详细解答

STEP II 2018，第 12 题

概率与统计 • 最优停止的期望值

STEP II 2018，第 12 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/18-s2-q12

关键思路

目标定为 h 次正面（Heads）时，我只有在游戏结束前达到目标，才会赢得 £h，否则分文不得。所以期望收益（expected winnings）就是 h 乘以达到 h 次正面的概率（probability），整道题的重点在于求出这个概率。要在整数 h 上求最大值，不要用微分，而应考虑比值（ratio）E_{h+1}/E_h，比较相邻的项。

解答

全题设 $q = 1 - p = \frac{1}{N+1}$ 为反面（Tails）的概率，并设 E_h 为目标定为 h（其中 $h \geq 1$）时我的期望收益。由于我达到 h 次正面便赢得 £h，否则赢得 £0，

$$E_h = h \times \mathrm{P}(\text{我在游戏结束前达到}h\text{ 次正面}).$$

第 (i) 部分

在版本 1 中，只要出现一次反面，游戏便结束，所以我达到 h 次正面，当且仅当首 h 次掷币全是正面。各次掷币互相独立（independent），因此这事件的概率为 p^h，而

$$E_h = hp^h.$$

要找出最佳的 h，可比较相邻的值。由于 $E_h > 0$，

$$\frac{E_{h+1}}{E_h} = \frac{(h+1)p^{h+1}}{hp^h} = \frac{(h+1)N}{h(N+1)},$$

而

$$\frac{E_{h+1}}{E_h} \geq 1 \iff (h+1)N \geq h(N+1) \iff N \geq h,$$

且等号只在 $h = N$ 时成立。所以

$$E_1 < E_2 < \cdots < E_N = E_{N+1} > E_{N+2} > \cdots .$$

这个数列（sequence）在 $h = N$ 之前严格递增，下一个值与它相等，之后便严格递减。所以 E_h 的最大值是 E_N，选取 $h = N$ 可使我的期望收益最大，正如所求。（选取 $h = N+1$ 效果同样好。）最大值为 $N(\frac{N}{N+1})^N$。

第 (ii) 部分

现在游戏只会在连续两次反面时结束。把直至第 h 次正面（包括该次）为止的掷币结果看成 h 个区块，每个区块都以一次正面结尾。如果我安全地达到 h 次正面，那么每个区块必定是

$$\text{H} \quad \text{或} \quad \text{TH},$$

因为连续两次反面会令游戏结束。（一开始就掷出反面是没有问题的：它前面甚么也没有。）反过来，任何由 h 个这类区块组成的序列，都不会含有连续两次反面。序列中每次反面，要么紧接在一次正面之后，要么位于开头，而且必定紧接在一次正面之前。所以我达到 h 次正面，当且仅当掷币结果可分拆为 h 个属于这两类的区块。

> ? **想一想** 为甚么可以直接把各区块的概率相乘?

每次掷出正面后，游戏的处境与开始时完全相同：没有待定的反面，而且将来的掷币与过去的结果互相独立。所以每个区块，与其他区块独立地，安全的概率为

$$\text{P(H)} + \text{P(TH)} = p + qp = p(1+q) = \frac{N}{N+1} \cdot \frac{N+2}{N+1} = \frac{N(N+2)}{(N+1)^2}.$$

（这两个事件互斥（disjoint），因为它们在第一次掷币已经不同。）因此达到 h 次正面的概率为 $(\frac{N(N+2)}{(N+1)^2})^h$，而

$$E_h = \frac{hN^h(N+2)^h}{(N+1)^{2h}},$$

正如所求。

$N = 2$ 的情况。此时

$$E_h = h\left(\frac{2 \cdot 4}{9}\right)^h = h\left(\frac{8}{9}\right)^h.$$

与第 (i) 部分一样，

$$\frac{E_{h+1}}{E_h} = \frac{8(h+1)}{9h} \geq 1 \iff 8h + 8 \geq 9h \iff h \leq 8,$$

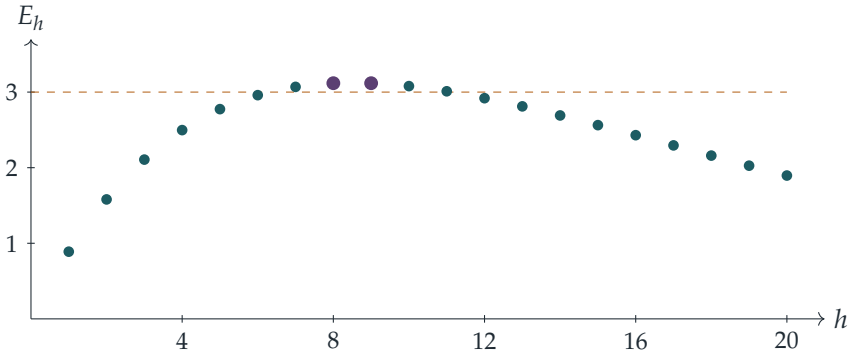
且等号只在 $h = 8$ 时成立。所以 E_h 在 $h = 8$ 之前严格递增，然后 $E_9 = E_8$，之后严格递减。最大值为

$$E_8 = 8\left(\frac{8}{9}\right)^8 = \frac{2^3 \cdot 2^{24}}{3^{16}} = \frac{2^{27}}{3^{16}}.$$

要估算这个值，可取以 3 为底的对数（logarithm），并利用 $\log_3 2 \approx 0.63$：

$$\log_3 E_8 = 27\log_3 2 - 16 \approx 27(0.63) - 16 = 17.01 - 16 = 1.01.$$

所以 $E_8 \approx 3^{1.01}$，非常接近 $3^1 = 3$。最大期望收益约为 £3。



版本 2 中 $N = 2$ 时的值 $E_h = h(8/9)^h$；在 $h = 8, 9$ 处并列的最大值刚好高于 3 的虚线。

验算

- 版本 2 只会比版本 1 对玩家更有利，因为每个全是正面的开局在两个版本中都是安全的。事实上 $\frac{N(N+2)}{(N+1)^2} = 1 - \frac{1}{(N+1)^2} > \frac{N}{N+1}$。
- 当 $N = 1$ 时，版本 1 给出 $E_1 = E_2 = \frac{1}{2}$：以概率 $\frac{1}{2}$ 得一次正面，或以概率 $\frac{1}{4}$ 得两次正面。
- 准确值为 $E_8 = 2^{27}/3^{16} = 134\,217\,728/43\,046\,721 \approx 3.118$，所以估计值 £3 相当准确。以电脑模拟（simulation）2×10^5 局 $N = 2$、$h = 8$ 的游戏，所得平均收益约为 £3.11。

常见错误

- 在版本 2 中，忘记开头可以有单独一次反面，或误以为第 h 次正面之后还可以出现反面。其实第 h 次正面一出现，我便停止。
- 把 hp^h 微分并令结果等于零。这样得出 $h = -1/\ln p$，它通常不是整数，而要由此找出最佳整数需要小心处理。比值方法更简洁，而且准确。
- 忽略了并列的情况：$h = N$ 与 $h = N + 1$ 给出相同的期望收益，所以题目的“证明”只要求证明 $h = N$ 能达到最大值。

STEP 之外

这是一个简单的最优停止（*optimal stopping*）问题。这里我预先定下目标，但你可以验证这样做并无损失：每当我手持 H 次正面而没有待定的反面时，多掷一次的增益为正，当且仅当 H 足够小。这类通过比较“现在停止”与“再走一步”来决定何时停止的问题，会引出鞅（martingale）和动态规划（dynamic programming）。第 (ii) 部分的区块论证，让你初尝马尔可夫链（Markov chain）中的更新（renewal）思想。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）是很好的下一步。

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。