詳細解答

STEP I 1999，第 13 題

概率與統計 • 指示變量

> **STEP I 1999，第 13 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/99-s1-q13

關鍵思路

不要直接數單元的數目，改為數間隙：單元的數目等於相鄰磁鐵彼此分開的位置數目再加一。某處會否出現間隙，只取決於相鄰兩塊磁鐵的指向是否相同，而這 $N-1$ 個事件原來是互相獨立（independent）的公平擲幣。因此間隙的數目服從二項分佈（binomial distribution），期望值（expectation）和方差（variance）便可立即得出。

解答

建立模型

我們把「隨機擺放」理解為：每塊磁鐵指向任何一方的可能性均等，且各磁鐵互相獨立。若一塊磁鐵的北極在右邊（SN），稱它為 A 型；若北極在左邊（NS），稱它為 B 型。設 $X_1, \ldots, X_N$ 為由左至右各磁鐵的類型。

考慮相鄰的兩塊磁鐵。互相面對的兩端，是第一塊磁鐵的右端和第二塊磁鐵的左端。

- 類型相同（先 A 後 A）：相對的兩端是 N 和 S，屬異極，所以它們會**相連**。（先 B 後 B 的情況相同，只是換成 S 和 N。）
- 類型不同（先 A 後 B）：相對的兩端是 N 和 N，所以它們會**分開**。（先 B 後 A 則是 S 和 S。）



因此，第 i 塊與第 $i+1$ 塊磁鐵之間出現間隙，當且僅當 $X_i \neq X_{i+1}$。對 $i = 1, \ldots, N-1$，定義

$$G_i = \begin{cases} 1 & \text{若} X_i \neq X_{i+1}\text{（有間隙）}, \\ 0 & \text{若} X_i = X_{i+1}\text{（相連）}. \end{cases}$$

每一個間隙都把整列磁鐵多分出一段，所以獨立單元的數目是

$$U = 1 + G_1 + G_2 + \cdots + G_{N-1}.$$

間隙的分佈

每個 G_i 取值 1 的概率都是 $\frac{1}{2}$：不論 X_i 是甚麼，X_{i+1} 與它不同的概率都是 $\frac{1}{2}$。

> ? **想一想** 「在 i 處有間隙」和「在 $i+1$ 處有間隙」這兩個事件都牽涉第 $i+1$ 塊磁鐵。各個 G_i 真的互相獨立嗎？

是的，儘管相鄰的 G 共用同一塊磁鐵。先固定第一塊磁鐵的類型 X_1。這樣，數列 $G_1, \ldots, G_{N-1}$ 便逐一決定了 $X_2, X_3, \ldots, X_N$（若 $G_i = 0$，X_{i+1} 等於 X_i；若 $G_i = 1$，則為另一類型）；反過來，各磁鐵的類型也決定了所有的 G。所以，不論 X_1 取哪一個值，$(X_2, \ldots, X_N)$ 的 2^{N-1} 種等可能的選擇，與 $(G_1, \ldots, G_{N-1})$ 的 2^{N-1} 個可能數列一一對應（one-to-one）。因此每一個間隙數列出現的概率都是 $1/2^{N-1}$，這正好表示各個 G_i 互相獨立，並且各以概率 $\frac{1}{2}$ 取值 0 或 1。

由此可得 $G_1 + \cdots + G_{N-1} \sim \mathrm{B}\left(N-1, \frac{1}{2}\right)$。

期望值與方差

服從 $\mathrm{B}(n, p)$ 的變量，其平均值為 np，方差為 $np(1-p)$。取 $n = N-1$、$p = \frac{1}{2}$，並留意加上常數 1 只會令平均值平移，不會改變方差，於是

$$\mathrm{E}(U) = 1 + \tfrac{1}{2}(N-1) = \frac{N+1}{2},$$
$$\mathrm{Var}(U) = (N-1) \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = \frac{N-1}{4}.$$

如果你不想直接引用二項分佈的結果：$\mathrm{E}(G_i) = \frac{1}{2}$，而且由於 $G_i^2 = G_i$，可得 $\mathrm{Var}(G_i) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$。由期望值的線性性質（linearity）可得平均值，而獨立性則讓我們可以把 $N-1$ 個方差相加。

驗算

- $N = 1$：一塊磁鐵必定是一個單元，而公式給出平均值 1 和方差 0。

- $N = 2$：兩塊磁鐵組成 1 個或 2 個單元，概率各為 $\frac{1}{2}$，所以平均值是 $\frac{3}{2}$，方差是 $\frac{5}{2} - \frac{9}{4} = \frac{1}{4}$，與公式相符。
- 對所有 2^N 種排列逐一窮舉，在 $N \leq 12$ 的情況下所得結果與 $\frac{N+1}{2}$ 和 $\frac{N-1}{4}$ 完全一致；另外以 $N = 20$ 進行模擬，得到的平均值約為 10.5，方差約為 4.73（預測值為 10.5 和 4.75）。

常見錯誤

- 以為兩塊指向相同的磁鐵會互相排斥。其實它們相對的是異極（N 對 S），所以會互相吸引並相連。
- 沒有說明獨立性的理由便把方差相加。這裏的獨立性確實成立，但並不明顯，因為相鄰的間隙共用同一塊磁鐵。
- 把間隙數目當作 N 個而非 $N - 1$ 個，因而寫出 $\mathrm{Var}(U) = \frac{N}{4}$；又或者在平均值中漏掉額外的 1。

STEP 之外

藉着數單元之間的邊界來數單元，這種方法就是指示變量（indicator variables）法，是概率論中最有用的工具之一：把一個計數寫成若干個 0–1 變量之和，然後運用期望值的線性性質，而這完全不需要獨立性。這裏的 U 所數的，是由兩種符號組成的隨機序列中的游程（runs）數目。游程數目是統計學中 Wald–Wolfowitz 游程檢驗（runs test）的基礎，這個檢驗用來判斷一個序列看來是否隨機。關於指示變量及其方差的更多內容，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門）。

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。