解答例

STEP II 2013、第 13 問

確率・統計 ・ 硬貨投げにおける連

> **STEP II 2013、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/13-s2-q13

鍵となる考え方

事象 $S = n$ は、最初の $n+1$ 回の投げの結果をちょうど指定している。すなわち、同じ結果が n 回続き、次の 1 回でその連（run）が途切れる。$A = n$ についても同様で、n 回交互に出たあと、同じ結果が繰り返される。したがって、それぞれの確率（probability）は、最初の投げの結果の 2 通りに対応する 2 つの積の和になる。S と A の比較は、4 つの項の差を因数分解（factorisation）することに帰着する。そしてその因数分解は、q を $1-p$ で置き換えるよりも、p と q を独立した文字のまま扱う方が見つけやすい。

解答

以下では、各投げは独立（independent）であり、それぞれ確率 p で表、確率 q で裏が出るとし、$p, q > 0$ かつ $p \neq q$ とする。

一般の公式

$S = n$ が起こるのは、ちょうど $1, \ldots, n$ 回目の投げがすべて 1 回目と一致し、$n+1$ 回目が異なるときである。1 回目が表ならばこれは $\mathrm{H}^n\mathrm{T}$ であり、その確率は $p^n q$ である。1 回目が裏ならば $\mathrm{T}^n\mathrm{H}$ であり、その確率は $q^n p$ である。これらの場合は互いに排反（disjoint）であるから、

$$\mathrm{P}(S = n) = p^n q + q^n p = pq\left(p^{n-1} + q^{n-1}\right).$$

$A = n$ が起こるのは、ちょうど $1, \ldots, n$ 回目の投げが交互になっていて、$n+1$ 回目が n 回目と同じになり、交互の並びが終わるときである。

▶ $n = 2m$（偶数）のとき：並び $\mathrm{HTHT}\cdots\mathrm{HT}$ は表を m 回、裏を m 回含み、T で終わ

るので、次の投げは T でなければならず、確率は $p^m q^m \cdot q$ となる。裏から始まる場合は $q^m p^m \cdot p$ となる。よって

$$\mathrm{P}(A = 2m) = p^m q^m (p + q).$$

- $n = 2m + 1$（奇数）のとき：並び HTH$\cdots$H は表を $m + 1$ 回、裏を m 回含み、H で終わるので、次の投げは H でなければならず、確率は $p^{m+2} q^m$ となる。裏から始まる場合は $q^{m+2} p^m$ となる。よって

$$\mathrm{P}(A = 2m + 1) = p^m q^m \left(p^2 + q^2\right).$$

小問 (i)

$A = 1$ は 2 回目の投げが 1 回目と同じであること、すなわち HH または TT を意味する。よって $\mathrm{P}(A = 1) = p^2 + q^2$ である（上の $m = 0$ の場合）。$S = 1$ は 2 回目の投げが 1 回目と異なること、すなわち HT または TH を意味するので、

$$\mathrm{P}(S = 1) = pq + qp = 2pq.$$

すると

$$\mathrm{P}(A = 1) - \mathrm{P}(S = 1) = p^2 - 2pq + q^2 = (p - q)^2 > 0,$$

であり、$p \neq q$ であるから不等号は厳密に成り立つ。したがって $\mathrm{P}(S = 1) < \mathrm{P}(A = 1)$ である。

小問 (ii)

公式より、

$$\mathrm{P}(S = 2) = p^2 q + q^2 p = pq(p + q) = \mathrm{P}(A = 2),$$

となり、示すべきことが得られた。長さ 3 については、

$$\mathrm{P}(S = 3) = p^3 q + q^3 p = pq \left(p^2 + q^2\right) = \mathrm{P}(A = 3),$$

となるので、この 2 つも等しい。

小問 (iii)

偶数の長さ。$n > 1$ とする。両方の確率を $pq > 0$ で割ると、

$$\frac{\mathrm{P}(S = 2n)}{pq} = p^{2n-1} + q^{2n-1}, \qquad \frac{\mathrm{P}(A = 2n)}{pq} = p^{n-1} q^{n-1} (p + q) = p^n q^{n-1} + p^{n-1} q^n.$$

考えてみよう $p^{2n-1}+q^{2n-1}-p^nq^{n-1}-p^{n-1}q^n$ の符号はどうすれば判定できるか？

それぞれの純粋なべき乗を、同じ文字の最高次のべきを共有する交差項と組にする：

$$\begin{aligned}p^{2n-1}+q^{2n-1}-p^nq^{n-1}-p^{n-1}q^n &= p^n\left(p^{n-1}-q^{n-1}\right)-q^n\left(p^{n-1}-q^{n-1}\right)\\ &= \left(p^n-q^n\right)\left(p^{n-1}-q^{n-1}\right).\end{aligned}$$

$p,q>0$ かつ $p\neq q$ であるから、任意の $k\geq 1$ に対して数 p^k-q^k は 0 でなく、$p-q$ と同符号である（$x\mapsto x^k$ は $x>0$ において狭義単調増加（strictly increasing）であるため）。ここで $n\geq 2$ であるから、n と $n-1$ はともに 1 以上であり、2 つの括弧はどちらも 0 でなく同符号である。その積は正であり、$pq>0$ を掛け戻すと、

$$\mathrm{P}(S=2n)>\mathrm{P}(A=2n)\qquad (n>1).$$

（$n=1$ のときは 2 つ目の括弧が $p^0-q^0=0$ となり、これが小問 (ii) の等式である。）

奇数の長さ。同様に、やはり $n>1$ として、

$$\frac{\mathrm{P}(S=2n+1)}{pq}=p^{2n}+q^{2n},$$

$$\frac{\mathrm{P}(A=2n+1)}{pq}=p^{n-1}q^{n-1}\left(p^2+q^2\right)=p^{n+1}q^{n-1}+p^{n-1}q^{n+1},$$

であり、

$$\begin{aligned}p^{2n}+q^{2n}-p^{n+1}q^{n-1}-p^{n-1}q^{n+1} &= p^{n+1}\left(p^{n-1}-q^{n-1}\right)-q^{n+1}\left(p^{n-1}-q^{n-1}\right)\\ &= \left(p^{n+1}-q^{n+1}\right)\left(p^{n-1}-q^{n-1}\right).\end{aligned}$$

同じ議論により、2 つの括弧はどちらも 0 でなく $p-q$ と同符号であるから、

$$\mathrm{P}(S=2n+1)>\mathrm{P}(A=2n+1)\qquad (n>1).$$

ここでも因数 $p^{n-1}-q^{n-1}$ は $n=1$ で 0 となり、これは $\mathrm{P}(S=3)=\mathrm{P}(A=3)$ と一致する。

まとめると、長さ 1 では同じ結果が続く連の方が交互の連より起こりにくく、長さ 2 と 3 では両者は等しく起こりやすく、長さ 4 以上ではどの長さでも同じ結果が続く連の方が起こりやすい。

検算

- $p=\frac{1}{3}$、$q=\frac{2}{3}$ とする。すると $\mathrm{P}(S=4)=pq(p^3+q^3)=\frac{2}{9}\cdot\frac{1}{3}=\frac{2}{27}\approx 0.074$、$\mathrm{P}(A=4)=p^2q^2=\frac{4}{81}\approx 0.049$ であり、確かに $S=4$ の方が起こりやすい。

- すべての $n \geq 1$ について $\mathrm{P}(S = n)$ を足し合わせると $q \cdot \frac{p}{1-p} + p \cdot \frac{q}{1-q} = p + q = 1$ となり、当然そうなるべき値になっている。
- 各公式は、いくつかの p の値について、14 回の投げのすべての列を正確な分数で列挙するコンピュータ計算により、長さ 13 まで確認した。

よくある間違い

- 連を終わらせる投げを忘れること。$S = n$ には $n+1$ 回目の投げが異なることが必要であり、これがないと $\mathrm{P}(S \geq n)$ を計算していることになる。
- 交互の連を誤った文字で終わらせること。奇数の長さの連は始まりと同じ文字で終わり、偶数の長さの連はもう一方の文字で終わる。
- 小問 (iii) で $q = 1 - p$ を代入すること。式が p の多項式になり、その因数分解ははるかに見つけにくくなる。

STEP のその先へ

S は幾何分布型の確率変数（geometric-type random variable）である。最初の投げの結果が与えられれば、連の長さは最初に結果が変わるまでの待ち時間（waiting time）になる。このような連は、結果の列がランダムに見えるかどうかを判定するために用いられる統計学の連検定（runs test）の基礎となっており、また母関数（generating function）を使うと最長連の分布全体をうまく求めることができる。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、母関数と待ち時間をちょうどよい水準で扱っている。

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。