詳細解答

STEP I 1990，第 15 題

概率與統計 • 負二項分佈

> **STEP I 1990，第 15 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/90-s1-q15

關鍵思路

第一部分的做法，是把事件拆成兩個獨立（independent）的部分：最後一次擲幣的結果，以及之前所有擲幣的結果。至於其餘部分，關鍵在於硬幣沒有記憶。每當剛擲出一個正面，等待下一個正面的過程便重新開始，所以我們只需知道一個服從幾何分佈（geometric distribution）的等待時間有多大機會是偶數或奇數。

解答

以下一律記 $q = 1 - p$。各次擲幣互相獨立，每次擲出正面的概率為 p。

第 k 個正面的分佈

設 T_k 為第 k 個正面出現時的擲幣次數。事件 $\{T_k = n\}$ 發生，當且僅當

- 第 n 次擲幣是正面，並且
- 在第 $1, 2, \ldots, n-1$ 次擲幣之中恰好有 $k-1$ 個正面（因而有 $n-k$ 個反面）。

兩個條件缺一不可。假如第 n 次擲幣是反面，第 k 個正面便會在第 n 次擲幣之前或之後出現，而不是在這一次。假如頭 $n-1$ 次擲幣中的正面數目不同，第 n 次擲幣便不會給出第 k 個正面。反過來，兩個條件合起來亦已足夠。

這兩個事件涉及的擲幣次數互不重疊，所以它們是獨立的。第一個事件的概率是 p。至於第二個事件，那 $k-1$ 個正面在頭 $n-1$ 次擲幣中的位置共有 $\binom{n-1}{k-1}$ 種選法。每一種排列都是由 $k-1$ 個正面和 $n-k$ 個反面組成的一個特定序列，根據獨立性，其概率為

$p^{k-1}q^{n-k}$，而不同的排列是互斥（mutually exclusive）的。因此

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

這正是所要證明的結果。

由偶數到偶數

設 $G = T_k - T_{k-1}$ 為第 $(k-1)$ 個正面出現之後還需要再擲的次數（其中 $T_0 = 0$）。第 T_{k-1} 次擲幣之後的各次擲幣，與此前發生的一切互相獨立。所以 G 與 T_{k-1} 獨立，並且服從幾何分佈

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

（即第一部分中 $k = 1$ 的情況）。

? **想一想** 如果 T_{k-1} 是偶數，T_k 在甚麼情況下是偶數？

由於 $T_k = T_{k-1} + G$，如果 T_{k-1} 是偶數，則 T_k 是偶數當且僅當 G 是偶數。因為 G 與 T_{k-1} 獨立，以 T_{k-1} 為偶數作為條件（conditioning）並不會改變 G 的分佈。因此

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

把等比級數（geometric series）求和，

$$\mathrm{P}(G \text{ 為偶數}) = p\left(q + q^3 + q^5 + \cdots\right) = \frac{pq}{1-q^2} = \frac{pq}{(1-q)(1+q)} = \frac{q}{1+q},$$
$$\mathrm{P}(G \text{ 為奇數}) = p\left(1 + q^2 + q^4 + \cdots\right) = \frac{p}{1-q^2} = \frac{1}{1+q}.$$

所以所求的概率是

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

答案與 k 無關。

恰好有兩個正面落在偶數次擲幣

現在一直擲到第三個正面出現為止。設 $G_1 = T_1$、$G_2 = T_2 - T_1$ 和 $G_3 = T_3 - T_2$ 為三段等待時間。根據同樣的無記憶性（memoryless）論證，它們互相獨立，而且各自服從上述的幾何分佈。記

$$e = \mathrm{P}(G_i \text{ 為偶數}) = \frac{q}{1+q}, \qquad o = \mathrm{P}(G_i \text{ 為奇數}) = \frac{1}{1+q}.$$

我們從第 0 次擲幣開始，這是偶數。間隔為偶數時，當前位置的奇偶性（parity）保持不變；間隔為奇數時，奇偶性便會改變。下表列出 (G_1, G_2, G_3) 的八種可能奇偶組合，以及相應的 (T_1, T_2, T_3) 的奇偶性：

G_1	G_2	G_3	T_1	T_2	T_3	偶數次的正面數目
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

其中有三種情況恰好有兩個正面落在偶數次擲幣，它們的概率分別是 e^2o、eo^2 和 o^2e。由於各段間隔互相獨立，所求的概率是

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

驗算

- 驗算第一部分：把 $\mathrm{P}(T_k = n)$ 對 $n \geq k$ 求和，結果應該是 1。這就是負二項級數 $\sum_{n\geq k}\binom{n-1}{k-1}q^{n-k} = (1-q)^{-k}$。
- 兩個奇偶概率 $\frac{q}{1+q}$ 與 $\frac{1}{1+q}$ 之和為 1。當 p 接近 1 時，第一個正面幾乎必定立即出現，即落在奇數次擲幣，所以 $\mathrm{P}(G \text{ 為偶數}) \to 0$，與公式給出的結果一致。
- 對於公平硬幣（$p = \frac{1}{2}$），最後的答案是 $\frac{(1/2)(5/2)}{(3/2)^3} = \frac{10}{27} \approx 0.370$。一次進行了 2×10^5 輪的模擬得出 0.369。直接對三段等待時間求和，所得結果在 $p = 0.3$、0.5 和 0.8 時均與公式吻合，誤差在 10^{-9} 以內。

常見錯誤

- 在第一部分使用 $\binom{n}{k}$。最後一次擲幣必定是正面，所以只有頭 $n-1$ 次擲幣可以自由安排。
- 在第二部分把負二項分佈對偶數的 n 求和。利用無記憶的間隔，便可完全避開這一步。
- 在最後一部分，計算落在偶數位置的正面時，把各個位置當作互相獨立。它們並不獨立，因為每個位置都是由前一個位置推算出來的。

STEP 之外

第 k 個正面出現的時間服從負二項（negative binomial）分佈。把它寫成若干個獨立幾何間隔之和，正是泊松（Poisson）過程的離散版本：泊松過程的到達時間，同樣是由獨立的指數等待時間構成的。最後一部分記錄奇偶性的做法，其實是一個雙狀態的馬可夫鏈（Markov chain），這是大學一年級概率課程中很早便會遇到的課題。Grimmett and Welsh 的 *Probability: An Introduction*（概率入門教科書）涵蓋了以上所有內容。

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。