解答例

STEP I 1990、第 15 問

確率・統計 ・ 負の二項分布

> **STEP I 1990、問 15**
>
> 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/90-s1-q15

鍵となる考え方

最初の部分では、事象を二つの独立な（independent）部分に分ける。すなわち、最後の投げで何が起こるかと、それ以前のすべての投げで何が起こるかである。残りの部分で重要なのは、硬貨には記憶がないという事実、すなわち無記憶性（memoryless property）である。表が出た直後から、次の表が出るまでの待ち時間は新たに始まる。したがって、幾何分布（geometric distribution）に従う待ち時間が偶数になる確率と奇数になる確率さえわかればよい。

解答

以下では一貫して $q = 1 - p$ とおく。各回の投げは独立であり、それぞれ確率 p で表が出る。

k 回目の表の分布

k 回目の表が出る投げの番号を T_k とする。事象 $\{T_k = n\}$ が起こるのは、次の二つが成り立つときであり、かつそのときに限る。

- n 回目の投げが表であり、かつ
- $1, 2, \ldots, n-1$ 回目の投げのうち、表がちょうど $k-1$ 回（したがって裏が $n-k$ 回）である。

両方の条件が必要である。n 回目が裏であれば、k 回目の表は n 回目より前か後に出ることになり、n 回目には出ない。最初の $n-1$ 回に含まれる表の個数が異なれば、そ

の回の投げは k 回目の表にはならない。また、二つの条件を合わせれば十分でもある。

これら二つの事象は互いに重ならない投げの集合に関するものなので、独立である。前者の確率は p である。後者については、$k-1$ 個の表は最初の $n-1$ 回のうち $\binom{n-1}{k-1}$ 通りの位置の組のどれにあってもよい。そのような配置はそれぞれ $k-1$ 個の表と $n-k$ 個の裏からなる特定の列であり、独立性によりその確率は $p^{k-1}q^{n-k}$ である。また、異なる配置は互いに排反である。よって

$$\mathrm{P}(T_k = n) = p \cdot \binom{n-1}{k-1} p^{k-1} q^{n-k} = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \qquad n \geq k,$$

となり、示すべき式が得られた。

偶数から偶数へ

$(k-1)$ 回目の表の後、さらに必要な投げの回数を $G = T_k - T_{k-1}$ とする（ただし $T_0 = 0$ とする）。T_{k-1} 回目より後の投げは、それまでに起こったすべてのことと独立である。したがって G は T_{k-1} と独立であり、幾何分布

$$\mathrm{P}(G = g) = q^{g-1}p, \qquad g = 1, 2, \ldots$$

に従う（これは最初の部分で $k = 1$ とした場合である）。

考えてみよう T_{k-1} が偶数のとき、T_k が偶数になるのはどのような場合か？

$T_k = T_{k-1} + G$ であるから、T_{k-1} が偶数ならば、T_k が偶数になるのは G が偶数のときであり、かつそのときに限る。G は T_{k-1} と独立なので、T_{k-1} が偶数であるという条件をつけても G の分布は変わらない。したがって

$$\mathrm{P}(T_k \text{ が偶数} \mid T_{k-1} \text{ が偶数}) = \mathrm{P}(G \text{ が偶数}).$$

幾何級数の和を計算すると、

$$\mathrm{P}(G \text{ が偶数}) = p\left(q + q^3 + q^5 + \cdots\right) = \frac{pq}{1-q^2} = \frac{pq}{(1-q)(1+q)} = \frac{q}{1+q},$$
$$\mathrm{P}(G \text{ が奇数}) = p\left(1 + q^2 + q^4 + \cdots\right) = \frac{p}{1-q^2} = \frac{1}{1+q}.$$

よって求める確率は

$$\frac{q}{1+q} = \frac{1-p}{2-p}.$$

である。この答えは k によらない。

偶数回目の投げでちょうど二回表が出る確率

今度は三回目の表が出るまで投げる。三つの待ち時間を $G_1 = T_1$、$G_2 = T_2 - T_1$、$G_3 = T_3 - T_2$ とする。同じ無記憶性の議論により、これらは独立であり、いずれも上

と同じ幾何分布に従う。

$$e = \mathrm{P}(G_i \text{ が偶数}) = \frac{q}{1+q}, \qquad o = \mathrm{P}(G_i \text{ が奇数}) = \frac{1}{1+q}.$$

とおく。出発点は 0 回目であり、これは偶数である。偶数の間隔は現在位置の偶奇（parity）を保ち、奇数の間隔は偶奇を入れ替える。次の表は、(G_1, G_2, G_3) の偶奇の八通りのパターンと、その結果として得られる (T_1, T_2, T_3) の偶奇を並べたものである。

G_1	G_2	G_3	T_1	T_2	T_3	偶数回目の表
偶	偶	偶	偶	偶	偶	3
偶	偶	奇	偶	偶	奇	**2**
偶	奇	偶	偶	奇	奇	1
偶	奇	奇	偶	奇	偶	**2**
奇	偶	偶	奇	奇	奇	0
奇	偶	奇	奇	奇	偶	1
奇	奇	偶	奇	偶	偶	**2**
奇	奇	奇	奇	偶	奇	1

偶数回目の投げでちょうど二回表が出るのは三つの場合であり、それらの確率は e^2o、eo^2、o^2e である。間隔の独立性により、求める確率は

$$\begin{aligned} e^2o + 2eo^2 = eo(e+2o) &= \frac{q}{(1+q)^2} \cdot \frac{q+2}{1+q} \\ &= \frac{q(2+q)}{(1+q)^3} = \frac{(1-p)(3-p)}{(2-p)^3}. \end{aligned}$$

検算

- 最初の部分の検算として、$\mathrm{P}(T_k = n)$ を $n \geq k$ について足し合わせると 1 になるはずである。これは負の二項級数（negative binomial series）$\sum_{n \geq k} \binom{n-1}{k-1} q^{n-k} = (1-q)^{-k}$ である。
- 偶奇の二つの確率 $\frac{q}{1+q}$ と $\frac{1}{1+q}$ の和は 1 である。p が 1 に近いとき、最初の表はほぼ必ずすぐに、すなわち奇数回目に出るので、$\mathrm{P}(G \text{ が偶数}) \to 0$ となり、公式と一致する。
- 公平な硬貨（$p = \frac{1}{2}$）のとき、最終的な答えは $\frac{(1/2)(5/2)}{(3/2)^3} = \frac{10}{27} \approx 0.370$ である。2×10^5 回のシミュレーションでは 0.369 が得られた。また、三つの待ち時間について直接和をとった値は、$p = 0.3$、0.5、0.8 のいずれでも公式と 10^{-9} の精度で一致した。

よくある間違い

- 最初の部分で $\binom{n}{k}$ を使ってしまう。最後の投げは表と決まっているので、自由に選べるのは最初の $n-1$ 回だけである。
- 二番目の部分で、負の二項分布を偶数の n について足し合わせようとする。無記憶性による間隔の考え方を使えば、これはまったく必要ない。
- 最後の部分で、偶数回目の位置にある表を、位置どうしが独立であるかのように数えてしまう。各位置は一つ前の位置をもとに決まるので、独立ではない。

STEP のその先へ

k 回目の表が出る時刻は**負の二項分布**（negative binomial distribution）に従う。これを独立な幾何分布の間隔の和として表すことは、ポアソン過程（Poisson process）の到着時刻が独立な指数分布（exponential distribution）の待ち時間から組み立てられることの離散版である。最後の部分の偶奇の記録は二状態のマルコフ連鎖（Markov chain）であり、大学一年次の確率論の講義の早い段階で学ぶ話題である。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、これらすべてを扱っている。

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。