解答例

STEP II 2019、第 12 問

確率・統計 ・ 連続分布と分位数

> **STEP II 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 II の問題用紙にある：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

鍵となる考え方

ここでの累積分布関数（cumulative distribution function）は $[0,1]$ 上で単に x^n であるから、どの分位数（quantile）も正確に書き下すことができる。分布のうち割合 p がそれ以下にあるような値は $p^{1/n}$ である。すると平均（mean）と分位数の比較はべき乗どうしの比較となり、両辺を n 乗すれば $\left(1+\frac{1}{n}\right)^n$ に関する主張に帰着する。これはまさに二項展開（binomial expansion）が評価するのに適した量である。

解答

小問 (i)

確率密度関数（probability density function）をかけて積分すると、

$$\mu = \mathrm{E}(X) = \int_0^1 x \cdot nx^{n-1}\,\mathrm{d}x = \int_0^1 nx^n\,\mathrm{d}x = \frac{n}{n+1},$$
$$\mathrm{E}(X^2) = \int_0^1 nx^{n+1}\,\mathrm{d}x = \frac{n}{n+2}.$$

したがって

$$\sigma^2 = \mathrm{E}(X^2) - \mu^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n\left[(n+1)^2 - n(n+2)\right]}{(n+1)^2(n+2)}$$
$$= \frac{n\left[n^2+2n+1-n^2-2n\right]}{(n+1)^2(n+2)} = \frac{n}{(n+1)^2(n+2)},$$

となり、示すべき式が得られた。

小問 (ii)

$0 \le x \le 1$ に対して、分布関数（distribution function）は

$$\mathrm{F}(x) = \mathrm{P}(X \le x) = \int_0^x nt^{n-1}\,\mathrm{d}t = x^n.$$

である。$n = 2$ のとき $\mathrm{F}(x) = x^2$ である。下側四分位数（lower quartile）q_1 は $q_1^2 = \frac{1}{4}$ を満たし、上側四分位数（upper quartile）q_3 は $q_3^2 = \frac{3}{4}$ を満たす。どちらも $[0,1]$ に含まれるので、正の平方根をとって

$$q_1 = \tfrac{1}{2}, \qquad q_3 = \tfrac{\sqrt{3}}{2}, \qquad q_3 - q_1 = \frac{\sqrt{3}-1}{2}.$$

を得る。小問 (i) より $\sigma^2 = \dfrac{2}{9\cdot 4} = \dfrac{1}{18}$ であるから、$\sigma = \dfrac{1}{3\sqrt{2}} = \dfrac{\sqrt{2}}{6}$ であり、$2\sigma = \dfrac{\sqrt{2}}{3}$ である。

示したいのは $\dfrac{\sqrt{3}-1}{2} < \dfrac{\sqrt{2}}{3}$ である。両辺に 6 をかければ、これは

$$3\sqrt{3} - 3 < 2\sqrt{2} \iff 3\sqrt{3} < 3 + 2\sqrt{2}.$$

と同値である。最後の不等式の両辺はともに正であるから、これは両辺を 2 乗した不等式と同値である：

$$27 < 9 + 12\sqrt{2} + 8 = 17 + 12\sqrt{2} \iff 10 < 12\sqrt{2},$$

そしてこれは、$\sqrt{2} > 1$ より $12\sqrt{2} > 12 > 10$ となるので正しい。どの段階も同値変形であったから、この連鎖を逆にたどることで、四分位範囲（interquartile range）は 2σ より小さいことがわかる。

考えてみよう ここで両辺を 2 乗してよいのはなぜか。また、一般には 2 乗してはいけないのはなぜか。

2 乗によって不等式 $a < b$ が保たれるのは、a と b が非負のときに限られる（たとえば $-3 < 2$ であるが $9 > 4$ である）。そのため、引き算を含んだままの $3\sqrt{3} - 3 < 2\sqrt{2}$ を 2 乗するのではなく、まず両辺が明らかに正である $3\sqrt{3} < 3 + 2\sqrt{2}$ の形に変形したのである。

小問 (iii)

二項定理（binomial theorem）より

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2}\,x^2 + \cdots,$$

であり、第 $(k+1)$ 項は

$$\binom{n}{k} x^k = \frac{n(n-1)\cdots(n-k+1)}{k!}\,x^k.$$

である。

中央値と下側四分位数。 $\mathrm{F}(x) = x^n$ であるから、中央値（median）m は $m^n = \frac{1}{2}$ を満たし、下側四分位数 q は $q^n = \frac{1}{4}$ を満たす。よって

$$m = \left(\tfrac{1}{2}\right)^{1/n}, \qquad q = \left(\tfrac{1}{4}\right)^{1/n}.$$

である。μ、m、q はいずれも正であり、$t \mapsto t^n$ は $t > 0$ で狭義単調増加（strictly increasing）であるから、これらの大小は n 乗どうしを比べることで比較できる。また

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

である。$A = \left(1+\frac{1}{n}\right)^n$ とおく。すると

$$\mu < m \iff \mu^n < \tfrac{1}{2} \iff A > 2, \qquad \mu > q \iff \mu^n > \tfrac{1}{4} \iff A < 4.$$

となる。

μ は中央値より小さい。 展開式に $x = \frac{1}{n}$ を代入する。どの項も非負であるから、A は最初の 3 項の和以上である：

$$A \geq 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{2} \cdot \frac{1}{n^2} = 2 + \frac{n-1}{2n} > 2,$$

ここで最後の不等式が真の不等号になるのは $n > 1$ だからである。（$n \geq 2$ のとき、第 3 項は確かに展開式に現れる。）したがって $\mu < m$ である。

μ は下側四分位数より大きい。 $x = \frac{1}{n}$ のとき、第 $(k+1)$ 項は

$$\binom{n}{k}\frac{1}{n^k} = \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \leq \frac{1}{k!},$$

である。k 個の分数 $\frac{n-j}{n}$（$0 \leq j \leq k-1$）はどれも 0 と 1 の間にあるからである。（$k = 0$ のとき、この項は $1 = \frac{1}{0!}$ である。）$k = 0, 1, \ldots, n$ について和をとると、

$$A \leq 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots < 4,$$

となる。中央の不等式は無限級数がさらに正の項を含むことから成り立ち、最後の不等式は仮定してよいとされた結果である。したがって $A < 4$ であり、$\mu > q$ である。

以上を合わせると、$n > 1$ を満たすすべての整数について、μ は下側四分位数より大きく、中央値より小さい。

考えてみよう n 乗どうしを比べることで、μ と m の大小が決まるのはなぜか。

正の数 a と b について、t^n は $t>0$ で狭義単調増加であるから、$a<b$ であることと $a^n<b^n$ であることは同値である。ここで扱う 3 つの数はすべて正であるから、n 乗しても何も失われず、しかもそうすることで m と q に含まれる扱いにくい n 乗根を取り除くことができる。

検算

- $n=2$ のとき：$\mu=\frac{2}{3}$ であり、$\sigma^2=\frac{2}{9\cdot 4}=\frac{1}{18}$ は $\mathrm{E}(X^2)-\mu^2=\frac{1}{2}-\frac{4}{9}$ と一致する。
- 数値的には、$n=2$ のとき四分位範囲は 0.366、$2\sigma=0.471$ である。
- $n=2$ のとき 3 つの値は $q=0.5$、$\mu\approx 0.667$、$m\approx 0.707$ であり、$n=10$ のときは $q\approx 0.871$、$\mu\approx 0.909$、$m\approx 0.933$ である。大小関係 $q<\mu<m$ は、2 から 2000 までのすべての n についてコンピュータで確認した。
- $n\to\infty$ のとき $A=\left(1+\frac{1}{n}\right)^n\to \mathrm{e}\approx 2.718$ であり、これは 2 と 4 の間に余裕をもって収まっている。

よくある間違い

- 両辺が正であることを確かめずに、$3\sqrt{3}-3<2\sqrt{2}$ をそのまま 2 乗してしまう。
- 逆数（reciprocal）をとるときに不等号の向きを間違える：$\mu^n<\frac{1}{2}$ に対応するのは $A>2$ であって、$A<2$ ではない。
- 第 $(k+1)$ 項が $\frac{1}{k!}$ 以下であるのではなく、$\frac{1}{k!}$ に等しいと主張してしまう。

STEP のその先へ

量 $\left(1+\frac{1}{n}\right)^n$ は n とともに増加して極限 e に収束（converge）する。そして $\sum 1/k!$ と項ごとに比べて示した評価は、$\mathrm{e}=\sum_{k\geq 0}1/k!$ を示すための第一歩であり、これは解析学（analysis）の最初の講義で扱われる標準的な結果である。この分布そのものはベータ分布（Beta distribution）$\mathrm{Beta}(n,1)$ である。また、$\mathrm{P}(\max\leq x)=x^n$ であることから、これは $[0,1]$ 上の n 個の独立な一様確率変数（uniform random variables）の最大値の分布でもある。これら両方の話題についてさらに知るには、Blitzstein と Hwang による *Introduction to Probability*（確率論入門）、および Abbott による *Understanding*

Analysis（解析学の理解）を参照されたい。

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。