答

在整个解答中，“随机抽出”是指袋中每一组 k 枚筹码被抽出的可能性均等。我们反复用到的事实是：若从一个装有 m 枚筹码的袋中随机抽出 k 枚，则根据对称性（symmetry），某一枚指定的筹码在其中的概率为 k/m（这 m 枚筹码各自被抽中的可能性均等，而平均有 k 枚被抽中）。

第一次转移之后，袋 P 有 $n-k$ 枚筹码，袋 Q 有 $n+k$ 枚。第二次转移时，从这 $n+k$ 枚中抽出 k 枚，移到 P。

第 (i) 部分

设 B 为那枚黑色筹码，它起初在 P 中。按 B 在第一次转移时有没有移动来分情况。

- B 没有从 P 中被抽出：概率为 $\dfrac{n-k}{n}$。这时 B 自始至终留在 P 中，因为第二次转移只会从 Q 取出筹码。
- B 从 P 中被抽出：概率为 $\dfrac{k}{n}$。这时 B 是 Q 中 $n+k$ 枚筹码之一，而它回到 P 当且仅当它在被抽回的 k 枚之中，其概率为 $\dfrac{k}{n+k}$。

这两种情况互斥（disjoint），并且涵盖了所有可能，所以

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

这个概率随 k 严格递减，所以在 $0 \le k \le n$ 的范围内，它只在 $k = 0$ 时取得最大值，此时其值为 1。（这并不意外：如果甚么都不移动，黑色筹码必定留在原处。）

第 (ii) 部分

现在设 B_1 为起初在 P 中的黑色筹码，B_2 为起初在 Q 中的那一枚。设 S 为两者最终在同一个袋中的事件。同样按 B_1 在第一次转移时有没有移动来分情况。

情况 1：B_1 留在 P 中（概率为 $\frac{n-k}{n}$）。这时无论之后发生甚么，B_1 最终都在 P 中，所以 S 发生当且仅当 B_2 在从 $n+k$ 枚筹码中抽出的 k 枚之中，这些筹码都在 Q 内：

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

情况 2：B_1 移到 Q（概率为 $\frac{k}{n}$）。这时两枚黑色筹码都在 Q 的 $n+k$ 枚筹码之中，而 S 发生的条件是第二次抽取把它们两枚都取走，或者两枚都不取走。

> **想一想** 从 $n+k$ 枚筹码中随机抽出一个 k 元子集，它同时包含两枚指定筹码，或者两枚都不包含的概率是多少？

计数子集的数目。共有 $\binom{n+k}{k}$ 种可能性均等的选法。同时包含两枚黑色筹码的选法，是从余下的 $n+k-2$ 枚中选出另外 $k-2$ 枚；两枚都不包含的选法，则是从 $n+k-2$ 枚白色筹码中选出全部 k 枚。相除可得

$$\begin{aligned}\mathrm{P}(\text{两枚都有}) &= \binom{n+k-2}{k-2}\Big/\binom{n+k}{k} = \frac{k(k-1)}{(n+k)(n+k-1)},\\ \mathrm{P}(\text{两枚都没有}) &= \binom{n+k-2}{k}\Big/\binom{n+k}{k} = \frac{n(n-1)}{(n+k)(n+k-1)}.\end{aligned}$$

（这些公式在 $k = 0, 1$ 时同样成立，此时 $\mathrm{P}(\text{两枚都有}) = 0$。）所以

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

把各情况合并。根据全概率定律（law of total probability），

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

在方括号内, $(n-k)(n+k-1) = n^2-k^2-n+k$, 所以方括号等于 $2n^2-2n = 2n(n-1)$。因此

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

求最大值。当 $k=0$ 时概率为 0，所以取 $1 \le k \le n$，并求其倒数的最小值。撇开正常数 $2(n-1)$ 不计，

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

所以我们需要在整数范围内求 $g(k) = k + \dfrac{n(n-1)}{k}$ 的最小值。比较相邻两项:

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

当 $k(k+1) < n(n-1)$，即 $k < n-1$ 时，这个差为负数；而当 $k = n-1$ 时，它恰好为零，因为此时 $k(k+1) = (n-1)n$。所以 g 由 $k=1$ 到 $k = n-1$ 严格递减，并且 $g(n) = g(n-1)$。因此概率取得最大值之处为

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

由于 $n \ge 2$，两者都在容许的范围内。当 $k = n-1$ 时，最大值为

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

而 $k=n$ 给出相同的值，$\dfrac{2n(n-1)}{2n(2n-1)} = \dfrac{n-1}{2n-1}$。（把 g 看成连续函数时，其最小值点为 $\sqrt{n(n-1)}$，严格位于 $n-1$ 与 n 之间；这两个整数处的值是完全相等，而非近似相等。）

验算

- 第 (i) 部分取 $k=n$：得 $\frac{1}{2}$。这时全部 $2n$ 枚筹码都经过 Q，其中 n 枚返回，所以黑色筹码最终落在两个袋中的可能性均等。吻合。
- 第 (ii) 部分取 $k=0$ 得 0，这是正确的，因为两枚黑色筹码永远不会相遇。取 $n=2$、$k=1$ 得 $\frac{2}{3\cdot 2} = \frac{1}{3}$，你可以亲手验证。
- 我们对 $2 \le n \le 7$ 及每一个 k，以分数精确地列举了每一对可能的抽取结果；两条公式在所有情况下都吻合，而 (ii) 中取得最大值之处总是 $\{n-1, n\}$，例如最大值为 $\frac{3}{7} \approx 0.4286$，对应 $n=4$。

常见错误

- 在第二次抽取时用了 k/n 而不是 $k/(n+k)$：袋 Q 已增至 $n+k$ 枚筹码。

- 在第 (ii) 部分的情况 2 中，忘记了“两枚都没有被抽出”的可能：如果两枚黑色筹码都留在 Q 中，它们仍然在一起。
- 用微积分求最大值，再把 $\sqrt{n(n-1)}$ 舍入为单一个整数。离散的比较显示，有两处同样取得最大值。

STEP 之外

两部分背后的工具是超几何分布（hypergeometric distribution）：在不放回的随机抽样中，样本内带记号物件的数目。先以第一阶段的结果为条件（conditioning），再把它用于第二阶段，这就是全概率定律；同样的两阶段推理，也是容器之间混合与扩散模型的基础，例如埃伦费斯特瓮模型（Ehrenfest urn）。Blitzstein and Hwang, *Introduction to Probability*（概率论入门教材）正是以这种方式处理超几何分布与条件概率。

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。