2022.02.01(火２)

ZOOM

1.4. 最小 2 乗解の計算法

最小 2 乗解　$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$

1^o) $\widehat{\boldsymbol{\beta}}$ を定義式によって計算するのはマズイことがある。（計算誤差が大きい恐れがある）

2^o) $\widehat{\boldsymbol{\beta}}$ を計算するための、もっとよい方法がある。

3^o) 定数項がある場合には、平均を先に引いておく方がよい。

ことを述べる。

Def.1.4.1.　$A : n \times p$

$$||A|| := \max_{\boldsymbol{x}:p\times 1,\ ||\boldsymbol{x}||=1} ||A\boldsymbol{x}||$$

を、A の <u>ユークリッドノルム</u> という。

Prop.1.4.1.

(i) $||A|| = \rho_1(A)$　第 1 特異値

(ii) $||A\boldsymbol{x}|| \leqq ||A|| \cdot ||\boldsymbol{x}||$

(iii) $||AB|| \leqq ||A|| \cdot ||B||$

(i) の証明：$||A||^2 = \max_{||\boldsymbol{x}||=1} ||A\boldsymbol{x}||^2 = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \dfrac{\boldsymbol{x}'A'A\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_1(A'A) = \rho_1^2(A)$

(ii) の証明：$\dfrac{||A\boldsymbol{x}||^2}{||\boldsymbol{x}||^2} = \dfrac{\boldsymbol{x}'A'A\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \leqq ||A||^2$

(iii) の証明：$B\boldsymbol{x} = \boldsymbol{y}$ とおく。

$$\frac{||AB\boldsymbol{x}||^2}{||\boldsymbol{x}||^2} = \frac{\boldsymbol{x}'B'A'AB\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \frac{\boldsymbol{y}'A'A\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} \times \frac{\boldsymbol{x}'B'B\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \leqq ||A||^2 \times ||B||^2$$

Def.1.4.2.　$A : n \times p$

$$\mathrm{cond}(A) := \frac{\max_{||\boldsymbol{x}||=1} ||A\boldsymbol{x}||}{\min_{||\boldsymbol{x}||=1} ||A\boldsymbol{x}||}$$

を、 A の <u>条件数</u> という。

Prop.1.4.2.

(i) $\mathrm{cond}(A) = \rho_1(A)/\rho_p(A)$

(ii) $\mathrm{cond}(A'A) = [\mathrm{cond}(A)]^2$

(iii) A が正則　$\Longrightarrow$　$\mathrm{cond}(A) = ||A|| \cdot ||A^{-1}||$

(i) の証明： $\min_{||\boldsymbol{x}||=1} ||A\boldsymbol{x}||^2 = \min_{\boldsymbol{x}\neq\boldsymbol{0}} \dfrac{\boldsymbol{x}'A'A\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_p(A'A) = \rho_p^2(A)$ と Prop.1.4.1(i) より分かる。

(ii) の証明： $\mathrm{cond}(A'A) = \rho_1(A'A)/\rho_p(A'A)$
$\rho_i(A'A) = \lambda_i(A'A) = [\rho_i(A)]^2 \quad (i = 1, \cdots, p)$

(iii) の証明： $A : n \times n$ 正則のとき

特異分解　$A = P\rho Q', \quad P, Q \in O(n),\ \rho \in D^+(n)$

$\therefore \quad A^{-1} = Q\rho^{-1}P'$

$\therefore \quad ||A^{-1}|| = \rho_1(A^{-1}) = \frac{1}{\rho_n(A)}$

$\therefore \quad \mathrm{cond}(A) = \rho_1(A)/\rho_p(A) = ||A|| \cdot ||A^{-1}||$

Prop.1.4.3. 方程式 $A\boldsymbol{x} = \boldsymbol{b}$ にて、$A : n \times n$ は正則とする。

(i) A：固定、$\boldsymbol{b} \to \boldsymbol{b} + \delta\boldsymbol{b}$ のとき、$\boldsymbol{x} \to \boldsymbol{x} + \delta\boldsymbol{x}$ とすれば、

$$\frac{||\delta\boldsymbol{x}||}{||\boldsymbol{x}||} \leqq \mathrm{cond}(A)\frac{||\delta\boldsymbol{b}||}{||\boldsymbol{b}||}$$

(ii) $\boldsymbol{b}$：固定、$A \to A + \delta A$ のとき、$\boldsymbol{x} \to \boldsymbol{x} + \delta\boldsymbol{x}$ とすれば、

$$\frac{||\delta\boldsymbol{x}||}{||\boldsymbol{x} + \delta\boldsymbol{x}||} \leqq \mathrm{cond}(A)\frac{||\delta A||}{||A||}$$

(i) の証明： $A\boldsymbol{x} = \boldsymbol{b}$ (1)

より、$||\boldsymbol{b}|| \leqq ||A|| \cdot ||\boldsymbol{x}||$ (2)

$$A(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}$$

と (1) 式より、$A\delta\boldsymbol{x} = \delta\boldsymbol{b} \quad \therefore \quad \delta\boldsymbol{x} = A^{-1}\delta\boldsymbol{b}$

$\therefore \quad ||\delta\boldsymbol{x}|| \leqq ||A^{-1}|| \cdot ||\delta\boldsymbol{b}||$ (3)

(2),(3) 式を辺々かけると、

$$\frac{||\delta\boldsymbol{x}||}{||\boldsymbol{x}||} \leqq ||A|| \cdot ||A^{-1}|| \cdot \frac{||\delta\boldsymbol{b}||}{||\boldsymbol{b}||}$$

(ii) の証明： $\boldsymbol{x} = A^{-1}\boldsymbol{b}$

$$\boldsymbol{x} + \delta\boldsymbol{x} = (A + \delta A)^{-1}\boldsymbol{b}$$

$\therefore \quad \delta\boldsymbol{x} = [(A + \delta A)^{-1} - A^{-1}]\boldsymbol{b}$

恒等式：$B^{-1} - A^{-1} = A^{-1}(A - B)B^{-1}$　を使うと

$$\begin{aligned}\delta\boldsymbol{x} &= A^{-1}[A - (A + \delta A)](A + \delta A)^{-1}\boldsymbol{b} \\ &= -A^{-1}(\delta A)(\boldsymbol{x} + \delta\boldsymbol{x})\end{aligned}$$

$$\begin{aligned}\therefore \quad ||\delta\boldsymbol{x}|| &\leqq ||A^{-1}(\delta A)|| \cdot ||\boldsymbol{x} + \delta\boldsymbol{x}|| \\ &\leqq ||A^{-1}|| \cdot ||\delta A|| \cdot ||\boldsymbol{x} + \delta\boldsymbol{x}||\end{aligned}$$

$$\therefore \quad \frac{||\delta\boldsymbol{x}||}{||\boldsymbol{x} + \delta\boldsymbol{x}||} \leqq ||A|| \cdot ||A^{-1}|| \cdot \frac{||\delta A||}{||A||}$$

注： $\mathrm{cond}(A)$ が大きいと、解が不安定になる。このとき A は悪条件 (ill-conditioned) という。

例

$$\begin{cases} 10x_1 + 9x_2 = b_1 \\ 9x_1 + 8x_2 = b_2 \end{cases}$$

$\forall \boldsymbol{b} = (b_1, b_2)'$　に対し　解 $\boldsymbol{x} = (x_1, x_2)'$　は存在

$\boldsymbol{x}_0 = (1,1)'$　に対し　$\boldsymbol{b}_0 = (19, 17)'$

$\boldsymbol{x}_1 = (3, -1.2)'$　に対し　$\boldsymbol{b}_1 = (19.2, 17.4)'$

絶対誤差：

$$\delta\boldsymbol{b} = \boldsymbol{b}_1 - \boldsymbol{b}_0 = (0.2, 0.4)'$$
$$\delta\boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_0 = (2, -2.2)'$$

相対誤差：

$$\frac{||\delta\boldsymbol{b}||}{||\boldsymbol{b}||} = \left[\frac{(0.2)^2 + (0.4)^2}{(19)^2 + (17)^2}\right]^{\frac{1}{2}} = \left[\frac{0.2}{650}\right]^{\frac{1}{2}} = 0.01754$$

$$\frac{||\delta\boldsymbol{x}||}{||\boldsymbol{x}||} = \left[\frac{2^2 + (-2.2)^2}{(1)^2 + (1)^2}\right]^{\frac{1}{2}} = \sqrt{4.42} = 2.102$$

$$\frac{||\delta\boldsymbol{x}||}{||\boldsymbol{x}||} \Big/ \frac{||\delta\boldsymbol{b}||}{||\boldsymbol{b}||} = 120 \qquad (*)$$

行列 $A = \begin{bmatrix} 10 & 9 \\ 9 & 8 \end{bmatrix}$　において　$\lambda(A) = 9 \pm \sqrt{82}$

$$\left.\begin{array}{l} \rho_1(A) = \sqrt{82} + 9 \\ \rho_2(A) = \sqrt{82} - 9 \end{array}\right\} \quad \therefore \quad \mathrm{cond}(A) = \frac{\sqrt{82}+9}{\sqrt{82}-9} = (\sqrt{82}+9)^2 = 326$$

$(*)$ は限界から離れている。$\boldsymbol{x}_2 = (3, -1.235)'$ を使うと近づく。

1^o) $\widehat{\boldsymbol{\beta}}$ を $(X'X)^{-1}X'\boldsymbol{y}$ によって計算するのはマズイ理由

理由：$\widehat{\boldsymbol{\beta}}$ は $(X'X)\widehat{\boldsymbol{\beta}} = X'\boldsymbol{y}$ （正規方程式）の解
ここで、$A = X'X,\ \boldsymbol{x} = \widehat{\boldsymbol{\beta}},\ \boldsymbol{b} = X'\boldsymbol{y}$ とおくと

$$A\boldsymbol{x} = \boldsymbol{b}$$

となる。ここで、

$$\mathrm{cond}(A) = \mathrm{cond}(X'X) = [\mathrm{cond}(X)]^2$$

もし、$\mathrm{cond}(X)$ が大きいと、$\mathrm{cond}(A)$ は更に大きくなり、
$\widehat{\boldsymbol{\beta}}$ は、$\boldsymbol{y},\ X$ の小さい変化（丸め誤差）に対して、非常に不安定である。　理由終

Prop.1.4.4.(直交三角分解：QR 分解)　$A : n \times p \quad (n \geqq p)$

$$\Longrightarrow \quad \exists Q \in O(n \times p),\ \exists R : \text{上三角行列}$$
$$\text{s.t.} \quad A = QR$$

証明：$A=[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_p],\ Q=[\boldsymbol{q}_1,\cdots,\boldsymbol{q}_p],\ R=(r_{ij})$ と書くと

$$A=QR \quad\Longleftrightarrow\quad \begin{cases} \boldsymbol{a}_1=\boldsymbol{q}_1r_{11} \\ \boldsymbol{a}_2=\boldsymbol{q}_1r_{12}+\boldsymbol{q}_2r_{22} \\ \vdots \\ \boldsymbol{a}_p=\boldsymbol{q}_1r_{1p}+\boldsymbol{q}_2r_{2p}+\cdots+\boldsymbol{q}_pr_{pp} \end{cases}$$

$$\therefore\quad \boldsymbol{q}_1=\frac{\boldsymbol{a}_1}{r_{11}},\ \ r_{11}=||\boldsymbol{a}_1|| \quad \text{のとき} \quad ||\boldsymbol{q}_1||=1$$

$$\boldsymbol{q}_2=\frac{1}{r_{22}}(\boldsymbol{a}_2-\boldsymbol{q}_1r_{12})$$

$$0=\boldsymbol{q}_1'\boldsymbol{q}_2=\frac{1}{r_{22}}(\boldsymbol{q}_1'\boldsymbol{a}_2-\boldsymbol{q}_1'\boldsymbol{q}_1r_{12})$$

$$\therefore\quad r_{12}=\boldsymbol{q}_1'\boldsymbol{a}_2,\ \ r_{22}=||\boldsymbol{a}_2-\boldsymbol{q}_1r_{12}||$$

以下、同様にして、

$$\boldsymbol{q}_p=\frac{1}{r_{pp}}(\boldsymbol{a}_p-\boldsymbol{q}_1r_{1p}-\cdots-\boldsymbol{q}_{p-1}r_{p-1,p})$$

$$0=\boldsymbol{q}_i'\boldsymbol{q}_p=\frac{1}{r_{pp}}(\boldsymbol{q}_i'\boldsymbol{a}_p-\boldsymbol{q}_i'\boldsymbol{q}_ir_{ip})$$

$$\therefore\quad r_{ip}=\boldsymbol{q}_i'\boldsymbol{a}_p\ \ (i=1,\cdots,p-1)$$

$$r_{pp}=||\boldsymbol{a}_p-\boldsymbol{q}_1r_{1p}-\cdots-\boldsymbol{q}_{p-1}r_{p-1,p}||$$

注：これは、Gram-Schumidt 法といいます。

Prop.1.4.5. X はデザイン行列とする。$X=QR$（直交三角分解）

$\Longrightarrow$

(i) 最小二乗解 $\widehat{\boldsymbol{\beta}}$ は、$R\widehat{\boldsymbol{\beta}}=Q'\boldsymbol{y}$ の解である。

(ii) $\mathrm{cond}(R)=\mathrm{cond}(X)$

証明：$X'X=(QR)'(QR)=R'R$

(i) の証明：$\widehat{\boldsymbol{\beta}}=(X'X)^{-1}X'\boldsymbol{y}=(R'R)^{-1}R'Q'\boldsymbol{y}=R^{-1}Q'\boldsymbol{y}$

$\therefore\quad R\widehat{\boldsymbol{\beta}}=Q'\boldsymbol{y}$ **(i)** の証明終

(ii) の証明：$\mathrm{cond}(R)=\sqrt{\mathrm{cond}(R'R)}=\sqrt{\mathrm{cond}(X'X)}=\mathrm{cond}(X)$ **(ii)** の証明終

証明終

2^o) $\widehat{\boldsymbol{\beta}}$ を計算するより良い方法は、

$$R\widehat{\boldsymbol{\beta}} = Q'\boldsymbol{y} \qquad (*)$$

を解くこと。（条件数が小さくなった！）

$(*)$ の解法を、２つ考える。

解法 a : $\widehat{\boldsymbol{\beta}} = R^{-1}Q'\boldsymbol{y}$

$$R^{-1} = \begin{bmatrix} r^{11} & \boldsymbol{r}^1 \\ \mathbf{0} & R^1 \end{bmatrix}, \quad R = \begin{bmatrix} r_{11} & \boldsymbol{r}_1 \\ \mathbf{0} & R_1 \end{bmatrix} \quad \text{と書くと}$$

$$R^{-1}R = \begin{bmatrix} r^{11}r_{11} & r^{11}\boldsymbol{r}_1 + \boldsymbol{r}^1 R_1 \\ \mathbf{0} & R^1 R_1 \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & I_{p-1} \end{bmatrix}$$

$$\therefore \quad r^{11} = \frac{1}{r_{11}}, \quad R^1 = R_1^{-1}, \quad \boldsymbol{r}^1 = -r^{11}\boldsymbol{r}_1 R_1^{-1} \qquad (1)$$

これは、帰納的に求められる。
こうして求めた R^{-1} に $Q'\boldsymbol{y}$ を掛ける。

解法 b（逆代入法）: $R\widehat{\boldsymbol{\beta}} = \boldsymbol{z},\ \boldsymbol{z} := Q'\boldsymbol{y}$

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\boldsymbol{\beta}}_1 \end{bmatrix} \text{と書くと、} \quad R\widehat{\boldsymbol{\beta}} = \begin{bmatrix} r_{11} & \boldsymbol{r}_1 \\ \mathbf{0} & R_1 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\boldsymbol{\beta}}_1 \end{bmatrix} = \begin{bmatrix} r_{11}\widehat{\beta}_1 + \boldsymbol{r}_1\widehat{\boldsymbol{\beta}}_1 \\ R_1\widehat{\boldsymbol{\beta}}_1 \end{bmatrix} = \begin{bmatrix} z_1 \\ \boldsymbol{z}_1 \end{bmatrix}$$

$$\therefore \quad \begin{cases} \widehat{\boldsymbol{\beta}}_1 = \dfrac{1}{r_{11}}(z_1 - \boldsymbol{r}_1\widehat{\boldsymbol{\beta}}_1) \\ R_1\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{z}_1 \end{cases} \qquad (2)$$

これを帰納的に解けば良い。

２つの解法の比較
（解析的には、全く同じ手法である。）

理由：解法 a のとき

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\boldsymbol{\beta}}_1 \end{bmatrix} = R^{-1}\boldsymbol{z} = \begin{bmatrix} r^{11} & \boldsymbol{r}^1 \\ \mathbf{0} & R^1 \end{bmatrix} \begin{bmatrix} z_1 \\ \boldsymbol{z}_1 \end{bmatrix}$$

$$\therefore \quad \begin{cases} \widehat{\beta}_1 = r^{11}z_1 + \boldsymbol{r}^1\boldsymbol{z}_1 \\ \widehat{\boldsymbol{\beta}}_1 = R^1\boldsymbol{z}_1 \end{cases}$$

(1) 式を代入すると、

$$\begin{cases} \widehat{\beta}_1 = \dfrac{1}{r_{11}}(z_1 - \boldsymbol{r}_1 R_1^{-1}\boldsymbol{z}_1) \\ \widehat{\boldsymbol{\beta}}_1 = R_1^{-1}\boldsymbol{z}_1 \end{cases}$$

これは、(2) 式と同じである。 **理由終**

補足：計算機の上では、解法 b は解法 a より良い。R^{-1} を計算せずにすむから。

3^o) 定数項がある場合には、平均を先に引いておく方がよい。

Prop.1.4.6. $X = \overset{\quad 1 \qquad p}{[\mathbf{1}_n : X_1]}$, $\boldsymbol{\beta} = \begin{array}{cc} \begin{bmatrix} \alpha \\ \boldsymbol{\beta}_1 \end{bmatrix} & \begin{array}{c} 1 \\ p \end{array} \end{array}$ のとき、

$$\begin{cases} C = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' \\ \overline{y} = \frac{1}{n}\mathbf{1}_n'\boldsymbol{y} \\ \overline{\boldsymbol{x}}_1 = \frac{1}{n}\mathbf{1}_n' X_1 \end{cases}$$ とおけば、

$\Longrightarrow$

(i) $C\boldsymbol{y} = \boldsymbol{y} - \mathbf{1}_n\overline{\boldsymbol{y}}$
$CX_1 = X_1 - \mathbf{1}_n\overline{\boldsymbol{x}}_1$

(ii) $||\boldsymbol{y} - X\boldsymbol{\beta}||^2 = n(\overline{y} - \alpha - \overline{\boldsymbol{x}}_1\boldsymbol{\beta}_1)^2 + ||C\boldsymbol{y} - CX_1\boldsymbol{\beta}_1||^2$

(iii) $\mathrm{cond}(CX_1) \leqq \mathrm{cond}(X)$

(i) の証明：略

(ii) の証明：略

(iii) の証明：略（Poincare の分離公理（問題 B5）を用いる）