解答例

STEP II 2020、第 12 問

確率・統計 ・ 偏りのあるさいころと不等式

> **STEP II 2020、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 2020 年の STEP 過去問一式（zip ファイル）のうち、STEP II の問題用紙にある：
>
> www.ocr.org.uk/Images/696589-2020-past-papers.zip

鍵となる考え方

確率（probability）の総和は 1 でなければならないから、偏り（bias）は $\sum \varepsilon_i = 0$ を満たす。これを書き出してしまえば、$\sum(\frac{1}{n} + \varepsilon_i)^2$ を展開したときに 1 次の項はすべて消え、明らかに非負であるものが残る。小問 (ii) は姿を変えた小問 (i) にすぎない。正の数をその総和で割ると、それらはあるさいころの各目の確率になるのである。

解答

$p_i = \mathrm{P}(X = i) = \frac{1}{n} + \varepsilon_i$ とおく。$\sum_{i=1}^{n} p_i = 1$ であるから、

$$1 = n \cdot \frac{1}{n} + \sum_{i=1}^{n} \varepsilon_i \quad \Longrightarrow \quad \sum_{i=1}^{n} \varepsilon_i = 0.$$

また、各 $p_i \geq 0$ であるから、すべての i について $\varepsilon_i \geq -\frac{1}{n}$ である。以下の和はすべて $i = 1, \ldots, n$ にわたってとる。

小問 (i)

2 回の試行は独立（independent）であるから、2 回とも i が出る確率は p_i^2 であり、異なる i に対するこれらの事象は互いに排反（disjoint）である。したがって

$$\begin{aligned}\mathrm{P}(\text{同じ目が 2 回続けて出る}) = \sum p_i^2 &= \sum \left(\frac{1}{n^2} + \frac{2\varepsilon_i}{n} + \varepsilon_i^2 \right) \\ &= \frac{1}{n} + \frac{2}{n} \sum \varepsilon_i + \sum \varepsilon_i^2 = \frac{1}{n} + \sum_{i=1}^{n} \varepsilon_i^2.\end{aligned}$$

公平なさいころ（fair die）ではすべての $\varepsilon_i = 0$ であり、確率は $\frac{1}{n}$ となる。偏ったさいころ（biased die）では ε_i のすべてが 0 というわけではないから、$\sum \varepsilon_i^2 > 0$ である。よって、偏ったさいころは公平なさいころよりも、2 回続けて同じ目が出る確率が**大きい**。

小問 (ii)

まず二重和を書き直す。これは $j < i$ を満たすすべての組について $x_i x_j$ を足し合わせたもの、すなわち相異なる添字の順序のない組それぞれについてちょうど 1 回ずつ足したものである。$S = \sum x_i$ を 2 乗すると、各 x_i^2 が 1 回ずつ、そのような組がそれぞれ 2 回ずつ現れるから、

$$\sum_{i=2}^{n} \sum_{j=1}^{i-1} x_i x_j = \frac{1}{2}\left(S^2 - \sum x_i^2\right).$$

考えてみよう x_i が単なる任意の正の数であるとき、さいころはどこから出てくるのだろうか。

$p_i = x_i / S$ とおく。これらは正で総和が 1 であるから、ある n 面さいころの各目の確率である。小問 (i) より、$\sum p_i^2 = \frac{1}{n} + \sum \varepsilon_i^2 \geq \frac{1}{n}$ である（等号が成り立つのはさいころが公平なとき、すなわちすべての x_i が等しいときに限る）。$S^2 > 0$ を掛けて、

$$\sum x_i^2 \geq \frac{S^2}{n}.$$

したがって

$$\sum_{i=2}^{n} \sum_{j=1}^{i-1} x_i x_j = \frac{1}{2}\left(S^2 - \sum x_i^2\right) \leq \frac{1}{2}\left(S^2 - \frac{S^2}{n}\right) = \frac{n-1}{2n}\left(\sum_{i=1}^{n} x_i\right)^2,$$

となり、これが求める不等式である。等号はすべての x_i が等しいときに成り立つ。

小問 (iii)

今度は確率は $\sum p_i^3$ である。展開して、再び $\sum \varepsilon_i = 0$ を用いると、

$$\begin{aligned}\sum p_i^3 &= \sum \left(\frac{1}{n^3} + \frac{3\varepsilon_i}{n^2} + \frac{3\varepsilon_i^2}{n} + \varepsilon_i^3\right) \\ &= \frac{1}{n^2} + \sum_{i=1}^{n}\left(\frac{3\varepsilon_i^2}{n} + \varepsilon_i^3\right) = \frac{1}{n^2} + \sum_{i=1}^{n} \varepsilon_i^2 \left(\frac{3}{n} + \varepsilon_i\right).\end{aligned}$$

公平なさいころでは、これは $\frac{1}{n^2}$ である。

考えてみよう 3 乗の項 ε_i^3 は負になりうる。これが 2 乗の項を上回ることはありうるだろうか。

ありえない。偏りは大きく負になることはできないからである。$p_i \geq 0$ より $\varepsilon_i \geq -\frac{1}{n}$ であるから、

$$\frac{3}{n} + \varepsilon_i \geq \frac{2}{n} > 0.$$

よって各項 $\varepsilon_i^2\left(\frac{3}{n} + \varepsilon_i\right)$ は ≥ 0 であり、$\varepsilon_i \neq 0$ のときは常に正である。偏ったさいころでは、少なくとも 1 つの i についてそうなる。ゆえに

$$\sum p_i^3 > \frac{1}{n^2},$$

となり、偏ったさいころはやはり公平なさいころよりも、3 回続けて同じ目が出る確率が大きい。

検算

- $n = 4$ とし、確率を $\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$ とすると、$\varepsilon = \frac{1}{4}, -\frac{1}{12}, -\frac{1}{12}, -\frac{1}{12}$ である。直接計算すると $\sum p_i^2 = \frac{1}{4} + \frac{3}{36} = \frac{1}{3}$ であり、$\frac{1}{4} + \sum \varepsilon_i^2 = \frac{1}{4} + \frac{1}{16} + \frac{3}{144} = \frac{1}{3}$ となって、公平な場合の $\frac{1}{4}$ を上回る。3 回の場合は $\sum p_i^3 = \frac{1}{8} + \frac{3}{216} = \frac{5}{36} \approx 0.139$ であり、公平な場合の $\frac{1}{16} = 0.0625$ を大きく上回る。
- 最も極端なさいころは常に 1 の目が出るものである。このとき両方の確率は 1 となり、とりうる最大の値である。
- 小問 (ii) で $n = 2$ とすると、不等式は $x_1 x_2 \leq \frac{1}{4}(x_1 + x_2)^2$ となり、これは相加相乗平均の不等式（AM–GM）である。小問 (i)、(ii)、(iii) の不等式は、何千ものランダムな場合についてコンピュータでも確かめた。

よくある間違い

- $\sum \varepsilon_i = 0$ を忘れ、符号のわからない 1 次の項が残ってしまう。
- 小問 (iii) で、$\varepsilon_i \geq -\frac{1}{n}$ を用いずに $\sum \varepsilon_i^3$ を「小さい」とか、どちらの符号もとりうるなどと述べてしまう。この評価こそが議論を成り立たせるものである。
- 小問 (ii) で、小問 (i) を $x_i / \sum x_j$ ではなく x_i そのものに適用してしまう。x_i の総和は 1 になるとは限らない。

STEP のその先へ

量 $\sum p_i^2$ は分布の**衝突確率**（collision probability）であり、小問 (i) はそれが一様分布（uniform distribution）のときに最小になることを述べている。小問 (iii) は $\sum p_i^3$ についての同じ主張である。どちらも $t \mapsto t^k$ の凸性（convexity）（イェンセン（Jensen）の不等式）、あるいは冪平均不等式（power-mean inequality）の特別な場合であり、$\sum p_i^k$ の対数をとると、情報理論（information theory）や暗号理論（cryptography）で用いられるレニー（Rényi）エントロピーが得られる。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）は、これらの考え方に確率論の文脈で出会うのによい本である。

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。