详细解答

STEP III 1993，第 15 题

概率与统计 • 等候时间的分布

STEP III 1993，第 15 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/93-s3-q15

关键思路

把注意力放在最后一次投掷上。实验恰好在第 n 次投掷停止，当且仅当这一次投掷补上了仍然欠缺的那一面的第二次出现，因此首 $n-1$ 次投掷必须恰好有一次是该面，其余全部是另一面。这样，分布（distribution）就是两个简单项之和。期望值（expectation）于是化为两个形如 $\sum n(n-1)x^{n-2}$ 的级数，而这类级数可由等比级数（geometric series）求导两次得到。

解答

我们假设 $0 < p < 1$（否则实验永远不会结束），并假设各次投掷互相独立（independent）。

N 的分布

我们至少需要两次正面和两次反面，所以 $N \geq 4$。固定 $n \geq 4$。在事件 $N = n$ 上，第 n 次投掷不是正面就是反面，而这两种情况互斥（disjoint）。

*第 n 次投掷是正面。*在此之前目标尚未达到，而第 n 次投掷添上一次正面后便达到目标。所以首 $n-1$ 次投掷已经包含至少两次反面，但只有一次正面：恰好一次正面和 $n-2$ 次反面（而且 $n-2 \geq 2$，因为 $n \geq 4$）。反过来说，任何这样的序列后面接上一次正面，都恰好在第 n 次停止。那唯一的一次正面可以出现在 $n-1$ 个位置中的任何一个，所以这种情况的概率是

$$(n-1)\,p\,q^{n-2} \cdot p = (n-1)p^2q^{n-2}.$$

*第 n 次投掷是反面。*把正面和反面对调，用同样的论证可知概率是 $(n-1)q^2p^{n-2}$。

两者相加，

$$\boxed{\mathrm{P}(N=n)=(n-1)(p^2q^{n-2}+q^2p^{n-2}),\qquad n=4,5,6,\ldots}$$

而当 $n\le 3$ 时 $\mathrm{P}(N=n)=0$。

> **想一想** 当最后一次投掷是正面时，为甚么首 $n-1$ 次投掷不可能包含两次正面?

假如包含两次正面，那么它们已经有两次正面和（停止条件所需的）两次反面，实验便会在第 n 次投掷之前结束。正是这一点把正面的次数确定为恰好一次。

期望值

对于 $|x|<1$，把 $\sum_{n\ge 0}x^n=(1-x)^{-1}$ 求导两次，得

$$\sum_{n=2}^{\infty}n(n-1)x^{n-2}=\frac{2}{(1-x)^3}.$$

我们的和式由 $n=4$ 开始，所以要减去 $n=2$ 和 $n=3$ 的两项，即 2 和 $6x$：

$$\sum_{n=4}^{\infty}n(n-1)x^{n-2}=\frac{2}{(1-x)^3}-2-6x.$$

现在

$$\mathrm{E}(N)=\sum_{n=4}^{\infty}n\,\mathrm{P}(N=n)=p^2\sum_{n=4}^{\infty}n(n-1)q^{n-2}+q^2\sum_{n=4}^{\infty}n(n-1)p^{n-2}.$$

先取 $x=q$（于是 $1-x=p$），再取 $x=p$（于是 $1-x=q$）代入上述级数；由于 $0<p,q<1$，两者皆成立：

$$\begin{aligned}\mathrm{E}(N)&=p^2\Big(\frac{2}{p^3}-2-6q\Big)+q^2\Big(\frac{2}{q^3}-2-6p\Big)\\&=\frac{2}{p}+\frac{2}{q}-2(p^2+q^2)-6pq(p+q).\end{aligned}$$

由于 $p+q=1$，我们有 $\frac{1}{p}+\frac{1}{q}=\frac{p+q}{pq}=\frac{1}{pq}$ 以及 $p^2+q^2=(p+q)^2-2pq=1-2pq$。因此

$$\begin{aligned}\mathrm{E}(N)&=\frac{2}{pq}-2(1-2pq)-6pq\\&=\frac{2}{pq}-2-2pq=2\Big(\frac{1}{pq}-1-pq\Big),\end{aligned}$$

这正是所要证明的结果。

验算

- 各概率之和为 1：同一个级数少一个因子的版本给出 $\sum_{n\geq 4}(n-1)p^2q^{n-2} = 1-p^2-2p^2q$，加上对调后的版本，余下 $1-(p^2+q^2)-2pq(p+q) = 1-1+2pq-2pq = 1$，当中用到 $p^2+q^2 = 1-2pq$。
- 对于公平的硬币，$pq = \frac{1}{4}$，而 $\mathrm{E}(N) = 2(4-1-\frac{1}{4}) = \frac{11}{2}$。直接计算：$\mathrm{P}(N=4) = \frac{3}{8}$，$\mathrm{P}(N=5) = \frac{1}{4}$，…，而 $n\,\mathrm{P}(N=n)$ 的部分和趋近 5.5。
- 当 $p \to 0$ 时，$\mathrm{E}(N) \approx 2/p$：我们要等待两次正面出现，每次大约需要 $1/p$ 次投掷。
- 列举所有长度不超过 10 的序列，所得结果与这个分布完全吻合；另外以 $p = 0.3$ 进行模拟，得出的平均值为 7.10，而公式给出的是 7.1038。

常见错误

- 把 $n-1$ 写成 $\binom{n}{2}$：最后一次投掷的位置是固定的，所以只有首 $n-1$ 次投掷中那唯一的一次正面（或反面）可以自由选择位置。
- 使用标准级数时，忘记减去 $n=2$ 和 $n=3$ 的两项。

STEP 之外

N 是两个经平移的负二项（negative binomial）等待时间的混合，而对等比级数求导这个技巧，其实是概率生成函数（probability generating function）的手工版本。设 $G(s) = \mathrm{E}(s^N)$，便可同样快捷地由 $G'(1)$ 得到平均值，并由 $G''(1)$ 得到方差（variance）。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门）详细讲述生成函数，以及像本题这类的等待时间问题。

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。