線形回帰モデル（最小二乗法とその性質）

1　二乗基準での最小化

一つの対象を何回か測定することを考える。このとき，同じ対象を測っているので，得られた観測値は同じであることが期待されるが，一般に，すべての観測値が同じ値をとるわけではない。これは，測定における「誤差」が存在するからである。真の値は存在するが，一般に，それはわならない（真の値がわかっていれば，そもそも測定する必要はない）。そのため，得られた観測値から，その真の値を「推定」する必要がある。真の値を「推定」するための方法の 1 つが**最小二乗法**である。

測定方法が正しければ，観測値は真の値よりも大きかったり，小さかったりするが，真の値に近い値が多く，真の値と遠く離れた値は稀であることが想定できる。つまり，観測された測定値は真の値の周りに分布すると考えられる。真の値と各観測値との差が全体的に小さくなるような値を真の値の推定値として考えるのが理に かなっている。

真の値として「もっともらしい値（もっとも確からしい値，最確値）」を「真の値と各測定値との差を全体的に最小にする」という考え方で求める方法が**最小二乗基準**を使うものである。

2　統計手法への応用

最小二乗法は，想定される関数をいくつかの観測値を用いて近似するとき，対応する関数値と観測値の差の二乗和は最小することにより，もっとも確からしい関数を求める方法である。この方法は，統計手法の多くに利用されている。

2.1　平均値

n 個の観測値 x_i $(i = 1, \ldots, n)$ がある。観測値 x_i が一つの対象を n 回測定した値であるとしたとき，真の値を推定する，あるいはもっともらしい値を求めることを考える。もとめる値を a として，最小二乗基準で推定する。つまり，次式を最小にする a を求めることになる。

$$f(a) = \sum_{i=1}^{n}(x_i - a)^2$$

$f(a)$ は次のように表すことができる。

$$\begin{aligned}
f(a) &= \sum_{i=1}^{n}(x_i^2 - 2x_i a + a^2) \\
&= \sum_{i=1}^{n} x_i^2 - 2a\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} a^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2a\sum_{i=1}^{n} x_i + na^2 \\
&= n\left(a - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 + \sum_{i=1}^{n} x_i^2
\end{aligned}$$

したがって，$f(a)$ が最小になるのは，$a = \frac{1}{n}\sum_{i=1}^{n} x_i$ のときである。つまり，平均値（算術平均）$\bar{x}$ が求める値となる。
(別方法)

$$\frac{\partial f(a)}{\partial a} = -2\sum_{i=1}^{n}(x_i - a)$$

$\frac{\partial f(a)}{\partial a} = 0$ として，a について方程式を解くと，次式を得る。

$$\begin{aligned}
&\sum_{i=1}^{n}(x_i - a) = 0 \\
&\Leftrightarrow \sum_{i=1}^{n} x_i - na = 0 \\
&\Leftrightarrow a = \frac{1}{n}\sum_{i=1}^{n} x_i
\end{aligned}$$

2.2　回帰直線

n 個の観測値の組 (x_i, y_i) $(i = 1, \ldots, n)$ がある。この n 組の観測値が一つの線形モデルから得られたとして，（真の直線上に乗っているべき点が測定による誤差で直線の周りにずれたと想定し）その真の直線を求めたい場合を考える。

求める直線を $y = \beta_0 + \beta_1 x$ とする。もっともらしい直線を求めるために，最小二乗基準，つまり次式の最小化を考える。

$$f(\beta_0, \beta_1) = \sum_{i=1}^{n}\{y_i - (\beta_0 + \beta_1 x)\}^2$$

$$\frac{\partial f(\beta_0, \beta_1)}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \quad (\equiv 0) \tag{1}$$

$$\frac{\partial f(\beta_0, \beta_1)}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i \quad (\equiv 0) \tag{2}$$

(1) と (2) 式を整理すると，次式を得る。

$$\beta_0 n + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{3}$$

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{4}$$

(3) 式より，次式を得る。

$$\beta_0 + \beta_1 \bar{x} = \bar{y} \Leftrightarrow \beta_0 = \bar{y} - \beta_1 \bar{x} \quad \left(\because \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\right) \tag{5}$$

(5) 式を (4) 式に代入することで，次式を得る。

$$
\begin{aligned}
&(\bar{y}-\beta_1\bar{x})\sum_{i=1}^{n}x_i+\beta_1\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}x_iy_i\\
&\Leftrightarrow \left(\sum_{i=1}^{n}x_i^2-\bar{x}\sum_{i=1}^{n}x_i\right)\beta_1=\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\bar{y}\\
&\Leftrightarrow \beta_1=\frac{\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\bar{y}}{\sum_{i=1}^{n}x_i^2-\bar{x}\sum_{i=1}^{n}x_i}=\frac{\sum_{i=1}^{n}x_i(y_i-\bar{y})}{\sum_{i=1}^{n}x_i(x_i-\bar{x})}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}
\end{aligned}
$$

したがって，求める直線のパラメータ β_0 と β_1 は次のようになる。

$$
\begin{aligned}
\hat{\beta}_0&=\bar{y}-\hat{\beta}_1\bar{x}\\
\hat{\beta}_1&=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{S_{xy}}{S_{xx}}=\frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}\frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}}=r\frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}}
\end{aligned}
$$

ここで，S_{xx} は x_i の分散，S_{yy} は y_i の分散，S_{xy} は x_i と y_i の共分散，r は標本相関係数である。

（解法 2）

$$
\begin{aligned}
f(\beta_0,\beta_1)&=\sum_{i=1}^{n}\{y_i-(\beta_0+\beta_1x_i)\}^2\\
&=\sum_{i=1}^{n}\left[\{(y_i-\bar{y})-\beta_1(x_i-\bar{x})\}+(\bar{y}-\beta_0-\beta_1\bar{x})\right]^2\\
&=\sum_{i=1}^{n}\left[(y_i-\bar{y})-\beta_1(x_i-\bar{x})\right]^2+2(\bar{y}-\beta_0-\beta_1\bar{x})\sum_{i=1}^{n}\left[(y_i-\bar{y})-\beta_1(x_i-\bar{x})\right]\\
&\quad+n(\bar{y}-\beta_0-\beta_1\bar{x})^2
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=1}^{n}\left[(y_i-\bar{y})-\beta_1(x_i-\bar{x})\right]^2&=\sum_{i=1}^{n}(y_i-\bar{y})^2-2\beta_1\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+\beta_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2\\
&=n(S_{yy}-2\beta_1S_{xy}+\beta_1^2S_{xx})
\end{aligned}
$$

$$
\sum_{i=1}^{n}\left[(y_i-\bar{y})-\beta_1(x_i-\bar{x})\right]=n\bar{y}-n\bar{y}-\beta_1(n\bar{x}-n\bar{x})=0
$$

したがって，

$$
\begin{aligned}
f(\beta_0,\beta_1)&=n(S_{yy}-2\beta_1S_{xy}+\beta_1^2S_{xx})+n(\bar{y}-\beta_0-\beta_1\bar{x})\\
&=nS_{xx}\left(\beta_1-\frac{S_{xy}}{S_{xx}}\right)^2-n\frac{S_{xy}^2}{S_{xx}}+nS_{yy}+n(\bar{y}-\beta_0-\beta_1\bar{x})^2
\end{aligned}
$$

となることから，$f(\beta_0, \beta_1)$ を最小にする β_0 と β_1 は次のようになる。

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

3 回帰分析

最小二乗法は，統計学では多くの場合，回帰分析を学ぶときに初めて出てくる（皆さんは統計学 2 または多変量解析の講義で初めて学んでいるはず）。回帰分析では，説明変数が二つ以上の場合を想定することも多く，より多くの変数に対しては行列の形で扱うのがよい。

3.1 線形回帰モデル

次の線形回帰モデルを考える。

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

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。また，$\boldsymbol{\varepsilon}$ に関して，

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

を仮定する。ただし，$\boldsymbol{I}_n$ は $n \times n$ 単位行列である。このとき，$\boldsymbol{Y}$ を目的変数，$\boldsymbol{X}$ を説明変数とし，$\boldsymbol{\beta}$ の推測問題を考える。$f(\boldsymbol{\beta}) = \|\boldsymbol{\varepsilon}\|^2 = \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2$ の最小化を**最小二乗問題**といい，その解 $\hat{\boldsymbol{\beta}}$ を求める方法を**最小二乗法**という。最小二乗問題の解 $\hat{\boldsymbol{\beta}}$ は,

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \dfrac{\partial f(\boldsymbol{\beta})}{\partial \beta_0} \\ \dfrac{\partial f(\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{\beta})}{\partial \beta_p} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$$

から得られる連立方程式を $\boldsymbol{\beta}$ について解くことで求めることができる。$f(\boldsymbol{\beta})$ を $\beta_0, \beta_1, \dots, \beta_p$ の一つ一つで偏微分して，それぞれを 0 とおくことにより得られた $p+1$ 個の方程式を連立させたものである。

$$\begin{aligned} f(\boldsymbol{\beta}) = \|\boldsymbol{\varepsilon}\|^2 = \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{Y}^\top \boldsymbol{Y} - \boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{Y} + \boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{Y}^\top \boldsymbol{Y} - 2\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{Y} + \boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} \quad (\because \text{補足 1 を参照}) \end{aligned}$$

補足 1

$$
\begin{aligned}
\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{\beta} &= \underbrace{\begin{pmatrix} Y_1 & Y_2 & \cdots & Y_n \end{pmatrix}}_{1\times n\ \text{行列}} \underbrace{\begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}}_{n\times(p+1)\ \text{行列}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{(p+1)\times 1\ \text{行列}} \\
&= \underbrace{\begin{pmatrix} \sum_{i=1}^n Y_i & \sum_{i=1}^n Y_i x_{1i} & \cdots & \sum_{i=1}^n Y_i x_{pi} \end{pmatrix}}_{1\times(p+1)\ \text{行列}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{(p+1)\times 1\ \text{行列}} \\
&= \sum_{i=1}^n Y_i\beta_0 + \sum_{i=1}^n \sum_{j=1}^p Y_i x_{ji}\beta_j \\
\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{Y} &= \underbrace{\begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_p \end{pmatrix}}_{1\times(p+1)\ \text{行列}} \underbrace{\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & x_{p3} & \dots & x_{pn} \end{pmatrix}}_{(p+1)\times n\ \text{行列}} \underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{n\times 1\ \text{行列}} \\
&= \underbrace{\begin{pmatrix} \sum_{i=1}^n Y_i & \sum_{i=1}^n Y_i x_{1i} & \cdots & \sum_{i=1}^n Y_i x_{pi} \end{pmatrix}}_{1\times(p+1)\ \text{行列}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{(p+1)\times 1\ \text{行列}} \\
&= \sum_{i=1}^n Y_i\beta_0 + \sum_{i=1}^n \sum_{j=1}^p Y_i x_{ji}\beta_j
\end{aligned}
$$

.

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

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

であることから，次式を得る。

$$
\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{Y} \tag{6}
$$

この連立方程式を**正規方程式**という。このとき，$\boldsymbol{X}^\top \boldsymbol{X}$ が正則であれば，(6) 式の解 $\hat{\boldsymbol{\beta}}$ は，次式で与えられる。

$$
\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \tag{7}
$$

これを**最小二乗推定量**という。

問 1

$\boldsymbol{X}^\top \boldsymbol{X}$ が正則であるとき，$\mathrm{rank}(\boldsymbol{X})$ を求めよ。

3.2 最小二乗推定量の性質

最小二乗推定量 $\hat{\boldsymbol{\beta}}$ の期待値は次式で与えられる。

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= E[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}] \\ &= E[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})] \\ &= E[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}] + E[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top E[\boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta} \end{aligned}$$

したがって，最小二乗推定量 $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の不偏推定量である。また，最小二乗推定量 $\hat{\boldsymbol{\beta}}$ の分散 $V[\hat{\boldsymbol{\beta}}]$ は次式で与えられる。

$$\begin{aligned} V[\hat{\boldsymbol{\beta}}] &= V[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}] \\ &= V[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})] \\ &= V[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon}] \\ &= V[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon}] \\ &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top V[\boldsymbol{\varepsilon}]((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\ &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\sigma^2\boldsymbol{I}_n\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \end{aligned}$$

$\boldsymbol{Y}$ の線形式で表される $\boldsymbol{\beta}$ の線形推定量 $\tilde{\boldsymbol{\beta}}$ は次式で与えられる。

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{C}\boldsymbol{Y} \tag{8}$$

ここで，$\boldsymbol{C}$ は $(p+1)\times n$ の定数行列とする。次の等式を満たす線形推定量 $\tilde{\boldsymbol{\beta}}$ を線形不偏推定量という。

$$E[\tilde{\boldsymbol{\beta}}] = E[\boldsymbol{C}\boldsymbol{Y}] = E[\boldsymbol{C}(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})] = E[\boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta}] = \boldsymbol{\beta}$$

したがって，$\tilde{\boldsymbol{\beta}}$ が不偏推定量になるためには，$\boldsymbol{C}\boldsymbol{X} = \boldsymbol{I}_{p+1}$ となる必要がある。また，$\tilde{\boldsymbol{\beta}}$ の分散共分散行列は次式で与えられる。

$$V[\tilde{\boldsymbol{\beta}}] = V[\boldsymbol{C}\boldsymbol{Y}] = \boldsymbol{C}V[\boldsymbol{Y}]\boldsymbol{C}^\top = \boldsymbol{C}V[\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}]\boldsymbol{C}^\top = \boldsymbol{C}V[\boldsymbol{\varepsilon}]\boldsymbol{C}^\top = \sigma^2\boldsymbol{C}\boldsymbol{C}^\top$$

(7) と (8) 式から，最小二乗推定量 $\hat{\boldsymbol{\beta}}$ は，$\boldsymbol{C} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ とした線形不偏推定量であることがわかる。最小二乗推定量 $\hat{\boldsymbol{\beta}}$ は，線形不偏推定量 $\tilde{\boldsymbol{\beta}}$ のなかで，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。

ここで，$\boldsymbol{C}^* = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ とし，$\boldsymbol{C} = \boldsymbol{D} + \boldsymbol{C}^*$ とする。$\boldsymbol{C}\boldsymbol{X} = \boldsymbol{I}_{p+1}$ であることから，次式より $\boldsymbol{D}\boldsymbol{X} = \boldsymbol{0}$ を得る。

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

したがって，次式が成り立つ。

$$\begin{aligned} \boldsymbol{C}\boldsymbol{C}^\top &= (\boldsymbol{D}+\boldsymbol{C}^*)(\boldsymbol{D}+\boldsymbol{C}^*)^\top \\ &= (\boldsymbol{D}+(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)(\boldsymbol{D}+(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\ &= \boldsymbol{D}\boldsymbol{D}^\top + \boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^\top + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &= \boldsymbol{D}\boldsymbol{D}^\top + (\boldsymbol{X}^\top\boldsymbol{X})^{-1} \quad \left(\because \boldsymbol{D}\boldsymbol{X} = (\boldsymbol{D}\boldsymbol{X})^\top = \boldsymbol{0}\right) \end{aligned} \tag{9}$$

(9) 式より，$\tilde{\boldsymbol{\beta}}$ の分散共分散行列は次式のように表すことができる。

$$\begin{aligned} V[\tilde{\boldsymbol{\beta}}] &= \sigma^2\boldsymbol{C}\boldsymbol{C}^\top \\ &= \sigma^2(\boldsymbol{D}\boldsymbol{D}^\top + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}) \\ &= \sigma^2\boldsymbol{D}\boldsymbol{D}^\top + \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &= \sigma^2\boldsymbol{D}\boldsymbol{D}^\top + V[\hat{\boldsymbol{\beta}}] \end{aligned} \tag{10}$$

次に，$\boldsymbol{D}\boldsymbol{D}^\top$ が非負定値行列であることを示す。(9) 式より，$\boldsymbol{C}\boldsymbol{C}^\top - (\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \boldsymbol{D}\boldsymbol{D}^\top$ が成り立つ。次式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。

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

ここで，$\mathrm{rank}(\boldsymbol{D}\boldsymbol{D}^\top) = \mathrm{rank}(\boldsymbol{D}) = p+1$ となることから，$\boldsymbol{D}\boldsymbol{D}^\top$ は非負定値行列である。

非負定値行列の性質

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

$\boldsymbol{D}\boldsymbol{D}^\top$ は非負定値行列であることから，0 ベクトルではない任意の $p+1$ 次元ベクトル $\boldsymbol{a} = (a_0, a_1, \ldots, a_{p+1})^\top$ に対して次式が成り立つ。

$$\boldsymbol{a}^\top(\sigma^2\boldsymbol{D}\boldsymbol{D}^\top)\boldsymbol{a} = \boldsymbol{a}^\top(V[\tilde{\boldsymbol{\beta}}] - V[\hat{\boldsymbol{\beta}}])\boldsymbol{a} \geq 0 \quad (\because (10)\ \text{式})$$

したがって，次式が成り立つ。

$$\boldsymbol{a}^\top V[\tilde{\boldsymbol{\beta}}]\boldsymbol{a} \geq \boldsymbol{a}^\top V[\hat{\boldsymbol{\beta}}]\boldsymbol{a} \Leftrightarrow V[\boldsymbol{a}^\top\tilde{\boldsymbol{\beta}}] \geq V[\boldsymbol{a}^\top\hat{\boldsymbol{\beta}}]$$

ここで，$\boldsymbol{a} = (1, \ldots, 1)^\top$ とすると，次式が成り立つ。

$$V[\tilde{\boldsymbol{\beta}}] \geq V[\hat{\boldsymbol{\beta}}]$$

以上のことから，最小二乗推定量 $\hat{\boldsymbol{\beta}}$ は，$\boldsymbol{Y}$ の線形式で表される $\boldsymbol{\beta}$ の線形推定量 $\tilde{\boldsymbol{\beta}}$ 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**であることがわかる。さらに，条件 1 と 2 のもとで，最小二乗推定量 $\hat{\boldsymbol{\beta}}$ が最良線形不偏推定量となることを**ガウス・マルコフの定理**という。

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．