解答例

STEP I 1999、第 13 問

確率・統計 ・ 指示変数

> **STEP I 1999、問 13**
>
> 先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：
>
> step.maths.org/questions/99-s1-q13

鍵となる考え方

まとまり（unit）の個数を直接数えるのではない。代わりに隙間（gap）を数えるのである。まとまりの個数は、隣り合う磁石が離れている箇所の個数より 1 だけ多い。隙間ができるかどうかは、隣り合う 2 つの磁石が同じ向きかどうかだけで決まり、これら $N-1$ 個の事象は、実は公平な硬貨投げを独立に繰り返したものになっている。したがって隙間の個数は二項分布（binomial distribution）に従い、期待値（expectation）も分散（variance）もただちに求まる。

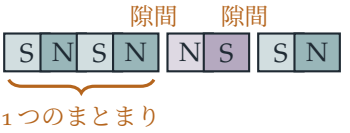
解答

モデルを立てる

「無作為に置く」とは、各磁石がどちら向きになるのも等確率であり、他の磁石とは独立（independent）であることと解釈する。N 極が右側にある磁石（SN）を A 型、N 極が左側にある磁石（NS）を B 型と呼ぶ。左から順に読んだ型を $X_1, \ldots, X_N$ とする。

隣り合う 2 つの磁石を見る。向かい合う端は、1 つ目の磁石の右端と 2 つ目の磁石の左端である。

- 同じ型（A の次に A）のとき：向かい合う端は N と S で、異なる極なので**くっつく**。（B の次に B の場合も、S と N で同様である。）
- 異なる型（A の次に B）のとき：向かい合う端は N と N なので**離れる**。（B の次に A の場合は S と S になる。）



よって、磁石 i と磁石 $i+1$ の間に隙間ができるのは、ちょうど $X_i \neq X_{i+1}$ のときである。$i = 1, \ldots, N-1$ に対して

$$G_i = \begin{cases} 1 & X_i \neq X_{i+1} \text{ のとき（隙間）}, \\ 0 & X_i = X_{i+1} \text{ のとき（くっつく）}. \end{cases}$$

と定める。隙間が 1 つあるごとに列は 1 つ多くの部分に分かれるので、別々のまとまりの個数は

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

である。

隙間の分布

各 G_i は確率 $\frac{1}{2}$ で 1 となる。X_i が何であっても、X_{i+1} がそれと異なる確率は $\frac{1}{2}$ だからである。

> **考えてみよう** 事象「i に隙間がある」と「$i+1$ に隙間がある」はどちらも磁石 $i+1$ に関わっている。G_i は本当に独立なのだろうか？

隣り合う G が磁石を共有しているにもかかわらず、実際に独立である。1 つ目の磁石の型 X_1 を固定する。すると、列 $G_1, \ldots, G_{N-1}$ によって $X_2, X_3, \ldots, X_N$ が 1 つずつ決まる（$G_i = 0$ ならば X_{i+1} は X_i と等しく、$G_i = 1$ ならばもう一方の型である）。逆に、型の列から G が決まる。したがって、X_1 がどちらの値であっても、等確率な 2^{N-1} 通りの $(X_2, \ldots, X_N)$ の選び方は、2^{N-1} 通りの可能な列 $(G_1, \ldots, G_{N-1})$ と 1 対 1 に対応する。ゆえに隙間のどの列も確率 $1/2^{N-1}$ をもつ。これはまさに、G_i が独立で、それぞれ確率 $\frac{1}{2}$ で 0 または 1 をとることを意味している。

よって $G_1 + \cdots + G_{N-1} \sim \mathrm{B}\left(N-1, \frac{1}{2}\right)$ である。

期待値と分散

$\mathrm{B}(n, p)$ に従う確率変数の平均は np、分散は $np(1-p)$ である。$n = N-1$、$p = \frac{1}{2}$ とし、定数 1 を加えると平均はずれるが分散は変わらないことに注意すれば、

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

を得る。

二項分布の結果を引用したくなければ、次のようにすればよい。$\mathrm{E}(G_i) = \frac{1}{2}$ であり、$G_i^2 = G_i$ より $\mathrm{Var}(G_i) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$ である。期待値の線形性（linearity of expectation）から平均が得られ、独立性により $N-1$ 個の分散を足し合わせることができる。

検算

- $N = 1$：磁石が 1 つならつねに 1 つのまとまりであり、公式からも平均 1、分散 0 が得られる。
- $N = 2$：2 つの磁石はそれぞれ確率 $\frac{1}{2}$ で 1 個または 2 個のまとまりになるので、平均は $\frac{3}{2}$、分散は $\frac{5}{2} - \frac{9}{4} = \frac{1}{4}$ であり、予測どおりである。
- $N \le 12$ について 2^N 通りのすべての配置を厳密に数え上げたところ、$\frac{N+1}{2}$ と $\frac{N-1}{4}$ が正確に再現された。また $N = 20$ のシミュレーションでは、平均は約 10.5、分散は約 4.73 であった（予測値は 10.5 と 4.75）。

よくある間違い

- **同じ**向きの 2 つの磁石が反発すると考えてしまう。それらは異なる極どうし（N と S）で向かい合うので、引き合ってくっつく。
- 独立性を示さずに分散を足し合わせる。ここでは独立性は成り立つが、隣り合う隙間が磁石を共有しているので自明ではない。
- 隙間を $N-1$ 個ではなく N 個と数えて $\mathrm{Var}(U) = \frac{N}{4}$ と書いてしまう。あるいは、平均で余分な 1 を忘れる。

STEP のその先へ

まとまりの個数を、その間の境界の個数によって数えるのは、**指示変数**（indicator variable）の方法であり、確率論で最も役に立つ道具の 1 つである。個数を 0–1 変数の和として書き、期待値の線形性を使う。これには独立性はまったく必要ない。ここで U は、2 種類の記号からなる無作為な列における**連**（run）の個数を数えている。連の個数は、統計学における Wald–Wolfowitz の連検定（runs test）の基礎であり、この検定は列が無作為に見えるかどうかを調べるものである。指示変数とその分散についてさらに知りたければ、Blitzstein and Hwang, *Introduction to Probability*（確率論入門）を参照されたい。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。