

まず、この資料では (ベクトルを含む) 行列  $A$  の転置を  $A^T$  で表す. また、特に断らない限り、実数値は斜体の小文字 (例えば  $x$ ), ベクトルは太字の小文字 ( $\mathbf{x}$ ), 行列は斜体の大文字 ( $X$ ) を用いて表す. さらに、確率変数は、1 次元は斜体の大文字 ( $X$ ), ベクトルは太字の大文字 ( $\mathbf{X}$ ) を用いる. 表記の使い分けに注意しよう.

## 1 実対称行列の対角化と二次形式

以下の議論では必ずしも  $A$  が  $n$  次実対称行列である必要はない部分があるが、簡単のため  $A$  は  $n$  次実対称行列であると仮定する.

**定義 1.1**  $A$  について零ベクトルでないベクトル  $\mathbf{x}$  とスカラー  $\lambda$  に対して

$$A\mathbf{x} = \lambda\mathbf{x} \quad (1.1)$$

が成り立つとき、 $\lambda$  を  $A$  の固有値、 $\mathbf{x}$  を  $\lambda$  に対する  $A$  の固有ベクトルという.

**問題 1.1** 実対称行列  $A$  の固有値は実数になることを示せ.

**問題 1.2** 実対称行列  $A$  の異なる固有値に対する固有ベクトルは直交することを示せ.

**問題 1.3**  $A$  の固有値  $\lambda_1, \dots, \lambda_k$  とそれぞれに対する固有ベクトル  $\mathbf{h}_1, \dots, \mathbf{h}_k$  に対して、

$$AH = HD \quad (1.2)$$

が成り立つことを示せ. ただし、 $H = (\mathbf{h}_1 \cdots \mathbf{h}_k)$  で  $D$  は  $\lambda_1, \dots, \lambda_k$  を対角成分にもつ対角行列である.

**定理 1.1** 実対称行列  $A$  に対して、(1.2) を満たす  $H$  を直交行列となるように選ぶことができ、

$$H^T A H = D \quad (1.3)$$

と対角化できる.

**問題 1.4** 実対称行列  $A$  は固有値  $\lambda_1, \dots, \lambda_n$  とそれぞれに対する固有ベクトル  $\mathbf{h}_1, \dots, \mathbf{h}_n$  により次のように分解されることを示せ.

$$A = \sum_{i=1}^n \lambda_i \mathbf{h}_i \mathbf{h}_i^T \quad (1.4)$$

**問題 1.5** 実対称行列  $A$  の  $n$  個の固有値が  $\lambda_1, \dots, \lambda_n$  のとき次が成り立つことを示せ.

$$(1) \operatorname{tr} A = \lambda_1 + \cdots + \lambda_n$$

$$(2) |A| = \lambda_1 \cdots \lambda_n$$

**定義 1.2**  $n$  次実対称行列  $A = (a_{ij})$  と  $n$  次元ベクトル  $\mathbf{x} = (x_1, \dots, x_n)^T$  に対して

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \quad (1.5)$$

を  $A$  によって定まる二次形式という.

**問題 1.6**  $n$  次実対称行列  $A$  について (1.3) を満たす直交行列  $H$  に対して、 $\mathbf{y} = H^T \mathbf{x}$  と置くととき  $\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^n \lambda_i y_i^2$  と標準化されることを示せ. ただし  $\mathbf{y} = (y_1, \dots, y_n)^T$  である.

**問題 1.7**  $n$  次実対称行列  $A$  がすべての  $\mathbf{x} \in \mathbb{R}^n$  に対して  $\mathbf{x}^T A \mathbf{x} \geq 0$  を満たすための必要十分条件は、すべての  $i = 1, \dots, n$  に対して  $\lambda_i \geq 0$  であることを示せ。

**定義 1.3** 実対称行列  $A$  の固有値が全て正のとき、 $A$  を正定値行列といい、実対称行列  $A$  の固有値が全て非負のとき、 $A$  を非負定値行列という。

**問題 1.8**  $A$  が非負定値行列であるとき、 $\tilde{A} = H D^{\frac{1}{2}} H^T$  と定めるとき、 $\tilde{A}$  は非負定値行列であり、 $\tilde{A}^2 = A$  が成り立つことを示せ。ただし、 $D^{\frac{1}{2}}$  は  $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$  を対角成分にもつ対角行列である。

**問題 1.9**  $A$  が零行列でない非負定値行列であるとき、 $\mathbf{x} \in \mathbb{R}^n$  が  $x_1^2 + \dots + x_n^2 = 1$  を満たすときの  $\mathbf{x}^T A \mathbf{x}$  の最大値が  $\lambda_1, \dots, \lambda_n$  の最大値に等しいことを示せ。

## 2 多次元確率変数の基礎

統計的推測では、各手法の精度などについて確率的な評価が必要であり、そのためデータの背景にある確率分布 (母集団分布とみなせる) を考える。ベクトル値変数では、正規分布を多次元化した多変量正規分布を使うことが多いが、今回はそれには触れない。

### 2.1 多次元確率変数の平均分散

まず、多次元の連続型<sup>\*1</sup>確率変数の密度関数や平均・分散などの定義などについてまとめておこう。

**定義 2.1 (密度関数)**  $d$  次元連続型確率変数ベクトル  $\mathbf{X} = (X_1, \dots, X_d)^T$  に対して  $k$  変数関数  $f(x_1, \dots, x_d)$  ( $= f(\mathbf{x})$  と略記) が存在して、任意の事象  $\{\mathbf{X} \in A\}$  ( $A \subset \mathbb{R}^d$ ) の確率が

$$P(\mathbf{X} \in A) = \int_A f(\mathbf{x}) d\mathbf{x} \quad (2.1)$$

で求まるとき、 $f(x_1, \dots, x_d)$  を  $\mathbf{X}$  の (同時) 密度関数という。また、

$$f_i(x_i) = \int_{\mathbb{R}^{d-1}} f(\mathbf{x}) d\mathbf{x}^{(i)} \quad (2.2)$$

を満たす  $f_i(x_i)$  を  $X_i$  の (周辺) 密度関数という。ここで、 $d\mathbf{x}$  は  $d$  個の変数  $x_1, \dots, x_d$  での  $d$  重積分を表し、 $d\mathbf{x}^{(i)}$  は  $x_1, \dots, x_d$  から  $x_i$  のみを除いた  $d-1$  個の変数での  $d-1$  重積分を表す。

**定義 2.2 (平均 (期待値))**  $\mathbf{X}$  の関数  $g(\mathbf{X})$  の平均  $E[g(\mathbf{X})]$  を

$$E[g(\mathbf{X})] = \int_{\mathbb{R}^d} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \quad (2.3)$$

で定義する。

$g(\mathbf{X}) = X_i$  のとき、

$$E[X_i] = \int_{\mathbb{R}^d} x_i f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}} x_i \left( \int_{\mathbb{R}^{d-1}} f(\mathbf{x}) d\mathbf{x}^{(i)} \right) d\mathbf{x}_i = \int_{\mathbb{R}} x_i f_i(x_i) dx_i \quad (2.4)$$

が成り立つので、 $E[X_i]$  は  $X_i$  の周辺分布に関する平均に他ならない。

<sup>\*1</sup> 簡単のために連続型のみを考えることとする。

**定義 2.3** ベクトル値 (行列値の場合もある) 確率変数  $\mathbf{X} = (X_1, X_2, \dots, X_d)^T$  ( $\mathbf{X} = (X_{ij})_{i,j}$ ) の平均  $E[\mathbf{X}]$  は成分毎の期待値で定義する. すなわち

$$E[\mathbf{X}] = (E[X_1], E[X_2], \dots, E[X_d])^T \quad (E[\mathbf{X}] = (E[X_{ij}])_{i,j})$$

**定義 2.4** ベクトル値確率変数  $\mathbf{X} = (X_1, X_2, \dots, X_d)^T$  に対して

$$V[\mathbf{X}] = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T] = (E[(X_i - E[X_i])(X_j - E[X_j])])_{i,j}$$

で定義される行列を  $\mathbf{X}$  の分散共分散行列という (転置の位置に注意).

$V[\mathbf{X}]$  は行列でその  $(i, j)$  成分は  $X_i$  と  $X_j$  の共分散 ( $i = j$  であれば  $X_i$  の分散) である. 1 次元確率変数  $X$  の場合, 定数  $c, d$  に対して期待値と分散に関して

$$E[cX + d] = cE[X] + d \quad (\text{期待値の線形性})$$

$$V[cX + d] = c^2 V[X]$$

が成り立つが, これらはベクトル値確率変数の場合に次のように拡張される.

**定理 2.1** ベクトル値 (行列値) 確率変数  $\mathbf{X}$ , 行列  $A$ , ベクトル  $\mathbf{b}$  に対して

$$E[A\mathbf{X}] = AE[\mathbf{X}], E[\mathbf{X}A] = E[\mathbf{X}]A, E[\mathbf{X} + \mathbf{b}] = E[\mathbf{X}] + \mathbf{b} \quad (2.5)$$

$$V[A\mathbf{X} + \mathbf{b}] = AV[\mathbf{X}]A^T \quad (2.6)$$

分散の公式 (2.6) では, 右から転置したものを掛けていることに注意する. これが 1 次元であれば, 転置したものを右からかけることとそのまま左から掛けることが同じ事となるので, 1 変数の場合の一般化となっている.

**問題 2.1**  $A = (a_{ij})$  を  $k \times d$  行列,  $\mathbf{X} = (X_1, \dots, X_d)^T$  を  $d$  次元確率変数ベクトルとして,  $E[A\mathbf{X}] = AE[\mathbf{X}]$  を示せ. ただし,  $\mathbf{X} = (X_1, \dots, X_d)^T$  の同時密度関数は  $f(\mathbf{x})$  とせよ.

**問題 2.2** (2.5) の関係を用いて (2.6) を示せ.

### 3 多次元ベクトルデータのハンドリング

この資料で扱うデータは原則,  $d$  次元のベクトルが  $n$  個集まったもの

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$$

と考える.  $\mathbf{x}_i$  は  $d$  次元ベクトルであるから,  $\mathbf{x}_i = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{di} \end{pmatrix}$  と表せる.  $n$  個をまとめて  $d \times n$  行列

$$X = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n) = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{d1} & \cdots & x_{dn} \end{pmatrix}$$

とおく.  $X$  をデータ行列と呼ぶ.

この節では, データ処理の上で基本となる, 平均, 分散 (より一般には線形変換と二次形式) の計算に関する, 代数的な表現や性質について整理する.

### 3.1 標本平均と標本分散の表現

データ  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  に対して,

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad (3.1)$$

を標本平均ベクトルという.  $\bar{\mathbf{x}}$  の第  $i$  成分はデータ行列  $X$  の第  $i$  行の成分の平均値  $\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$  に等しい. また,

$$S_X = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (3.2)$$

を標本分散共分散行列という.

$S_X$  の  $(i, j)$  成分  $s_{ij}$  は  $s_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$  に等しく, これはデータ行列  $X$  の第  $i$  行と第  $j$  行の標本共分散を表している.

**問題 3.1**  $S_X$  の  $(i, j)$  成分  $s_{ij}$  は  $s_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$  に等しいことを確認せよ.

**問題 3.2**  $\mathbf{x}_i$  が  $k$  次元ベクトルのとき  $(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$  および  $(\mathbf{x}_i - \bar{\mathbf{x}})^T(\mathbf{x}_i - \bar{\mathbf{x}})$  の (行列としての) サイズを答えよ. (転置  $^T$  の位置の違いに注意する. )

以上の計算をベクトル・行列の演算として表そう. まず, (3.1) は  $n$  次元ベクトル  $\mathbf{p} = \frac{1}{n} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$  を使って,

$$\bar{\mathbf{x}} = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n) \mathbf{p} = X\mathbf{p} \quad (\bar{\mathbf{x}} = \mathbf{p}X \text{ ではない})$$

と表すことができるので,  $\bar{\mathbf{x}}$  を  $n$  列並べた行列 (標本平均行列と呼ぶ) は

$$\bar{X} = (\bar{\mathbf{x}} \quad \cdots \quad \bar{\mathbf{x}}) = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n) (\mathbf{p} \quad \cdots \quad \mathbf{p}) = XP \quad (3.3)$$

と表せる. ただし,  $P$  はすべての成分が  $\frac{1}{n}$  の  $n$  次正方行列  $\frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$  である. 一方, (3.2) については,

$$\begin{aligned} S_X &= \frac{1}{n} (\mathbf{x}_1 - \bar{\mathbf{x}} \quad \cdots \quad \mathbf{x}_n - \bar{\mathbf{x}}) (\mathbf{x}_1 - \bar{\mathbf{x}} \quad \cdots \quad \mathbf{x}_n - \bar{\mathbf{x}})^T = (X - \bar{X})(X - \bar{X})^T \\ &= (X - XP)(X - XP)^T = X(I - P)(I - P)^T X^T = X(I - P)X^T \end{aligned} \quad (3.4)$$

と表すことができる.

**問題 3.3**  $I - P$  は対称行列かつべき等行列である, すなわち  $(I - P)^T = I - P$ ,  $(I - P)^2 = I - P$  が成り立つことを示せ.

このようにデータの平均や分散共分散は, データ行列  $X$  と単位行列  $I$ , そして  $P$  を用いて簡潔にあらわされることが分かる.

**定理 3.1** データ行列  $X$  で与えられるデータについて, 平均行列  $\bar{X}$  と分散共分散行列  $S_X$  はそれぞれ

$$\bar{X} = XP, \quad S_X = X(I - P)X^T \quad (3.5)$$

で与えられる.

### 3.2 データの線形変換

$d$  次元ベクトル  $\mathbf{x}_1, \dots, \mathbf{x}_n$  を  $p$  次元ベクトル  $\mathbf{y}_1, \dots, \mathbf{y}_n$  に変換する一次関係

$$\mathbf{y}_i = A\mathbf{x}_i + \mathbf{b} \quad (i = 1, \dots, n)$$

を考える. ただし,  $A$  は  $p \times d$  行列,  $\mathbf{b}$  は  $p$  次元ベクトルである. データ行列  $Y = (\mathbf{y}_1 \ \dots \ \mathbf{y}_n)$  について

$$Y = (A\mathbf{x}_1 + \mathbf{b} \ \dots \ A\mathbf{x}_n + \mathbf{b}) = A(\mathbf{x}_1 \ \dots \ \mathbf{x}_n) + (\mathbf{b} \ \dots \ \mathbf{b}) = AX + B \quad (3.6)$$

と表される. ただし,  $B$  はすべての列がベクトル  $\mathbf{b}$  の  $p \times n$  行列である. 変換後のデータ行列  $Y$  に対して平均  $\bar{Y}$ , 分散共分散行列  $S_Y$  を定理 2.1 より求めると

$$\begin{aligned} Y &= YP = (AX + B)P = AXP + BP = AX + B \\ S_Y &= (AX + B)(I - P)(AX + B)^T = (AX + B)(I - P)(X^T A^T + B^T) \\ &= AX(I - P)X^T A^T + AX(I - P)B^T + B(I - P)X^T A^T + B(I - P)B^T \\ &= AX(I - P)X^T A^T = AS_X A^T \end{aligned}$$

を得る.

**問題 3.4**  $BP = B, B(I - P) = O$  が成り立つことを確認せよ.

**定理 3.2** データ行列の変換  $Y = AX + B$  を考えるとき, 変換前後の平均および分散共分散行列について次が成り立つ.

$$\bar{Y} = A\bar{X}, \quad S_Y = AS_X A^T \quad (3.7)$$