详细解答

STEP I 2007，第 12 题

概率与统计 · 条件概率

STEP I 2007，第 12 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/07-s1-q12

关键思路

抽第二粒糖果时，我们并不知道袋中的情况，因为那取决于之前发生了甚么事。所以我们要考虑条件概率（conditional probability）：按第二次抽取之前发生的所有情况分类，求出每种情况下抽到红色的机会，再用全概率定理（law of total probability）加起来。第 (ii) 部分有一条好用的捷径：一个袋刚收到一粒随机糖果后，袋中红色糖果的期望数目（expected number）很容易写出来，所以我们不必就转移的那粒糖果的颜色重复分类。

解答

第 (i) 部分

设 R_1 为第一粒糖果是红色的事件，R_2 为第二粒糖果是红色的事件。每粒糖果被首先抽中的可能性均等，所以

$$\mathrm{P}(R_1) = \frac{a}{N}.$$

至于 R_2，按第一粒糖果的颜色分类。若第一粒是红色，则余下 $N-1$ 粒糖果，其中 $a-1$ 粒是红色；若不是红色，则余下 $N-1$ 粒，其中 a 粒是红色。（这里 $N \geq 2$ 保证了第二粒糖果存在。）由全概率定理，

$$\begin{aligned}\mathrm{P}(R_2) &= \mathrm{P}(R_1)\,\mathrm{P}(R_2 \mid R_1) + \mathrm{P}(R_1')\,\mathrm{P}(R_2 \mid R_1') \\ &= \frac{a}{N}\cdot\frac{a-1}{N-1} + \frac{N-a}{N}\cdot\frac{a}{N-1} \\ &= \frac{a[(a-1)+(N-a)]}{N(N-1)} = \frac{a(N-1)}{N(N-1)} = \frac{a}{N}.\end{aligned}$$

所以 $\mathrm{P}(R_2) = \mathrm{P}(R_1)$，正是所求。（用对称性（symmetry）也可得出同一结论：由两粒不同糖果组成的每个有序对出现的可能性均等，所以第二粒糖果是某一粒指定糖果的机会，与第一粒是该糖果的机会相同。）

第 (ii) 部分

这里的“第一粒糖果”是第一次掷硬币后转移的那一粒，“第二粒糖果”是第二次掷硬币后选出的那一粒。设 R_1、R_2 分别为这两粒糖果是红色的事件。两次掷硬币互相独立（independent），而每次掷硬币的结果亦与选中哪粒糖果无关。

第一粒糖果。 按第一次掷硬币的结果分类，

$$\mathrm{P}(R_1) = p \cdot \frac{a}{N} + q \cdot \frac{b}{N} = \frac{pa + qb}{N}.$$

第二粒糖果。 两次掷硬币共有四种情况。

正面、正面。先从第 1 袋取出一粒糖果，然后再从第 1 袋取出另一粒。这正是把第 (i) 部分的情况套用于第 1 袋（有 N 粒糖果，其中 a 粒红色），所以第二粒糖果是红色的概率为 a/N。

正面、反面。一粒糖果由第 1 袋移到第 2 袋，第 2 袋现在有 $N+1$ 粒糖果；然后我们从第 2 袋抽取。第 2 袋有 $b+1$ 粒红色糖果的概率为 a/N，有 b 粒红色糖果的概率为 $1 - a/N$，所以

$$\mathrm{P}(R_2 \mid \mathrm{HT}) = \frac{a}{N} \cdot \frac{b+1}{N+1} + \left(1 - \frac{a}{N}\right) \frac{b}{N+1} = \frac{b + \frac{a}{N}}{N+1} = \frac{Nb + a}{N(N+1)}.$$

? **想一想** 为甚么第二行可以化简得如此干净俐落?

分子就是转移之后第 2 袋中红色糖果的期望数目：原有的 b 粒，加上新来的糖果是红色的机会 a/N。由于在这种情况下第 2 袋的大小固定为 $N+1$，抽到红色的机会就是这个期望数目除以 $N+1$。

反面、正面。把两个袋对调，用同样的论证可知，第 1 袋现在有 $N+1$ 粒糖果，其中红色糖果的期望数目为 $a + b/N$，所以

$$\mathrm{P}(R_2 \mid \mathrm{TH}) = \frac{Na + b}{N(N+1)}.$$

反面、反面。把第 (i) 部分套用于第 2 袋，得概率为 b/N。

现在把四种情况合并，它们的概率分别为 p^2、pq、qp、q^2：

$$\begin{aligned}\mathrm{P}(R_2) &= p^2\frac{a}{N} + pq\,\frac{(Nb+a)+(Na+b)}{N(N+1)} + q^2\frac{b}{N}\\ &= p^2\frac{a}{N} + pq\,\frac{(N+1)(a+b)}{N(N+1)} + q^2\frac{b}{N}\\ &= \frac{p^2a + pqa + pqb + q^2b}{N}\\ &= \frac{pa(p+q)+qb(p+q)}{N} = \frac{pa+qb}{N},\end{aligned}$$

最后一步用了 $p+q=1$。这等于 $\mathrm{P}(R_1)$，正是我们要证明的结果。

验算

- 若 $p=1$，一切都在第 1 袋中发生，两粒糖果的答案 a/N 正好就是第 (i) 部分。若 $p=0$，则得到第 2 袋的第 (i) 部分。
- 若 $a=b=0$，则没有红色糖果，两个概率都是 0；若 $a=b=N$，则每粒糖果都是红色，两个概率都是 1。
- 取 $N=5$、$a=2$、$b=4$、$p=0.3$，公式给出 $(0.6+2.8)/5=0.68$。把整个过程模拟 200 000 次，第一粒糖果的结果为 0.679，第二粒为 0.681；而对所有较小的 N、a、b 逐一穷举计算，结果亦吻合。

常见错误

- 忘记了转移之后，收到糖果的袋有 $N+1$ 粒糖果而不是 N 粒，而送出糖果的袋则有 $N-1$ 粒。
- 在 HH 和 TT 两种情况中，把第二次抽取当作与第一次独立。其实两者并不独立，但第 (i) 部分告诉我们答案始终不变，所以应直接引用，而不必重新计算。

STEP 之外

这两部分都是同一个深刻想法的例子：如果你从瓮中抽取一件东西而不看抽到的是甚么，你对瓮的认识并没有改变，所以下一次抽取的分布（distribution）依然相同。严格来说，这一连串抽取是可交换的（exchangeable），而在第 (ii) 部分中，$p\times$（第 1 袋中红色所占比例）$+$ $q\times$（第 2 袋中红色所占比例）这个量，即下一粒抽出的糖果是红色的机会，是一个鞅（martingale）：它未来的期望值等于它现在的值。鞅是现代概率论的核心，由赌博到金融都有它的身影。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）是

认识条件期望（conditional expectation）和这些想法的好起点。

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。