解答例

STEP I 2008、第 12 問

確率・統計 ・ 離散一様分布の最大値

STEP I 2008、問 12

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/08-s1-q12

鍵となる考え方

X は 2 つの得点のうち大きい方にすぎない。「大きい方が r に等しい」確率を直接求めようとすると、いくつかの場合を数え上げる必要があるが、「大きい方が r **以下**である」確率は簡単に求まる。それはちょうど**両方**の得点が r 以下であるときに起こるからである。そこで、まず累積分布（cumulative distribution）$\mathrm{P}(X \leq r)$ を考え、その差をとる。同じ累積分布を使えば、中央値（median）もほとんどすぐに求まる。

解答

X_1 と X_2 を独立（independent）な 2 つの得点とし、$X = \max(X_1, X_2)$ とする（$X_1 = X_2$ のときは定義中の 2 つの場合が一致するので、X はきちんと定義されている）。

小問 (i)

$r = 0, 1, \ldots, N$ に対し、事象 $\{X \leq r\}$ は事象 $\{X_1 \leq r \text{ かつ} X_2 \leq r\}$ である。それぞれの確率は r/N であり、X_1 と X_2 が独立であるからこれらの事象も独立である。よって

$$\mathrm{P}(X \leq r) = \frac{r}{N} \cdot \frac{r}{N} = \frac{r^2}{N^2}.$$

これは $r = 0$ のときも成り立ち、値は 0 となる。したがって $r = 1, 2, \ldots, N$ に対し

$$\mathrm{P}(X = r) = \mathrm{P}(X \leq r) - \mathrm{P}(X \leq r - 1) = \frac{r^2 - (r-1)^2}{N^2} = \frac{2r - 1}{N^2},$$

となり、題意が示された。（確認として、$\sum_{r=1}^{N}(2r - 1) = N^2$ であるから、これらの和は 1 になる。）

小問 (ii)

与えられた 2 つの和の公式を用いると

$$\begin{aligned}\mu = \sum_{r=1}^{N} r\,\frac{2r-1}{N^2} &= \frac{1}{N^2}\left(2\sum_{r=1}^{N} r^2 - \sum_{r=1}^{N} r\right)\\ &= \frac{1}{N^2}\left(\frac{N(N+1)(2N+1)}{3} - \frac{N(N+1)}{2}\right)\\ &= \frac{N+1}{N}\cdot\frac{2(2N+1)-3}{6} = \frac{(N+1)(4N-1)}{6N}.\end{aligned}$$

$N = 100$ のとき

$$\mu = \frac{101\times 399}{600} = \frac{40\,299}{600} = 67.165,$$

となり、題意が示された。

小問 (iii)

小問 (i) より $\mathrm{P}(X \le m) = m^2/N^2$ であり、$\mathrm{P}(X \ge m) = 1 - \mathrm{P}(X \le m-1) = 1-(m-1)^2/N^2$ である。中央値に関する 2 つの条件は

$$\mathrm{P}(X \le m) \ge \tfrac{1}{2} \iff m^2 \ge \tfrac{1}{2}N^2 \iff m \ge \frac{N}{\sqrt{2}},$$
$$\mathrm{P}(X \ge m) \ge \tfrac{1}{2} \iff (m-1)^2 \le \tfrac{1}{2}N^2 \iff m-1 \le \frac{N}{\sqrt{2}},$$

となる。ここでは正の平方根をとったが、$m \ge 1$ より m と $m-1$ はともに非負であるから、これは正当である。よって m は

$$\frac{N}{\sqrt{2}} \le m \le \frac{N}{\sqrt{2}} + 1.$$

を満たす整数でなければならない。

考えてみよう 長さ 1 の区間には整数が 2 つ含まれることがある。中央値が 2 つになることはあり得るだろうか？

それが起こるのは両端がともに整数である場合に限られ、そのためには $N/\sqrt{2}$ が整数でなければならない。しかし $N/\sqrt{2} = k$ とすると $\sqrt{2} = N/k$ となり、$\sqrt{2}$ が有理数（rational number）になってしまうので、これは不可能である。よって 2 つの不等式はいずれも等号が成り立たず、長さ 1 の開区間にはちょうど 1 つの整数が含まれる。したがって

$$m = \left\lceil \frac{N}{\sqrt{2}} \right\rceil,$$

すなわち m は $N/\sqrt{2}$ より大きい最小の整数である。

$N=100$ のとき、$N/\sqrt{2}=50\sqrt{2}\approx 70.71$ であるから $m=71$ である。直接確かめると、$71^2=5041\geq 5000=\frac{1}{2}\times 100^2$ かつ $70^2=4900\leq 5000$ であるから、2 つの条件はともに成り立つ。

小問 (iv)

小問 (iii) より、ある $0<\theta<1$ を用いて $m=N/\sqrt{2}+\theta$ と表せるので

$$\frac{m}{N}=\frac{1}{\sqrt{2}}+\frac{\theta}{N}\to\frac{1}{\sqrt{2}}\quad(N\to\infty.$$

小問 (ii) より

$$\frac{\mu}{N}=\frac{(N+1)(4N-1)}{6N^2}=\frac{1}{6}\left(1+\frac{1}{N}\right)\left(4-\frac{1}{N}\right)\to\frac{4}{6}=\frac{2}{3}.$$

したがって、N が大きいとき

$$\frac{\mu}{m}=\frac{\mu/N}{m/N}\approx\frac{2/3}{1/\sqrt{2}}=\frac{2\sqrt{2}}{3}.$$

検算

- $N=2$ の場合：同様に確からしい 4 つの組のうち、$X=1$ となるのは 1 回、$X=2$ となるのは 3 回であり、小問 (i) の $\frac{1}{4}$ と $\frac{3}{4}$ に一致する。また $\mu=\frac{7}{4}$ であり、$\frac{3\times 7}{12}$ に一致する。
- $N\leq 40$ についてコンピュータで N^2 個の組をすべて列挙すると、小問 (i) と (ii) が厳密に正しいことが確かめられる。また $N\leq 2000$ について、すべての整数を中央値の定義に照らして調べると、中央値は常にちょうど 1 つであり、$\lceil N/\sqrt{2}\rceil$ に等しい。
- $N=10^5$ のとき $\mu/m\approx 0.942812$ であり、$2\sqrt{2}/3\approx 0.942809$ とよく一致する。

よくある間違い

- $\mathrm{P}(X=r)$ を $2\times\frac{r}{N}\cdot\frac{1}{N}$ と数えてしまう。これでは $X_1=X_2=r$ の場合を二重に数えている。正しい個数は $2(r-1)+1=2r-1$ である。
- $\mathrm{P}(X\geq m)=1-m^2/N^2$ と書いてしまう。これは $\mathrm{P}(X>m)$ であって、$\mathrm{P}(X\geq m)$ ではない。
- $N=100$ のとき、70.71 を切り上げずに四捨五入して $m=70$ と答えてしまう。

STEP のその先へ

X は順序統計量（order statistic）の最も簡単な例、すなわち標本の最大値である。$N \to \infty$ のとき、X/N は $[0,1]$ 上の確率密度（density）$2x$ をもつ連続確率変数（random variable）に近づく。この確率変数の平均は $\frac{2}{3}$、中央値は $1/\sqrt{2}$ であり、まさにこれが小問 (iv) の比が現れる理由である。より大きな標本の順序統計量を考えると、ベータ分布（beta distribution）に行き着く。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）は、順序統計量をわかりやすく扱っている。

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。