13.7 制造相关性

工作纸

题目

STEP III 2010，第 13 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/10-s3-q13

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 Z_1 和 Z_2 互相独立。它们的协方差是多少？相关系数又是多少？

阶段 1 两个变量

涵盖：Z_1 与 Z_2 的相关系数（correlation），以及变量 Y_2

想一想 独立（independent）变量的协方差（covariance）为 0。对于 Y_2，利用协方差的线性性质（linearity）以及 Z_1 与 Z_2 的独立性，求出 $\mathrm{Var}(Y_2)$ 和 $\mathrm{Cov}(Y_1, Y_2)$。

阶段 2 第三个变量

涵盖：变量 Y_3

停一停，试一试 把 $\mathrm{Var}(Y_3) = 1$、$\mathrm{Cov}(Y_1, Y_3) = \rho_{13}$ 和 $\mathrm{Cov}(Y_2, Y_3) = \rho_{23}$ 写成关于 a、b、c 的方程，然后按 a、b、c 的次序逐一求解。

阶段 3　任意的平均值与方差

涵盖：最后一部分

> **想一想**　把一个变量平移和缩放，对它的平均值（mean）、方差（variance）和相关系数各有甚么影响？试试 $X_i = \mu_i + \sigma_i Y_i$。

背后的数学

科列斯基分解

用矩阵形式来写，这个构造就是 $\mathbf{Y} = L\mathbf{Z}$，其中的矩阵是下三角（lower triangular）矩阵

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

而 $\mathbf{Y}$ 的相关矩阵（correlation matrix）是 $R = LL^{\mathsf{T}}$。把一个对称矩阵写成 LL^{T} 的形式，就是*科列斯基分解*，由法国军事工程师科列斯基（André-Louis Cholesky）约于 1910 年发现，用来求解大地测量中的方程组（他在第一次世界大战中阵亡，这项成果在他身后才发表）。只要 R 是*正定*（positive definite）的，这个分解就行得通，而计算过程正是我们刚才用手做的：逐行求解。在金融、工程和气候模型中，这是模拟相关变量的标准方法。

哪些相关系数是可能的？

c 的平方根内的数必须非负。展开后，条件是

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

这正是相关矩阵的行列式（determinant）非负的条件。从几何上看：三个单位向量两两之间的夹角为 θ_{ij}（其中 $\cos\theta_{ij} = \rho_{ij}$），它们必须真的能在空间中拼合起来。如果 $\rho_{12} = \rho_{13} = \rho_{23} = -0.6$，各个夹角都约为 127°，而三个两两相隔 127° 的向量，夹角之和就要超

过 360°：这是不可能的。一般而言，一个矩阵要成为相关矩阵，必须是半正定（positive semi-definite）的，即所有特征值（eigenvalue）均非负（Section 7.2）。如果各个相关系数是根据不完整的数据分别估计出来的，这个条件就可能不成立，统计学家于是要找出最接近的有效相关矩阵，这个问题由 Nicholas Higham 在 2002 年解决。

高尔顿、皮尔逊与相关系数

高尔顿（Francis Galton）在 1888 年研究父母与子女身高的关系时引入了相关的概念，而皮尔逊（Karl Pearson）在 1896 年给出了本题所用的积矩（product moment）公式。高尔顿还注意到均值回归（regression to the mean）的现象（Section 13.1）。皮尔逊于 1911 年在伦敦大学学院（University College London）创立了全世界第一个大学统计学系。“相关不等于因果”这句警告，几乎与相关系数本身一样古老。

延伸练习

1 **不可能的相关系数。**证明：除非 $\rho \geq -\frac{1}{2}$，否则三个随机变量不可能两两之间的相关系数都是 ρ。对于 n 个变量，相应的结论是甚么？

2 **检查行列式。**证明 $1 - \rho_{13}^2 - b^2 \geq 0$ 等价于 $\det R \geq 0$，其中 R 是 3×3 相关矩阵。

3 **投资组合的风险。**两只股票回报的标准差（standard deviation）分别为 $\sigma_1 = 0.2$ 和 $\sigma_2 = 0.3$，相关系数为 ρ。若投资组合中两只股票各占一半，求 $\rho = -1$、0 和 1 时该组合的标准差。分散投资达到了甚么效果？

4 **还原材料。**给定题目中的 Y_1、Y_2，试以 Y_1 和 Y_2 表示 Z_2。（这就是格拉姆–施密特过程中的一步。）

延伸阅读

- **Gilbert Strang, *Introduction to Linear Algebra***（Wellesley-Cambridge），讲述正定矩阵和科列斯基分解。
- **Paul Glasserman, *Monte Carlo Methods in Financial Engineering***（Springer）。讲述实际应用中如何模拟相关变量；内容较深，但开头几章不难读。
- **Theodore Porter, *Karl Pearson: The Scientific Life in a Statistical Age***（Princeton University Press）。

完整的引导式解答见第 13.7 节：

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