データ解析

線形回帰モデル

東京理科大学
創域理工学部情報計算科学科
安藤宗司

2023年10月5日

Contents

線形回帰モデル

次の線形回帰モデルを考える.

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

ここで, $p + 1 < n$とし,

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$n \times 1$行列　$n \times (p+1)$行列　$(p+1) \times 1$行列　$n \times 1$行列

とする．ただし，$E[\boldsymbol{\varepsilon}] = \boldsymbol{0}, V[\boldsymbol{\varepsilon}] = \sigma^2 \boldsymbol{I_n}, \mathrm{rank}(\boldsymbol{X}) = p + 1$ とする.

このとき，$\|\boldsymbol{\varepsilon}\|^2 = (\boldsymbol{Y} - \boldsymbol{X\beta})^\top(\boldsymbol{Y} - \boldsymbol{X\beta})$ を最小にする $\boldsymbol{\beta}$ として,

$\boldsymbol{\beta}$の最小二乗推定量$\widetilde{\boldsymbol{\beta}}$は次式で与えられる.

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

最小二乗法

$\|\boldsymbol{\varepsilon}\|^2 = (\boldsymbol{Y}-\boldsymbol{X\beta})^\top(\boldsymbol{Y}-\boldsymbol{X\beta})$を最小にする$\boldsymbol{\beta}$を求める.

$$\begin{aligned}\|\boldsymbol{\varepsilon}\|^2 &= (\boldsymbol{Y}-\boldsymbol{X\beta})^\top(\boldsymbol{Y}-\boldsymbol{X\beta})\\ &= \boldsymbol{Y}^\top\boldsymbol{Y} - \boldsymbol{Y}^\top\boldsymbol{X\beta} - \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X\beta}\\ &= \boldsymbol{Y}^\top\boldsymbol{Y} - 2\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{\beta}^\top(\boldsymbol{X}^\top\boldsymbol{X})\boldsymbol{\beta}\end{aligned}$$

$$\boldsymbol{Y}^\top\boldsymbol{X\beta} = \underbrace{\begin{pmatrix}Y_1 & Y_2 & \cdots & Y_n\end{pmatrix}}_{1\times n\text{行列}}\underbrace{\begin{pmatrix}1 & x_{11} & x_{21} & \cdots & x_{p1}\\ 1 & x_{12} & x_{22} & \cdots & x_{p2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn}\end{pmatrix}}_{n\times(p+1)\text{行列}}\underbrace{\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{pmatrix}}_{(p+1)\times 1\text{行列}} = \begin{pmatrix}\sum_{i=1}^{n} Y_i & \sum_{i=1}^{n} Y_i x_{1i} & \cdots & \sum_{i=1}^{n} Y_i x_{pi}\end{pmatrix}\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{pmatrix} = \sum_{i=1}^{n} Y_i\beta_0 + \sum_{i=1}^{n}\sum_{j=1}^{p} Y_i x_{ji}\beta_j$$

$$\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{Y} = \underbrace{\begin{pmatrix}\beta_0 & \beta_1 & \cdots & \beta_p\end{pmatrix}}_{1\times(p+1)\text{行列}}\underbrace{\begin{pmatrix}1 & 1 & 1 & \cdots & 1\\ x_{11} & x_{12} & x_{13} & \cdots & x_{1n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ x_{p1} & x_{p2} & x_{p3} & \cdots & x_{pn}\end{pmatrix}}_{(p+1)\times n\text{行列}}\underbrace{\begin{pmatrix}Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix}}_{n\times 1\text{行列}} = \begin{pmatrix}\beta_0 + \sum_{j=1}^{p} x_{j1}\beta_j & \beta_0 + \sum_{j=1}^{p} x_{j2}\beta_j & \cdots & \beta_0 + \sum_{j=1}^{p} x_{jn}\beta_j\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix}$$

$$= \sum_{i=1}^{n} Y_i\beta_0 + \sum_{i=1}^{n}\sum_{j=1}^{p} Y_i x_{ji}\beta_j$$

最小二乗推定量

$$\|\boldsymbol{\varepsilon}\|^2 = \boldsymbol{Y}^\top\boldsymbol{Y} - 2\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{Y} + \boldsymbol{\beta}^\top(\boldsymbol{X}^\top\boldsymbol{X})\boldsymbol{\beta}$$

$\dfrac{\partial\|\boldsymbol{\varepsilon}\|^2}{\partial\boldsymbol{\beta}} = \boldsymbol{0}$を解くと

$$\frac{\partial\|\boldsymbol{\varepsilon}\|^2}{\partial\boldsymbol{\beta}} = -2\boldsymbol{X}^\top\boldsymbol{Y} + 2(\boldsymbol{X}^\top\boldsymbol{X})\boldsymbol{\beta} \qquad \left(\because \frac{\partial\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{Y}}{\partial\boldsymbol{\beta}} = \boldsymbol{X}^\top\boldsymbol{Y}, \frac{\partial\boldsymbol{\beta}^\top(\boldsymbol{X}^\top\boldsymbol{X})\boldsymbol{\beta}}{\partial\boldsymbol{\beta}} = 2(\boldsymbol{X}^\top\boldsymbol{Y})\boldsymbol{\beta}\right)$$

正規方程式

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

$\mathrm{rank}(\boldsymbol{X}) = p + 1$より，$\mathrm{rank}(\boldsymbol{X}^\top\boldsymbol{X}) = p + 1$となることから,

$(\boldsymbol{X}^T\boldsymbol{X})$は正則であるため，$\boldsymbol{\beta}$の最小二乗推定量$\tilde{\boldsymbol{\beta}}$は次式で与えられる.

$(p+1)\times(p+1)$行列

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

最小二乗推定量の性質 (1)

$\boldsymbol{Y}$ の線形式で表される $\boldsymbol{\beta}$ の線形推定量 $\ddot{\boldsymbol{\beta}}$ を考える.

$$\ddot{\boldsymbol{\beta}} = \boldsymbol{C}\boldsymbol{Y} \qquad \boldsymbol{C}: (p+1) \times n \text{定数行列}$$

次の等式を満たす線形推定量 $\ddot{\boldsymbol{\beta}}$ を線形不偏推定量という.

$$E(\ddot{\boldsymbol{\beta}}) = E(\boldsymbol{C}\boldsymbol{Y}) = E(\boldsymbol{C}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})) = \underbrace{\boldsymbol{C}\boldsymbol{X}}_{(p+1)\times(p+1)\text{行列}} E(\boldsymbol{\beta}) = \boldsymbol{\beta}$$

つまり, $\boldsymbol{C}\boldsymbol{X} = \boldsymbol{I}_{p+1}$ となるとき線形推定量 $\ddot{\boldsymbol{\beta}}$ は線形不偏推定量となる.

$\boldsymbol{\beta}$ の最小二乗推定量 $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$ は線形不偏推定量となる.

$$\boldsymbol{C}\boldsymbol{X} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{I}_{p+1}$$

$\ddot{\boldsymbol{\beta}}$ の分散共分散行列は次式で与えられる.

$$V(\ddot{\boldsymbol{\beta}}) = V(\boldsymbol{C}\boldsymbol{Y}) = \boldsymbol{C}V(\boldsymbol{Y})\boldsymbol{C}^\top = \boldsymbol{C}\sigma^2 \boldsymbol{I}_n \boldsymbol{C}^\top = \sigma^2 \boldsymbol{C}\boldsymbol{C}^\top$$

最小二乗推定量の性質 (2)

$\boldsymbol{C}^* = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$とし，$\boldsymbol{C} = \boldsymbol{D} + \boldsymbol{C}^*$とする.

$\boldsymbol{C}\boldsymbol{X} = \boldsymbol{I}_{p+1}$となるとき$\ddot{\boldsymbol{\beta}}$は線形不偏推定量になることに注意すると

$$\boldsymbol{I}_{p+1} = \boldsymbol{C}\boldsymbol{X} = (\boldsymbol{D} + \boldsymbol{C}^*)\boldsymbol{X} = \boldsymbol{D}\boldsymbol{X} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{D}\boldsymbol{X} + \boldsymbol{I}_{p+1}$$

となることから，$\boldsymbol{D}\boldsymbol{X} = \boldsymbol{0}$を得る.

$\boldsymbol{X}^\top \boldsymbol{X}$は対称行列であるため，
$(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$も対称行列

$$(\boldsymbol{X}^\top \boldsymbol{X})^\top = \boldsymbol{X}^\top \boldsymbol{X}$$

$$\begin{aligned}
\boldsymbol{C}\boldsymbol{C}^T &= (\boldsymbol{D} + \boldsymbol{C}^*)(\boldsymbol{D} + \boldsymbol{C}^*)^\top \\
&= (\boldsymbol{D} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)(\boldsymbol{D} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\
&= \boldsymbol{D}\boldsymbol{D}^\top + \boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{D}^\top + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \\
&= \boldsymbol{D}\boldsymbol{D}^\top + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \qquad (\because \boldsymbol{D}\boldsymbol{X} = (\boldsymbol{D}\boldsymbol{X})^\top = \boldsymbol{0})
\end{aligned}$$

最小二乗推定量の性質 (3)

$\ddot{\boldsymbol{\beta}}$の分散共分散行列は次式となる．

$$V\left(\ddot{\boldsymbol{\beta}}\right) = \sigma^2 \boldsymbol{C}\boldsymbol{C}^\top = \sigma^2 (\boldsymbol{D}\boldsymbol{D}^\top + (\boldsymbol{X}^\top \boldsymbol{X})^{-1})$$
$$= \sigma^2 \boldsymbol{D}\boldsymbol{D}^\top + \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$
$$= \sigma^2 \boldsymbol{D}\boldsymbol{D}^\top + V\left(\tilde{\boldsymbol{\beta}}\right) \qquad \left(\because V\left(\tilde{\boldsymbol{\beta}}\right) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\right)$$

練習問題 (4)と同様に導出可能

$V\left(\ddot{\boldsymbol{\beta}}\right) \geq V\left(\tilde{\boldsymbol{\beta}}\right)$が成り立つことを示す．

つまり，最小二乗推定量 $\tilde{\boldsymbol{\beta}}$は，線形不偏推定量の中で最小の

分散共分散行列をもつ，最良線形不偏推定量であることを示す．

この性質をガウス・マルコフの定理という．

補足資料：行列の大小関係

大小関係

対称行列$\boldsymbol{A}$に対して，$\boldsymbol{A} \geq \boldsymbol{0}$を$\boldsymbol{A}$が非負定値行列であることを表すとする．

半順序関係

対称行列$\boldsymbol{A}, \boldsymbol{B}$に対して，$\boldsymbol{A} - \boldsymbol{B}$が非負定値行列であるとき，

非負定値行列上の半順序関係を$\boldsymbol{A} \geq \boldsymbol{B}$と表す．

(1)反射律，(2)反対称律，(3)推移律を満たす．

(1)反射律

$$\boldsymbol{A} - \boldsymbol{A} = \boldsymbol{0} \geq \boldsymbol{0}$$
$$\Rightarrow \boldsymbol{A} \geq \boldsymbol{A}$$

(2)反対称律

$$\boldsymbol{A} \geq \boldsymbol{B}, \boldsymbol{B} \geq \boldsymbol{A}$$
$$\Rightarrow \boldsymbol{A} = \boldsymbol{B}$$

(3)推移律

$$\boldsymbol{A} \geq \boldsymbol{B}, \boldsymbol{B} \geq \boldsymbol{C}$$
$$\Rightarrow \boldsymbol{A} \geq \boldsymbol{C}$$

最小二乗推定量の性質 (4)

$\boldsymbol{C}\boldsymbol{C}^{\top} = \boldsymbol{D}\boldsymbol{D}^{\top} + (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}$ より $\boldsymbol{C}\boldsymbol{C}^{\top} - (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1} = \boldsymbol{D}\boldsymbol{D}^{\top}$ が成り立つ.

$$(\boldsymbol{C}\boldsymbol{C}^{\top} - (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1})^{\top} = \boldsymbol{C}\boldsymbol{C}^{\top} - (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}$$

が成り立つことから， $\boldsymbol{C}\boldsymbol{C}^{\top} - (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}$ は対称行列である.

したがって，$\boldsymbol{D}\boldsymbol{D}^{\top}$ は非負定値行列である.

ただし，$\mathrm{rank}(\boldsymbol{D}) = \mathrm{rank}(\boldsymbol{D}\boldsymbol{D}^{\top}) = p + 1$ である.

非負定値行列の性質

$m \times m$ 対称行列 $\boldsymbol{A}$ が非負定値行列であるための必要十分条件は,

$\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^{T}$ となる $m \times k$ 行列 $\boldsymbol{B}$ が存在することである.

ただし，$\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{B}) = k$ である.

最小二乗推定量の性質 (5)

任意の$p+1$次元ベクトル$\boldsymbol{a} = \left(a_1, \dots, a_{p+1}\right)^\top$に対して次式が成り立つ.

$$\boldsymbol{a}^\top(\sigma^2\boldsymbol{D}\boldsymbol{D}^\top)\boldsymbol{a} = \boldsymbol{a}^\top(\sigma^2\boldsymbol{C}\boldsymbol{C}^\top - \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1})\boldsymbol{a}$$

$$= \boldsymbol{a}^\top\left(V(\ddot{\boldsymbol{\beta}}) - V(\tilde{\boldsymbol{\beta}})\right)\boldsymbol{a}$$

$$\geq 0$$

したがって，次式が成り立つ.

$$\boldsymbol{a}^\top V(\ddot{\boldsymbol{\beta}})\boldsymbol{a} \geq \boldsymbol{a}^\top V(\tilde{\boldsymbol{\beta}})\boldsymbol{a} \iff V(\boldsymbol{a}^\top\ddot{\boldsymbol{\beta}}) \geq V(\boldsymbol{a}^\top\tilde{\boldsymbol{\beta}})$$

$\boldsymbol{a} = (1, \dots, 1)^\top$とすると，次式が成り立つ.

$$V(\ddot{\boldsymbol{\beta}}) \geq V(\tilde{\boldsymbol{\beta}})$$

練習問題1

次の線形回帰モデルを考える.

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

ここで, $p+1<n$とし,

$$\boldsymbol{Y} = \underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{n\times 1\text{行列}}, \boldsymbol{X} = \underbrace{\begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{pmatrix}}_{n\times(p+1)\text{行列}}, \boldsymbol{\beta} = \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{(p+1)\times 1\text{行列}}, \boldsymbol{\varepsilon} = \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}_{n\times 1\text{行列}}$$

とする. ただし, $E[\boldsymbol{\varepsilon}] = \boldsymbol{0}, V[\boldsymbol{\varepsilon}] = \sigma^2\boldsymbol{I_n}, \mathrm{rank}(\boldsymbol{X}) = p+1$ とする.

このとき, $\|\boldsymbol{\varepsilon}\|^2 = (\boldsymbol{Y}-\boldsymbol{X\beta})^T(\boldsymbol{Y}-\boldsymbol{X\beta})$ を最小にする $\boldsymbol{\beta}$ として,

$\boldsymbol{\beta}$の最小二乗推定量$\widetilde{\boldsymbol{\beta}}$は次式で与えられる.

$$\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

以下の各問では, 誤差ベクトル$\boldsymbol{\varepsilon}$が多変量正規分布$N(\boldsymbol{0}, \sigma^2\boldsymbol{I_n})$ に従うと仮定して答えよ.

(1) $\boldsymbol{\beta}$ と σ^2 の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と $\hat{\sigma}^2$ を求めよ．
(2) $\boldsymbol{\beta}$ の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と最小二乗推定量 $\widetilde{\boldsymbol{\beta}}$ の関係を述べよ．

$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_n})$であることから，$\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I_n})$となる．

$\boldsymbol{Y}$の確率密度関数

$$f(\boldsymbol{Y}) = \frac{1}{(2\pi)^{n/2} |\sigma^2 \boldsymbol{I_n}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X\beta})^\top (\sigma^2 \boldsymbol{I_n})^{-1} (\boldsymbol{Y} - \boldsymbol{X\beta})\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X\beta})^\top (\boldsymbol{Y} - \boldsymbol{X\beta})\right]$$

対数尤度関数

$$\log L = \log \frac{1}{(2\pi\sigma^2)^{n/2}} - \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X\beta})^\top (\boldsymbol{Y} - \boldsymbol{X\beta})$$

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

(1) $\boldsymbol{\beta}$ と $\boldsymbol{\sigma}^2$ の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と $\hat{\boldsymbol{\sigma}}^2$ を求めよ．
(2) $\boldsymbol{\beta}$ の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と最小二乗推定量 $\widetilde{\boldsymbol{\beta}}$ の関係を述べよ．

対数尤度関数$\log L$を$\boldsymbol{\beta}$と σ^2に関して最大化することを考える．

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

対数尤度関数を$\boldsymbol{\beta}$に関して最大化する$\Leftrightarrow$ $(\boldsymbol{Y}-\boldsymbol{X\beta})^\top(\boldsymbol{Y}-\boldsymbol{X\beta})$を$\boldsymbol{\beta}$に関して最小化する

最小二乗法は$(\boldsymbol{Y}-\boldsymbol{X\beta})^\top(\boldsymbol{Y}-\boldsymbol{X\beta})$ を $\boldsymbol{\beta}$ に関して最小化する方法であることから，

$\boldsymbol{\beta}$の最小二乗推定量$\widetilde{\boldsymbol{\beta}}$と最尤推定量$\widehat{\boldsymbol{\beta}}$は一致することがわかる．

実際に，$\boldsymbol{\beta}$ と $\boldsymbol{\sigma}^2$ の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と $\hat{\boldsymbol{\sigma}}^2$を求める．

(1) $\boldsymbol{\beta}$ と σ^2 の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と $\hat{\sigma}^2$ を求めよ．
(2) $\boldsymbol{\beta}$ の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と最小二乗推定量 $\widetilde{\boldsymbol{\beta}}$ の関係を述べよ．

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2}\left(-\boldsymbol{X}^\top \boldsymbol{Y} + (\boldsymbol{X}^\top \boldsymbol{X})\boldsymbol{\beta}\right) = \mathbf{0} \qquad \cdots ①$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = 0 \qquad \cdots ②$$

①式より

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

②式より

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

したがって

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

(3) 最尤推定量 $\hat{\boldsymbol{\beta}}$ の期待値を求めよ.

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

(4) 最尤推定量 $\hat{\boldsymbol{\beta}}$ の分散共分散行列を求めよ.

$$
\begin{aligned}
V(\hat{\boldsymbol{\beta}}) &= V((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}) \\
&= \left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)V(\boldsymbol{Y})\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)^\top \\
&= \left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)V(\boldsymbol{Y})(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}) \\
&= \left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)\sigma^2\boldsymbol{I}_{\boldsymbol{n}}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}) \\
&= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}
\end{aligned}
$$

(4) 最尤推定量 $\hat{\boldsymbol{\beta}}$ の分散共分散行列を求めよ．

別解

$$V(\hat{\boldsymbol{\beta}}) = E\left(\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)^\top\right)$$

$\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} - \boldsymbol{\beta} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$が成り立つことから

$$\begin{aligned}
V(\hat{\boldsymbol{\beta}}) &= E\left(((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}))((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}))^\top\right) \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top E\left((\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top\right)((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\
&= \left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)V(\boldsymbol{Y})\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)^\top \\
&= \left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)\sigma^2\boldsymbol{I}_{\boldsymbol{n}}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}) \\
&= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}
\end{aligned}$$

(5) 最尤推定量$\widehat{\boldsymbol{\beta}}$が従う分布を示せ．

最尤推定量$\widehat{\boldsymbol{\beta}}$は，$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$である．

$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I_n})$ であることから，最尤推定量$\widehat{\boldsymbol{\beta}}$は正規分布に従う．

$$\widehat{\boldsymbol{\beta}} \sim N\left(E(\widehat{\boldsymbol{\beta}}), V(\widehat{\boldsymbol{\beta}})\right)$$

ただし，

$$E\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$$

$$V\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

練習問題2

次の線形回帰モデルを考える．

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

ここで，$p + 1 < n$とし，

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

とする．ただし，$\mathrm{rank}(\boldsymbol{X}) = p + 1$，誤差ベクトル$\boldsymbol{\varepsilon}$が

多変量正規分布$N(\boldsymbol{0}, \sigma^2 \boldsymbol{I_n})$ に従うとする．

このとき，$\boldsymbol{\beta}$ と$\boldsymbol{\sigma}^2$の最尤推定量 $\widehat{\boldsymbol{\beta}}$ と $\hat{\boldsymbol{\sigma}}^2$ は次式のようになる．

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

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

(1)

残差ベクトル $\boldsymbol{e}$ を $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ 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.

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

（解答）

$$\begin{aligned}
\boldsymbol{e} &= \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} \\
&= (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{Y} \\
&= (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y} \\
&= (\boldsymbol{I}_n - \boldsymbol{H})(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{\varepsilon} \qquad (\because (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}) \\
&= (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{\varepsilon}
\end{aligned}$$

(2) $E(\boldsymbol{e}^\top \boldsymbol{e})$を求めよ.

$$E(\boldsymbol{e}^\top \boldsymbol{e}) = E\left(\left((\boldsymbol{I_n} - \boldsymbol{H})\boldsymbol{\varepsilon}\right)^\top ((\boldsymbol{I_n} - \boldsymbol{H})\boldsymbol{\varepsilon})\right)$$

$$= E\left(\boldsymbol{\varepsilon}^\top (\boldsymbol{I_n} - \boldsymbol{H})^\top ((\boldsymbol{I_n} - \boldsymbol{H})\boldsymbol{\varepsilon})\right)$$

$$= E(\underset{(1\times n)}{\boldsymbol{\varepsilon}^\top} \underset{(n\times n)}{(\boldsymbol{I_n} - \boldsymbol{H})} \underset{(n\times 1)}{\boldsymbol{\varepsilon}}) \Leftarrow \text{スカラー}$$

$$= E\left(\mathrm{tr}(\boldsymbol{\varepsilon}^\top (\boldsymbol{I_n} - \boldsymbol{H})\boldsymbol{\varepsilon})\right)$$

$$= E\left(\mathrm{tr}((\boldsymbol{I_n} - \boldsymbol{H})\, \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)\right)$$

$$= \mathrm{tr}\left((\boldsymbol{I_n} - \boldsymbol{H}) E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)\right)$$

$$= \mathrm{tr}\left((\boldsymbol{I_n} - \boldsymbol{H}) V(\boldsymbol{\varepsilon})\right) \qquad (\because \boldsymbol{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0})$$

$$= \mathrm{tr}\left((\boldsymbol{I_n} - \boldsymbol{H})\sigma^2 \boldsymbol{I_n}\right)$$

補足

$$(\boldsymbol{I_n} - \boldsymbol{H})^\top = \boldsymbol{I_n} - \boldsymbol{H}^\top$$
$$= \boldsymbol{I_n} - (\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top$$
$$= \boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$$
$$= \boldsymbol{I_n} - \boldsymbol{H}$$

$$\boldsymbol{H}^2 = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{H}$$

$$(\boldsymbol{I_n} - \boldsymbol{H})^\top (\boldsymbol{I_n} - \boldsymbol{H}) = (\boldsymbol{I_n} - \boldsymbol{H})^2$$
$$= \boldsymbol{I_n} - 2\boldsymbol{H} + \boldsymbol{H}^2$$
$$= \boldsymbol{I_n} - \boldsymbol{H}$$

(2) $E(\boldsymbol{e}^\top \boldsymbol{e})$を求めよ．

$$
\begin{aligned}
E(\boldsymbol{e}^\top \boldsymbol{e}) &= E\left(\left((\boldsymbol{I_n} - \boldsymbol{H})\boldsymbol{\varepsilon}\right)^\top ((\boldsymbol{I_n} - \boldsymbol{H})\boldsymbol{\varepsilon}\ )\right) \\
&= \mathrm{tr}\left((\boldsymbol{I_n} - \boldsymbol{H})\sigma^2 \boldsymbol{I_n}\right) \\
&= \sigma^2 \mathrm{tr}\left((\boldsymbol{I_n} - \boldsymbol{H})\right) \\
&= \sigma^2\left(\mathrm{tr}(\boldsymbol{I_n}) - \mathrm{tr}(\boldsymbol{H})\right) \\
&= \sigma^2(n - (p+1)) \\
&= \sigma^2(n - p - 1)
\end{aligned}
$$

補足

$$
\begin{aligned}
\mathrm{tr}(\boldsymbol{H}) &= \mathrm{tr}(\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top) \\
&= \mathrm{tr}\left((\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}\right) \\
&= \mathrm{tr}\left(\boldsymbol{I_{p+1}}\right) \\
&= p + 1
\end{aligned}
$$

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

$$
\begin{aligned}
E(\hat{\sigma}^2) &= E\left(\frac{1}{n}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\right) \\
&= \frac{1}{n}E\left((\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\right) \\
&= \frac{1}{n}E(\boldsymbol{e}^\top\boldsymbol{e}) \qquad (\because \boldsymbol{e}=\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\
&= \frac{1}{n}\sigma^2(n-p-1) \\
&\neq \sigma^2
\end{aligned}
$$

したがって， $\hat{\boldsymbol{\sigma}}^2$は$\boldsymbol{\sigma}^2$の不偏推定量ではない

σ^2の不偏推定量　$\tilde{\sigma}^2 = \dfrac{1}{n-p-1}\boldsymbol{e}^T\boldsymbol{e} = \dfrac{1}{n-p-1}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})$