解答例

STEP II 2023、第 12 問

確率・統計 ・ 最大値の分布

> **STEP II 2023、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 2023 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/707102-2023-past-papers.zip

鍵となる考え方

いくつかの数の最大値が t 以下であるのは、それらの**すべて**が t 以下であるとき、かつそのときに限る。したがって、Y の累積分布関数（cumulative distribution function）は、1 つの X_i の累積分布関数のべき乗になる。残りはすべて、密度関数（density）ではなく累積分布関数を用いて考えることから導かれる。中央値（median）は $\mathrm{P}(Y \le m) = \frac{1}{2}$ を直接解けば得られる。平均（mean）は、部分積分（integration by parts）によって積分の中に密度関数ではなく分布関数が現れるようにすると、すっきりと求まる。

解答

小問 (i)

$F(t) = \mathrm{P}(X_1 \le t)$ とおく。$0 \le t \le \pi$ に対して

$$F(t) = \int_0^t \tfrac{1}{2} \sin x \, \mathrm{d}x = \tfrac{1}{2}(1 - \cos t).$$

事象 $\{Y \le t\}$ が起こるのは、$X_1 \le t$, $X_2 \le t$, ..., $X_n \le t$ がすべて起こるとき、かつそのときに限る。実際、最大のものが t 以下ならば他のものもすべてそうであり、逆も成り立つ。X_i は独立（independent）かつ同一分布（identically distributed）であるから、この共通部分の確率は個々の確率の積になる：

$$\mathrm{P}(Y \le t) = \mathrm{P}(X_1 \le t)\,\mathrm{P}(X_2 \le t) \cdots \mathrm{P}(X_n \le t) = [\mathrm{P}(X_1 \le t)]^n.$$

よって、Y の累積分布関数を G と書くと

$$G(t) = \frac{(1 - \cos t)^n}{2^n} \qquad (0 \le t \le \pi),$$

であり、$t < 0$ では $G(t) = 0$、$t > \pi$ では $G(t) = 1$ である。連鎖律（chain rule）を用いて微分すると、Y の確率密度関数（probability density function）は

$$g(t) = \frac{n\,(1-\cos t)^{n-1}\sin t}{2^n} \quad \text{ただし} 0 \le t \le \pi,$$

であり、それ以外の範囲では o である。

小問 (ii)

中央値 $m = m(n)$ は $G(m) = \frac{1}{2}$ を満たす：

$$\left(\frac{1-\cos m}{2}\right)^n = \frac{1}{2} \iff \frac{1-\cos m}{2} = 2^{-1/n}$$
$$\iff \cos m = 1 - 2^{1-1/n},$$

ここで、$(1-\cos m)/2 \ge 0$ であるから正の実数の n 乗根をとった。m は $[0, \pi]$ に属さなければならず、この範囲で余弦は一対一（one-to-one）であるから

$$m(n) = \arccos(1 - 2^{1-1/n}).$$

n が増加すると $1/n$ は減少するので、$2^{1-1/n}$ は（2 に向かって）増加し、$1 - 2^{1-1/n}$ は（-1 に向かって）減少する。arccos は減少関数（decreasing function）であるから、$m(n)$ は n とともに**増加**し、$n \to \infty$ のとき $m(n) \to \arccos(-1) = \pi$ となる。（$n = 1$ のときは $m(1) = \arccos 0 = \pi/2$ となり、$\frac{1}{2}\sin x$ が $\pi/2$ に関して対称であることから当然要求される値と一致する。）これは納得のいく結果である。ますます多くの値の最大値をとれば、それは範囲の上端へと押しやられるのである。

小問 (iii)

定義により $\mu(n) = \int_0^\pi t\,g(t)\,\mathrm{d}t$ である。$g = G'$ であるから、$u = t$、$\mathrm{d}v = g(t)\,\mathrm{d}t$、$v = G(t)$ として部分積分を行う：

$$\mu(n) = \Big[\,t\,G(t)\Big]_0^\pi - \int_0^\pi G(t)\,\mathrm{d}t$$
$$= \pi \cdot 1 - 0 - \frac{1}{2^n}\int_0^\pi (1-\cos x)^n\,\mathrm{d}x,$$

ここで $G(\pi) = (\frac{1+1}{2})^n = 1$ を用いた。これが求める式である。

考えてみよう $\int t\,(1-\cos t)^{n-1}\sin t\,\mathrm{d}t$ を直接計算しようとせず、部分積分を用いるのはなぜか。

目標の式には $(1-\cos x)^n$ が含まれている。これは $2^n G(x)$、すなわち密度関数ではなく**分布**関数である。部分積分はまさに $t\,g(t)$ を $G(t)$ に取り替える操作であり、同時に厄介な因子 t も取り除いてくれる。

(a) $h(x)=\frac{1}{2}(1-\cos x)$ とおくと、$[0,\pi]$ 上で $0\le h(x)\le 1$ であり、$0<x<\pi$ では $0<h(x)<1$ である。式は $\mu(n)=\pi-\int_0^\pi h(x)^n\,\mathrm{d}x$ と書けるので

$$\mu(n+1)-\mu(n)=\int_0^\pi (h(x)^n-h(x)^{n+1})\,\mathrm{d}x=\int_0^\pi h(x)^n(1-h(x))\,\mathrm{d}x.$$

$0<x<\pi$ では $h(x)>0$ かつ $1-h(x)>0$ であるから、被積分関数は連続で真に正である。よってこの積分は真に正であり、すべての $n\ge 1$ に対して $\mu(n+1)>\mu(n)$ が成り立つ。すなわち $\mu(n)$ は n とともに増加する。

(b) $n=2$ として展開し、$\cos^2 x=\frac{1}{2}(1+\cos 2x)$ を用いると

$$\int_0^\pi (1-\cos x)^2\,\mathrm{d}x=\int_0^\pi \Big(1-2\cos x+\tfrac{1}{2}+\tfrac{1}{2}\cos 2x\Big)\mathrm{d}x=\tfrac{3}{2}\pi,$$

となる。$\cos x$ と $\cos 2x$ はいずれも $[0,\pi]$ 上で積分すると零になるからである。よって

$$\mu(2)=\pi-\tfrac{1}{4}\cdot\tfrac{3}{2}\pi=\frac{5\pi}{8}.$$

小問 (ii) より $m(2)=\arccos(1-\sqrt{2})$ である。$5\pi/8$ と $m(2)$ はともに $[0,\pi]$ に属し、この範囲で余弦は狭義単調減少であるから

$$\mu(2)<m(2)\iff \cos\tfrac{5\pi}{8}>\cos m(2)=1-\sqrt{2}.$$

? **考えてみよう** $\cos\frac{5\pi}{8}$ の正確な値はどうすれば求められるか。

$\cos\frac{5\pi}{8}=\cos(\frac{\pi}{2}+\frac{\pi}{8})=-\sin\frac{\pi}{8}$ と書き、半角の公式（half-angle formula）$\sin^2\frac{\pi}{8}=\frac{1}{2}(1-\cos\frac{\pi}{4})=\frac{2-\sqrt{2}}{4}$ を用いる。$\sin\frac{\pi}{8}>0$ であるから

$$\cos\tfrac{5\pi}{8}=-\tfrac{1}{2}\sqrt{2-\sqrt{2}}.$$

したがって $-\frac{1}{2}\sqrt{2-\sqrt{2}}>1-\sqrt{2}$、すなわち $\sqrt{2}-1>\frac{1}{2}\sqrt{2-\sqrt{2}}$ を示せばよい。両辺はともに正であるから、これは両辺を 2 乗して 4 倍した不等式と同値である：

$$\begin{aligned}4(3-2\sqrt{2})>2-\sqrt{2}&\iff 10>7\sqrt{2}\\&\iff 100>98,\end{aligned}$$

ここで最後の段階では 2 つの正の数を 2 乗した。これは正しく、各段階はすべて同値変形であったから、$\cos\frac{5\pi}{8}>1-\sqrt{2}$ であり、したがって $\mu(2)<m(2)$ である。

検算

- $n=1$ のとき、$Y=X_1$ は $\pi/2$ に関して対称であり、実際 $\mu(1)=\pi-\frac{1}{2}\int_0^\pi(1-\cos x)\,\mathrm{d}x=\pi/2=m(1)$ となる。

- 数値的には $\mu(2) = 5\pi/8 \approx 1.9635$、$m(2) = \arccos(1-\sqrt{2}) \approx 1.9979$ であり、平均は中央値のわずかに下にある。1 から 7 までの各 n について 10^5 個の最大値をコンピュータでシミュレーションしたところ、どちらの式とも 0.01 以内で一致し（例えば $\mu(3) \approx 2.1598$、$m(3) \approx 2.1986$）、どちらの数列も増加していた。
- $n = 2$ のときの Y の密度関数は $\frac{1}{2}(1-\cos t)\sin t$ であり、これは左に歪んで（skewed）いる（0 に向かって長い裾をもつ）。左に歪んだ分布では、平均が中央値より小さくなるのが普通である。

よくある間違い

- $\mathrm{P}(Y \le t) = [F(t)]^n$ の根拠を述べ忘れる。「すべてが t 以下」という同値性と独立性の両方が必要である。
- 部分積分を行ったのに境界項（boundary term）$\left[tG(t)\right]_0^{\pi} = \pi$ を忘れる。先頭の π はここから来ている。
- 余弦が $[0, \pi]$ 上で**減少**することに注意せずに余弦を通して角を比較し、不等号の向きを逆にしてしまう。

STEP のその先へ

標本の最大値は**順序統計量**（order statistic）の一例であり、小問 (i) の技法によって任意の標本最大値の分布が得られる。恒等式 $\mathrm{E}[Y] = \pi - \int_0^{\pi} G(t)\,\mathrm{d}t$ は、非負の確率変数（random variable）に対する一般的な裾の積分公式（tail-integral formula）$\mathrm{E}[Y] = \int_0^{\infty} \mathrm{P}(Y > t)\,\mathrm{d}t$ の特別な場合であり、大学の確率論の講義の早い段階で出会うことになる。どちらの考え方も Grimmett and Welsh, *Probability: An Introduction*（確率論入門）で展開されている。

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。