解答例

STEP III 2008、第 13 問

確率・統計 ・ 指示変数の和

> **STEP III 2008、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/08-s3-q13

この問題は、*Beyond STEP* の第 12.10 節「ひもの端を結ぶ」で詳しく解説されている：
https://beyondstep.keithmcnulty.org/ja/ch12/loose-ends.html

鍵となる考え方

各ステップで何が起こっても、ひもの本数は必ずちょうど 1 本減る。輪が 1 つできてゲームから取り除かれるか、2 本のひもが 1 本になるかのどちらかだからである。したがって、それまでに何が起こったかに関係なく、各ステップの後には本数が 1 本少ないだけの同じ状況に戻る。このことから、輪の総数を独立な指示変数（indicator variable）の和として表すことができ、平均と分散は単にそれぞれの和として求まる。

解答

最初のステップ

ひもが n 本あるとき、自由な端は $2n$ 個あるので、2 つの端の選び方は

$$\binom{2n}{2} = n(2n-1)$$

通りあり、これらは同様に確からしい。輪ができるのは 2 つの端が同じひもに属するときであり、そのような組はちょうど n 組（ひも 1 本につき 1 組）ある。よって

$$\mathrm{P}(\text{最初のステップで輪ができる}) = \frac{n}{n(2n-1)} = \frac{1}{2n-1}.$$

言い換えると、最初にどの端を選んでも、残りの $2n-1$ 個の端のうちその相方はちょうど 1 つである。最初のステップで輪ができるとき 1、そうでないとき 0 をとる確率

変数を X_n とすると、最初のステップでできる輪の個数の期待値（expectation）は

$$\mathrm{E}(X_n) = \frac{1}{2n-1}.$$

輪の個数の期待値

最初のステップの後、ひもは $n-1$ 本になり、それぞれに自由な端が 2 つずつある。これはどちらの場合にも成り立つ。輪ができればひもが 1 本ゲームから取り除かれ、異なる 2 本をつなげば 2 本が 1 本になるからである。ひもの長さは関係なく、関係するのは端だけである。

> **考えてみよう** いくつかのひもが他より長くなっていることや、それ以前に輪ができたかどうかは影響するだろうか？

影響しない。どの端の組が選ばれるかに影響するのは、残っているひも（したがって端）の本数だけである。よって、ひもが k 本残っているとき、そのステップで輪ができる確率は、それ以前のステップに関係なく $\frac{1}{2k-1}$ である。

各ステップを、その直前にあるひもの本数 $k = n, n-1, \ldots, 1$ で番号づけし、そのステップで輪ができることの指示変数を X_k とする。すると $\mathrm{P}(X_k = 1) = \frac{1}{2k-1}$ であり、この操作は $k = 1$ のステップの後で終わる（このステップでは必ず輪ができるので $X_1 = 1$ である）。輪の総数は

$$R = X_1 + X_2 + \cdots + X_n,$$

であるから、期待値の線形性（linearity of expectation）により

$$\mathrm{E}(R) = \sum_{k=1}^{n} \frac{1}{2k-1} = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1}.$$

分散

分散（variance）を求めるには線形性だけでは足りず、X_k が独立（independent）であることが必要である。実際に独立である。それ以前のステップがどのような結果であっても、ひもが k 本のステップで輪ができる確率は $\frac{1}{2k-1}$ のままである。したがって、それ以前のすべての X を与えたときの X_k の条件付き分布は、条件なしの分布と同じであり、よって X_k は独立である。

各 X_k は $p_k = \frac{1}{2k-1}$ のベルヌーイ分布（Bernoulli distribution）に従うので $\mathrm{Var}(X_k) =$

$p_k(1-p_k)$ であり、独立な確率変数については分散を足し合わせることができる：

$$\begin{aligned}\mathrm{Var}(R) &= \sum_{k=1}^{n} \frac{1}{2k-1}\left(1-\frac{1}{2k-1}\right) = \sum_{k=1}^{n} \frac{2k-2}{(2k-1)^2} \\ &= \sum_{k=1}^{n} \frac{1}{2k-1} \;-\; \sum_{k=1}^{n} \frac{1}{(2k-1)^2}.\end{aligned}$$

$n = 40\,000$ の場合

$H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$ とおく。$2n-1$ までの奇数の逆数の和は、$2n$ までのすべての逆数の和から偶数の逆数の和を引いたものである：

$$\sum_{k=1}^{n} \frac{1}{2k-1} = H_{2n} - \left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2n}\right) = H_{2n} - \tfrac{1}{2}H_n.$$

m が大きいとき $H_m \approx \ln m$ であることを用いると

$$\mathrm{E}(R) \approx \ln(2n) - \tfrac{1}{2}\ln n = \tfrac{1}{2}\left(2\ln 2n - \ln n\right) = \tfrac{1}{2}\ln(4n).$$

$n = 40\,000$ のとき $4n = 160\,000 = 20^4$ であるから

$$\mathrm{E}(R) \approx \tfrac{1}{2}\ln\left(20^4\right) = 2\ln 20 \approx 6.$$

したがって、40 000 本のひもからできる輪は、期待値でわずか 6 個ほどである。

検算

- $n = 1$：ひもが 1 本なら、輪は確実に 1 つできる。公式は $\mathrm{E}(R) = 1$、$\mathrm{Var}(R) = 0$ を与え、正しい。
- $n = 2$：最初のステップで輪ができる確率は $\frac{1}{3}$ であり、最後のステップでは必ず輪ができるので、R は確率 $\frac{1}{3}$ で 2、それ以外では 1 である。よって平均は $\frac{4}{3}$、分散は $\frac{2}{9}$ となり、公式から得られる $1 + \frac{1}{3}$ および $\frac{2}{9}$ と一致する。
- $n \le 5$ について選び方の列をすべて厳密に数え上げた結果、および $n = 6$ のシミュレーションの結果は、どちらの公式とも一致する。$n = 40\,000$ のとき、厳密な和は $\mathrm{E}(R) \approx 6.28$ である。上の見積もりより 0.29 大きいのは、粗い近似 $H_m \approx \ln m$ によるものである（下記参照）。

よくある間違い

- 輪ができる確率を、$\frac{1}{2n-1}$（2 つの**端**を選ぶ）ではなく $\frac{1}{n}$（2 本の**ひも**を選ぶ）としてしまう。

- X_k が独立であることを先に示さずに分散を足し合わせてしまう。平均には線形性で十分だが、分散にはそれでは足りない。
- 偶数の項を分けずに、$\sum \frac{1}{2k-1}$ を $\ln n$ や $\frac{1}{2}\ln n$ で近似してしまう。

STEP のその先へ

より精密な評価は $H_m = \ln m + \gamma + o(1)$ である。ここで $\gamma \approx 0.5772$ はオイラー・マスケローニ定数（Euler–Mascheroni constant）である。これにより $\mathrm{E}(R) \approx \frac{1}{2}\ln(4n) + \frac{1}{2}\gamma$ が得られ、さらに $\sum 1/(2k-1)^2 = \pi^2/8$ であることから $\mathrm{Var}(R) \approx \frac{1}{2}\ln(4n) + \frac{1}{2}\gamma - \frac{\pi^2}{8}$ も得られる。同じ「確率 $1/k$ の独立な指示変数の和」という構造は、ランダムな置換（permutation）の巡回（cycle）の個数や、ランダムな数列における記録（record）の個数を数える際にも現れ、どちらの場合も個数は $\ln n$ のように増加する。指示変数やこうしたランダムな構造についてさらに学ぶには、Grimmett and Welsh, *Probability: An Introduction*（確率論の入門書）がよい出発点である。

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。