12.24 中心极限定理

工作纸

题目

STEP III 2008，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/08-s3-q12

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 密度函数 $\frac{1}{2}e^{-|x|}$ 是对称的。它的图像是甚么样子？它的平均值（mean）是多少？

阶段 1 矩母函数

涵盖：草图与矩母函数（moment generating function）

? **想一想** 把 $\mathrm{E}(e^{\theta X})$ 的积分在 0 处拆开。为甚么需要 $|\theta| < 1$?

? **想一想** 把 $\frac{1}{1-\theta^2}$ 展开成级数(series),从中读出平均值(mean)和方差(variance)。

阶段 2 青蛙的位移

涵盖：$Y/\sqrt{2n}$ 的矩母函数

停一停，试一试 各次跳跃互相独立（independent），那么它们的总和 Y 的矩母函数是甚么？把 Y 除以 $\sqrt{2n}$，对 θ 有甚么影响？

阶段 3 极限

涵盖：极限（limit）

想一想 取对数，并利用当 u 很小时 $\ln(1-u) = -u - \frac{1}{2}u^2 - \cdots$。哪一个分布的矩母函数正是这个极限？

阶段 4 **青蛙的机会**

涵盖：最少的跳跃次数

想一想 当 n 很大时，Y 近似服从 $\mathrm{N}(0, 2n)$。对标准正态（standard normal）变量 Z，有 $\mathrm{P}(|Z| \geq 1.96) = 0.05$。由此解出 n。

背后的数学

一般的定理

同样的论证适用于任何平均值为 0、方差 σ^2 有限的分布（还须有矩母函数，不过这个条件可以避开）。这时 $\mathrm{M}(\theta) = 1 + \frac{1}{2}\sigma^2\theta^2 + o(\theta^2)$，而且

$$\left(\mathrm{M}\left(\frac{\theta}{\sigma\sqrt{n}}\right)\right)^n = \left(1 + \frac{\theta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{\theta^2/2}.$$

取极限之后只有方差留下来：原来分布的形状已被遗忘。这就是正态分布具有普遍性的原因。在大学里，这个定理是用*特征函数*（characteristic function）$\mathrm{E}(e^{i\theta X})$ 来证明的，特征函数必定存在，而分布的收敛则由莱维连续性定理（Lévy's continuity theorem）赋予严格的意义。

有多快?

中心极限定理说明极限情况下会发生甚么，却没有说明收敛得有多快。*Berry–Esseen 定理*（Berry–Esseen theorem）给出误差的上界：一个常数乘以 $\frac{\mathrm{E}|X|^3}{\sigma^3\sqrt{n}}$。对于像青蛙跳跃那样的对称分布，跳了几十次之后近似已相当准确。对于偏态（skewed）分布，或者在极端的尾部，即使 n 很大，近似也可能很差：正态近似会低估极罕见事件发生的机会，这个事实对保险和金融都有影响。

从棣美弗到 Lindeberg

棣美弗（De Moivre）在 1733 年发现了二项分布（binomial distribution）的正态近似（Section 12.23），拉普拉斯在 1810 年把它推广至更一般的变量之和，并首先提出量度误差是许多微小效应之和这个想法。严格的证明要到很久之后才出现：李亚普诺夫（Aleksandr Lyapunov）在 1901 年给出证明，Jarl Lindeberg 则在 1922 年给出证明，并找到了基本上是令定理成立的最弱条件。“中心极限定理”这个名称是波利亚（George Pólya）在 1920 年起的，意思是这个定理在概率论中处于中心地位。

延伸练习

1 **两次跳跃的精确结果。**证明 $Y = X_1 + X_2$ 的密度函数为 $\frac{1}{4}(1 + |y|)e^{-|y|}$，并验证其方差为 4。

2 **从二项到正态。**设 $X \sim \mathrm{B}(1, \frac{1}{2})$，并把它平移使平均值为 0（即取值 $\pm\frac{1}{2}$），求出 $\mathrm{M}_X(\theta) = \cosh\frac{\theta}{2}$，并证明 n 个这样的变量的标准化和的矩母函数趋向 $e^{\theta^2/2}$。

3 **青蛙的精确结果。**利用电脑（或以 $n = 2$ 时的精确密度函数作检验），看看当 $n = 82$ 时正态近似有多准确。

4 **为何柯西不行？**柯西分布（Cauchy distribution）（Section 12.22）没有矩母函数。论证中哪一步失效？n 个柯西变量的平均值又会怎样？

延伸阅读

- **Geoffrey Grimmett & Dominic Welsh,** ***Probability: An Introduction***（Oxford University Press），关于生成函数和极限定理的一章。
- **Hans Fischer,** ***A History of the Central Limit Theorem***（Springer）。从拉普拉斯到 Lindeberg。
- **David Spiegelhalter,** ***The Art of Statistics***（Pelican），讲述为何正态分布无处不在。

完整的引导式解答见第 12.24 节：
https://beyondstep.keithmcnulty.org/zh-hans/ch12/clt.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