详细解答

STEP II 1992，第 16 题

概率与统计 • 条件概率

STEP II 1992，第 16 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/92-s2-q16

本题在 *Beyond STEP* 第 12.17 节巴拿赫的火柴盒中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/matchboxes.html

关键思路

这是巴拿赫（Banach）火柴盒问题的一个变体。题目告诉了我们何时发现糖果不够（第 N 次伸手），所以这是一个条件概率（conditional probability）问题。因此我们要计算两个联合概率（joint probability）：第一次失败发生在第 N 次伸手而且拿的是拖肥糖，以及第一次失败发生在第 N 次伸手而且拿的是薄荷糖。两者都是“头 $N-1$ 次伸手中每种糖恰好各有若干次，然后下一次是指定的一种”这类计算，属于二项（binomial）概率的计算。

解答

假设各次伸手互相独立（independent），每次拿拖肥糖的概率为 p，拿薄荷糖的概率为 q，其中 $0 < p < 1$。设 T 为以下事件（event）：她在第 N 次伸手时想拿拖肥糖，却发现拖肥糖那包已经空了，而且之前各次伸手都没有发现任何一包是空的。设 M 为把拖肥糖换成薄荷糖后的相应事件。我们要求的是

$$\mathrm{P}(\text{拖肥糖} \mid \text{首次发现时的伸手次数为} N) = \frac{\mathrm{P}(T)}{\mathrm{P}(T) + \mathrm{P}(M)}.$$

T 的概率

要 T 发生，拖肥糖那包必须在 $N-1$ 次伸手之后已经空了。到目前为止没有一次伸手失败，所以这些伸手每一次都拿到了一颗糖。因此头 $N-1$ 次伸手中恰好有 n 次是拿拖肥糖（多一次就会失败），其余 $N-1-n$ 次是拿薄荷糖。然后第 N 次伸手是拿拖肥糖。

反过来，假设头 $N-1$ 次伸手中恰好有 n 次拿拖肥糖，以及 $N-1-n \le m$ 次拿薄荷糖，次序不限。那么之前不可能有任何一次伸手失败：在任何阶段，她伸手拿拖肥糖都不超过 n 次，拿薄荷糖都不超过 m 次。这样，如果第 N 次伸手是拿拖肥糖，T 就会发生。

> **想一想** 头 $N-1$ 次伸手的次序难道不要紧吗？不是要确保她不会提早把糖拿光吗？

不要紧。她只有在某一包已空而又伸手进去的时候，才会发现那一包是空的。提早把拖肥糖拿光（即拿第 n 颗拖肥糖之时）并无问题，只要她在第 N 次伸手之前不再伸手拿拖肥糖便可。这一点自动成立，因为头 $N-1$ 次伸手中只有 n 次是拿拖肥糖。所以全部 $\binom{N-1}{n}$ 种次序都可以，于是

$$\mathrm{P}(T) = \binom{N-1}{n} p^n q^{N-1-n} \cdot p = \binom{N-1}{n} p^{n+1} q^{N-1-n}, \qquad n \le N-1 \le n+m,$$

其他情况下 $\mathrm{P}(T)=0$。如果 $N-1<n$，她不可能已经把拖肥糖拿光。如果 $N-1>n+m$，她就需要多于 m 颗薄荷糖。

M 的概率

把两种糖的角色对调，用完全相同的方法可得

$$\mathrm{P}(M) = \binom{N-1}{m} q^{m+1} p^{N-1-m}, \qquad m \le N-1 \le n+m,$$

其他情况下 $\mathrm{P}(M)=0$。

答案

这次发现必定发生在第 $\min(n,m)+1$ 次至第 $n+m+1$ 次伸手之间。第 $n+m+2$ 次伸手是不可能的，因为到那时两包糖都已经空了，早就应该有一次伸手失败。

- 如果 $N \le m$（即 $\mathrm{P}(M)=0$ 但 $\mathrm{P}(T)>0$），她肯定是在拿拖肥糖：概率为 1。
- 如果 $N \le n$（即 $\mathrm{P}(T)=0$），概率为 0。

- 如果 $\max(n,m)+1 \le N \le n+m+1$，两种情况都有可能，而

$$\mathrm{P}(\text{拖肥糖}) = \frac{\binom{N-1}{n} p^{n+1} q^{N-1-n}}{\binom{N-1}{n} p^{n+1} q^{N-1-n} + \binom{N-1}{m} q^{m+1} p^{N-1-m}}.$$

在最后一种情况，我们可以把分子和分母同时除以 $\mathrm{P}(T)$ 来化简。二项式系数（binomial coefficient）中的公因数 $(N-1)!$ 互相约去，而各个幂则合并成 $(q/p)^{n+m+2-N}$。由此得出

$$\mathrm{P}(\text{拖肥糖}) = \left[1 + \frac{n!\,(N-1-n)!}{m!\,(N-1-m)!} \left(\frac{q}{p}\right)^{n+m+2-N}\right]^{-1}.$$

验算

- 对称性：如果 $n=m$ 而且 $p=q=\frac{1}{2}$，两个式子相等，答案是 $\frac{1}{2}$，正如所料。
- 取 $n=2$、$m=3$、$p=\frac{1}{2}$、$N=5$。那么 $\mathrm{P}(T) = \binom{4}{2}/2^5 = \frac{3}{16}$，而 $\mathrm{P}(M) = \binom{4}{3}/2^5 = \frac{1}{8}$，所以答案是 $\frac{3/16}{5/16} = \frac{3}{5}$。拖肥糖较少，所以先把拖肥糖拿光的可能性较大。
- 对 $N = 1, \ldots, n+m+1$ 求和，有 $\sum_N \big(\mathrm{P}(T) + \mathrm{P}(M)\big) = 1$，因为终究必有一次伸手会失败。我们已就所有 $n, m \le 6$ 及若干个 p 值，以逐步的精确枚举核对过这一点以及上述公式。以 $n=4$、$m=3$、$N=7$、$p=0.55$ 进行 2×10^5 次试验的模拟（simulation），结果是 0.527，而精确值是 0.528。

常见错误

- 漏掉第 N 次伸手本身所对应的最后一个因数 p（或 q）。
- 答了 $\mathrm{P}(T)$，而不是条件概率 $\mathrm{P}(T)/(\mathrm{P}(T)+\mathrm{P}(M))$。题目已告诉我们，发现是在第 N 次伸手时发生的。
- 以为所指的事件是“拖肥糖在第 N 次伸手时拿光”。拿走最后一颗拖肥糖并不算发现，下一次再去拿才是发现。

STEP 之外

第一次失败发生的时间，其分布是两个负二项（negative binomial）分布的混合（mixture）。负二项分布计算的是取得指定次数的成功所需的试验次数。巴拿赫原本的火柴盒问题（$n=m$，$p=\frac{1}{2}$）问的是：在发现的一刻，另一个盒子里还剩下多少根火柴。其答案可以用斯特灵公式（Stirling's formula）来近似，得出平均值（mean）约为 $2\sqrt{n/\pi}$。关

于负二项分布，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教材）。费勒（Feller）的经典著作 *An Introduction to Probability Theory and Its Applications* 第 1 卷（概率论及其应用导论）中，有一段关于火柴盒问题的著名讨论。

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。