解答例

STEP III 1990、第 15 問

確率・統計 ・ 幾何分布

> **STEP III 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-s3-q15

鍵となる考え方

各回の投げで A が出る確率は $\frac{1}{4}$、B が出る確率は $\frac{3}{4}$ であり、それぞれ独立である。したがって事象 $\{M = m,\ N = n\}$ はある特定の投げの列 1 つに対応し、その確率は積で表される。この積は m のみに依存する因子と n のみに依存する因子の積に分解でき、これがほとんどすべての仕事をしてくれる。すなわち、周辺分布（marginal distribution）は幾何分布（geometric distribution）となり、2 つの確率変数は独立（independent）であり、以降のすべての確率は等比級数（geometric series）になる。

解答

1 回の投げについて $p = \mathrm{P}(A) = \frac{3}{12} = \frac{1}{4}$、$q = \mathrm{P}(B) = \frac{9}{12} = \frac{3}{4}$ とおく。各回の投げは独立である。

事象 $\{M = m,\ N = n\}$ が起こるのは、最初の $m + n + 1$ 回の投げが

$$\underbrace{A\,A\cdots A}_{m}\ \underbrace{B\,B\cdots B}_{n}\ A,$$

となるとき、かつそのときに限る。よって独立性より、$m \geq 0$ かつ $n \geq 1$ に対して

$$\mathrm{P}(M = m,\ N = n) = \left(\tfrac{1}{4}\right)^{m} \left(\tfrac{3}{4}\right)^{n} \tfrac{1}{4} = \left(\tfrac{1}{4}\right)^{m+1} \left(\tfrac{3}{4}\right)^{n}.$$

である。（確率 1 で、列はいずれ B を出し、その後に A を出すので、これらの確率の総和は 1 である。このことは以下の周辺分布によっても確かめられる。）

小問 (i)

M の周辺分布を求めるには、$\sum_{n\geq 1}\left(\frac{3}{4}\right)^n = \dfrac{3/4}{1-3/4} = 3$ を用いて $n \geq 1$ について和をとる。

$$\mathrm{P}(M=m) = \left(\tfrac{1}{4}\right)^{m+1}\cdot 3 = \tfrac{3}{4}\left(\tfrac{1}{4}\right)^m, \qquad m = 0,1,2,\ldots$$

N の周辺分布を求めるには、$\sum_{m\geq 0}\left(\frac{1}{4}\right)^{m+1} = \dfrac{1/4}{1-1/4} = \frac{1}{3}$ を用いて $m \geq 0$ について和をとる。

$$\mathrm{P}(N=n) = \tfrac{1}{3}\left(\tfrac{3}{4}\right)^n = \tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1}, \qquad n = 1,2,3,\ldots$$

どちらも直接的に解釈できる。M は最初の B より前に出る A の個数を数えたものであり、N は B の連が始まってから最初の A が出るまでの待ち時間である。いずれも総和は 1 になる。

小問 (ii)

$|x| < 1$ のとき $\sum_{k\geq 1} kx^{k-1} = \dfrac{1}{(1-x)^2}$ が成り立つこと（等比級数を微分すればよい）を用いる。すると

$$\mathrm{E}(M) = \tfrac{3}{4}\sum_{m\geq 1} m\left(\tfrac{1}{4}\right)^m = \tfrac{3}{4}\cdot\tfrac{1}{4}\cdot\frac{1}{(3/4)^2} = \tfrac{1}{3},$$

$$\mathrm{E}(N) = \tfrac{1}{4}\sum_{n\geq 1} n\left(\tfrac{3}{4}\right)^{n-1} = \tfrac{1}{4}\cdot\frac{1}{(1/4)^2} = 4.$$

となる。

独立性

周辺分布どうしを掛け合わせると、すべての $m \geq 0, n \geq 1$ に対して

$$\mathrm{P}(M=m)\,\mathrm{P}(N=n) = \tfrac{3}{4}\left(\tfrac{1}{4}\right)^m\cdot\tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1} = \left(\tfrac{1}{4}\right)^{m+1}\left(\tfrac{3}{4}\right)^n = \mathrm{P}(M=m,\ N=n)$$

が成り立つ。したがって M と N は**独立である**。

> ? **考えてみよう** A の連の長さは、B の連に影響を与えるはずではないか？　なぜ影響しないのか？

A の連は最初の B で終わり、その投げ以降、過程は新たに始まり直す。すなわち、残りの投げはそれ以前のすべてと独立である。B の連の長さはその後の投げのみに依存するので、M を知っても N については何もわからない。同時確率の因数分解は、このことを代数的に表現したものである。

N と M に関する確率

整数 $k \geq 1$ に対して、幾何分布の裾の部分の和をとると

$$\mathrm{P}(N > k) = \sum_{n \geq k+1} \tfrac{1}{4} \left(\tfrac{3}{4}\right)^{n-1} = \tfrac{1}{4} \cdot \frac{(3/4)^k}{1 - 3/4} = \left(\tfrac{3}{4}\right)^k.$$

となる。この式は $\mathrm{P}(N > 0) = 1$ も与えるが、$N \geq 1$ であるからこれは正しい。

$\mathrm{P}(N > M)$ については、M で条件付けして独立性を用いる。

$$\begin{aligned}\mathrm{P}(N > M) &= \sum_{m \geq 0} \mathrm{P}(M = m)\,\mathrm{P}(N > m) = \sum_{m \geq 0} \tfrac{3}{4} \left(\tfrac{1}{4}\right)^m \left(\tfrac{3}{4}\right)^m \\ &= \tfrac{3}{4} \sum_{m \geq 0} \left(\tfrac{3}{16}\right)^m = \tfrac{3}{4} \cdot \tfrac{16}{13} = \tfrac{12}{13}.\end{aligned}$$

$\mathrm{P}(N = M)$ については、$N \geq 1$ であるから共通の値は 1 以上でなければならない。

$$\begin{aligned}\mathrm{P}(N = M) &= \sum_{m \geq 1} \tfrac{3}{4} \left(\tfrac{1}{4}\right)^m \cdot \tfrac{1}{4} \left(\tfrac{3}{4}\right)^{m-1} = \tfrac{3}{64} \sum_{m \geq 1} \left(\tfrac{3}{16}\right)^{m-1} \\ &= \tfrac{3}{64} \cdot \tfrac{16}{13} = \tfrac{3}{52}.\end{aligned}$$

最後に、3 つの事象 $N > M, N = M, N < M$ は互いに排反であり、かつすべての場合を尽くしているので、

$$\mathrm{P}(N < M) = 1 - \tfrac{12}{13} - \tfrac{3}{52} = \tfrac{52 - 48 - 3}{52} = \tfrac{1}{52},$$

となり、示すべきことが示された。

検算

- 平均値は直観に合っている。A は出にくいので、最初の A の連は短く（平均 $\frac{1}{3}$）、一方 B の連を終わらせる A が出るまでには平均 $1/p = 4$ 回待つことになる。
- 直接計算すると $\mathrm{P}(M > n) = \left(\tfrac{1}{4}\right)^{n+1}$ であるから、$\mathrm{P}(N < M) = \sum_{n \geq 1} \tfrac{1}{4} \left(\tfrac{3}{4}\right)^{n-1} \left(\tfrac{1}{4}\right)^{n+1} = \tfrac{1}{64} \sum_{n \geq 1} \left(\tfrac{3}{16}\right)^{n-1} = \tfrac{1}{64} \cdot \tfrac{16}{13} = \tfrac{1}{52}$ となり、引き算で求めた結果と一致する。$\mathrm{P}(N > k) = \left(\tfrac{3}{4}\right)^k$ の直接的な理由は、最初の B の後の k 回の投げがすべて B でなければならないことである。
- 同時分布を厳密に足し合わせると $\mathrm{P}(N > M) = 0.92308$、$\mathrm{P}(N = M) = 0.05769$、$\mathrm{P}(N < M) = 0.01923$ となる。200 000 回の投げの系列をシミュレーションしたところ、0.923、0.058、0.019 が得られ、標本平均は 0.334 と 3.99 であった。

よくある間違い

- $\mathrm{P}(N = M)$ の和に $m = 0$ を含めてしまうこと。N が 0 になることはない。
- M が 0 になりうること（最初の投げが B かもしれない）を忘れること。これにより M の周辺分布がずれ、誤った平均が得られてしまう。

STEP のその先へ

ここで連の長さが独立になるのは、独立試行の列が、最初の B のように過去のみによって決まる任意の時刻において「再出発」するからである。これは**強マルコフ性**（strong Markov property）と呼ばれ、再生理論（renewal theory）やマルコフ連鎖（Markov chain）の解析の基礎となっている。これらは大学 1、2 年次で学ぶ内容である。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、幾何分布、独立性、マルコフ連鎖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。