详细解答

STEP III 2007，第 13 题

概率与统计 · 随机漫步的期望值

STEP III 2007，第 13 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：

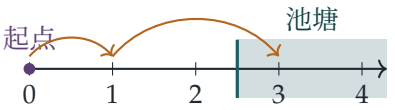
step.maths.org/questions/07-s3-q13

关键思路

那半米的作用，是消除“刚好落在池边”的任何疑问：青蛙由距离池边 $(n-\frac{1}{2})\,\mathrm{m}$ 处出发，牠身处池中，当且仅当牠已跳的总距离至少为 n。因此整个问题的核心，就是由 1 和 2 组成的累计总和首次达到 n 的时刻。当 n 较小时，我们可以直接列出所有可能情况；至于一般公式，u_n 的三个值可给出三条方程，用以求出三个常数。

解答

设 S_j 为首 j 次跳跃所跳的总距离（以米为单位），并设 T 为到达池塘所需的跳跃次数。青蛙由距离池边 $(n-\frac{1}{2})\,\mathrm{m}$ 处出发，跳了 j 次后身处池中，当且仅当 $S_j > n-\frac{1}{2}$，亦即 $S_j \geq n$，因为 S_j 是整数。所以 $p_n(j) = \mathrm{P}(T=j)$，其中 T 是首个满足 $S_j \geq n$ 的 j。



图中所示为 $n=3$ 的情况（池边距离起点 2.5 m），以及“短跳、长跳”这个跳跃序列，青蛙跳两次便到达池塘。

第 (i) 部分

当 $n=2$ 时，池边距离起点 1.5 m。若第一跳是长跳（概率（probability）为 q），青蛙便直接跳进池中，所以 $p_2(1)=q$。否则第一跳是短跳（概率为 p），青蛙距离池边 0.5 m，而

第二跳无论长短都会把牠带进池中。各次跳跃互相独立（independent），所以

$$p_2(2) = p \times 1 = p,$$

这正是所要证明的结果。（此外，当 $j \geq 3$ 时 $p_2(j) = 0$，而 $p + q = 1$ 则确认了这个分布（distribution）是完整的。）

第 (ii) 部分

由距离池边 $\frac{1}{2}$ m 处出发，任何一跳都会落入池中，所以 $u_1 = 1$。

利用第 (i) 部分，

$$u_2 = 1 \cdot p_2(1) + 2 \cdot p_2(2) = q + 2p = 2 - q.$$

当 $n = 3$ 时，我们需要 $S_j \geq 3$。

- 只跳一次永远不够，因为 $S_1 \leq 2$：$p_3(1) = 0$。
- 跳两次仍不够，只会在两次都是短跳时发生（$S_2 = 2$）；其余每一种组合都有 $S_2 \geq 3$。所以 $p_3(2) = 1 - p^2$。
- 两次短跳之后，青蛙距离池边 0.5 m，第三跳必定落入池中：$p_3(3) = p^2$。

因此

$$\begin{aligned} u_3 &= 2(1 - p^2) + 3p^2 = 2 + p^2 \\ &= 2 + (1 - q)^2 = 3 - 2q + q^2, \end{aligned}$$

这正是所要证明的结果。

? **想一想** 有没有更快的方法看出这些值？这个方法能否推广至所有 n？

有：以第一跳的结果作条件（conditioning）来考虑。第一跳之后，青蛙的处境与原来相同，只是 n 减少了 1 或 2，所以当 $n \geq 2$ 时

$$u_n = 1 + p\,u_{n-1} + q\,u_{n-2}, \qquad u_0 = 0,$$

其中 $u_0 = 0$ 对应由 1.5 m 处出发而第一跳是长跳的情况。由此再次得出 $u_2 = 1 + p$ 和 $u_3 = 1 + p(1 + p) + q = 2 + p^2$。这也解释了第 (iii) 部分的形式为何正确：辅助方程（auxiliary equation）$\lambda^2 - p\lambda - q = 0$ 可因式分解为 $(\lambda - 1)(\lambda + q) = 0$，由此得出 $(-q)^{n-1}$ 这一项。而重根 1（源自那个“+1”）则解释了 $B + Cn$ 的出现。

第 (iii) 部分

把 $n = 1, 2, 3$ 代入 $u_n = A(-q)^{n-1} + B + Cn$，并利用第 (ii) 部分：

$$A + B + C = 1, \tag{1}$$

$$-qA + B + 2C = 2 - q, \tag{2}$$

$$q^2A + B + 3C = 3 - 2q + q^2. \tag{3}$$

两式相减以消去 B：

$$(2) - (1): \quad -(1+q)A + C = 1 - q, \tag{4}$$

$$(3) - (2): \quad q(1+q)A + C = 1 - q + q^2. \tag{5}$$

现在 $q \times (4) + (5)$ 可消去 A：

$$(1+q)C = q - q^2 + 1 - q + q^2 = 1, \qquad \text{所以} \qquad C = \frac{1}{1+q},$$

这正是所要证明的结果。然后由 (4) 得

$$(1+q)A = \frac{1}{1+q} - (1-q) = \frac{1-(1-q^2)}{1+q} = \frac{q^2}{1+q}, \qquad \text{所以} \qquad A = \frac{q^2}{(1+q)^2},$$

再由 (1) 得

$$B = 1 - \frac{q^2}{(1+q)^2} - \frac{1}{1+q} = \frac{(1+q)^2 - q^2 - (1+q)}{(1+q)^2} = \frac{q}{(1+q)^2}.$$

所以

$$u_n = \frac{q^2(-q)^{n-1} + q + (1+q)n}{(1+q)^2}.$$

n 很大时。 由于 $0 \le q \le 1$，$A(-q)^{n-1}$ 和 B 这两项都是有界（bounded）的（各自的绝对值最多为 $\frac{1}{4}$），而 Cn 则无止境地增大。因此当 n 很大时

$$u_n \approx Cn = \frac{n}{1+q} = \frac{n}{p+2q},$$

这里用到 $1 = p + q$，故 $1 + q = p + 2q$。

为何这在意料之中。 每一跳的平均长度是 $1 \cdot p + 2 \cdot q = p + 2q$ 米。青蛙须跳过大约 n 米，而根据大数定律（law of large numbers），当 N 很大时，N 次跳跃的总长度接近 $N(p+2q)$。所以所需的跳跃次数接近 $n/(p+2q)$。这个近似相当准确，因为起点那半米的偏移，以及终点处多跳出的距离（青蛙落点最多超出池边 1.5 m）都是有界的。它们只会令次数改变一个有界的量，与 n 相比微不足道。这个有界的修正，正正就是 $A(-q)^{n-1} + B$ 这部分。

验算

- 在公式中取 $n = 0$，得 $\left(q^2 \cdot (-1/q) + q\right)/(1+q)^2 = 0 = u_0$，与应有的结果相符，而且公式满足上面的递推关系（recurrence）。
- $q = 0$（所有跳跃都是短跳）：$u_n = n$，这是正确的，因为青蛙恰好需要 n 次单位长度的跳跃。
- 以分数精确枚举所有跳跃序列，结果在 $n = 1, \ldots, 12$ 及若干个 p 值下均与公式吻合。以 $q = 0.7$、$n = 200$ 进行模拟，得出平均跳跃次数为 117.90，而公式给出 117.89，$n/(p+2q) \approx 117.65$。

常见错误

- 把“落入池中”当作恰好 $S_j = n$。由于有那半米，青蛙可以跳过头，凡是满足 $S_j \geq n$ 的序列都要计算在内。
- 写成 $p_3(2) = 2pq + q^2$ 时，忘记了长跳在先的情况。事实上 $pq + qp + q^2 = 1 - p^2$ 已涵盖两种次序。

STEP 之外

u_n 的递推关系是一条更新方程（renewal equation）。结果 $u_n \sim n/\mu$（其中 μ 为每步的平均长度）就是基本更新定理（elementary renewal theorem）。由沃尔德恒等式（Wald's identity）$\mathrm{E}(S_T) = \mu\,\mathrm{E}(T)$ 可得出一个精确的版本：这里 S_T 不是 n 就是 $n+1$，所以 $(1+q)u_n = n + \mathrm{P}(\text{跳过头})$。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）以同样的思路讲述随机漫步（random walk）和母函数（generating function）。

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。