解答例

# STEP II 2022、第 12 問

確率・統計 ・ 平均値・中央値・最頻値

**STEP II 2022、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 2022 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/696587-2022-past-papers.zip

## 鍵となる考え方

すべての出発点は累積分布関数（cumulative distribution function）F であり、これは今の場合、単純な多項式である。F は $[0,1]$ 上で増加するから、数 $a$ が中央値（median）より小さいのは $\mathrm{F}(a) < \frac{1}{2}$ のとき、かつそのときに限る。したがって、平均（mean）や最頻値（mode）を中央値と比較するために、中央値そのものを求める必要はまったくない。平均または最頻値における F の値を計算し、それを $\frac{1}{2}$ と比べればよいのである。

## 解答

### 小問 (i)

確率の総和は 1 であるから、

$$1 = k\int_0^1 \left(x^n - x^{n+1}\right)\mathrm{d}x = k\left(\frac{1}{n+1} - \frac{1}{n+2}\right) = \frac{k}{(n+1)(n+2)},$$

よって、示すべきとおり $k = (n+1)(n+2)$ である。さらに

$$\mu = k\int_0^1 \left(x^{n+1} - x^{n+2}\right)\mathrm{d}x = (n+1)(n+2)\cdot\frac{1}{(n+2)(n+3)} = \frac{n+1}{n+3}.$$

### 小問 (ii)

$0 \le x \le 1$ に対して、

$$\mathrm{F}(x) = (n+1)(n+2)\left(\frac{x^{n+1}}{n+1} - \frac{x^{n+2}}{n+2}\right) = x^{n+1}((n+2) - (n+1)x).$$

$(0,1)$ 上で $f > 0$ であるから、F はそこで狭義単調増加（strictly increasing）であり、中央値 $m$ は $F(m) = \frac{1}{2}$ を満たす唯一の点である。したがって

$$\mu < m \iff F(\mu) < \tfrac{1}{2}.$$

ここで $\mu = \frac{n+1}{n+3}$ とすると、

$$(n+2) - (n+1)\mu = \frac{(n+2)(n+3) - (n+1)^2}{n+3} = \frac{3n+5}{n+3},$$

であるから、

$$F(\mu) = \left(\frac{n+1}{n+3}\right)^{n+1} \frac{3n+5}{n+3}.$$

よって条件 $F(\mu) < \frac{1}{2}$ は、両辺に正の量 $2\left(\frac{n+3}{n+1}\right)^{n+1}$ を掛けると次のようになる。

$$\frac{2(3n+5)}{n+3} < \left(\frac{n+3}{n+1}\right)^{n+1}.$$

$\frac{2(3n+5)}{n+3} = \frac{6(n+3)-8}{n+3} = 6 - \frac{8}{n+3}$ かつ $\frac{n+3}{n+1} = 1 + \frac{2}{n+1}$ であるから、これはまさに与えられた不等式（inequality）である。したがって、その不等式が成り立つときには必ず、$\mu$ は中央値より小さい。

この不等式を証明するために $N = n+1$ とおくと、$N \geq 3$ は整数である。$\left(1 + \frac{2}{N}\right)^N$ の二項展開（binomial expansion）の各項はすべて正であり、項数は $N+1 \geq 4$ であるから、和は最初の 4 項の和以上である。

$$\begin{aligned}\left(1 + \frac{2}{N}\right)^N &\geq 1 + N \cdot \frac{2}{N} + \frac{N(N-1)}{2} \cdot \frac{4}{N^2} + \frac{N(N-1)(N-2)}{6} \cdot \frac{8}{N^3} \\ &= 3 + 2\left(1 - \frac{1}{N}\right) + \frac{4}{3}\left(1 - \frac{3}{N} + \frac{2}{N^2}\right) \\ &= \frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2}.\end{aligned}$$

**考えてみよう** この 4 項の和は下界にすぎない。これが $6 - \frac{8}{N+2}$ より大きいことはどのように示せばよいだろうか。

差をとり、正の分母 $3N^2(N+2)$ を払う。

$$\begin{aligned}&3N^2(N+2)\left[\frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2} - 6 + \frac{8}{N+2}\right] \\ &\quad = N^2(N+2) - 18N(N+2) + 8(N+2) + 24N^2 \\ &\quad = N^3 + 8N^2 - 28N + 16.\end{aligned}$$

$N \geq 3$ のとき $N^2 + 8N - 28 \geq 9 + 24 - 28 = 5 > 0$ であるから、$N^3 + 8N^2 - 28N + 16 = N(N^2 + 8N - 28) + 16 > 0$ である。したがって

$$6 - \frac{8}{n+3} < \frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2} \leq \left(1 + \frac{2}{n+1}\right)^{n+1},$$

となり、与えられた不等式はすべての整数 $n > 1$ について成り立つ。よって $X$ の中央値は $\mu$ より大きい。

### 小問 (iii)

$(0,1)$ 上で

$$f'(x) = k\left(nx^{n-1} - (n+1)x^n\right) = kx^{n-1}(n-(n+1)x),$$

であり、これは $x < \frac{n}{n+1}$ で正、$x > \frac{n}{n+1}$ で負である。したがって f は増加した後に減少し、最頻値は $M = \frac{n}{n+1}$ である。

小問 (ii) と同様に、$M > m$ となるのは $F(M) > \frac{1}{2}$ のとき、かつそのときに限る。$(n+2) - (n+1)M = (n+2) - n = 2$ であるから、

$$F(M) = 2\left(\frac{n}{n+1}\right)^{n+1},$$

となり、$\left(\frac{n}{n+1}\right)^{n+1} > \frac{1}{4}$、すなわち次が成り立てばよい。

$$\left(1+\frac{1}{n}\right)^{n+1} < 4.$$

**考えてみよう** $n$ が 2 以上の任意の値をとりうるとき、$\left(1+\frac{1}{x}\right)^{x+1}$ についての与えられた事実はどのように役立つだろうか。

この関数は正の $x$ に対して減少するから、任意の整数 $n \geq 2$ における値は $x = 2$ における値以下である。

$$\left(1+\frac{1}{n}\right)^{n+1} \leq \left(\frac{3}{2}\right)^3 = \frac{27}{8} < 4.$$

したがって $F(M) > \frac{1}{2}$ であり、$X$ の最頻値は中央値より大きい。

以上をまとめると、すべての $n > 1$ について 平均 < 中央値 < 最頻値 が成り立つ。

## 検算

- $n = 2$ とすると、平均は $\frac{3}{5} = 0.6$、最頻値は $\frac{2}{3} \approx 0.667$ である。$4x^3 - 3x^4 = \frac{1}{2}$ を数値的に解くと中央値は $\approx 0.6143$ となり、ちょうど両者の間にある。また $F(0.6) = 0.4752 < \frac{1}{2}$、$F(\frac{2}{3}) = \frac{16}{27} \approx 0.593 > \frac{1}{2}$ である。
- $n = 2$（$N = 3$）のとき展開はちょうど 4 項からなるので、下界は等号で成り立つ。両辺はともに $\frac{125}{27} \approx 4.63$ であり、$6 - \frac{8}{5} = 4.4$ を十分に上回っている。
- 平均 < 中央値 < 最頻値という大小関係は、2 から 79 までのすべての $n$ についてコンピュータで確認した。

## よくある間違い

- 中央値を具体的に求めようとすること。方程式 $x^{n+1}(n+2-(n+1)x) = \frac{1}{2}$ にはきれいな解はなく、また解を求める必要もない。
- 不等号の向きを取り違えること。F は増加関数であるから、$\mu < m$ は $\mathrm{F}(\mu) < \frac{1}{2}$ に対応する。
- 二項展開の後ろの項を捨ててよい理由（それらはすべて正であること）を述べ忘れること。

## STEP のその先へ

この確率密度はベータ分布（Beta distribution）Beta$(n+1, 2)$ であり、$[0,1]$ 上の分布族のなかで最も重要なものの一つで、ベイズ統計学（Bayesian statistics）の中心をなす。平均 < 中央値 < 最頻値という大小関係は、左側に長い裾をもつ分布（負の歪度（negative skew））に典型的なものであるが、普遍的な法則ではない。だからこそ、ここで証明する必要があったのである。小問 (iii) の減少関数は $x \to \infty$ のとき $e$ に収束（converge）する。これは古典的な極限であり、解析学できちんと学ぶことになる。Blitzstein と Hwang による *Introduction to Probability*（確率論入門）には、ベータ分布についての優れた解説がある。

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。