詳細解答

STEP II 2018，第 12 題

概率與統計 • 最優停止的期望值

STEP II 2018，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/18-s2-q12

關鍵思路

目標定為 h 次正面（Heads）時，我只有在遊戲結束前達到目標，才會贏得 £h，否則分文不得。所以期望收益(expected winnings)就是 h 乘以達到 h 次正面的概率(probability)，整道題的重點在於求出這個概率。要在整數 h 上求最大值，不要用微分，而應考慮比值（ratio）E_{h+1}/E_h，比較相鄰的項。

解答

全題設 $q = 1 - p = \frac{1}{N+1}$ 為反面（Tails）的概率，並設 E_h 為目標定為 h（其中 $h \geq 1$）時我的期望收益。由於我達到 h 次正面便贏得 £h，否則贏得 £0，

$$E_h = h \times \mathrm{P}(\text{我在遊戲結束前達到}h\text{ 次正面}).$$

第 (i) 部分

在版本 1 中，只要出現一次反面，遊戲便結束，所以我達到 h 次正面，當且僅當首 h 次擲幣全是正面。各次擲幣互相獨立（independent），因此這事件的概率為 p^h，而

$$E_h = hp^h.$$

要找出最佳的 h，可比較相鄰的值。由於 $E_h > 0$，

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

而

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

且等號只在 $h = N$ 時成立。所以

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

這個數列（sequence）在 $h = N$ 之前嚴格遞增，下一個值與它相等，之後便嚴格遞減。所以 E_h 的最大值是 E_N，選取 $h = N$ 可使我的期望收益最大，正如所求。（選取 $h = N+1$ 效果同樣好。）最大值為 $N(\frac{N}{N+1})^N$。

第 (ii) 部分

現在遊戲只會在連續兩次反面時結束。把直至第 h 次正面（包括該次）為止的擲幣結果看成 h 個區塊，每個區塊都以一次正面結尾。如果我安全地達到 h 次正面，那麼每個區塊必定是

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

因為連續兩次反面會令遊戲結束。（一開始就擲出反面是沒有問題的：它前面甚麼也沒有。）反過來，任何由 h 個這類區塊組成的序列，都不會含有連續兩次反面。序列中每次反面，要麼緊接在一次正面之後，要麼位於開頭，而且必定緊接在一次正面之前。所以我達到 h 次正面，當且僅當擲幣結果可分拆為 h 個屬於這兩類的區塊。

想一想 為甚麼可以直接把各區塊的概率相乘？

每次擲出正面後，遊戲的處境與開始時完全相同：沒有待定的反面，而且將來的擲幣與過去的結果互相獨立。所以每個區塊，與其他區塊獨立地，安全的概率為

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

（這兩個事件互斥（disjoint），因為它們在第一次擲幣已經不同。）因此達到 h 次正面的概率為 $(\frac{N(N+2)}{(N+1)^2})^h$，而

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

正如所求。

$N = 2$ 的情況。此時

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

與第 (i) 部分一樣，

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

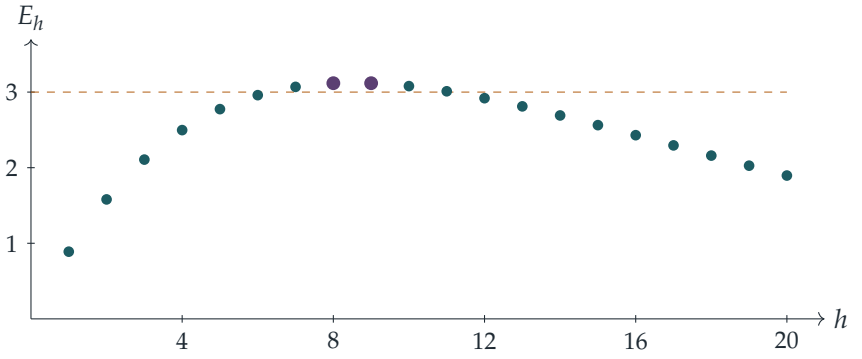
且等號只在 $h = 8$ 時成立。所以 E_h 在 $h = 8$ 之前嚴格遞增，然後 $E_9 = E_8$，之後嚴格遞減。最大值為

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

要估算這個值，可取以 3 為底的對數（logarithm），並利用 $\log_3 2 \approx 0.63$：

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

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



版本 2 中 $N = 2$ 時的值 $E_h = h(8/9)^h$；在 $h = 8, 9$ 處並列的最大值剛好高於 3 的虛線。

驗算

- 版本 2 只會比版本 1 對玩家更有利，因為每個全是正面的開局在兩個版本中都是安全的。事實上 $\frac{N(N+2)}{(N+1)^2} = 1 - \frac{1}{(N+1)^2} > \frac{N}{N+1}$。
- 當 $N = 1$ 時，版本 1 給出 $E_1 = E_2 = \frac{1}{2}$：以概率 $\frac{1}{2}$ 得一次正面，或以概率 $\frac{1}{4}$ 得兩次正面。
- 準確值為 $E_8 = 2^{27}/3^{16} = 134\,217\,728/43\,046\,721 \approx 3.118$，所以估計值 £3 相當準確。以電腦模擬（simulation）2×10^5 局 $N = 2$、$h = 8$ 的遊戲，所得平均收益約為 £3.11。

常見錯誤

- 在版本 2 中，忘記開頭可以有單獨一次反面，或誤以為第 h 次正面之後還可以出現反面。其實第 h 次正面一出現，我便停止。
- 把 hp^h 微分並令結果等於零。這樣得出 $h = -1/\ln p$，它通常不是整數，而要由此找出最佳整數需要小心處理。比值方法更簡潔，而且準確。
- 忽略了並列的情況：$h = N$ 與 $h = N + 1$ 給出相同的期望收益，所以題目的「證明」只要求證明 $h = N$ 能達到最大值。

STEP 之外

這是一個簡單的最優停止（*optimal stopping*）問題。這裏我預先定下目標，但你可以驗證這樣做並無損失：每當我手持 H 次正面而沒有待定的反面時，多擲一次的增益為正，當且僅當 H 足夠小。這類透過比較「現在停止」與「再走一步」來決定何時停止的問題，會引出鞅（martingale）和動態規劃（dynamic programming）。第 (ii) 部分的區塊論證，讓你初嘗馬爾可夫鏈（Markov chain）中的更新（renewal）思想。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）是很好的下一步。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。