12.25 众多之中的最大值

工作纸

题目

STEP III 2025，第 11 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2025 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

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

提示

请先自行尝试解题。如果遇到困难，请逐一阅读提示，每读完一个便回到题目再试。标有停一停，试一试的提示是关键步骤。每个提示之后均留有空位供你演算。

想一想 $X_1, \ldots, X_n$ 中的最大者不超过 y，恰好等同于发生了甚么事？

阶段 1 最大值的分布（distribution）

涵盖：第 (i) 部分

想一想 各个 X_i 之中最大的一个不超过 y，当且仅当每个 X_i 都不超过 y。利用独立性（independence）。

阶段 2 分位数（quantiles）

涵盖：第 (ii) 部分

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

阶段 3 n 很大的情况

涵盖：第 (iii) 部分

停一停，试一试 写成 $\alpha^{1/n} = e^{(\ln \alpha)/n}$。当 n 很大时，指数很小，因此可用 $e^t \approx 1 + t$。

阶段 4 增长与离散程度

涵盖：第 (iv) 部分

想一想 中位数（median）就是 $L(\frac{1}{2})$：当 $n \to \infty$ 时它会怎样变化？把你对 U 和 L 的近似式相减。哪些项会互相抵消？

阶段 5 90% **区间**

涵盖：第 (v) 部分

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

背后的数学

耿贝尔分布

由于 $\mathrm{P}(\lambda Y - \ln n \le x) = (1 - \frac{e^{-x}}{n})^n \to \exp(-e^{-x})$，平移后的最大值 $\lambda Y - \ln n$ 在极限下服从耿贝尔分布，其分布函数（distribution function）为 $\exp(-e^{-x})$。它的平均值是欧拉常数（Euler's constant）$\gamma \approx 0.577$。你其实早已遇过它：集齐全部 n 款优惠券所需的时间（Section 9.11）就是 n 个（接近独立的）等待时间中的最大值，这正是它约为 $n \ln n + \gamma n$ 的原因，而其波动属耿贝尔型，数量级为 n。

极值理论

费雪（Fisher）、Tippett 和格涅坚科（Gnedenko）有一个非凡的定理（1928–1943）：许多独立同分布（independent identically distributed）的变量，其最大值经适当平移和缩放后，极限形状只可能有三种：耿贝尔型（适用于轻尾分布，例如指数分布或正态分布）、弗雷歇（Fréchet）型（适用于重尾分布，例如柯西（Cauchy）分布）和韦布尔（Weibull）型（适用于有硬性上限的变量）。这是中心极限定理（central limit theorem）（Section 12.24）在极值方面的对应结果，所属的学问称为极值理论（extreme value theory）。工程师用它来设计防洪设施和海上平台，以抵御“百年一遇的巨浪”；保险公司用它为巨灾保障定价；气候科学家则用它来估计热浪正在怎样变化。

耿贝尔与洪水

Emil Julius Gumbel（耿贝尔）是德国统计学家，直言不讳地反对纳粹，1932 年被迫离开大学教席，最终逃往美国。他在 1930 和 1940 年代把极值理论应用于河流泛滥的研究。他的著作 *Statistics of Extremes*（1958）使这门学问成为实用的工具。荷兰的三角洲工程（Delta Works）是在 1953 年北海洪灾之后为保护荷兰而兴建的，设计时就运用了极值统计学，以抵御预计约 10 000 年一遇的风暴潮。

延伸练习

1 **最大值的期望值。**证明 $\mathrm{E}(Y) = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n})$。（提示：根据无记忆性（memorylessness），n 个寿命中的第一个结束之后，余下的 $n-1$ 个会重新开始。）

2 **最小值。**证明 $X_1, \ldots, X_n$ 中的最小者服从参数为 $n\lambda$ 的指数分布。

3 **中位数的精确值。**求 Y 的精确中位数，并在 $n = 10$ 时与近似值 $\frac{1}{\lambda}(\ln n - \ln\ln 2)$ 比较。

4 **均匀分布的最大值。**若 $X_1, \ldots, X_n$ 在 $[0,1]$ 上均匀分布（uniform），求它们的最大值 M 的分布，并证明当 n 很大时，$n(1-M)$ 近似服从指数分布。这属于三种极值形状中的哪一种？

延伸阅读

- **Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values***（Springer）。标准的入门书，附有洪水和海平面的真实数据。
- **Emil Julius Gumbel, *Statistics of Extremes***（Dover）。经典之作。
- **David Spiegelhalter, *The Art of Uncertainty***（Pelican），谈罕见事件以及应该怎样思考它们。

完整的引导式解答见第 12.25 节：

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

每星期一道 STEP 题目，完整讲解：The STEP Up，以电子邮件免费发送，订阅网址：keithmcnulty.substack.com