解答例

STEP III 2019、第 12 問

確率・統計 ・ ランダムな部分集合

> **STEP III 2019、問 12**
>
> 先に進む前に、まず問題を読んでおこう。問題は、OCR が配布している 2019 年の STEP 過去問一式（zip ファイル）のうち、STEP III の問題用紙にある：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

この問題は、*Beyond STEP* の第 9.12 節「ランダムな部分集合」で詳しく解説されている：
https://beyondstep.keithmcnulty.org/ja/ch09/random-subsets.html

鍵となる考え方

S の部分集合（subset）を一様に無作為に（uniformly at random）選ぶことは、整数 $1, 2, \ldots, n$ を一つずつ順に見ていき、それぞれについて公平なコイン（fair coin）を投げて、その整数を入れるかどうかを決めることと同じである。このことに気づけば、問題に現れるすべての事象（event）は各整数ごとの条件に言い換えられる。しかも各整数は独立（independent）にふるまうので、答えは一つの小さな確率を n 乗したものになる。

解答

部分集合を数える

S の部分集合は、n 個の整数それぞれについて、入れるか入れないかを決めることで完全に定まる。各整数について選び方は 2 通りあり、それらは独立に選べて、選び方のパターンが異なれば異なる部分集合が得られる。よって、$\varnothing$（すべて入れない）と S（すべて入れる）も含めて、部分集合の個数は

$$2 \times 2 \times \cdots \times 2 = 2^n.$$

である。

考えてみよう なぜ各整数を独立なコイン投げとして扱ってよいのだろうか。

A を T 上の一様分布に従うものとする。2^n 通りの入る・入らないのパターンはそれぞれ一つの部分集合に対応するので、各パターンの確率は $2^{-n} = (\frac{1}{2})^n$ である。これは、n 枚の独立な公平なコインを投げたときに特定のパターンが出る確率とちょうど一致する。したがって、事象「$1 \in A$」、「$2 \in A$」、...、「$n \in A$」は独立であり、それぞれの確率は $\frac{1}{2}$ である。さらに $A_1, \ldots, A_m$ も互いに独立に選ばれるので、mn 個の事象「$i \in A_k$」はすべて独立であり、それぞれの確率は $\frac{1}{2}$ である。

小問 (i)

2^n 個の部分集合のうち、ちょうど半分が 1 を含む（1 を含む各部分集合を、そこから 1 を取り除いた部分集合と対にすればよい）。よって

$$\mathrm{P}(1 \in A_1) = \tfrac{1}{2}.$$

である。

小問 (ii)

整数 i に対し、i が A_1 と A_2 の両方には**属さない**という事象を E_i とする。このとき $A_1 \cap A_2 = \varnothing$ となるのは、すべての E_i が起こるとき、かつそのときに限る。独立性より

$$\mathrm{P}(i \in A_1 \text{ かつ} i \in A_2) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \qquad \text{よって} \qquad \mathrm{P}(E_i) = \tfrac{3}{4}.$$

である。事象 E_i は整数 i に対するコインだけで決まるので、$E_1, \ldots, E_n$ は独立であり、

$$\mathrm{P}(A_1 \cap A_2 = \varnothing) = \mathrm{P}(E_1) \cdots \mathrm{P}(E_n) = \left(\tfrac{3}{4}\right)^n,$$

となって、示すべきことが得られた。

同じ議論は集合がいくつあっても成り立つ。整数 i が共通部分（intersection）$A_1 \cap A_2 \cap A_3$ に属する確率は $(\frac{1}{2})^3 = \frac{1}{8}$ なので、共通部分に属さない確率は $\frac{7}{8}$ であり、

$$\mathrm{P}(A_1 \cap A_2 \cap A_3 = \varnothing) = \left(\tfrac{7}{8}\right)^n.$$

となる。一般に、i が m 個の集合すべてに属する確率は 2^{-m} なので、

$$\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_m = \varnothing) = \left(1 - \frac{1}{2^m}\right)^n.$$

である。

小問 (iii)

ここで $A_1 \subseteq A_2$ とは、A_1 に属するが A_2 には属さない整数が一つもないということである。各整数 i について、(A_1 に属するか？、A_2 に属するか？) の組には等確率の四つの可能性があり、そのうち禁止されるのは「A_1 に属し、A_2 に属さない」の一つだけである。よって各整数は独立に確率 $\frac{3}{4}$ で条件を満たし、

$$\mathrm{P}(A_1 \subseteq A_2) = \left(\tfrac{3}{4}\right)^n.$$

となる。(これは小問 (ii) と整合する。$A_1 \subseteq A_2$ となるのは、A_1 と A_2 の補集合 (complement) が互いに素 (disjoint) であるとき、かつそのときに限り、さらに一様に無作為な部分集合の補集合もまた一様に分布するからである。)

$A_1 \subseteq A_2 \subseteq A_3$ については、一つの整数 i の (A_1, A_2, A_3) への所属パターンを、「属する」を 1、「属さない」を 0 と書いて調べる。鎖 (chain) の条件は、i がある A_k に属したら、それ以降のすべての集合にも属するということなので、パターンは

$$000, \quad 001, \quad 011, \quad 111.$$

のいずれかでなければならない。これは等確率な 8 通りのパターンのうち 4 通りなので、各整数が条件を満たす確率は $\frac{4}{8} = \frac{1}{2}$ であり、

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3) = \left(\tfrac{1}{2}\right)^n.$$

となる。

m 個の集合からなる鎖については、許されるパターンは 0 の並びの後に 1 の並びが続くものである。これは 1 が始まる位置によって決まり、その位置は $1, 2, \ldots, m$ のいずれかか、または「始まらない」である。よって 2^m 通りのうち許されるパターンは $m+1$ 通りであり、

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq \cdots \subseteq A_m) = \left(\frac{m+1}{2^m}\right)^n.$$

である。

検算

- $m = 2$ のとき一般の公式は $(1 - \frac{1}{4})^n$ と $(\frac{3}{4})^n$ を与え、$m = 3$ のときは $(\frac{7}{8})^n$ と $(\frac{4}{8})^n$ を与えるので、特別な場合と一致する。$m = 1$ のときは $(\frac{1}{2})^n$ ($A_1 = \varnothing$ の場合だけが条件を満たす) と 1 (集合が一つなら常に鎖である) を与え、これも当然の結果である。
- $n = 1$ のときはすべてを手で書き出せる。たとえば、四つの選び方のうち共通部分が空でないのは $A_1 = A_2 = \{1\}$ だけなので、確率は $\frac{3}{4}$ となる。

- 集合の選び方をすべて列挙するコンピュータによる検算で、小さな n と m についてすべての公式が正しいことを確認した。たとえば $n = 3$ のとき、$\mathrm{P}(A_1 \cap A_2 = \varnothing) = \frac{27}{64}$ であり、$\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3) = \frac{1}{8}$ である。

よくある間違い

- A_1 の各要素が A_2 に属さなければならないという議論から、A_1 自身の大きさも無作為であることを考慮せずに、$\mathrm{P}(A_1 \subseteq A_2)$ を $\left(\frac{1}{2}\right)^n$ と答えてしまう。
- 整数がどの集合にも属さないパターンを忘れて、鎖のパターンの数を $m+1$ ではなく m と数えてしまう。

STEP のその先へ

無作為な部分集合を独立なコイン投げの列（その**指示ベクトル**（indicator vector））として表すことは、組合せ論（combinatorics）の標準的な道具であり、ここでの鎖の数え上げは**半順序集合**（poset）とブール束（Boolean lattice）の理論の入り口である。同じ考え方を使えば、鎖の個数の期待値（expected number）や、無作為な集合たちが S を覆う確率も、わずかな追加の手間で求められる。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、このような精神で独立性と指示変数（indicator variables）を展開している。

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。