12.22 没有平均值的分布

工作纸

题目

STEP III 1995，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/95-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**　若 (U, V) 是一个随机点，其密度只取决于它与原点的距离，那么由原点到该点的直线与坐标轴所成的角 Θ，你能说些甚么？

阶段 1　联合密度

涵盖：联合密度（joint density）与径向对称（radial symmetry）

> **想一想**　X 与 Y 互相独立（independent），所以把它们的密度相乘。所得的结果取决于甚么？

阶段 2 运用对称性

涵盖：$\mathrm{P}(U/V < k)$ 的恒等式（identity）

> **停一停，试一试** $\frac{U}{V} < k$ 可以在 $V > 0$ 时发生，也可以在 $V < 0$ 时发生。把两种情况各写成一个区域。旋转（rotation）$(u, v) \mapsto (-u, -v)$ 会把这两个区域变成甚么？

阶段 3 **一个角**

涵盖：U/V 的密度

> **想一想** 画出区域 $\{V > 0, U < kV\}$。它是一个以原点为顶点的楔形（wedge）：它的角度是多少？若密度是径向对称的，楔形的概率就是它的角度除以 2π。然后求导（differentiate）。

背后的数学

没有平均值

柯西分布的平均值本应是 $\int_{-\infty}^{\infty} \frac{k}{\pi(1+k^2)}\,\mathrm{d}k$，但 $\int_0^K \frac{k}{1+k^2}\,\mathrm{d}k = \frac{1}{2}\ln(1+K^2) \to \infty$：这个积分并不收敛（converge），所以平均值不存在。（对称性或会令人以为答案是 0，但正负两部分都是无穷大，而 $\infty - \infty$ 并没有值。）方差（variance）同样是无穷大。

更糟的是，n 个独立柯西变量的平均仍然是柯西变量，而且分布完全相同。（其中一个看出这点的方法，是利用特征函数（characteristic function）$\mathrm{E}(e^{itX}) = e^{-|t|}$。）所以取平均完全无补于事：一百万次量度的平均，并不比单独一次可靠。图 12.22.1 显示了实际情况：正态样本的累计平均会稳定下来，而柯西样本的累计平均却永远东歪西倒，不时被偶然出现的极大数值撞得偏离原位。

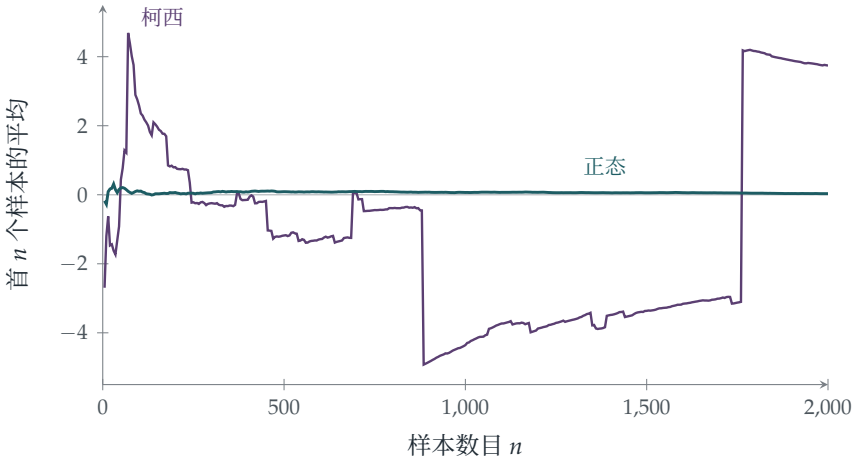

图 12.22.1 2000 个独立样本的累计平均。对于标准正态样本（蓝绿色），平均值一如大数定律所预测，稳定下来并趋向 0。对于柯西样本（紫红色），它永远不会稳定：每隔一段时间，单单一个极大的数值就会把它抛到另一个位置。

柯西变量从何而来

拿一支电筒，朝一个均匀随机的方向照向一面墙：光点在墙上的位置便服从柯西分布。在物理学中，同一个密度称为*劳仑兹分布*（*Lorentzian*），用来描述谱线和共振峰的形状（比较 Section 11.1）。在金融学中，像柯西分布这类厚尾（heavy-tailed）分布，会用来为偶尔出现的股灾建立模型，而正态模型往往严重低估这类事件。这对统计学是一个严肃的教训：建基于平均值和方差的方法，遇上厚尾数据便会失效，这时就需要稳健的方法，例如使用中位数（median）（柯西分布的中位数完全没有问题）。

泊松、柯西与一场论战

这个分布在 1824 年由泊松（Siméon Denis Poisson）研究过，他注意到它违反大数定律。柯西（Augustin-Louis Cauchy）则在 1853 年与 Irénée-Jules Bienaymé 就最小二乘法（method of least squares）公开论战时遇上它：柯西用它来说明，当误差属于厚尾时，最小二乘法可以是一个很差的方法。这个分布最终冠上了柯西的名字，正是 Stephen Stigler 所称的“史蒂格勒命名定律”（Stigler's law of eponymy）的一例：没有任何科学发现是以其最初发现者命名的。

延伸练习

1 **电筒。**一支电筒与一面长而直的墙相距 1，所指的方向与垂直线成角 Θ，其中 Θ 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上呈均匀分布。证明光点的位置 $\tan\Theta$ 服从柯西分布。

2 **倒数。**证明若 C 服从柯西分布，则 $\frac{1}{C}$ 亦然。（提示：$\frac{U}{V}$ 与 $\frac{V}{U}$。）

3 **中位数行得通。**证明柯西分布的中位数是 0，而它的四分位数（quartiles）是 ± 1。

4 **半径与角度。**设 X、Y 为独立的标准正态变量，通过转换成极坐标（polar coordinates），证明 $R^2 = X^2 + Y^2$ 服从平均值为 2 的指数分布（exponential distribution），并且与角度独立。（这是用来产生正态随机数的 Box–Muller 方法的基础。）

延伸阅读

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），论随机变量的变换。
- **Nassim Nicholas Taleb, *The Black Swan***（Penguin）。讲述厚尾现象及其重要性，适合一般读者。
- **Stephen Stigler, *The History of Statistics***（Harvard University Press）。书中记述了柯西与 Bienaymé 的论战。

完整的引导式解答见第 12.22 节：

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