13.7 相関をつくり出す

ワークシート

問題

STEP III 2010、問 13

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

step.maths.org/questions/10-s3-q13

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

? **考えてみよう**　Z_1 と Z_2 は独立である。その共分散はいくらか。したがって相関はいくらか。

段階 1　2 つの確率変数

扱う内容：Z_1 と Z_2 の相関、および確率変数 Y_2

? **考えてみよう**　独立な確率変数（independent random variables）の共分散（covariance）は 0 である。Y_2 については、共分散の線形性（linearity）と Z_1, Z_2 の独立性を用いて $\mathrm{Var}(Y_2)$ と $\mathrm{Cov}(Y_1, Y_2)$ を求めよ。

段階 2　3 つ目の確率変数

扱う内容：確率変数 Y_3

手を止めて、やってみよう　$\mathrm{Var}(Y_3) = 1$、$\mathrm{Cov}(Y_1, Y_3) = \rho_{13}$、$\mathrm{Cov}(Y_2, Y_3) = \rho_{23}$ を a, b, c についての方程式として書き表せ。それらを a, b, c の順に解け。

段階 3　任意の平均と分散

扱う内容：最後の小問

> **考えてみよう**　確率変数を平行移動したり定数倍したりすると、その平均、分散（variance）、相関（correlation）はどう変わるか。$X_i = \mu_i + \sigma_i Y_i$ とおいてみよ。

背後にある数学

コレスキー分解

行列の形で書くと、この構成は $\mathbf{Y} = L\mathbf{Z}$ と表される。ここで L は**下三角**（lower triangular）行列

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

であり、$\mathbf{Y}$ の相関行列（correlation matrix）は $R = LL^{\mathsf{T}}$ となる。対称行列を LL^{T} の形に書くことを**コレスキー分解**という。これはフランスの軍の技術者アンドレ＝ルイ・コレスキー（André-Louis Cholesky）が 1910 年ごろ、測地測量の方程式を解くために見いだしたものである（第一次世界大戦で本人が戦死した後に発表された）。この分解は R が**正定値**（positive definite）であればいつでも可能であり、その計算は私たちが手で行ったこととまったく同じである。すなわち、各行を順に解いていくのである。これは、金融、工学、気候モデリングにおいて相関のある変数をシミュレーションする標準的な方法である。

どのような相関が可能か

c の根号の中の数は負であってはならない。展開すると、その条件は

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

となり、これは相関行列の行列式（determinant）が負でないという条件である。幾何学的にいえば、2 つずつのなす角が θ_{ij}（ただし $\cos\theta_{ij} = \rho_{ij}$）である 3 つの単位ベクトルが、実際に空間の中で両立しなければならないということである。もし $\rho_{12} = \rho_{13} = \rho_{23} = -0.6$ ならば、角はすべて約 127° であり、互いに 127° をなす 3 つのベクトルには、和が 360° を超える角が必要になってしまう。これは不可能である。一般に、ある行列が相関行列となりうるのは、それが半正定値（positive semi-definite）、すなわちすべての固有値（eigenvalue）が負でない場合に限られる（Section 7.2）。相関を不完全なデータから個別に推定すると、この条件が成り立たないことがあり、そのとき統計学者は最も近い有効な相関行列を求めなければならない。この問題は 2002 年に Nicholas Higham によって解決された。

ゴルトン、ピアソンと相関

フランシス・ゴルトン（Francis Galton）は 1888 年、親と子の身長がどのように関係しているかを研究する中で相関を導入し、1896 年にはカール・ピアソン（Karl Pearson）がこの問題で用いられている積率（product moment）の公式を与えた。ゴルトンは平均への回帰（regression to the mean）にも気づいていた（Section 13.1）。ピアソンは 1911 年、ユニバーシティ・カレッジ・ロンドン（University College London）に世界初の大学の統計学科を創設した。「相関は因果を意味しない」という警句は、相関そのものとほとんど同じくらい古い。

次の一歩

1. **不可能な相関。** 3 つの確率変数が 2 つずつすべて相関 ρ をもつことは、$\rho \geq -\tfrac{1}{2}$ でない限り不可能であることを示せ。n 個の変数の場合、対応する主張はどうなるか。

2. **行列式を確かめる。** R を 3×3 の相関行列とするとき、$1-\rho_{13}^2-b^2 \geq 0$ が $\det R \geq 0$ と同値であることを示せ。

3. **ポートフォリオのリスク。** 2 つの株の収益率の標準偏差（standard deviation）が $\sigma_1 = 0.2$ と $\sigma_2 = 0.3$ で、相関が ρ であるとする。それぞれに半分ずつ投資したポートフォリオの標準偏差を、$\rho = -1, 0, 1$ のそれぞれについて求めよ。分散投資（diversification）によって何が得られるか。

4. **材料を復元する。** 問題文のように Y_1, Y_2 が与えられたとき、Z_2 を Y_1 と Y_2 を用いて表せ。（これがグラム–シュミットの 1 ステップである。）

さらに学ぶために

- **Gilbert Strang,** ***Introduction to Linear Algebra***（Wellesley-Cambridge）。正定値行列とコレスキー分解について。
- **Paul Glasserman,** ***Monte Carlo Methods in Financial Engineering***（Springer）。相関のある変数を実際にシミュレーションする方法。高度だが、最初のほうの章は読みやすい。
- **Theodore Porter,** ***Karl Pearson: The Scientific Life in a Statistical Age***（Princeton University Press）。

詳しい解説付きの解答は第 13.7 節にある：

https://beyondstep.keithmcnulty.org/ja/ch13/correlation.html

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com