解答例

STEP III 2025、第 11 問

確率・統計 ・ 最大値の分布

STEP III 2025、問 11

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 2025 年の STEP 過去問一式（zip ファイル）のうち、STEP III の問題用紙にある：

www.ocr.org.uk/Images/750486-2025-past-papers.zip

この問題は、*Beyond STEP* の第 12.25 節「**多数の中の最大値**」で詳しく解説されている：
https://beyondstep.keithmcnulty.org/ja/ch12/extremes.html

鍵となる考え方

X_i の最大値（maximum）が y 以下であるのは、**すべての** X_i が y 以下であるときであり、かつそのときに限る。そして独立性（independence）により、これは積に書き換えられる。あとは G から分位点（quantile）を読み取り、n が大きいときにそれらがどうなるかを調べることに尽きる。決め手となる近似（approximation）はただ一つ、$\alpha^{1/n} = \mathrm{e}^{(\ln\alpha)/n} \approx 1 + \frac{\ln\alpha}{n}$ である。これは n が大きいとき $(\ln\alpha)/n$ が小さいことによる。

解答

小問 (i)

$x \geq 0$ に対して

$$F(x) = \int_0^x \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = 1 - \mathrm{e}^{-\lambda x}.$$

$Y = \max(X_1, \ldots, X_n)$ であるから、事象 $\{Y \leq y\}$ は、すべての i について $X_i \leq y$ となる事象と同じである。X_i は互いに独立であるから、$y \geq 0$ に対して

$$G(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(X_1 \leq y) \cdots \mathrm{P}(X_n \leq y) = F(y)^n = \left(1 - \mathrm{e}^{-\lambda y}\right)^n,$$

となり、示すべき式が得られた。

小問 (ii)

Y は連続型確率変数（continuous random variable）であるから、$\mathrm{P}(Y < y) = \mathrm{P}(Y \leq y) = G(y)$ である。したがって $L(\alpha)$ に関する条件は $G(L(\alpha)) = \alpha$ となる。

$$\begin{aligned}\left(1 - \mathrm{e}^{-\lambda L}\right)^n = \alpha &\iff 1 - \mathrm{e}^{-\lambda L} = \alpha^{1/n} \qquad \text{（正の } n \text{ 乗根をとる）} \\ &\iff \mathrm{e}^{-\lambda L} = 1 - \alpha^{1/n} \\ &\iff L(\alpha) = -\frac{1}{\lambda}\ln\left(1 - \alpha^{1/n}\right).\end{aligned}$$

$0 < \alpha < 1$ より $0 < \alpha^{1/n} < 1$ であるから、この対数は定義される。

$U(\alpha)$ については、条件 $\mathrm{P}(Y > U) = \alpha$ は $G(U) = 1 - \alpha$ を意味するので、α を $1 - \alpha$ に置き換えて同じ手順をたどると

$$U(\alpha) = -\frac{1}{\lambda}\ln\left(1 - (1-\alpha)^{1/n}\right).$$

が得られる。$\alpha \leq \frac{1}{2} \leq 1 - \alpha$ であり G は増加関数であるから、$L(\alpha) \leq U(\alpha)$ であり、等号が成り立つのは $\alpha = \frac{1}{2}$ のときに限る。

小問 (iii)

$\alpha^{1/n} = \mathrm{e}^{(\ln\alpha)/n}$ と書く。α を固定して n を大きくすると $t = (\ln\alpha)/n$ は小さいので、$\mathrm{e}^t \approx 1 + t$ より

$$1 - \alpha^{1/n} \approx -\frac{\ln\alpha}{n} = \frac{1}{n}\ln\left(\frac{1}{\alpha}\right).$$

$\alpha < 1$ であるから、これは正であることに注意する。小問 (ii) に代入すると

$$\lambda L(\alpha) = -\ln\left(1 - \alpha^{1/n}\right) \approx -\ln\left(\frac{1}{n}\ln\frac{1}{\alpha}\right) = \ln(n) - \ln\left(\ln\frac{1}{\alpha}\right).$$

考えてみよう 近似式の対数をとってもよいのだろうか。$1 - \alpha^{1/n}$ 自体はごく小さいので、$1 - \alpha^{1/n}$ の誤差がそれに比べて大きくなってしまうおそれがある。

重要なのは**相対誤差**（relative error）である。$t = (\ln\alpha)/n$ とおくと $1 - \mathrm{e}^t = -t\left(1 + \frac{t}{2} + \cdots\right)$ であるから、真の値と近似値の比は $1 + O(1/n)$ である。対数をとると、$\lambda L(\alpha)$ の誤差は $\ln(1 + O(1/n))$ となり、これは 0 に収束する。したがって、この近似は n が大きいときによい近似であるというだけでなく、両辺の差が実際に 0 に収束するのである。

小問 (iv)

Y の中央値（median）m は $G(m) = \frac{1}{2}$ を満たすので、$m = L(\frac{1}{2})$（これは $U(\frac{1}{2})$ でもある）である。小問 (iii) より

$$\lambda m \approx \ln n - \ln(\ln 2).$$

$\ln(\ln 2)$ は定数（約 -0.37）であり $\ln n \to \infty$ であるから、n が増加すると中央値は無限大に発散する。

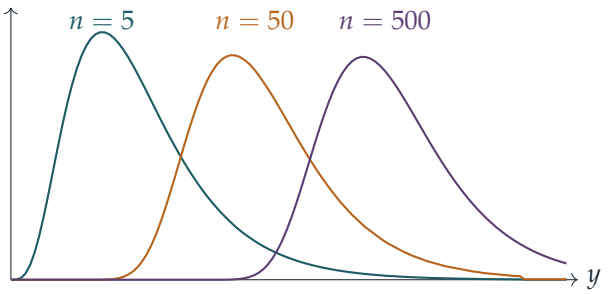
$U(\alpha)$ については、小問 (iii) で α を $1-\alpha$ に置き換えると

$$\lambda U(\alpha) \approx \ln n - \ln\left(\ln \frac{1}{1-\alpha}\right).$$

差をとると $\ln n$ の項が打ち消し合う。

$$U(\alpha) - L(\alpha) \approx \frac{1}{\lambda}\left[\ln\left(\ln \frac{1}{\alpha}\right) - \ln\left(\ln \frac{1}{1-\alpha}\right)\right].$$

（上の囲みで見たように）両方の近似の誤差は 0 に収束するので、区間の幅はこの値に収束する。この値は α と λ には依存するが n には依存しない。つまり n が大きくなると、Y の分布は広がることなく、$\frac{1}{\lambda}\ln n$ のように無限大へと移動していく。



図は $\lambda = 1$ のときの Y の確率密度関数（density）を示している。n が 10 倍になるごとに曲線は右へ約 $\ln 10 \approx 2.3$ だけ移動し、その形は一定のものに落ち着いていく。

小問 (v)

$\alpha = 0.05$ とする。このとき

$$\mathrm{P}\big(L(\alpha) \le Y \le U(\alpha)\big) = 1 - \mathrm{P}(Y < L) - \mathrm{P}(Y > U) = 1 - 2(0.05) = 0.9,$$

であるから、n の値によらず、Y はちょうど確率 0.9 で区間 $[L(0.05), U(0.05)]$ に含まれる。n が大きいとき、小問 (iv) よりこの区間の幅はおよそ

$$\frac{1}{\lambda}\left[\ln(\ln 20) - \ln\left(\ln \frac{1}{0.95}\right)\right].$$

である。ここで、与えられた二つの事実を用いる。

- $e^3 \approx 20$ であるから $\ln 20 \approx 3$ となり、$\ln(\ln 20) \approx \ln 3 \approx 1.1$ である。
- $t = -0.05$ として $\ln(1+t) \approx t$ を用いると $\ln \frac{1}{0.95} = -\ln(1-0.05) \approx 0.05 = \frac{1}{20}$ となるので、$\ln\left(\ln \frac{1}{0.95}\right) \approx \ln \frac{1}{20} = -\ln 20 \approx -3$ である。

よって幅はおよそ

$$\frac{1}{\lambda}(\ln 3 + 3) \approx \frac{4.1}{\lambda} \approx 4\lambda^{-1},$$

となり、示すべきことが得られた。

検算

- 小問 (ii) で $L(\alpha)$ を G に代入し直すと、ちょうど $\left(1-(1-\alpha^{1/n})\right)^n = \alpha$ となる。
- $\alpha = 0.05$、$n = 10^4$ のとき、$\lambda L(\alpha)$ の真の値は 8.11330 であり、近似式 $\ln n - \ln\ln 20$ の値は 8.11315 である。$n = 10$ のときはそれぞれ 1.35 と 1.21 であり、「十分大きい」（sufficiently large）という条件が本当に重要であることがわかる。
- 小問 (v) の極限における幅の正確な値は $4.067/\lambda$ であり、$4/\lambda$ に十分近い。また、5 個の指数分布に従う確率変数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、$G(y)$ と小数第 3 位まで一致した。

よくある間違い

- $G(y) = 1-(e^{-\lambda y})^n$ と書いてしまう。これは最大値ではなく**最小値**（minimum）の累積分布関数（cumulative distribution function）である。
- $P(Y > U) = \alpha$ を $G(U) = 1-\alpha$ ではなく $G(U) = \alpha$ としてしまう。
- 小問 (v) で $\alpha = 0.1$ を選んでしまう。区間の外側には両側それぞれに 0.1 ではなく 0.05 を残さなければならない。

STEP のその先へ

これは**極値理論**（extreme value theory）への最初の入り口である。小問 (iii) が実際に示しているのは、$\lambda Y - \ln n$ が極限分布をもち、$P(\lambda Y - \ln n \le x) \to \exp(-e^{-x})$ となることである。これはガンベル分布（Gumbel distribution）と呼ばれ、記録的な洪水、1 年間で最も強い突風、最大の保険金請求額などのモデルとして用いられる。適切に尺度を変換した最大値の極限分布は三つの分布族のいずれかに限られる、という驚くべき定理がある。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）は順序統計量（order statistics）を扱っており、最初に読む本として適している。

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。