详细解答

STEP III 1990，第 15 题

概率与统计 • 几何分布

STEP III 1990，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/90-s3-q15

关键思路

每次投掷出现 A 的概率（probability）是 $\frac{1}{4}$，出现 B 的概率是 $\frac{3}{4}$，而且各次投掷互相独立（independent）。因此事件 $\{M = m,\ N = n\}$ 就是某一串特定的投掷结果，其概率是一个乘积。这个乘积可以拆成只与 m 有关的因子乘以只与 n 有关的因子，而这一点几乎包办了所有工作：边际分布（marginal distribution）是几何分布（geometric distribution），两个变量互相独立，其后每一个概率都是一个等比级数（geometric series）。

解答

设单次投掷中 $p = \mathrm{P}(A) = \frac{3}{12} = \frac{1}{4}$，$q = \mathrm{P}(B) = \frac{9}{12} = \frac{3}{4}$。各次投掷互相独立。

事件 $\{M = m,\ N = n\}$ 发生，当且仅当最初的 $m + n + 1$ 次投掷为

$$\underbrace{A\,A\cdots A}_{m}\ \underbrace{B\,B\cdots B}_{n}\ A,$$

所以根据独立性，对于 $m \geq 0$ 及 $n \geq 1$，

$$\mathrm{P}(M = m,\ N = n) = \left(\tfrac{1}{4}\right)^m \left(\tfrac{3}{4}\right)^n \tfrac{1}{4} = \left(\tfrac{1}{4}\right)^{m+1} \left(\tfrac{3}{4}\right)^n.$$

（这串投掷结果以概率 1 终会出现一个 B，然后再出现一个 A，所以这些概率之和为 1；下面的边际分布可以印证这一点。）

第 (i) 部分

求 M 的边际分布时，对 $n \geq 1$ 求和，并利用 $\sum_{n\geq 1}\left(\frac{3}{4}\right)^n = \dfrac{3/4}{1-3/4} = 3$：

$$\mathrm{P}(M=m) = \left(\tfrac{1}{4}\right)^{m+1}\cdot 3 = \tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}, \qquad m = 0,1,2,\ldots$$

求 N 的边际分布时，对 $m \geq 0$ 求和，并利用 $\sum_{m\geq 0}\left(\frac{1}{4}\right)^{m+1} = \dfrac{1/4}{1-1/4} = \frac{1}{3}$：

$$\mathrm{P}(N=n) = \tfrac{1}{3}\left(\tfrac{3}{4}\right)^{n} = \tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1}, \qquad n = 1,2,3,\ldots$$

两个结果都可以直接理解：M 是第一个 B 出现之前 A 的个数，而 N 则是一连串 B 开始之后，等待第一个 A 出现所需的次数。两者各自的总和都是 1。

第 (ii) 部分

我们利用 $\sum_{k\geq 1} kx^{k-1} = \dfrac{1}{(1-x)^2}$，其中 $|x|<1$（把等比级数求导即得）。于是

$$\mathrm{E}(M) = \tfrac{3}{4}\sum_{m\geq 1} m\left(\tfrac{1}{4}\right)^{m} = \tfrac{3}{4}\cdot\tfrac{1}{4}\cdot\frac{1}{(3/4)^2} = \tfrac{1}{3},$$

$$\mathrm{E}(N) = \tfrac{1}{4}\sum_{n\geq 1} n\left(\tfrac{3}{4}\right)^{n-1} = \tfrac{1}{4}\cdot\frac{1}{(1/4)^2} = 4.$$

独立性

把两个边际分布相乘，

$$\mathrm{P}(M=m)\,\mathrm{P}(N=n) = \tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}\cdot\tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1} = \left(\tfrac{1}{4}\right)^{m+1}\left(\tfrac{3}{4}\right)^{n} = \mathrm{P}(M=m,\ N=n)$$

对所有 $m \geq 0$、$n \geq 1$ 都成立。所以 M 与 N 确实互相独立。

> ? **想一想** 一连串 A 的长度，难道不会影响一连串 B 的长度吗？为甚么不会？

一连串 A 在第一个 B 出现时结束，而由该次投掷起，整个过程重新开始：余下的投掷与之前发生的一切互相独立。一连串 B 的长度只取决于这些较后的投掷，所以知道 M 的值，并不能告诉我们任何关于 N 的资料。联合概率（joint probability）可以分解为两个因子，正是这个事实的代数形式。

涉及 N 与 M 的概率

对于整数 $k \geq 1$，把几何分布的尾部求和：

$$\mathrm{P}(N>k) = \sum_{n\geq k+1}\tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1} = \tfrac{1}{4}\cdot\frac{(3/4)^k}{1-3/4} = \left(\tfrac{3}{4}\right)^{k}.$$

这条公式同时给出 $\mathrm{P}(N>0)=1$，由于 $N\geq 1$，这是正确的。

求 $\mathrm{P}(N>M)$ 时，以 M 的值为条件，并利用独立性：

$$\begin{aligned}\mathrm{P}(N>M)&=\sum_{m\geq 0}\mathrm{P}(M=m)\,\mathrm{P}(N>m)=\sum_{m\geq 0}\tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}\left(\tfrac{3}{4}\right)^{m}\\&=\tfrac{3}{4}\sum_{m\geq 0}\left(\tfrac{3}{16}\right)^{m}=\tfrac{3}{4}\cdot\tfrac{16}{13}=\tfrac{12}{13}.\end{aligned}$$

至于 $\mathrm{P}(N=M)$，由于 $N\geq 1$，两者的共同值必须至少为 1：

$$\begin{aligned}\mathrm{P}(N=M)&=\sum_{m\geq 1}\tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}\cdot\tfrac{1}{4}\left(\tfrac{3}{4}\right)^{m-1}=\tfrac{3}{64}\sum_{m\geq 1}\left(\tfrac{3}{16}\right)^{m-1}\\&=\tfrac{3}{64}\cdot\tfrac{16}{13}=\tfrac{3}{52}.\end{aligned}$$

最后，$N>M$、$N=M$、$N<M$ 这三个事件互斥（exclusive）且穷尽（exhaustive）所有可能，所以

$$\mathrm{P}(N<M)=1-\tfrac{12}{13}-\tfrac{3}{52}=\tfrac{52-48-3}{52}=\tfrac{1}{52},$$

正是所求。

验算

- 两个平均值都符合直觉：A 很少出现，所以开头的一连串 A 很短（平均为 $\frac{1}{3}$），而我们平均要等 $1/p=4$ 次投掷，才有一个 A 出现来结束一连串 B。
- 直接计算，$\mathrm{P}(M>n)=\left(\frac{1}{4}\right)^{n+1}$，所以 $\mathrm{P}(N<M)=\sum_{n\geq 1}\frac{1}{4}\left(\frac{3}{4}\right)^{n-1}\left(\frac{1}{4}\right)^{n+1}=\frac{1}{64}\sum_{n\geq 1}\left(\frac{3}{16}\right)^{n-1}=\frac{1}{64}\cdot\frac{16}{13}=\frac{1}{52}$，与用减法所得的结果一致。$\mathrm{P}(N>k)=\left(\frac{3}{4}\right)^{k}$ 的直接解释是：第一个 B 之后，接着的 k 次投掷必须全部是 B。
- 把联合分布精确求和，得 $\mathrm{P}(N>M)=0.92308$，$\mathrm{P}(N=M)=0.05769$，$\mathrm{P}(N<M)=0.01923$；模拟 200 000 组投掷，所得结果为 0.923、0.058、0.019，样本平均值则为 0.334 和 3.99。

常见错误

- 在 $\mathrm{P}(N=M)$ 的求和中包括了 $m=0$：N 永远不可能是 0。
- 忘记 M 可以是 0（第一次投掷可能就是 B），以致 M 的边际分布移了位，算出错误的平均值。

STEP 之外

这里两段连串的长度互相独立，是因为一连串独立试验会在任何只由过去决定的时刻“重新开始”，例如第一个 B 出现的时刻。这就是强马可夫性质（strong Markov property），它是更新理论（renewal theory）和马可夫链（Markov chain）分析的基础，而这些都是大学一、二年级的课题。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教科书）以恰到好处的程度讲解几何分布、独立性和马可夫链。

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。