13.7 製造相關性

工作紙

題目

STEP III 2010，第 13 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/10-s3-q13

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 Z_1 和 Z_2 互相獨立。它們的協方差是多少？相關係數又是多少？

階段 1　兩個變量

涵蓋：Z_1 與 Z_2 的相關係數（correlation），以及變量 Y_2

想一想 獨立（independent）變量的協方差（covariance）為 0。對於 Y_2，利用協方差的線性性質（linearity）以及 Z_1 與 Z_2 的獨立性，求出 $\mathrm{Var}(Y_2)$ 和 $\mathrm{Cov}(Y_1, Y_2)$。

階段 2 第三個變量

涵蓋：變量 Y_3

> **停一停，試一試** 把 $\text{Var}(Y_3) = 1$、$\text{Cov}(Y_1, Y_3) = \rho_{13}$ 和 $\text{Cov}(Y_2, Y_3) = \rho_{23}$ 寫成關於 a、b、c 的方程，然後按 a、b、c 的次序逐一求解。

階段 3　任意的平均值與方差

涵蓋：最後一部分

> **想一想**　把一個變量平移和縮放，對它的平均值（mean）、方差（variance）和相關係數各有甚麼影響？試試 $X_i = \mu_i + \sigma_i Y_i$。

背後的數學

科列斯基分解

用矩陣形式來寫，這個構造就是 $\mathbf{Y} = L\mathbf{Z}$，其中的矩陣是下三角（lower triangular）矩陣

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho_{12}^2} & 0 \\ a & b & c \end{pmatrix},$$

而 $\mathbf{Y}$ 的相關矩陣（correlation matrix）是 $R = LL^{\mathsf{T}}$。把一個對稱矩陣寫成 LL^{T} 的形式，就是*科列斯基分解*，由法國軍事工程師科列斯基（André-Louis Cholesky）約於 1910 年發現，用來求解大地測量中的方程組（他在第一次世界大戰中陣亡，這項成果在他身後才發表）。只要 R 是*正定*（positive definite）的，這個分解就行得通，而計算過程正是我們剛才用手做的：逐行求解。在金融、工程和氣候模型中，這是模擬相關變量的標準方法。

哪些相關係數是可能的？

c 的平方根內的數必須非負。展開後，條件是

$$1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} \geq 0,$$

這正是相關矩陣的行列式（determinant）非負的條件。從幾何上看：三個單位向量兩兩之間的夾角為 θ_{ij}（其中 $\cos\theta_{ij} = \rho_{ij}$），它們必須真的能在空間中拼合起來。如果 $\rho_{12} =$

$\rho_{13} = \rho_{23} = -0.6$，各個夾角都約為 127°，而三個兩兩相隔 127° 的向量，夾角之和就要超過 360°：這是不可能的。一般而言，一個矩陣要成為相關矩陣，必須是半正定（positive semi-definite）的，即所有特徵值（eigenvalue）均非負（Section 7.2）。如果各個相關係數是根據不完整的數據分別估計出來的，這個條件就可能不成立，統計學家於是要找出最接近的有效相關矩陣，這個問題由 Nicholas Higham 在 2002 年解決。

高爾頓、皮爾遜與相關係數

高爾頓（Francis Galton）在 1888 年研究父母與子女身高的關係時引入了相關的概念，而皮爾遜（Karl Pearson）在 1896 年給出了本題所用的積矩（product moment）公式。高爾頓還注意到均值回歸（regression to the mean）的現象（Section 13.1）。皮爾遜於 1911 年在倫敦大學學院（University College London）創立了全世界第一個大學統計學系。「相關不等於因果」這句警告，幾乎與相關係數本身一樣古老。

延伸練習

1 **不可能的相關係數。**證明：除非 $\rho \geq -\frac{1}{2}$，否則三個隨機變量不可能兩兩之間的相關係數都是 ρ。對於 n 個變量，相應的結論是甚麼？

2 **檢查行列式。**證明 $1 - \rho_{13}^2 - b^2 \geq 0$ 等價於 $\det R \geq 0$，其中 R 是 3×3 相關矩陣。

3 **投資組合的風險。**兩隻股票回報的標準差（standard deviation）分別為 $\sigma_1 = 0.2$ 和 $\sigma_2 = 0.3$，相關係數為 ρ。若投資組合中兩隻股票各佔一半，求 $\rho = -1$、0 和 1 時該組合的標準差。分散投資達到了甚麼效果？

4 **還原材料。**給定題目中的 Y_1、Y_2，試以 Y_1 和 Y_2 表示 Z_2。（這就是格拉姆–施密特過程中的一步。）

延伸閱讀

- **Gilbert Strang, *Introduction to Linear Algebra***（Wellesley-Cambridge），講述正定矩陣和科列斯基分解。
- **Paul Glasserman, *Monte Carlo Methods in Financial Engineering***（Springer）。講述實際應用中如何模擬相關變量；內容較深，但開頭幾章不難讀。
- **Theodore Porter, *Karl Pearson: The Scientific Life in a Statistical Age***（Princeton University Press）。

完整的引導式解答見第 13.7 節：

https://beyondstep.keithmcnulty.org/zh-hant/ch13/correlation.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