詳細解答

STEP III 2008，第 12 題

概率與統計 • 矩生成函數

> **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

本題在 *Beyond STEP* 第 12.24 節中心極限定理中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch12/clt.html

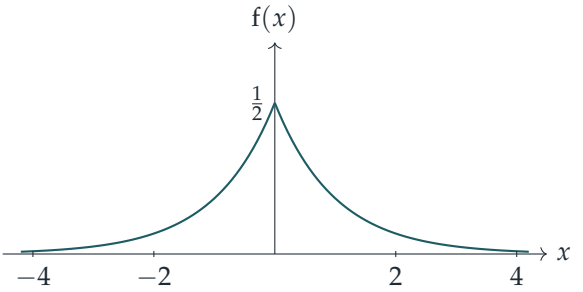
關鍵思路

密度函數中的絕對值（absolute value）把每個積分分成 $x < 0$ 和 $x > 0$ 兩半，每一半都是普通的指數積分。此後整道題只靠矩母函數（moment generating function，簡稱 m.g.f.）的兩個性質：獨立（independent）變量之和的矩母函數，等於各自矩母函數之積；以及 $\mathrm{M}_{aY}(\theta) = \mathrm{M}_Y(a\theta)$。取對數後，$n$ 次方就變成 n 乘以一個很小的量，這正是麥克勞林級數（Maclaurin series）最擅長處理的情況。這就是中心極限定理（central limit theorem），而且在這個例子中，每一步都可以親手算出來。

解答

略圖

密度函數是偶函數，在 $x = 0$ 處取得最大值 $\frac{1}{2}$，該處有一個尖角（左、右導數分別為 $+\frac{1}{2}$ 和 $-\frac{1}{2}$），並在兩側以指數形式遞減至 0。



矩母函數

根據定義，$\mathrm{M}_X(\theta) = \mathrm{E}(\mathrm{e}^{\theta X})$。在 0 處把積分分開，

$$\begin{aligned}\mathrm{M}_X(\theta) &= \tfrac{1}{2}\int_{-\infty}^{0} \mathrm{e}^{(1+\theta)x}\,\mathrm{d}x + \tfrac{1}{2}\int_{0}^{\infty} \mathrm{e}^{-(1-\theta)x}\,\mathrm{d}x \\ &= \frac{1}{2(1+\theta)} + \frac{1}{2(1-\theta)} = \frac{(1-\theta)+(1+\theta)}{2(1-\theta^2)} = \frac{1}{1-\theta^2}.\end{aligned}$$

第一個積分只在 $1+\theta > 0$ 時收斂（converge），第二個只在 $1-\theta > 0$ 時收斂，所以上式對 $-1 < \theta < 1$ 成立，而我們只需要 $\theta = 0$ 附近的情況，這已經足夠。這就是所求的結果。

至於方差（variance），當 $|\theta| < 1$ 時，把它展開為等比級數（geometric series）：

$$\mathrm{M}_X(\theta) = 1 + \theta^2 + \theta^4 + \cdots.$$

與 $\mathrm{M}_X(\theta) = 1 + \mathrm{E}(X)\,\theta + \frac{1}{2}\mathrm{E}(X^2)\,\theta^2 + \cdots$ 比較，可得 $\mathrm{E}(X) = 0$ 和 $\frac{1}{2}\mathrm{E}(X^2) = 1$。因此

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = 2.$$

（求導兩次再代入 $\theta = 0$ 也得到相同結果，但用級數較快。）

青蛙：$Y/\sqrt{2n}$ 的矩母函數

由於各次跳躍 $X_1, \ldots, X_n$ 互相獨立，$\mathrm{e}^{\theta X_1}, \ldots, \mathrm{e}^{\theta X_n}$ 也互相獨立，所以它們乘積的期望值（expectation）等於各期望值之積：

$$\mathrm{M}_Y(\theta) = \mathrm{E}\big(\mathrm{e}^{\theta X_1}\cdots\mathrm{e}^{\theta X_n}\big) = \prod_{i=1}^{n}\mathrm{M}_{X_i}(\theta) = (1-\theta^2)^{-n}.$$

設 $Z = Y/\sqrt{2n}$，則 $\mathrm{E}(\mathrm{e}^{\theta Z}) = \mathrm{E}\big(\mathrm{e}^{(\theta/\sqrt{2n})Y}\big) = \mathrm{M}_Y\big(\theta/\sqrt{2n}\big)$，所以

$$\mathrm{M}_Z(\theta) = \left(1 - \frac{\theta^2}{2n}\right)^{-n}, \qquad |\theta| < \sqrt{2n}.$$

以 $\sqrt{2n}$ 作縮放是很自然的：$\mathrm{Var}(Y) = 2n$，所以 Z 的平均值為 0，方差為 1。

極限

固定 θ。只要 $n > \frac{1}{2}\theta^2$，便有 $0 \le \theta^2/(2n) < 1$，於是可以取對數，並使用對 $|u| < 1$ 成立的 $\ln(1-u) = -u - \frac{1}{2}u^2 - \frac{1}{3}u^3 - \cdots$：

$$\begin{aligned}\ln \mathrm{M}_Z(\theta) &= -n\ln\left(1 - \frac{\theta^2}{2n}\right) = n\left(\frac{\theta^2}{2n} + \frac{\theta^4}{8n^2} + \frac{\theta^6}{24n^3} + \cdots\right)\\ &= \frac{\theta^2}{2} + \frac{\theta^4}{8n} + \frac{\theta^6}{24n^2} + \cdots.\end{aligned}$$

? **想一想** 我們怎樣知道，當 $n \to \infty$ 時，這個級數的尾部真的會消失？

設 $u = \theta^2/(2n)$，首項之後的所有項之和為 $n\left(\frac{1}{2}u^2 + \frac{1}{3}u^3 + \cdots\right) \le n \cdot \frac{1}{2}u^2(1 + u + u^2 + \cdots) = \dfrac{nu^2}{2(1-u)} = \dfrac{\theta^4}{8n(1-u)}$，由於 $u \to 0$，這個量趨向 0。所以 $\ln \mathrm{M}_Z(\theta) \to \frac{1}{2}\theta^2$，而 exp 是連續（continuous）函數，故

$$\mathrm{M}_{Y/\sqrt{2n}}(\theta) \to \exp\left(\tfrac{1}{2}\theta^2\right) \quad \text{當} n \to \infty,$$

這正是所要證明的。

最少跳躍次數

由於這個極限正是標準正態分佈（standard Normal）的矩母函數，當 n 很大時，我們把 $Z = Y/\sqrt{2n}$ 視為近似服從 $\mathrm{N}(0,1)$。「落在距離起點 25 cm 或以上的位置」是指 $|Y| \ge 25$，兩個方向都計算在內，所以

$$\mathrm{P}(|Y| \ge 25) = \mathrm{P}\left(|Z| \ge \frac{25}{\sqrt{2n}}\right) \approx 2\left(1 - \Phi\left(\frac{25}{\sqrt{2n}}\right)\right).$$

這個概率隨 n 增加而上升（青蛙散佈得更遠），所以我們要找使它達到 5% 的最小 n。由於 $\Phi(1.96) = 0.975$，我們需要

$$\frac{25}{\sqrt{2n}} \le 1.96 \iff 2n \ge \left(\frac{25}{1.96}\right)^2 \approx 162.7 \iff n \ge 81.3.$$

所以最少跳躍次數是 $n = 82$。（當 $n = 81$ 時，近似概率約為 4.95%；當 $n = 82$ 時，約為 5.09%。）

驗算

- $\mathrm{M}_X(0) = 1$，任何矩母函數都必須如此；而且 M_X 是偶函數，與略圖的對稱性吻合。
- 以數值積分（numerical quadrature）計算 $\int x^2 \mathrm{f}(x)\,\mathrm{d}x$ 得 2.0000，計算 $\int \mathrm{e}^{0.3x}\mathrm{f}(x)\,\mathrm{d}x$ 得 $1.0989 = 1/0.91$。

- 取 $\theta = 1$，$(1 - \frac{1}{2n})^{-n}$ 在 $n = 10$ 時為 1.6702，在 $n = 100$ 時為 1.6508，逐漸逼近 $e^{1/2} = 1.6487$。
- 模擬 10^5 隻青蛙，每隻作 82 次拉普拉斯（Laplace）跳躍，結果約有 5.1% 的青蛙最後落在距離起點至少 25 cm 的位置。

常見錯誤

- 使用單尾（one-tailed）數值 1.645：青蛙可以落在任何一側 25 cm 以外，所以兩邊尾部都要計算。
- 寫成 $M_{Y/\sqrt{2n}}(\theta) = M_Y(\theta)/\sqrt{2n}$。把變量縮放，縮放的是其矩母函數的自變量。
- 把 81.3 四捨五入至最接近的整數而答 $n = 81$：當 $n = 81$ 時，概率僅略低於 5%。

STEP 之外

你剛才證明的，就是步長服從拉普拉斯分佈時的中心極限定理，而一般情況的證明也循同一路線：把 $\ln M(\theta/\sqrt{n})$ 展開至二階，再證明餘下部分會消失。在大學一年級的課程中，你會遇到「連續性定理」（continuity theorem），它說明為何可以由矩母函數（或特徵函數（characteristic function），它必定存在）的收斂推出概率的收斂，這一步我們在這裏只是姑且相信。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）對兩者都有論述。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。