線形代数学 I/基礎 練習問題 4

講義担当者: 中村　知繁

問 1：共分散と相関行列に関する計算式の理解

以下の各問いに答えよ。

1. n 個のデータ組 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ が与えられ、変数 X の平均値を $\bar{x}$、変数 Y の平均値を $\bar{y}$ とするとき、変数 X と Y の共分散 $\mathrm{Cov}(X, Y)$ を計算するための定義式を、$\sum$ (シグマ) 記号を用いて記述せよ。

2. (1) で定義される共分散 $\mathrm{Cov}(X, Y)$ と、変数 X の標準偏差 σ_X および変数 Y の標準偏差 σ_Y を用いて、変数 X と Y の相関係数 ρ_{XY} を計算するための定義式を記述せよ。

3. 変数 X の偏差ベクトルを
$$\mathbf{x}_{dev} = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix},$$
変数 Y の偏差ベクトルを
$$\mathbf{y}_{dev} = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}$$
とするとき、共分散 $\mathrm{Cov}(X, Y)$ を、これらの偏差ベクトル $\mathbf{x}_{dev}, \mathbf{y}_{dev}$ とデータ数 n を用いて表す式を記述せよ。

4. 2 つの変数 X と Y に関する共分散行列 $\mathbf{\Sigma}$ は、一般的に以下のような 2×2 の対称行列で表される。行列の各成分 A, B, C, D が、$\mathrm{Var}(X)$ (変数 X の分散)、$\mathrm{Var}(Y)$ (変数 Y の分散)、$\mathrm{Cov}(X, Y)$ (変数 X と Y の共分散) のうち、それぞれどれに該当するかを答えよ。
$$\mathbf{\Sigma} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$
(ヒント: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ であり、変数の分散はその変数自身との共分散と等しい、すなわち $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ である。)

問 2

以下の 2 次元データ (X, Y) について考える。データ数は $n = 5$ である。

データ点	X	Y
1	1	5
2	2	4
3	3	4
4	4	3
5	5	4

1. 変数 X と Y の平均 $\bar{x}, \bar{y}$、分散 $\mathrm{Var}(X), \mathrm{Var}(Y)$、共分散 $\mathrm{Cov}(X, Y)$ をそれぞれ計算し、分数で答えよ。
2. (1) で計算した値を用いて、変数 X と Y の共分散行列 $\mathbf{\Sigma}_{XY}$ を記述し、相関係数 ρ_{XY} を計算せよ。相関係数は $\sqrt{\cdot}$ を含む形で答えてもよい。（共分散行列は $\mathbf{\Sigma}_{XY} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(Y, X) & \mathrm{Var}(Y) \end{bmatrix}$ の形で表される。）

次に、これらのデータをスケーリング（線形変換）して新しい変数 X' と Y' を作る。変換式は以下の通りである。

$$X' = 2X$$
$$Y' = -3Y + 10$$

3. 新しい変数 X' と Y' の各データ点を計算し、表にまとめよ。
4. 新しい変数 X' と Y' の平均 $\bar{x}', \bar{y}'$、分散 $\mathrm{Var}(X'), \mathrm{Var}(Y')$、共分散 $\mathrm{Cov}(X', Y')$ をそれぞれ計算し、分数で答えよ。
5. (4) で計算した値を用いて、新しい変数 X' と Y' の共分散行列 $\mathbf{\Sigma}_{X'Y'}$ を記述し、相関係数 $\rho_{X'Y'}$ を計算せよ。相関係数は $\sqrt{\cdot}$ を含む形で答えてもよい。
6. (2) で得られた $\mathrm{Cov}(X, Y)$, $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, ρ_{XY} と、(5) で得られた $\mathrm{Cov}(X', Y')$, $\mathrm{Var}(X')$, $\mathrm{Var}(Y')$, $\rho_{X'Y'}$ を比較せよ。スケーリング $X' = aX + b$, $Y' = cY + d$ (ここで本問題では $a = 2, b = 0, c = -3, d = 10$ である) が、共分散、分散、相関係数にそれぞれどのような影響を与えたか考察し、変換の係数 a, c と関連付けて説明せよ。（ヒント：一般的に $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y)$、$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ である。相関係数 $\rho_{aX+b,cY+d}$ と ρ_{XY} の関係性はどうなるか。）