# 線形回帰モデル（一般化最小二乗法，最尤法）

## 1　一般化最小二乗法

次の線形回帰モデルを考える。

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

ここで，$p < n$ として，

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

とする。ただし，$E(\boldsymbol{\varepsilon}) = \boldsymbol{0}, V(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega}$ とし，$\boldsymbol{\Omega}$ は $n \times n$ の正定値対称行列とする。また，前回の講義では，$\boldsymbol{\varepsilon}$ に関して，

条件 1: $E[\boldsymbol{\varepsilon}] = \boldsymbol{0}$
条件 2: $V[\boldsymbol{\varepsilon}] = \sigma^2\boldsymbol{I}_n$

を仮定したが，(1) 式の線形回帰モデルは，条件 2 を満たすとは限らない。つまり，ガウス・マルコフの定理は，条件 1 と 2 を仮定しているため，最小二乗推定量

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$$

は最良線形不偏推定量になるとは限らない。そのため，$\boldsymbol{I}_n \neq \boldsymbol{\Omega}$ の場合における最良線形不偏推定量を得ることを考える。

$\boldsymbol{\Omega}$ は正定値対称行列であることから，適当な直交行列 $\boldsymbol{W}$ により次式のように対角化可能である。

$$\boldsymbol{W}^\top\boldsymbol{\Omega}\boldsymbol{W} = \boldsymbol{\Lambda}$$

ここで，$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_0, \lambda_1, \dots, \lambda_p)$ $(\lambda_i > 0, i = 0, 1, \dots, p)$ であることから，次式が成り立つ。

$$\begin{aligned} \boldsymbol{\Omega} &= \boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{W}^\top \\ &= \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^\top \\ &= \boldsymbol{W}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{W}^\top \\ &= (\boldsymbol{W}\boldsymbol{\Lambda}^{1/2})(\boldsymbol{W}\boldsymbol{\Lambda}^{1/2})^\top \end{aligned}$$

$\boldsymbol{P} = (\boldsymbol{W}\boldsymbol{\Lambda}^{1/2})^{-1} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{W}^{-1}$ とすると，$\boldsymbol{P}\boldsymbol{\Omega}\boldsymbol{P}^\top = \boldsymbol{I}_n$ となる。また，$\boldsymbol{W}$ は直交行列であるので，$\boldsymbol{W}^\top = \boldsymbol{W}^{-1}$ となる。したがって，次式が成り立つ。

$$\begin{aligned} \boldsymbol{P}^\top &= ((\boldsymbol{W}\boldsymbol{\Lambda}^{1/2})^{-1})^\top \\ &= (\boldsymbol{\Lambda}^{-1/2}\boldsymbol{W}^{-1})^\top \\ &= (\boldsymbol{W}^{-1})^\top(\boldsymbol{\Lambda}^{-1/2})^\top \\ &= \boldsymbol{W}\boldsymbol{\Lambda}^{-1/2} \end{aligned}$$

$$\begin{aligned}\boldsymbol{P}^\top\boldsymbol{P} &= \boldsymbol{W}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{W}^{-1}\\ &= \boldsymbol{W}\boldsymbol{\Lambda}^{-1}\boldsymbol{W}^{-1}\\ &= (\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^{-1})^{-1}\\ &= (\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^\top)^{-1}\\ &= \boldsymbol{\Omega}^{-1}\end{aligned}$$

(1) 式の線形回帰モデルの両辺に左から $\boldsymbol{P}$ をかけると，次式が成り立つ。

$$\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{P}\boldsymbol{\varepsilon} \tag{2}$$

このとき，次式が成り立つため，(2) 式は，ガウス・マルコフの定理が成り立つための条件 1 と 2 を満たしている。

$$E[\boldsymbol{P}\boldsymbol{\varepsilon}] = \boldsymbol{P}E[\boldsymbol{\varepsilon}] = \boldsymbol{0}, \quad V[\boldsymbol{P}\boldsymbol{\varepsilon}] = \boldsymbol{P}V[\boldsymbol{\varepsilon}]\boldsymbol{P}^\top = \sigma^2\boldsymbol{P}\boldsymbol{\Omega}\boldsymbol{P}^\top = \sigma^2\boldsymbol{I}_n$$

したがって，(2) 式の線形回帰モデルに対して，最小二乗法を用いることで，最良線形不偏推定量を得ることを考える。つまり，次式の関数の最小化問題を考える。

$$\begin{aligned}f(\boldsymbol{\beta}) &= \|\boldsymbol{P}\boldsymbol{\varepsilon}\|^2\\ &= \|\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta}\|^2\\ &= (\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{P}\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta})\\ &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{P}^\top\boldsymbol{P}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\\ &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\\ &= (\boldsymbol{Y}^\top\boldsymbol{\Omega}^{-1} - \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\\ &= \boldsymbol{Y}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} - \boldsymbol{Y}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{\beta}\\ &= \boldsymbol{Y}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} - 2\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{\beta} \quad (\because \text{前回講義資料補足 1 と同様})\end{aligned}$$

$\dfrac{\partial f(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \boldsymbol{0}$ を解くと，

$$\begin{aligned}\frac{\partial f(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= -2\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} + 2\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{\beta}\\ &\left(\because \frac{\partial\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y}}{\partial\boldsymbol{\beta}} = \boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y}, \quad \frac{\partial\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{\beta}}{\partial\boldsymbol{\beta}} = 2\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{\beta}\right)\end{aligned}$$

であることから，次式を得る。

$$\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \tag{3}$$

この連立方程式 (3) の解を求める方法を**一般化最小二乗法**という。$\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}$ が正則であれば，**一般化最小二乗推定量**は次式で与えられる。

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \tag{4}$$

## 2 最尤法

線形回帰モデルにおける最小二乗法と最尤法の関係を考える。多変量確率変数 $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ の関数を $f(\boldsymbol{y}|\boldsymbol{\theta})$ $(\boldsymbol{\theta} \in \boldsymbol{\Theta})$ で定義する。ここで，$f(\boldsymbol{y}|\boldsymbol{\theta})$ は，$\boldsymbol{Y}$ が

連続型であれば確率密度関数，離散型であれば確率関数である。$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$ は，$p \times 1$ の未知のパラメータベクトルであり，$\boldsymbol{\Theta}$ はパラメータ空間である。

$\boldsymbol{Y}$ の観測値 $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ が得られたとき，$\boldsymbol{\theta}$ の関数とみなした $\boldsymbol{\theta}$ の尤度関数を次のように定義する。

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\theta}) \tag{5}$$

確率（密度）関数 $f(\boldsymbol{y}|\boldsymbol{\theta})$ は，$\boldsymbol{\theta}$ が与えられたとき，どのような $\boldsymbol{y}$ が得られやすいかを示す関数である。一方，尤度関数 $L(\boldsymbol{\theta}|\boldsymbol{y})$ は，$\boldsymbol{y}$ が与えられたとき，それが出現しやすいパラメータの値を示す関数である。

尤度関数 $L(\boldsymbol{\theta}|\boldsymbol{y})$ を最大化することにより，未知パラメータ $\boldsymbol{\theta}$ を推定する方法を**最尤法**という。最尤法により得られる $\boldsymbol{\theta}$ の推定量

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{Y}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta}|\boldsymbol{Y})$$

を**最尤推定量**といい，$\boldsymbol{Y}$ を $\boldsymbol{y}$ で置き換えた $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{y})$ を**最尤推定値**という。したがって，最尤法では，何かしらの統計モデルを仮定し，得られた $\boldsymbol{y}$ から尤度関数 $L(\boldsymbol{\theta}|\boldsymbol{y})$ を最大にする未知パラメータ $\boldsymbol{\theta}$ の推定問題を考えている。

最尤推定量を求める場合，(5) 式の尤度関数を対数変換した**対数尤度関数** $l(\boldsymbol{\theta}|\boldsymbol{y}) = \log L(\boldsymbol{\theta}|\boldsymbol{y})$ の最大化

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{Y}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l(\boldsymbol{\theta}|\boldsymbol{Y})$$

を考えることが多い。最尤推定量 $\hat{\boldsymbol{\theta}}$ は，次式から得られる方程式の解として求めることが一般的である。

$$\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \dfrac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_1} \\ \dfrac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_2} \\ \vdots \\ \dfrac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_p} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \boldsymbol{0}$$

この方程式を**尤度方程式**という。

## 3 線形回帰モデルの最尤推定問題

次の線形モデルの最尤推定問題を考える。

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n) \tag{6}$$

ここで，$p < n$ として，

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{p1} \\ 1 & x_{12} & x_{22} & \ldots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{pn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

とする。$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ であることから，$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$ となる（補足 2 参照)。

補足 2

多変量確率変数 $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$ の積率母関数は，$\boldsymbol{\theta}^\top = (\theta_1, \theta_2, \ldots, \theta_n)$ とするとき，次式で定義される。

$$\phi_{\boldsymbol{x}}(\boldsymbol{\theta}) = E[\exp(\boldsymbol{\theta}^\top \boldsymbol{x})]$$

$\phi_{\boldsymbol{x}}(\boldsymbol{\theta})$ を適当回数微分することにより，任意の次数のモーメントを求めることができる。つまり，$\sum_{i=1}^n k_i = k$ とすれば，次式が成り立つ。

$$\left[\frac{\partial^k \phi_{\boldsymbol{x}}(\boldsymbol{\theta})}{\partial\theta_1^{k_1} \partial\theta_2^{k_2} \cdots \partial\theta_n^{k_n}}\right] = E\left[x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}\right]$$

$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ のとき，$\boldsymbol{x}$ の確率密度関数は次式となる。

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right]$$

多変量正規分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ の積率母関数は次式によって与えられる。

$$\phi_{\boldsymbol{x}}(\boldsymbol{\theta}) = \exp\left(\boldsymbol{\theta}^\top \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}\right)$$

$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ のとき，$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{\lambda} \sim N(\boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{\lambda}, \boldsymbol{C}^\top \boldsymbol{\Sigma} \boldsymbol{C})$ となることを示す。ただし，$\boldsymbol{C}$ は $p \times n$ 行列，$\mathrm{rank}(\boldsymbol{C}) = p$，$\boldsymbol{\lambda}$ は $p$ 次元定数ベクトルである。$\boldsymbol{y}$ の積率母関数が $N(\boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{\lambda}, \boldsymbol{C}^\top \boldsymbol{\Sigma} \boldsymbol{C})$ の積率母関数であることを示せばよい。

$$\begin{aligned}
\phi_{\boldsymbol{y}}(\boldsymbol{\theta}) &= E\left[\exp\left(\boldsymbol{\theta}^\top (\boldsymbol{C}\boldsymbol{x} + \boldsymbol{\lambda})\right)\right] \\
&= E\left[\exp\left(\boldsymbol{\theta}^\top \boldsymbol{C}\boldsymbol{x} + \boldsymbol{\theta}^\top \boldsymbol{\lambda}\right)\right] \\
&= \exp\left(\boldsymbol{\theta}^\top \boldsymbol{\lambda}\right) E\left[\exp\left(\boldsymbol{\theta}^\top \boldsymbol{C}\boldsymbol{x}\right)\right] \\
&= \exp\left(\boldsymbol{\theta}^\top \boldsymbol{\lambda}\right) \phi_{\boldsymbol{x}}(\boldsymbol{C}^\top \boldsymbol{\theta}) \\
&= \exp\left(\boldsymbol{\theta}^\top \boldsymbol{\lambda}\right) \exp\left(\boldsymbol{\theta}^\top \boldsymbol{C}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top \boldsymbol{\theta}\right) \\
&= \exp\left(\boldsymbol{\theta}^\top (\boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{\lambda}) + \frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top \boldsymbol{\theta}\right)
\end{aligned}$$

以上より，$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ のとき，$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{\lambda} \sim N(\boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{\lambda}, \boldsymbol{C}^\top \boldsymbol{\Sigma} \boldsymbol{C})$ となる。

このとき，尤度関数は次式となる。

$$\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}) &= \frac{1}{(2\pi)^{n/2} |\sigma^2 \boldsymbol{I}_n|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\sigma^2 \boldsymbol{I}_{\boldsymbol{n}})^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\right] \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\right]
\end{aligned}$$

したがって，対数尤度関数は次式となる。

$$
\begin{aligned}
l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}) &= \log L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}) \\
&= \log \left[ \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \right] \right] \\
&= -\frac{n}{2} \log \left( 2\pi\sigma^2 \right) - \frac{1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \qquad (7)
\end{aligned}
$$

$\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)^\top$ とし，次式のように対数尤度関数 $l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y}) = l(\boldsymbol{\theta} | \boldsymbol{Y})$ の最大化を考える。

$$
\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{Y}) = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} l(\boldsymbol{\theta} | \boldsymbol{Y})
$$

(7) 式より，対数尤度関数を $\boldsymbol{\beta}$ に関して最大化することと，$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$ を $\boldsymbol{\beta}$ に関して最小化することは同値であることがわかる。つまり，最小二乗法は，$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$ を $\boldsymbol{\beta}$ に関して最小化する方法であることから，$\boldsymbol{\beta}$ の最尤推定量と最小二乗推定量は一致することがわかる。実際に，$\boldsymbol{\beta}$ と $\sigma^2$ の最尤推定量を求める。

$$
\frac{\partial l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y})}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2} (-\boldsymbol{X}^\top \boldsymbol{Y} + \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}) \quad (\equiv \boldsymbol{0}) \qquad (8)
$$

$$
\frac{\partial l(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{Y})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \quad (\equiv 0) \qquad (9)
$$

(8) 式より，次式を得る。

$$
\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{Y}
$$

したがって，$\boldsymbol{X}^\top \boldsymbol{X}$ が正則であれば，$\boldsymbol{\beta}$ の最尤推定量は次式となる。

$$
\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}
$$

以上のことから，$\boldsymbol{\beta}$ の最尤推定量と最小二乗推定量は一致する。

一方，(9) 式より，次式を得る。

$$
\begin{aligned}
\frac{n}{2\sigma^2} &= \frac{1}{2(\sigma^2)^2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\
\Leftrightarrow \sigma^2 &= \frac{1}{n} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})
\end{aligned}
$$

したがって，$\sigma^2$ の最尤推定量は次式となる。

$$
\hat{\sigma}^2 = \frac{1}{n} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})
$$

問 2

$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, $\mathrm{rank}(\boldsymbol{X}) = p+1$, $\hat{\sigma}^2 = \dfrac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$, $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$ とする。

(1) 残差ベクトル $\boldsymbol{e}$ を $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ として定義する。$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ としたとき，次式が成り立つことを示せ。

$$\boldsymbol{e} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{\varepsilon}$$

(2) $E[\boldsymbol{e}^\top\boldsymbol{e}]$ を求めよ。

(3) $\hat{\sigma}^2$ は $\sigma^2$ の不偏推定量であるかどうかを確認せよ。

問 3

次の線形モデルの最尤推定問題を考える。

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega})$$

ここで，$p < n$ として，

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

とする。このとき，$\boldsymbol{\beta}$ の最尤推定量を求めよ。

# 参考文献

[1] 佐和隆光．(2020)．回帰分析（新装版）．朝倉書店.

[2] 森裕一，黒田正博，足立浩平．(2017)．最小二乗法・交互最小二乗法．共立出版.