12.24 中心極限定理

工作紙

題目

STEP III 2008，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

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

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 密度函數 $\frac{1}{2}e^{-|x|}$ 是對稱的。它的圖像是甚麼樣子？它的平均值（mean）是多少？

階段 1 矩母函數

涵蓋：草圖與矩母函數（moment generating function）

想一想 把 $\mathrm{E}(e^{\theta X})$ 的積分在 0 處拆開。為甚麼需要 $|\theta| < 1$？

想一想 把 $\frac{1}{1-\theta^2}$ 展開成級數（series），從中讀出平均值（mean）和方差（variance）。

階段 2　青蛙的位移

涵蓋：$Y/\sqrt{2n}$ 的矩母函數

停一停，試一試　各次跳躍互相獨立（independent），那麼它們的總和 Y 的矩母函數是甚麼？把 Y 除以 $\sqrt{2n}$，對 θ 有甚麼影響？

階段 3　極限

涵蓋：極限（limit）

想一想　取對數，並利用當 u 很小時 $\ln(1-u) = -u - \frac{1}{2}u^2 - \cdots$。哪一個分佈的矩母函數正是這個極限？

階段 4 **青蛙的機會**

涵蓋：最少的跳躍次數

想一想 當 n 很大時，Y 近似服從 $\mathrm{N}(0, 2n)$。對標準正態（standard normal）變量 Z，有 $\mathrm{P}(|Z| \geq 1.96) = 0.05$。由此解出 n。

背後的數學

一般的定理

同樣的論證適用於任何平均值為 0、方差 σ^2 有限的分佈（還須有矩母函數，不過這個條件可以避開）。這時 $\mathrm{M}(\theta) = 1 + \frac{1}{2}\sigma^2\theta^2 + o(\theta^2)$，而且

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

取極限之後只有方差留下來：原來分佈的形狀已被遺忘。這就是正態分佈具有普遍性的原因。在大學裏，這個定理是用特徵函數（characteristic function）$\mathrm{E}(e^{i\theta X})$ 來證明的，特徵函數必定存在，而分佈的收斂則由萊維連續性定理（Lévy's continuity theorem）賦予嚴格的意義。

有多快？

中心極限定理說明極限情況下會發生甚麼，卻沒有說明收斂得有多快。*Berry–Esseen* 定理（Berry–Esseen theorem）給出誤差的上界：一個常數乘以 $\frac{\mathrm{E}|X|^3}{\sigma^3\sqrt{n}}$。對於像青蛙跳躍那樣的對稱分佈，跳了幾十次之後近似已相當準確。對於偏態（skewed）分佈，或者在極端的尾部，即使 n 很大，近似也可能很差：正態近似會低估極罕見事件發生的機會，這個事實對保險和金融都有影響。

從棣美弗到 Lindeberg

棣美弗（De Moivre）在 1733 年發現了二項分佈（binomial distribution）的正態近似（Section 12.23），拉普拉斯在 1810 年把它推廣至更一般的變量之和，並首先提出量度誤差是許多微小效應之和這個想法。嚴格的證明要到很久之後才出現：李亞普諾夫（Aleksandr Lyapunov）在 1901 年給出證明，Jarl Lindeberg 則在 1922 年給出證明，並找到了基本上是令定理成立的最弱條件。「中心極限定理」這個名稱是波利亞（George Pólya）在 1920 年起的，意思是這個定理在概率論中處於中心地位。

延伸練習

1 **兩次跳躍的精確結果。**證明 $Y = X_1 + X_2$ 的密度函數為 $\frac{1}{4}(1 + |y|)e^{-|y|}$，並驗證其方差為 4。

2 **從二項到正態。**設 $X \sim \mathrm{B}(1, \frac{1}{2})$，並把它平移使平均值為 0（即取值 $\pm\frac{1}{2}$），求出 $\mathrm{M}_X(\theta) = \cosh\frac{\theta}{2}$，並證明 n 個這樣的變量的標準化和的矩母函數趨向 $e^{\theta^2/2}$。

3 **青蛙的精確結果。**利用電腦（或以 $n = 2$ 時的精確密度函數作檢驗），看看當 $n = 82$ 時正態近似有多準確。

4 **為何柯西不行？**柯西分佈（Cauchy distribution）（Section 12.22）沒有矩母函數。論證中哪一步失效？n 個柯西變量的平均值又會怎樣？

延伸閱讀

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），關於生成函數和極限定理的一章。
- **Hans Fischer, *A History of the Central Limit Theorem***（Springer）。從拉普拉斯到 Lindeberg。
- **David Spiegelhalter, *The Art of Statistics***（Pelican），講述為何正態分佈無處不在。

完整的引導式解答見第 12.24 節：
https://beyondstep.keithmcnulty.org/zh-hant/ch12/clt.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