12.22 沒有平均值的分佈

工作紙

題目

STEP III 1995，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/95-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**　若 (U, V) 是一個隨機點，其密度只取決於它與原點的距離，那麼由原點到該點的直線與坐標軸所成的角 Θ，你能說些甚麼？

階段 1　聯合密度

涵蓋：聯合密度（joint density）與徑向對稱（radial symmetry）

> **想一想**　X 與 Y 互相獨立（independent），所以把它們的密度相乘。所得的結果取決於甚麼？

階段 2 運用對稱性

涵蓋：$\mathrm{P}(U/V < k)$ 的恆等式（identity）

> **停一停，試一試** $\frac{U}{V} < k$ 可以在 $V > 0$ 時發生，也可以在 $V < 0$ 時發生。把兩種情況各寫成一個區域。旋轉（rotation）$(u, v) \mapsto (-u, -v)$ 會把這兩個區域變成甚麼？

階段 3　一個角

涵蓋：U/V 的密度

想一想　畫出區域 $\{V > 0, U < kV\}$。它是一個以原點為頂點的楔形（wedge）：它的角度是多少？若密度是徑向對稱的，楔形的概率就是它的角度除以 2π。然後求導（differentiate）。

背後的數學

沒有平均值

柯西分佈的平均值本應是 $\int_{-\infty}^{\infty} \frac{k}{\pi(1+k^2)}\,\mathrm{d}k$，但 $\int_0^K \frac{k}{1+k^2}\,\mathrm{d}k = \frac{1}{2}\ln(1+K^2) \to \infty$：這個積分並不收斂（converge），所以平均值不存在。（對稱性或會令人以為答案是 0，但正負兩部分都是無窮大，而 $\infty - \infty$ 並沒有值。）方差（variance）同樣是無窮大。

更糟的是，n 個獨立柯西變量的平均仍然是柯西變量，而且分佈完全相同。（其中一個看出這點的方法，是利用特徵函數（characteristic function）$\mathrm{E}(e^{itX}) = e^{-|t|}$。）所以取平均完全無補於事：一百萬次量度的平均，並不比單獨一次可靠。圖 12.22.1 顯示了實際情況：正態樣本的累計平均會穩定下來，而柯西樣本的累計平均卻永遠東歪西倒，不時被偶然出現的極大數值撞得偏離原位。

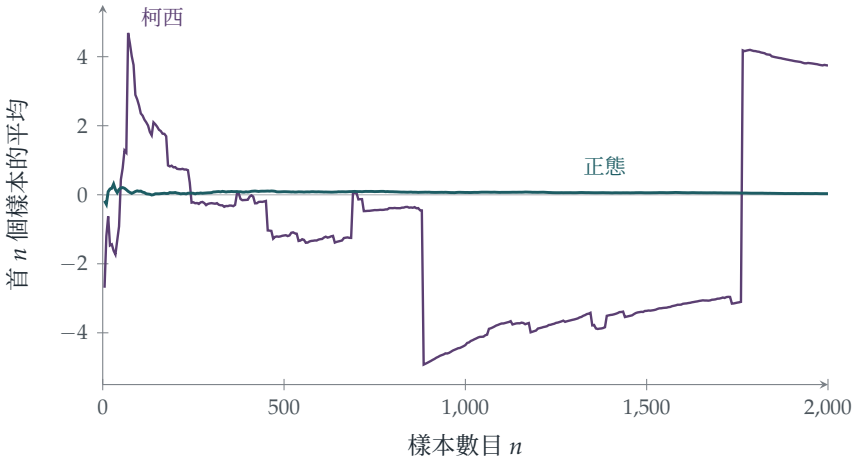

圖 12.22.1 2000 個獨立樣本的累計平均。對於標準正態樣本（藍綠色），平均值一如大數定律所預測，穩定下來並趨向 0。對於柯西樣本（紫紅色），它永遠不會穩定：每隔一段時間，單單一個極大的數值就會把它拋到另一個位置。

柯西變量從何而來

拿一支電筒，朝一個均勻隨機的方向照向一面牆：光點在牆上的位置便服從柯西分佈。在物理學中，同一個密度稱為勞侖茲分佈（*Lorentzian*），用來描述譜線和共振峯的形狀（比較 Section 11.1）。在金融學中，像柯西分佈這類厚尾（heavy-tailed）分佈，會用來為偶爾出現的股災建立模型，而正態模型往往嚴重低估這類事件。這對統計學是一個嚴肅的教訓：建基於平均值和方差的方法，遇上厚尾數據便會失效，這時就需要穩健的方法，例如使用中位數（median）（柯西分佈的中位數完全沒有問題）。

泊松、柯西與一場論戰

這個分佈在 1824 年由泊松（Siméon Denis Poisson）研究過，他注意到它違反大數定律。柯西（Augustin-Louis Cauchy）則在 1853 年與 Irénée-Jules Bienaymé 就最小二乘法（method of least squares）公開論戰時遇上它：柯西用它來說明，當誤差屬於厚尾時，最小二乘法可以是一個很差的方法。這個分佈最終冠上了柯西的名字，正是 Stephen Stigler 所稱的「史蒂格勒命名定律」（Stigler's law of eponymy）的一例：沒有任何科學發現是以其最初發現者命名的。

延伸練習

1 **電筒。**一支電筒與一面長而直的牆相距 1，所指的方向與垂直線成角 Θ，其中 Θ 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上呈均勻分佈。證明光點的位置 $\tan\Theta$ 服從柯西分佈。

2 **倒數。**證明若 C 服從柯西分佈，則 $\frac{1}{C}$ 亦然。（提示：$\frac{U}{V}$ 與 $\frac{V}{U}$。）

3 **中位數行得通。**證明柯西分佈的中位數是 0，而它的四分位數（quartiles）是 ± 1。

4 **半徑與角度。**設 X、Y 為獨立的標準正態變量，透過轉換成極坐標（polar coordinates），證明 $R^2 = X^2 + Y^2$ 服從平均值為 2 的指數分佈（exponential distribution），並且與角度獨立。（這是用來產生正態隨機數的 Box–Muller 方法的基礎。）

延伸閱讀

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），論隨機變量的變換。
- **Nassim Nicholas Taleb, *The Black Swan***（Penguin）。講述厚尾現象及其重要性，適合一般讀者。
- **Stephen Stigler, *The History of Statistics***（Harvard University Press）。書中記述了柯西與 Bienaymé 的論戰。

完整的引導式解答見第 12.22 節：

https://beyondstep.keithmcnulty.org/zh-hant/ch12/cauchy.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