解答例

STEP II 2011、第 13 問

確率・統計 ・ 歪度の尺度

STEP II 2011、問 13

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/11-s2-q13

鍵となる考え方

歪度（skewness）とは分布の偏りのことであり、ここで扱う 3 つの指標はそれを 3 通りの方法でとらえる。すなわち、3 次の中心モーメント（central moment）によるもの、2 つの十分位数（decile）の間で中央値（median）がどこに位置するかによるもの、そして平均（mean）と中央値の差によるものである。$[0,1]$ 上の確率密度関数（probability density function）$2x$ については、すべてを具体的に計算できる。モーメントは 1 行で済む積分であり、累積分布関数（cumulative distribution function）は x^2 なので、その逆関数は単なる平方根である。実質的な作業は根号を含む数の注意深い計算と、最後にそれらをすっきりと比較する方法だけである。

解答

歪度が測るもの。歪度は、分布の中心に関する**非対称性**（asymmetry）を測る。対称な分布の歪度は 0 である。長い裾が右側にあれば歪度は正であり、長い裾が左側にあれば負である。

小問 (i)

μ が定数であることに注意して、3 乗を展開し、期待値の線形性（linearity of expectation）を用いる。

$$\begin{aligned}\mathrm{E}((X-\mu)^3) &= \mathrm{E}(X^3) - 3\mu\,\mathrm{E}(X^2) + 3\mu^2\,\mathrm{E}(X) - \mu^3 \\ &= \mathrm{E}(X^3) - 3\mu(\sigma^2+\mu^2) + 3\mu^3 - \mu^3 \\ &= \mathrm{E}(X^3) - 3\mu\sigma^2 - \mu^3,\end{aligned}$$

ここで $\mathrm{E}(X) = \mu$ と $\mathrm{E}(X^2) = \sigma^2 + \mu^2$（分散（variance）の通常の公式を変形したもの）を用いた。これを σ^3 で割れば、γ について求める式が得られる。

確率密度関数 $2x$ をもつ分布。$k \geq 0$ に対して

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

よって $\mu = \frac{2}{3}$、$\mathrm{E}(X^2) = \frac{1}{2}$、$\mathrm{E}(X^3) = \frac{2}{5}$ である。したがって

$$\sigma^2 = \tfrac{1}{2} - \tfrac{4}{9} = \tfrac{1}{18}, \qquad \sigma = \frac{1}{3\sqrt{2}}, \qquad \sigma^3 = \frac{1}{27 \cdot 2\sqrt{2}} = \frac{1}{54\sqrt{2}}.$$

分子は

$$\mathrm{E}(X^3) - 3\mu\sigma^2 - \mu^3 = \frac{2}{5} - 3 \cdot \frac{2}{3} \cdot \frac{1}{18} - \frac{8}{27} = \frac{2}{5} - \frac{1}{9} - \frac{8}{27} = \frac{54 - 15 - 40}{135} = -\frac{1}{135}.$$

ゆえに

$$\gamma = -\frac{1}{135} \cdot 54\sqrt{2} = -\frac{54\sqrt{2}}{135} = -\frac{2\sqrt{2}}{5},$$

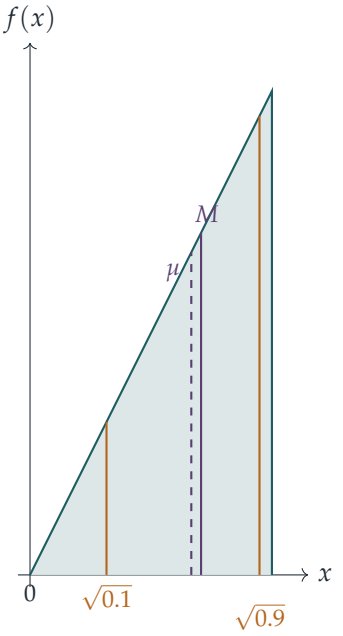
となり、示すべき式が得られた。負の符号は納得のいくものである。確率密度は $x = 1$ の近くに集中しており、左側の 0 に向かって長い裾が残るからである。

小問 (ii)

$0 \leq x \leq 1$ に対して、累積分布関数は

$$\mathrm{F}(x) = \int_0^x 2t \,\mathrm{d}t = x^2,$$

であり、これは $[0,1]$ 上で増加するので、$0 \leq p \leq 1$ に対してその逆関数は $\mathrm{F}^{-1}(p) = \sqrt{p}$ である。



十分位歪度。D の分子と分母に $\sqrt{10}$ を掛けると、根号は 1 つを残してすべて消える。

$$D=\frac{\sqrt{\frac{9}{10}}-2\sqrt{\frac{1}{2}}+\sqrt{\frac{1}{10}}}{\sqrt{\frac{9}{10}}-\sqrt{\frac{1}{10}}}=\frac{3-2\sqrt{5}+1}{3-1}=\frac{4-2\sqrt{5}}{2}=2-\sqrt{5},$$

ここで $\sqrt{10}\cdot\sqrt{\frac{1}{2}}=\sqrt{5}$ を用いた。これが示すべきことであった。

ピアソン（Pearson）の歪度。中央値は $\mathrm{F}(M)=\frac{1}{2}$ を満たすので、$M=\frac{1}{\sqrt{2}}$ である。$\frac{1}{\sigma}=3\sqrt{2}$ を用いると

$$P=3\left(\frac{2}{3}-\frac{1}{\sqrt{2}}\right)\cdot 3\sqrt{2}=9\sqrt{2}\cdot\frac{2}{3}-9=6\sqrt{2}-9.$$

数値的には $D\approx -0.236$、$P\approx -0.515$、$\gamma\approx -0.566$ であり、大小関係は正しそうである。それでは、これを証明しよう。

> **考えてみよう** $2-\sqrt{5}$ や $6\sqrt{2}-9$ のような数を、電卓を使わずにどう比較すればよいだろうか？

両辺がともに正になるように変形してから 2 乗すればよい。正の数については、$a<b$ であることと $a^2<b^2$ であることは同値である。以下の各段階はすべて逆にたどれるので、最後の主張が正しければ最初の主張が証明される。

$D>P$。示すべきことは $2-\sqrt{5}>6\sqrt{2}-9$、すなわち $\sqrt{5}+6\sqrt{2}<11$ である。両辺は

正なので、これは次と同値である。

$$\begin{aligned}(\sqrt{5}+6\sqrt{2})^2 = 77+12\sqrt{10} < 121 &\iff 12\sqrt{10} < 44\\ &\iff 3\sqrt{10} < 11\\ &\iff 90 < 121,\end{aligned}$$

ここで最後の段階では、再び 2 つの正の数を 2 乗した。$90 < 121$ は正しいので、$D > P$ である。

$P > \gamma$。示すべきことは $6\sqrt{2} - 9 > -\frac{2\sqrt{2}}{5}$ である。5 を掛けて整理すると、これは $32\sqrt{2} > 45$ となる。両辺は正なので、これは $32^2 \cdot 2 > 45^2$、すなわち $2048 > 2025$ と同値であり、これは正しい。よって $P > \gamma$ である。

以上より $D > P > \gamma$ である。3 つの指標はいずれも負であり、分布が左に歪んでいるという点では一致するが、その程度については食い違っている。

検算

- (i) の恒等式は、任意の対称な分布に対して 0 を与える。たとえば、X が等しい確率で ± 1 をとるならば、$\mu = 0$、$\sigma = 1$、$\mathrm{E}(X^3) = 0$ である。
- 数値的に、中点則（midpoint rule）によって $\mathrm{E}\big((X-\mu)^3\big)/\sigma^3$ を計算すると -0.56569 となり、$-2\sqrt{2}/5$ と一致する。
- U を $[0,1]$ 上の一様分布（uniform distribution）に従う確率変数として $X = \sqrt{U}$（これは $\mathrm{F}(x) = x^2$ をもつ）をシミュレーションすると、標本中央値は 0.707 に近く、十分位数は 0.316 と 0.949 に近くなり、図のとおりである。
- 平均 $\frac{2}{3}$ は中央値 $\frac{1}{\sqrt{2}} \approx 0.707$ の左にある。これは左に歪んだ分布に典型的なことであり、$P < 0$ となる理由でもある。

よくある間違い

- $(\frac{1}{18})^{3/2}$ とすべきところを $\sigma^3 = (\frac{1}{18})^3$ と書いてしまう。3 乗するのは分散ではなく標準偏差（standard deviation）である。
- 一方の辺が負になりうるのに不等式を 2 乗してしまう。まず両辺が正になるように変形すること。
- γ の式や P の分母で、$\sigma^2 = \frac{1}{18}$ を用いるべきところに $\mathrm{E}(X^2) = \frac{1}{2}$ を用いてしまう。

STEP のその先へ

量 $\mathrm{E}\big((X-\mu)^3\big)$ は 3 次の**中心モーメント**であり、(i) の展開はモーメントと**キュムラント**（cumulant）の関係の入口である。キュムラントは、独立な確率変数の和に対してうまく足し合わされるという性質をもつ。D のような分位点に基づく指標は、**頑健**（robust）であるために統計学で重宝される。極端な観測値が 1 つあっても十分位数はほとんど動かないが、3 次のモーメントは劇的に変わりうるのである。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、モーメントと積率母関数（moment generating function）をちょうどよい水準で展開している。

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。