详细解答

STEP I 2006，第 14 题

概率与统计 • 不放回抽样

STEP I 2006，第 14 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/06-s1-q14

关键思路

在第 r 次才取到红色糖果，即是先取到 $r-1$ 颗蓝色糖果，然后取到红色那一颗。若每次抽取后放回（with replacement），每次抽取的情况都相同，所以概率只是几个幂的简单乘积；之后把 n 当作连续变量，对其对数求导，便可求出使概率最大的 n。若抽取后不放回（without replacement），最简洁的看法是想像把*所有*糖果都抽出来：红色那一颗落在任何一个位置的可能性都相等。

解答

以下 $n \geq 0$ 均为整数，而 R 表示首次取到红色糖果时是第几次抽取。

第 (i) 部分

由于糖果会放回袋中并摇匀，各次抽取互相独立（independent），而每次抽取时，取到红色糖果的概率是 $\frac{1}{n+1}$，取到蓝色糖果的概率是 $\frac{n}{n+1}$。事件 $R=r$ 表示首 $r-1$ 次抽到蓝色，而第 r 次抽到红色，所以根据独立性，

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

（这就是成功概率为 $\frac{1}{n+1}$ 的几何分布（geometric distribution）。）

现在固定 r，把上式看成 n 的函数：

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

若 $r=1$，则 $f(n)=\frac{1}{n+1}$，它在 $n=0$ 时最大（没有蓝色糖果时，第一颗必定是红色）。

因此假设 $r \geq 2$。这时 $f(0) = 0$，而当 $n > 0$ 时 $f(n) > 0$，所以我们可以在 $n > 0$ 上取对数，并把 n 当作实变量：

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

想一想 n 必须是整数。那么求导是否合理?

合理，因为我们只是利用导数（derivative）来找出 f 在哪里递增、在哪里递减。当 $n > 0$ 时，分母 $n(n+1)$ 是正数，所以导数在 $0 < n < r-1$ 时为正，在 $n > r-1$ 时为负。因此 f 在 $[0, r-1]$ 上严格递增（strictly increasing），在 $[r-1, \infty)$ 上严格递减。由于 $r-1$ 本身是整数，它胜过其他所有整数，无须再比较相邻的整数。

所以在所有情况下（包括 $r = 1$），使概率最大的是

$$n = r - 1,$$

最大值为 $\dfrac{(r-1)^{r-1}}{r^r}$（当 $r = 1$ 时理解为 1）。例如，当 $r = 3$ 时，最理想的袋中有 $n = 2$ 颗蓝色糖果，所得概率为 $\frac{4}{27}$。

第 (ii) 部分

现在各次抽取不再独立，但有两个简单的做法。

直接相乘。要 $R = r$，首 $r-1$ 次每次都要抽到蓝色，然后抽到红色。这只有在蓝色糖果足够多时才可能发生，即 $r - 1 \leq n$。已知之前抽到的全是蓝色，第 k 次抽取时袋中有 $n + 2 - k$ 颗糖果，其中一颗是红色。所以

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

每个分子都与下一个分母约去，乘积逐项相消（telescope），得

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

对称性。抽到红色糖果便停止，并不影响它所在的位置。所以不妨想像一直抽下去，直至袋中空无一物。这些（可区分的）糖果共有 $(n+1)!$ 种排列次序，每种的可能性都相等，因此红色糖果落在 $n+1$ 个位置中任何一个的可能性都相等。这样得出相同的答案。

至于求最大值，先固定 r。若 $n < r-1$，概率为 0。若 $n \geq r-1$，概率为 $\frac{1}{n+1}$，它随 n 增大而严格递减。所以使概率最大的，是符合条件的最小的袋，

$$n = r - 1,$$

而最大概率为 $\frac{1}{r}$。（当 $n = r-1$ 时，红色糖果是 r 颗糖果中任何一颗的可能性都相等，而最后才抽到它，正是 r 个等可能结果的其中一个。）

验算

- 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，$n+1$ 个相等的值 $\frac{1}{n+1}$ 之和为 1。
- 在 (i) 中取 $r=3$：$n=1,2,3$ 分别给出 $\frac{1}{8}$、$\frac{4}{27}\approx 0.148$ 和 $\frac{9}{64}\approx 0.141$，所以 $n=2$ 胜出。
- 我们就 $n\leq 6$ 列出糖果的所有排列次序，验证了 (ii)。就 (i) 取 $n=3$、$r=2$ 进行模拟，得出 0.188，而准确值为 $\frac{3}{16}=0.1875$。

常见错误

- 在 (i) 中写成 $\left(\frac{n}{n+1}\right)^r\frac{1}{n+1}$：在红色之前只有 $r-1$ 颗蓝色。
- 在 (ii) 中，忘记了当 $r>n+1$ 时答案是 0。正是这一点，令“把 n 取得愈小愈好”不致一直滑落到 $n=0$。

STEP 之外

这两个求最大值的问题，其实都是*最大似然估计*（maximum likelihood estimation）。假如你看到红色糖果在第 r 次首次出现，却不知道 n 是多少，那么在两个实验中，令你的观察结果最有可能出现的 n 值都是 $n=r-1$。这是大学一年级统计学的基石之一。(ii) 中的对称性论证是*可交换性*（exchangeability）的一个例子，它令许多不放回抽样的问题变得容易。这两个概念在 Blitzstein 与 Hwang 所著的 *Introduction to Probability*（概率论入门教科书）中都有详细阐述。

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。