详细解答

STEP I 1990，第 15 题

概率与统计 • 负二项分布

> **STEP I 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-s1-q15

关键思路

第一部分的做法，是把事件拆成两个独立（independent）的部分：最后一次掷币的结果，以及之前所有掷币的结果。至于其余部分，关键在于硬币没有记忆。每当刚掷出一个正面，等待下一个正面的过程便重新开始，所以我们只需知道一个服从几何分布（geometric distribution）的等待时间有多大机会是偶数或奇数。

解答

以下一律记 $q = 1 - p$。各次掷币互相独立，每次掷出正面的概率为 p。

第 k 个正面的分布

设 T_k 为第 k 个正面出现时的掷币次数。事件 $\{T_k = n\}$ 发生，当且仅当

- 第 n 次掷币是正面，并且
- 在第 $1, 2, \ldots, n-1$ 次掷币之中恰好有 $k-1$ 个正面（因而有 $n-k$ 个反面）。

两个条件缺一不可。假如第 n 次掷币是反面，第 k 个正面便会在第 n 次掷币之前或之后出现，而不是在这一次。假如头 $n-1$ 次掷币中的正面数目不同，第 n 次掷币便不会给出第 k 个正面。反过来，两个条件合起来亦已足够。

这两个事件涉及的掷币次数互不重叠，所以它们是独立的。第一个事件的概率是 p。至于第二个事件，那 $k-1$ 个正面在头 $n-1$ 次掷币中的位置共有 $\binom{n-1}{k-1}$ 种选法。每一种排列都是由 $k-1$ 个正面和 $n-k$ 个反面组成的一个特定序列，根据独立性，其概率为 $p^{k-1}q^{n-k}$，

而不同的排列是互斥（mutually exclusive）的。因此

$$\mathrm{P}(T_k = n) = p \cdot \binom{n-1}{k-1} p^{k-1} q^{n-k} = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \qquad n \geq k,$$

这正是所要证明的结果。

由偶数到偶数

设 $G = T_k - T_{k-1}$ 为第 $(k-1)$ 个正面出现之后还需要再掷的次数（其中 $T_0 = 0$）。第 T_{k-1} 次掷币之后的各次掷币，与此前发生的一切互相独立。所以 G 与 T_{k-1} 独立，并且服从几何分布

$$\mathrm{P}(G = g) = q^{g-1} p, \qquad g = 1, 2, \ldots$$

（即第一部分中 $k = 1$ 的情况）。

> ? **想一想** 如果 T_{k-1} 是偶数，T_k 在甚么情况下是偶数?

由于 $T_k = T_{k-1} + G$，如果 T_{k-1} 是偶数，则 T_k 是偶数当且仅当 G 是偶数。因为 G 与 T_{k-1} 独立，以 T_{k-1} 为偶数作为条件（conditioning）并不会改变 G 的分布。因此

$$\mathrm{P}(T_k \text{ 为偶数} \mid T_{k-1} \text{ 为偶数}) = \mathrm{P}(G \text{ 为偶数}).$$

把等比级数（geometric series）求和，

$$\mathrm{P}(G \text{ 为偶数}) = p\left(q + q^3 + q^5 + \cdots\right) = \frac{pq}{1-q^2} = \frac{pq}{(1-q)(1+q)} = \frac{q}{1+q},$$
$$\mathrm{P}(G \text{ 为奇数}) = p\left(1 + q^2 + q^4 + \cdots\right) = \frac{p}{1-q^2} = \frac{1}{1+q}.$$

所以所求的概率是

$$\frac{q}{1+q} = \frac{1-p}{2-p}.$$

答案与 k 无关。

恰好有两个正面落在偶数次掷币

现在一直掷到第三个正面出现为止。设 $G_1 = T_1$、$G_2 = T_2 - T_1$ 和 $G_3 = T_3 - T_2$ 为三段等待时间。根据同样的无记忆性（memoryless）论证，它们互相独立，而且各自服从上述的几何分布。记

$$e = \mathrm{P}(G_i \text{ 为偶数}) = \frac{q}{1+q}, \qquad o = \mathrm{P}(G_i \text{ 为奇数}) = \frac{1}{1+q}.$$

我们从第 0 次掷币开始，这是偶数。间隔为偶数时，当前位置的奇偶性（parity）保持不变；间隔为奇数时，奇偶性便会改变。下表列出 (G_1, G_2, G_3) 的八种可能奇偶组合，以及相应的 (T_1, T_2, T_3) 的奇偶性：

G_1	G_2	G_3	T_1	T_2	T_3	偶数次的正面数目
偶	偶	偶	偶	偶	偶	3
偶	偶	奇	偶	偶	奇	**2**
偶	奇	偶	偶	奇	奇	1
偶	奇	奇	偶	奇	偶	**2**
奇	偶	偶	奇	奇	奇	0
奇	偶	奇	奇	奇	偶	1
奇	奇	偶	奇	偶	偶	**2**
奇	奇	奇	奇	偶	奇	1

其中有三种情况恰好有两个正面落在偶数次掷币，它们的概率分别是 e^2o、eo^2 和 o^2e。由于各段间隔互相独立，所求的概率是

$$\begin{aligned} e^2o + 2eo^2 = eo(e+2o) &= \frac{q}{(1+q)^2}\cdot\frac{q+2}{1+q} \\ &= \frac{q(2+q)}{(1+q)^3} = \frac{(1-p)(3-p)}{(2-p)^3}. \end{aligned}$$

验算

- 验算第一部分：把 $\mathrm{P}(T_k = n)$ 对 $n \geq k$ 求和，结果应该是 1。这就是负二项级数 $\sum_{n\geq k}\binom{n-1}{k-1}q^{n-k} = (1-q)^{-k}$。
- 两个奇偶概率 $\frac{q}{1+q}$ 与 $\frac{1}{1+q}$ 之和为 1。当 p 接近 1 时，第一个正面几乎必定立即出现，即落在奇数次掷币，所以 $\mathrm{P}(G \text{ 为偶数}) \to 0$，与公式给出的结果一致。
- 对于公平硬币（$p = \frac{1}{2}$），最后的答案是 $\frac{(1/2)(5/2)}{(3/2)^3} = \frac{10}{27} \approx 0.370$。一次进行了 2×10^5 轮的模拟得出 0.369。直接对三段等待时间求和，所得结果在 $p = 0.3$、0.5 和 0.8 时均与公式吻合，误差在 10^{-9} 以内。

常见错误

- 在第一部分使用 $\binom{n}{k}$。最后一次掷币必定是正面，所以只有头 $n-1$ 次掷币可以自由安排。
- 在第二部分把负二项分布对偶数的 n 求和。利用无记忆的间隔，便可完全避开这一步。
- 在最后一部分，计算落在偶数位置的正面时，把各个位置当作互相独立。它们并不独立，因为每个位置都是由前一个位置推算出来的。

STEP 之外

第 k 个正面出现的时间服从负二项（negative binomial）分布。把它写成若干个独立几何间隔之和，正是泊松（Poisson）过程的离散版本：泊松过程的到达时间，同样是由独立的指数等待时间构成的。最后一部分记录奇偶性的做法，其实是一个双状态的马可夫链（Markov chain），这是大学一年级概率课程中很早便会遇到的课题。Grimmett and Welsh 的 *Probability: An Introduction*（概率入门教科书）涵盖了以上所有内容。

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。