12.25 多数の中の最大値

ワークシート

問題

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

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

考えてみよう　$X_1, \ldots, X_n$ の最大値が y 以下になるのは、ちょうどどのようなときか。

段階 1　最大値の分布

扱う内容：小問 (i)

考えてみよう　X_i のうち最大のものが y 以下になるのは、すべての X_i が y 以下になるときであり、またそのときに限る。独立性（independence）を用いよ。

段階 2 分位点

扱う内容：小問 (ii)

考えてみよう $\mathrm{G}(L)=\alpha$ を L について、$\mathrm{G}(U)=1-\alpha$ を U について解け。

段階 3 n が大きい場合

扱う内容：小問 (iii)

手を止めて、やってみよう $\alpha^{1/n}=e^{(\ln\alpha)/n}$ と書け。n が大きいとき指数は小さいので、$e^t\approx 1+t$ を用いよ。

段階 4　増大と広がり

扱う内容：小問 (iv)

考えてみよう　中央値（median）は $L(\frac{1}{2})$ である。$n \to \infty$ のとき、これはどうなるか。U と L の近似式の差をとれ。どの項が打ち消し合うか。

段階 5 90% **区間**

扱う内容：小問 (v)

考えてみよう $\alpha = 0.05$ とせよ。$\ln 20$ と $\ln \frac{1}{0.95}$ の近似値を求めよ。

背後にある数学

ガンベル分布

$P(\lambda Y - \ln n \le x) = (1 - \frac{e^{-x}}{n})^n \to \exp(-e^{-x})$ であるから、ずらした最大値 $\lambda Y - \ln n$ は、極限において分布関数（distribution function）$\exp(-e^{-x})$ をもつ**ガンベル分布**に従う。その平均はオイラー（Euler）の定数 $\gamma \approx 0.577$ である。この分布にはすでに出会っている。n 種類すべてのクーポンを集めるまでの時間（Section 9.11）は、n 個の（ほぼ独立な）待ち時間の最大値であり、それがおよそ $n \ln n + \gamma n$ となり、n のオーダーのガンベル型のゆらぎをもつのはこのためである。

極値理論

フィッシャー（Fisher）、ティペット（Tippett）、グネジェンコ（Gnedenko）による注目すべき定理（1928–1943）によれば、独立同分布な多数の確率変数の最大値は、適切にずらし尺度を変えると、極限において 3 通りの形しかとりえない。すなわち、ガンベル型（指数分布や正規分布のような軽い裾（tail）をもつ場合）、フレシェ（Fréchet）型（コーシー（Cauchy）分布のような重い裾をもつ場合）、ワイブル（Weibull）型（明確な上限をもつ変数の場合）である。これは中心極限定理（central limit theorem）（Section 12.24）の極値版にあたる。技術者はこれを用いて「百年に一度の波」に耐える防潮施設や海洋プラットフォームを設計し、保険会社は巨大災害補償の価格を決め、気候科学者は熱波がどのように変化しているかを推定する。

ガンベルと洪水

エミール・ユリウス・ガンベル（Emil Julius Gumbel）はドイツの統計学者であり、ナチスに公然と反対したために 1932 年に大学の職を追われ、最終的にアメリカ合衆国へ亡命した人物である。彼は 1930 年代から 1940 年代にかけて、極値理論を河川の氾濫に応用した。その著書 *Statistics of Extremes*（『極値統計学』、1958 年）によって、この分野は実用的な道具となった。1953 年の北海大洪水の後にオランダを守るために建設されたデルタ計画の施設は、極値統計を用いて、およそ 10 000 年に一度と見込まれる高潮に耐えるよう設計された。

次の一歩

1. **最大値の期待値。** $\mathrm{E}(Y) = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n})$ を示せ。（**ヒント**：無記憶性（memorylessness）により、n 個の寿命のうち最初の一つが尽きた後、残りの $n-1$ 個は新たに始まったのと同じである。）

2. **最小値。** $X_1, \ldots, X_n$ の最小値が、パラメータ $n\lambda$ の指数分布に従うことを示せ。

3. **中央値を厳密に。** Y の厳密な中央値を求め、$n = 10$ の場合について近似値 $\frac{1}{\lambda}(\ln n - \ln\ln 2)$ と比較せよ。

4. **一様分布の最大値。** $X_1, \ldots, X_n$ が $[0,1]$ 上の一様分布（uniform distribution）に従うとき、その最大値 M の分布を求め、n が大きいとき $n(1-M)$ が近似的に指数分布に従うことを示せ。これは極値の 3 つの形のうちどれか。

さらに学ぶために

- **Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values***（Springer）。洪水や海面水位の実データを用いた標準的な入門書。
- **Emil Julius Gumbel, *Statistics of Extremes***（Dover）。古典的名著。
- **David Spiegelhalter, *The Art of Uncertainty***（Pelican）。まれな事象と、それについてどう考えるべきかを論じた本。

詳しい解説付きの解答は第 12.25 節にある：
https://beyondstep.keithmcnulty.org/ja/ch12/extremes.html

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com