详细解答

STEP I 1999，第 13 题

概率与统计 · 指示变量

> **STEP I 1999，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/99-s1-q13

关键思路

不要直接数单元的数目，改为数间隙：单元的数目等于相邻磁铁彼此分开的位置数目再加一。某处会否出现间隙，只取决于相邻两块磁铁的指向是否相同，而这 $N-1$ 个事件原来是互相独立（independent）的公平掷币。因此间隙的数目服从二项分布（binomial distribution），期望值（expectation）和方差（variance）便可立即得出。

解答

建立模型

我们把“随机摆放”理解为：每块磁铁指向任何一方的可能性均等，且各磁铁互相独立。若一块磁铁的北极在右边（SN），称它为 A 型；若北极在左边（NS），称它为 B 型。设 $X_1, \ldots, X_N$ 为由左至右各磁铁的类型。

考虑相邻的两块磁铁。互相面对的两端，是第一块磁铁的右端和第二块磁铁的左端。

- 类型相同（先 A 后 A）：相对的两端是 N 和 S，属异极，所以它们会**相连**。（先 B 后 B 的情况相同，只是换成 S 和 N。）
- 类型不同（先 A 后 B）：相对的两端是 N 和 N，所以它们会**分开**。（先 B 后 A 则是 S 和 S。）



因此，第 i 块与第 $i+1$ 块磁铁之间出现间隙，当且仅当 $X_i \neq X_{i+1}$。对 $i = 1, \ldots, N-1$，定义

$$G_i = \begin{cases} 1 & \text{若} X_i \neq X_{i+1}\text{（有间隙）}, \\ 0 & \text{若} X_i = X_{i+1}\text{（相连）}. \end{cases}$$

每一个间隙都把整列磁铁多分出一段，所以独立单元的数目是

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

间隙的分布

每个 G_i 取值 1 的概率都是 $\frac{1}{2}$：不论 X_i 是甚么，X_{i+1} 与它不同的概率都是 $\frac{1}{2}$。

> **想一想** “在 i 处有间隙”和“在 $i+1$ 处有间隙”这两个事件都牵涉第 $i+1$ 块磁铁。各个 G_i 真的互相独立吗?

是的，尽管相邻的 G 共用同一块磁铁。先固定第一块磁铁的类型 X_1。这样，数列 $G_1, \ldots, G_{N-1}$ 便逐一决定了 $X_2, X_3, \ldots, X_N$（若 $G_i = 0$，X_{i+1} 等于 X_i；若 $G_i = 1$，则为另一类型）；反过来，各磁铁的类型也决定了所有的 G。所以，不论 X_1 取哪一个值，$(X_2, \ldots, X_N)$ 的 2^{N-1} 种等可能的选择，与 $(G_1, \ldots, G_{N-1})$ 的 2^{N-1} 个可能数列一一对应（one-to-one）。因此每一个间隙数列出现的概率都是 $1/2^{N-1}$，这正好表示各个 G_i 互相独立，并且各以概率 $\frac{1}{2}$ 取值 0 或 1。

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

期望值与方差

服从 $\mathrm{B}(n, p)$ 的变量，其平均值为 np，方差为 $np(1-p)$。取 $n = N-1$、$p = \frac{1}{2}$，并留意加上常数 1 只会令平均值平移，不会改变方差，于是

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

如果你不想直接引用二项分布的结果：$\mathrm{E}(G_i) = \frac{1}{2}$，而且由于 $G_i^2 = G_i$，可得 $\mathrm{Var}(G_i) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$。由期望值的线性性质（linearity）可得平均值，而独立性则让我们可以把 $N-1$ 个方差相加。

验算

► $N = 1$：一块磁铁必定是一个单元，而公式给出平均值 1 和方差 0。

- $N = 2$: 两块磁铁组成 1 个或 2 个单元, 概率各为 $\frac{1}{2}$, 所以平均值是 $\frac{3}{2}$, 方差是 $\frac{5}{2} - \frac{9}{4} = \frac{1}{4}$, 与公式相符。
- 对所有 2^N 种排列逐一穷举，在 $N \leq 12$ 的情况下所得结果与 $\frac{N+1}{2}$ 和 $\frac{N-1}{4}$ 完全一致；另外以 $N = 20$ 进行模拟，得到的平均值约为 10.5，方差约为 4.73（预测值为 10.5 和 4.75）。

常见错误

- 以为两块指向相同的磁铁会互相排斥。其实它们相对的是异极（N 对 S），所以会互相吸引并相连。
- 没有说明独立性的理由便把方差相加。这里的独立性确实成立，但并不明显，因为相邻的间隙共用同一块磁铁。
- 把间隙数目当作 N 个而非 $N - 1$ 个，因而写出 $\mathrm{Var}(U) = \frac{N}{4}$；又或者在平均值中漏掉额外的 1。

STEP 之外

借着数单元之间的边界来数单元，这种方法就是指示变量（indicator variables）法，是概率论中最有用的工具之一：把一个计数写成若干个 0–1 变量之和，然后运用期望值的线性性质，而这完全不需要独立性。这里的 U 所数的，是由两种符号组成的随机序列中的游程（runs）数目。游程数目是统计学中 Wald–Wolfowitz 游程检验（runs test）的基础，这个检验用来判断一个序列看来是否随机。关于指示变量及其方差的更多内容，可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门）。

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。