12.25 眾多之中的最大值

工作紙

題目

STEP III 2025，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2025 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：

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

提示

請先自行嘗試解題。如果遇到困難，請逐一閱讀提示，每讀完一個便回到題目再試。標有停一停，試一試的提示是關鍵步驟。每個提示之後均留有空位供你演算。

> **想一想**　$X_1, \ldots, X_n$ 中的最大者不超過 y，恰好等同於發生了甚麼事？

階段 1　最大值的分佈（distribution）

涵蓋：第 (i) 部分

> **想一想**　各個 X_i 之中最大的一個不超過 y，當且僅當每個 X_i 都不超過 y。利用獨立性（independence）。

階段 2 分位數（quantiles）

涵蓋：第 (ii) 部分

> **想一想** 由 $\mathrm{G}(L) = \alpha$ 解出 L，並由 $\mathrm{G}(U) = 1 - \alpha$ 解出 U。

階段 3 n 很大的情況

涵蓋：第 (iii) 部分

> **停一停，試一試** 寫成 $\alpha^{1/n} = e^{(\ln \alpha)/n}$。當 n 很大時，指數很小，因此可用 $e^t \approx 1 + t$。

階段 4　增長與離散程度

涵蓋：第 (iv) 部分

想一想　中位數（median）就是 $L(\frac{1}{2})$：當 $n \to \infty$ 時它會怎樣變化？把你對 U 和 L 的近似式相減。哪些項會互相抵消？

階段 5 90% **區間**

涵蓋：第 (v) 部分

> **想一想** 取 $\alpha = 0.05$。求 $\ln 20$ 和 $\ln \frac{1}{0.95}$ 的近似值。

背後的數學

耿貝爾分佈

由於 $\mathrm{P}(\lambda Y - \ln n \le x) = (1 - \frac{e^{-x}}{n})^n \to \exp(-e^{-x})$，平移後的最大值 $\lambda Y - \ln n$ 在極限下服從耿貝爾分佈，其分佈函數（distribution function）為 $\exp(-e^{-x})$。它的平均值是歐拉常數（Euler's constant）$\gamma \approx 0.577$。你其實早已遇過它：集齊全部 n 款優惠券所需的時間（Section 9.11）就是 n 個（接近獨立的）等待時間中的最大值，這正是它約為 $n \ln n + \gamma n$ 的原因，而其波動屬耿貝爾型，數量級為 n。

極值理論

費雪（Fisher）、Tippett 和格涅堅科（Gnedenko）有一個非凡的定理（1928–1943）：許多獨立同分佈（independent identically distributed）的變量，其最大值經適當平移和縮放後，極限形狀只可能有三種：耿貝爾型（適用於輕尾分佈，例如指數分佈或正態分佈）、弗雷歇（Fréchet）型（適用於重尾分佈，例如柯西（Cauchy）分佈）和韋布爾（Weibull）型（適用於有硬性上限的變量）。這是中心極限定理（central limit theorem）（Section 12.24）在極值方面的對應結果，所屬的學問稱為極值理論（extreme value theory）。工程師用它來設計防洪設施和海上平台，以抵禦「百年一遇的巨浪」；保險公司用它為巨災保障定價；氣候科學家則用它來估計熱浪正在怎樣變化。

耿貝爾與洪水

Emil Julius Gumbel（耿貝爾）是德國統計學家，直言不諱地反對納粹，1932 年被迫離開大學教席，最終逃往美國。他在 1930 和 1940 年代把極值理論應用於河流氾濫的研究。他的著作 *Statistics of Extremes*（1958）使這門學問成為實用的工具。荷蘭的三角洲工程（Delta Works）是在 1953 年北海洪災之後為保護荷蘭而興建的，設計時就運用了極值統計學，以抵禦預計約 10 000 年一遇的風暴潮。

延伸練習

1 **最大值的期望值。**證明 $\mathrm{E}(Y) = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n})$。（提示：根據無記憶性（memorylessness），n 個壽命中的第一個結束之後，餘下的 $n-1$ 個會重新開始。）

2 **最小值。**證明 $X_1, \ldots, X_n$ 中的最小者服從參數為 $n\lambda$ 的指數分佈。

3 **中位數的精確值。**求 Y 的精確中位數，並在 $n = 10$ 時與近似值 $\frac{1}{\lambda}(\ln n - \ln\ln 2)$ 比較。

4 **均勻分佈的最大值。**若 $X_1, \ldots, X_n$ 在 $[0,1]$ 上均勻分佈（uniform），求它們的最大值 M 的分佈，並證明當 n 很大時，$n(1-M)$ 近似服從指數分佈。這屬於三種極值形狀中的哪一種？

延伸閱讀

- **Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values***（Springer）。標準的入門書，附有洪水和海平面的真實數據。
- **Emil Julius Gumbel, *Statistics of Extremes***（Dover）。經典之作。
- **David Spiegelhalter, *The Art of Uncertainty***（Pelican），談罕見事件以及應該怎樣思考它們。

完整的引導式解答見第 12.25 節：

https://beyondstep.keithmcnulty.org/zh-hant/ch12/extremes.html

每星期一道 STEP 題目，完整講解：The STEP Up，以電郵免費發送，訂閱網址：keithmcnulty.substack.com