統計数学特別講義第四

4月27日（第1回と2回）の講義内容 No. 1

第1章 線型回帰モデル

1.1 線型単回帰モデルと最小2乗推定量

1.2 線型重回帰モデル

1.3 変数選択の規準

1.3.2 Mallows C_p

1.3.3 交差検証法

1.3.4 AIC, 1.3.5 BIC

1.1 線形単回帰モデルと最小2乗推定量　No.2

$n \in \mathbb{N}$ $(n \geq 3)$ とし、n 個の観測

$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

各 y_j $(j = 1, 2, \ldots, n)$ は

$$\underbrace{y_j}_{\text{ランダム}} = \alpha^* + \beta^* x_j + \underbrace{\varepsilon_j}_{\text{ランダム}}$$

ただし、α^*, β^* は未知の定数（母数）

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ は i.i.d. 確率変数として

$E[\varepsilon_1] = 0$ かつ $E[\varepsilon_1^2] = \sigma^2$; $0 < \sigma < \infty$ は未知。

No.3

y_jを従属変数(応答変数),

x_jを独立変数(説明変数)といい,x_jはランダム

でないと仮定する.

No. 4

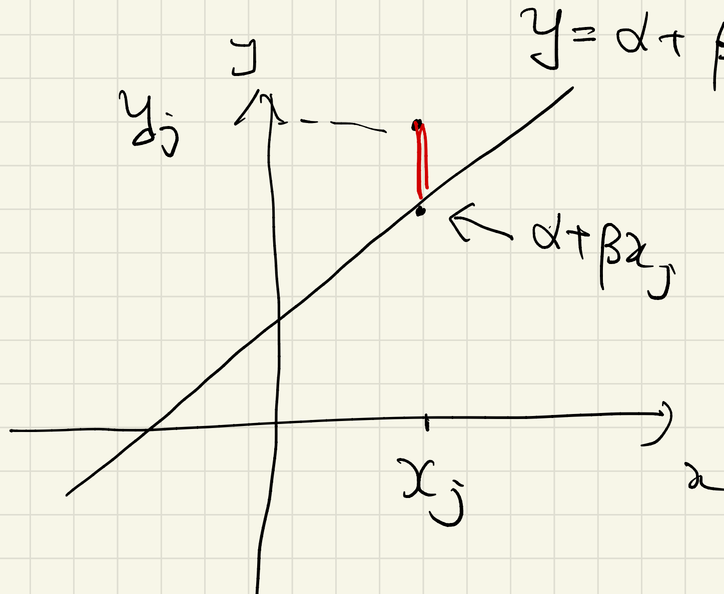
未知母数 (α^*, β^*) を推定するために

残差平方和

$$f(\alpha, \beta) = \sum_{j=1}^{n} \left\{ y_j - (\alpha + \beta x_j) \right\}^2$$

の最小化を考える。

解を最小2乗推定量(値)という。



記号の準備

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j \ ; \quad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \ ;$$

$$Q_{xy} = \sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y})$$

$$Q_{xx} = \sum_{j=1}^{n} (x_j - \bar{x})^2 \ ; \quad Q_{yy} = \sum_{j=1}^{n} (y_j - \bar{y})^2$$

$Q_{xx} > 0$ を仮定

代数的計算(平方完成)により

$$h(\alpha,\beta) = Q_{xx}\left\{\beta - \frac{Q_{xy}}{Q_{xx}}\right\}^2 + n\left\{\bar{y} - \alpha - \beta\bar{x}\right\}^2 + \frac{Q_{xx}Q_{yy} - Q_{xy}^2}{Q_{xx}}$$

よって

$$\hat{\alpha} := \bar{y} - \hat{\beta}\bar{x}\ ;\quad \hat{\beta} := \frac{Q_{xy}}{Q_{xx}}$$

のとき、hは最小となる。

すなわち

$$h(\alpha, \beta) \geq h(\hat{\alpha}, \hat{\beta}) \quad ({}^\forall \alpha, \beta \in \mathbb{R})$$

となる.

残差

$$e_j := y_j - (\hat{\alpha} + \hat{\beta} x_j), \quad j = 1, 2, \ldots n$$

残差平方和

$$RSS := \sum_{j=1}^{n} e_j^2 = \sum_{j=1}^{n} \{ y_j - (\hat{\alpha} + \hat{\beta} x_j) \}^2.$$

σ^2の推定量 ($n \geq 3$ のとき) No.8

$$\hat{\sigma}^2 = \frac{1}{n-2} RSS$$

命題1.1

(1) $E[\hat{\beta}] = \beta^*$; $\mathrm{Var}[\hat{\beta}] = \dfrac{\sigma^2}{Q_{xx}}$.

(2) $E[\hat{\alpha}] = \alpha^*$; $\mathrm{Var}[\hat{\alpha}] = \sigma^2\left\{\dfrac{1}{n} + \dfrac{\bar{x}^2}{Q_{xx}}\right\}$.

(3) $\mathrm{Cov}[\hat{\alpha}, \hat{\beta}] = -\dfrac{\bar{x}\,\sigma^2}{Q_{xx}}$.

$j=1,2,\ldots,n$.

(4) $E[e_j] = 0$; $\displaystyle\sum_{j=1}^{n} \mathrm{Var}[e_j] = (n-2)\sigma^2$.

証明の概略

代数的な計算により

i.i.d. 確率変数の1次結合

$$\hat{\beta} - \beta^* = \sum_{j=1} \frac{x_j - \bar{x}}{Q_{xx}} \varepsilon_j$$

$$\hat{\alpha} - \alpha^* = \sum_{j=1}^{n} \left\{ \frac{1}{n} - \frac{(x_j - \bar{x})\bar{x}}{Q_{xx}} \right\} \varepsilon_j$$

と書き直す。

$w_j, \tilde{w}_j$ $(j=1,2,\ldots n)$ を定数とする

$$E\left[\sum_{j=1}^{n} w_j \varepsilon_j\right] = \sum_{j=1}^{n} w_j \underbrace{E[\varepsilon_j]}_{=0}$$

$$Var\left[\sum_{j=1}^{n} w_j \varepsilon_j\right] = \sum_{j=1}^{n} w_j^{2} \underbrace{Var[\varepsilon_j]}_{=\sigma^2}$$

$$Cov\left[\sum_{j=1}^{n} w_j \varepsilon_j, \sum_{j=1}^{n} \tilde{w}_j \varepsilon_j\right] = \sum_{j=1}^{n} w_j \tilde{w}_j \underbrace{Cov[\varepsilon_j, \varepsilon_j]}_{= Var[\varepsilon_j] = \sigma^2}$$

(4) の証明の概略

$$e_j = y_j - (\hat{\alpha} + \hat{\beta} x_j)$$

$$= (\varepsilon_j - \bar{\varepsilon}) - (x_j - \bar{x})(\hat{\beta} - \beta^*) \quad (1.10)$$

$$\bar{\varepsilon} = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j , \quad j = 1, 2, \ldots, n$$

と書き直す。

すると

$$E[e_j] = \underbrace{E[\varepsilon_j]}_{=0} - \underbrace{E[\bar{\varepsilon}]}_{=0} - (x_j - \bar{x})\underbrace{E[\hat{\beta} - \beta^*]}_{=0}.$$

次に、(1.10)を展開すると

$(n-1)\sigma^2 - 2\sigma^2 + \sigma^2 = (n-2)\sigma^2$

$$\sum_{j=1}^{n} E[e_j^2] = \underbrace{\boxed{\sum_{j=1}^{n} E[(\varepsilon_j - \bar{\varepsilon})^2]}}_{=(n-1)\sigma^2}$$

$$- 2\underbrace{\boxed{\sum_{j=1}^{n} (x_j - \bar{x}) E[(\varepsilon_j - \bar{\varepsilon})(\hat{\beta} - \beta^*)]}}_{=\sigma^2}$$

$$+ \underbrace{\boxed{\sum_{j=1}^{n} (x_j - \bar{x})^2}}_{Q_x =} \boxed{E[(\hat{\beta} - \beta^*)^2]} = \frac{\sigma^2}{Q_{xx}}$$

1.2 線型重回帰モデル

$d, n \in \mathbb{N}$ とし、n 個の観測の組

$$(y_j, x_{j1}, x_{j2}, \dots, x_{jd}) \; ; \; j = 1, 2, \dots, n.$$

モデル

β_j^* は未知

$0 < \sigma < \infty$ は未知。

$$y_j = \underbrace{\beta_0^* + \beta_1^* x_{j1} + \cdots + \beta_d^* x_{jd}}_{\text{non-random}} + \underbrace{\varepsilon_j}_{\text{誤差項}}$$

- $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ は i.i.d. で $E[\varepsilon_j] = 0$, $Var[\varepsilon_j] = \sigma^2$

No. 15

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nd} \end{pmatrix} \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \vdots \\ \beta_d^* \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$= \underset{n\times 1}{\underline{y}} \qquad = \underset{n\times(d+1)}{\underline{X}} \qquad \underset{(d+1)\times 1}{\underline{\beta}^*} \qquad \underset{n\times 1}{\underline{\varepsilon}}$$

$$\underline{y} = \underline{X}\,\underline{\beta}^* + \underline{\varepsilon}$$

残差平方和の最小化

No. 16

$$h(\underset{\sim}{\beta}) = h(\beta_0, \beta_1, \dots, \beta_d)$$

$$= \sum_{j=1}^{n} \left\{ y_j - (\beta_0 + \beta_1 x_{j1} + \cdots + \beta_d x_{jd}) \right\}^2$$

$$= (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})^T (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})$$

ただし、$(\cdot)^T$ は $(\cdot)$ の転置

No.17

解は存在。さらに $\underline{X}^T\underline{X}$ が正則のとき

一意的に存在。

ここで

$$\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$$

とおく。

No.18

代数的計算(平方完成)により

$$h(\underline{\beta}) = (\underline{y} - \underline{X}\hat{\underline{\beta}})^T (\underline{y} - \underline{X}\hat{\underline{\beta}})$$

$$+ (\hat{\underline{\beta}} - \underline{\beta})^T \underline{X}^T \underline{X} (\hat{\underline{\beta}} - \underline{\beta})$$

$\Downarrow$

最小化

よって、

$$\hat{\underline{\beta}} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$$

$\underline{\beta} = \hat{\underline{\beta}}$ のとき、$h(\underline{\beta})$ は最小

は最小2乗推定を

注意　$d=1$ のとき、単回帰の場合と一致

$$\hat{\underline{\beta}} - \underline{\beta}^* = (\underline{X}^\top \underline{X})^{-1} \underline{X}^\top \{ \underbrace{\underline{X}\,\underline{\beta}^* + \underline{\varepsilon}}_{= \underline{y}} \} - \underline{\beta}^*$$

$$= (\underline{X}^\top \underline{X})^{-1} \underline{X}^\top \underline{\varepsilon}$$

b)

$$E[\hat{\underline{\beta}} - \underline{\beta}^*] = (\underline{X}^\top \underline{X})^{-1} X^\top E[\underline{\varepsilon}] = \underline{0}_n$$

$$\mathrm{Var}[\hat{\underline{\beta}}] = (\underline{X}^\top \underline{X})^{-1} \underline{X}^\top \underbrace{\mathrm{Var}[\underline{\varepsilon}]}_{= \sigma^2 I_n} \underline{X} (\underline{X}^\top \underline{X})^{-1}$$

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

$\underline{A}$: $n \times (d+1)$ 定数行列

No.20

$$E[\underline{A}\,\underline{\varepsilon}] = \underline{A}\,E[\underline{\varepsilon}]$$

$$\mathrm{Var}[\underline{A}\,\underline{\varepsilon}] = A^T\,\mathrm{Var}[\underline{\varepsilon}]\,A$$

No.21

$$RSS := h(\underline{\hat{\beta}})$$

$$= (\underline{y} - \underline{X}\underline{\hat{\beta}})^T (\underline{y} - \underline{X}\underline{\hat{\beta}})$$

$$= \left\{(I_n - \underline{X}(\underline{X}^T X)^{-1}\underline{X}^T)\underline{y}\right\}^T \left\{ \cdots \right\}$$

$$= \underline{y}^T \left\{ I_n - \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T \right\} \underline{y}$$

ここで.

さらに、$\tilde{P} = X(X^T X)^{-1} X^T$ とおくと

$$\tilde{P} = \tilde{P}^T, \quad \tilde{P}^2 = \tilde{P}, \quad \tilde{P} X = X$$

(命題1.13の証明で P が全て成立するので $\tilde{P}$ についても。

に注意すると

$$\boldsymbol{y}^T \{ I_n - \tilde{P} \} \boldsymbol{y} = \{ (I_n - \tilde{P}) \boldsymbol{y} \}^T \{ (I_n - \tilde{P}) \boldsymbol{y} \}$$

$$(I_n - \tilde{P}) \boldsymbol{y} = (I_n - \tilde{P})(X \boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$= (I - \tilde{P}) \boldsymbol{\varepsilon}$$

よって

$$E[RSS] = E[\mathrm{Tr}\{\underline{\varepsilon}^T (I_n - \underline{\tilde{P}}) \underline{\varepsilon}\}]$$

$$= E[\mathrm{Tr}\{(I_n - \underline{\tilde{P}}) \underline{\varepsilon}\, \underline{\varepsilon}^T\}]$$

$$= \mathrm{Tr}\{(I_n - \tilde{P}) \underbrace{E[\underline{\varepsilon}\, \underline{\varepsilon}^T]}_{= \sigma^2 I}\}$$

$$= \sigma^2 \mathrm{Tr}(I_n - \tilde{P})$$

$$= (n - d - 1)\sigma^2.$$

よって $\hat{\sigma}^2 = \frac{1}{n-d-1} RSS$ とおくと　　No.22

$$E[\hat{\sigma}^2] = \sigma^2$$

$\hat{\sigma}^2$ は σ^2 の不偏推定量

No. 23

定理1.3

$\underline{y} = \underline{X}\beta^* + \underline{\varepsilon}$

$\hat{\beta} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$, $\underline{C}$: $n \times (d+1)$の定数行列

$E[\underline{C}\,\underline{y}] = \beta^*$

すると

$$\mathrm{Var}[\underline{C}\,\underline{y}] \succsim \mathrm{Var}[\hat{\beta}]$$

ただし、$\underline{A} \succsim \underline{B} \iff \underline{A}-\underline{B}$は半正定値

$\iff \underline{x}^T(\underline{A}-\underline{B})\underline{x} \geq 0 \quad ({}^\forall \underline{x} \in \mathbb{R}^{d+1})$

証明の概略

$$E[\underline{C}\,\underline{y}] = \beta^* \Longleftrightarrow \underline{C}\,\underline{X} = I_{d+1}$$

さらに

$$\mathrm{Var}[\underline{C}\,\underline{y}] = \sigma^2 \underline{C}\,\underline{C}^T.$$

しかし、代数的な計算により、$X^+ := (X^T X)^{-1} X^T$

とおくと

$$\mathrm{Var}[\underline{C}\,\underline{y}] = \overbrace{\boxed{\sigma^2 X^+ (X^+)^T}}^{= V[\hat{\beta}]} + \sigma^2 \overbrace{\boxed{(C - X^+)(C - X^+)^T}}^{\text{半正定値}}$$

No.21

命題 1.4 $n \geq d+2$, $\underline{\varepsilon} \sim N_n(\underline{0}_n, \sigma^2 \underline{I}_n)$

を仮定する。以下が成立。

(1) $\hat{\underline{\beta}} \sim N_{d+1}(\underline{\beta}^*, \sigma^2 (\underline{X}^T \underline{X})^{-1})$

(2) $\dfrac{(n-d-1)}{\sigma^2} \hat{\sigma}^2 = \dfrac{RSS}{\sigma^2} \sim \chi^2_{n-d-1}$

(3) $\hat{\underline{\beta}}$ と $\hat{\sigma}^2$ は独立。

証明の概略

$$\underline{z} := \frac{1}{\sigma}(\underline{y} - \underline{X}\beta^*)$$

$$U = (\underline{X}^T\underline{X})^{-1/2}\underline{X}^T\underline{z}$$

$$V = \underline{z}^T\{\underline{I}_n - \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T\}\underline{z}$$

これらを

示せば...

とおく。命題の主張は (3a) U と V は独立で

(1a) $U \sim N_{d+1}(\underline{0}_{d+1}, \underline{I}_{d+1})$ (2a) $V \sim \chi^2_{n-d-1}$

特異値分解により

$$\underline{X} = \underline{P}\,\underline{\Lambda}\,\underline{O}^T ; \quad \underline{P}^T\underline{P} = \underline{I}_{d+1} ; \quad \Lambda \text{は対角}$$

$\underline{O}$ は $d+1$ 次の直交行列

と書ける。すると

$$(\underline{X}^T\underline{X})^{-1} = \underline{O}\,\underline{\Lambda}^{-2}\,\underline{O}^T \text{ より } (\underline{X}^T\underline{X})^{-1/2} = \underline{O}\,\underline{\Lambda}^{-1}\underline{O}^T$$

$$U = \underline{O}\,\underline{P}^T\underline{z} ; \quad V = \underline{z}^T(\underline{I}_n - \underline{P}\,\underline{P}^T)\underline{z}$$

$n \times (n-d-1)$ の行列 $\underline{Q}$ をうまく取れば

$$\underline{H} := \begin{bmatrix} \underline{P} \\ \underline{Q} \end{bmatrix}_{n} \text{ は直交行列}$$

$$\hphantom{\underline{H} := \,}\underbrace{\hphantom{\begin{bmatrix} \underline{P} \\ \underline{Q} \end{bmatrix}}}_{n}$$

とできる。すると

$$\underline{I}_n = \underline{H}\,\underline{H}^T = \underline{P}\,\underline{P}^T + \underline{Q}\,\underline{Q}^T \rightsquigarrow \underline{I}_n - \underline{P}\,\underline{P}^T = \underline{Q}\,\underline{Q}^T$$

No.29

よって

$$V = z^T \underline{Q}\,\underline{Q}^T \underline{z} = (\underline{Q}^T \underline{z})^T (\underline{Q}^T \underline{z})$$

$$H^T Z = \begin{bmatrix} \underline{P}^T \underline{z} \\ \underline{Q}^T \underline{z} \end{bmatrix} \sim N_n(\underline{0}_n, \underline{I}_n)$$

//

1.3 変数選択

$$E\left[\mathrm{Tr}\left\{(\underset{\sim}{X}\hat{\beta}-\underset{\sim}{X}\beta^*)^T(\underset{\sim}{X}\hat{\beta}-\underset{\sim}{X}\beta^*)\right\}\right]$$

$$\approx C\frac{\sigma^2 d}{n}$$

1.3.2. Mallows C_p

$$\underline{y} = \underline{X}\underline{\beta}^* + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 \underline{I}_n)$$

未来の観測

$$\tilde{\underline{y}} = \underline{X}\underline{\beta}^* + \tilde{\underline{\varepsilon}}, \quad \tilde{\underline{\varepsilon}} \sim N_n(0_n, \sigma^2 \underline{I}_n)$$

$\tilde{\underline{y}}$ を $\hat{\underline{y}} := \underline{X}\hat{\underline{\beta}} = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}\,\underline{y}$ で予測

No.32

$$MSE(\hat{\underline{y}}) = E\left[(\hat{y} - \tilde{y})^T (\hat{y} - \tilde{y})\right]$$

知りたい量

$$= (n+d+1)\sigma^2.$$

未知。

一方

$$E[RSS_d] = (n-d-1)\sigma^2$$

よって

$$MSE(\hat{y}) = E\left[RSS_d + 2(d+1)\sigma^2\right]$$

よって，$MSE[\hat{y}]$ の代用品として

$$RSS_d + 2Cd + 11\sigma^2$$

未知。

最大のモデル

No.34

$$(x_{j1}, x_{j2}, \dots, x_{jd}, \dots, x_{jk}) \quad (1 \leq d \leq k)$$

$$X_F = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2d} & \cdots & x_{2k} \\ \vdots & & & & & \\ x_{n1} & x_{n2} & \cdots & x_{nd} & \cdots & x_{nk} \end{pmatrix}$$

No.35

$$\hat{\sigma}_F^2 = \underline{y}^T \left\{ \underline{I}_n - \underline{X}_F \left(\underline{X}_F^T \underline{X}_F\right)^{-1} \underline{X}_F^T \right\} \underline{y}$$

とおくと

$$C_d := \frac{RSS_d}{\hat{\sigma}_F^2} + 2(d+1)$$

を Mallows の C_p

$$C_d = \boxed{\text{モデルの適合度}} + \boxed{\text{モデルの複雑さ}}$$

1.3.3. 交差検証法

$$\underline{x}_j^T = (x_{j1}, x_{j2}, \ldots, x_{jd})$$

とおくと

$$y_j = \underline{x}_j^T \beta^* + \varepsilon_j \quad (j = 1, 2, \ldots, n)$$

と書ける。

No.36

$$\underline{X}^{(j)} = \begin{bmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_{j-1}^T \\ \underline{x}_{j+1}^T \\ \vdots \\ \underline{x}^T \end{bmatrix}, \quad \underline{y}^{(j)} = \begin{bmatrix} y_1 \\ \vdots \\ y_{j-1} \\ y_{j+1} \\ \vdots \\ y_n \end{bmatrix}$$

とおく。すれば、$(\underline{y}^{(j)}, \underline{X}^{(j)})$ は $(\underline{y}, \underline{X})$

から (y_j, x_j) を除いたもの

$$\hat{\beta}^{(j)} = \{(\underline{X}^{(j)})^T \underline{X}^{(j)}\}^{-1}(\underline{X}^{(j)})^T \underline{y}^{(j)}$$

(No.3)

y_jの予測値として, $\hat{y}_j = x_j^T \hat{\beta}^{(j)}$

$$
\begin{tikzcd}
 & (\underline{y}, \underline{X}) \arrow[dl] \arrow[dr] & \\
\underline{y}^{(j)}, \underline{X}^{(j)} & & (y_j, \underline{x}_j)
\end{tikzcd}
$$

データ　　　　　　　　　　未知のデータ

$\{y_j - \underline{x}_j^T \hat{\beta}^{(j)}\}^2$: 誤差.

予測誤差

$$CV := \frac{1}{n}\sum_{j=1}^{n}\left\{y_j - \underline{x}_j^T \hat{\beta}^{(j)}\right\}^2$$

代数的な計算により

← 正しい.

$$CV = \frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j - \underline{x}_j^T \hat{\beta}}{1 - \underline{x}_j^T (\underline{X}^T\underline{X})^{-1}\underline{x}_j}\right\}^2$$

逆行列の計算は1回ですむ.

証明のポイント

No.39

$$\underline{X} = \begin{bmatrix} \underline{x}_1^\top \\ \hline \underline{X}^{(1)} \end{bmatrix} \left.\vphantom{\underline{X}^{(1)}}\right\} n-1 \text{ 行}$$

$$\underline{X}^\top \underline{X} = \underline{x}_1 \underline{x}_1^\top + (\underline{X}^{(1)})^\top \underline{X}^{(1)} =: \underline{x}_1 \underline{x}_1^\top + \underline{A}$$

$$(X^\top X)^{-1} = \underline{A}^{-1} - \frac{\underline{A}^{-1} \underline{x}_1 \underline{x}_1^\top A_1^{-1}}{1 + \underline{x}_1^\top \underline{A}^{-1} \underline{x}_1}$$

代数的計算から

No.40

$$\underline{x}_1^T \hat{\underline{\beta}} = \frac{\underline{x}_1^T \underline{A}^{-1} \underline{x}_1 y_1 + \underline{x}_1^T \hat{\underline{\beta}}^{(1)}}{1 + \underline{x}_1^T \underline{A}^{-1} \underline{x}_1} \qquad (1.21)$$

$$\frac{y_1 - \underline{x}_1^T \hat{\underline{\beta}}}{1 - \underline{x}_1^T (\underline{X}^T \underline{X})^{-1} \underline{x}_1} \underset{(1.20)}{=} (1 + \underline{x}_1^T \underline{A} x)(y_1 - \underline{x}_1^T \underline{\beta})$$

$$\underset{(1.21)}{=} y_1 - \underline{x}_1^T \underline{\beta}^{(1)} \qquad \square$$

1.3.4 AIC

定義 1.6 p, q : densities

$$KL(p, q) = \int \log\left(\frac{p}{q}\right) p \, dm$$

注 $x \log x \geq x - 1 \ (x > 0)$ より

$$KL(p, q) = \int \frac{p}{q} \log\left(\frac{p}{q}\right) q \, dm \geq \int \left(\frac{p}{q} - 1\right) q \, dm = 0.$$

AIC の導出

No.42

$\boxed{p(\cdot \mid \underline{X}\beta^*, \sigma^2)}$: density — p とかく

$\boxed{p(\cdot \mid \underline{X}\hat{\beta}(\underline{y}), \hat{\sigma}^2(\underline{y}))}$: estimated density — $\hat{p}$ とかく.

$$\hat{\beta}(\underline{y}) = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$$

$$\hat{\sigma}^2(\underline{y}) = \underline{y}^T(I_n - \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T)\underline{y}$$

No.43

$$KL(p, \hat{p})$$

$$= \int \log\left(\frac{p(\tau)}{\hat{p}(\tau)}\right) p(\tau)\, dm$$

$E[KL(p, \hat{p})]$ で計算

期待値を $\underline{y} \sim N(\underline{X}\vec{\beta}, \sigma^2 I_n)$ に関して.

$$E[KL(p, \hat{p})]$$

$$= E\left[\int \log\{p(\tau)\} \, p(\tau) \, dm(\tau)\right]$$

← yには依存しない

$$- E\left[\int \log\{\hat{p}(\tau)\} \, p(\tau) \, dm(\tau)\right]$$

$$AI(\beta^*, \sigma^2) := -2E\left[\int \log\{\hat{p}(\tau)\} \, p(\tau) \, dm(\tau)\right]$$

$AI(\beta^*, \sigma^2)$ の評価

まず

$$-2\log \hat{p}(\underline{t}) = n\log(2\pi\hat{\sigma}^2(\underline{y})) + \frac{(\underline{t} - \underline{X}\hat{\beta}(\underline{y}))^T(\underline{t} - X\hat{\beta}(\underline{y}))}{\hat{\sigma}^2(\underline{y})}$$

に注意して

No. 46

$$\int \log \left\{ -2 \log \hat{p}(z \mid \underline{y}) \right\} p(z) \, dm(z)$$

$$= n \log \left(2\pi \hat{\sigma}^2(\underline{y}) \right)$$

$$+ \frac{n\sigma^2 + (\hat{\underline{\beta}}(\underline{y}) - \beta^*)^T X X^T (\hat{\underline{\beta}}(\underline{y}) - \beta^*)}{\hat{\sigma}^2(\underline{y})}$$

No.67

$$E[(\hat{\beta}(y) - \beta^*) X^T X (\hat{\beta}(y) - \beta^*)]$$

$$= (d+1)\sigma^2$$

$$E\left[\frac{\sigma^2}{\hat{\sigma}^2(y)}\right] = \frac{n-d-1}{n-d-3}$$

No.48

$$AI(\beta^*, \sigma^2) = \boxed{E\left[n \log (2\pi \hat{\sigma}^2(y)) \right]}$$

$$+ \frac{n\sigma^2 + E\left[(\hat{\beta}(y) - \beta^*)^T X X^T (\hat{\beta}(y) - \beta^*) \right]}{\sigma^2} \times E\left[\frac{\sigma^2}{\hat{\sigma}^2(y)} \right]$$

$$= \frac{(n+d+1)(n-d-1)}{n-d-3}$$

No. 48

$-\frac{1}{2}$

$$E\left[-\log \hat{p}(\underline{y} \mid X\hat{\beta}(y), \hat{\sigma}^2(y))\right]$$

対数尤度

$$= E\left[n \log(2\pi\hat{\sigma}^2(y))\right] + (n-d-1)$$

$$AI(\beta^*, \sigma^2) = E\left[-2\log(\hat{p}(y \mid X\hat{\beta}(\underline{y}), \hat{\sigma}^2(y)) + 2(d+1)\frac{n-d-1}{n-d-3}\right]$$

No.50

$$-2\log\left(p(\underline{y} \mid \underline{X}\hat{\beta}(\underline{y}), \hat{\sigma}^2(\underline{y}))\right)$$

$$+2(d+1)\boxed{\frac{n-d-1}{n-d-3}} \approx 1$$

$$AIC = -2\log\left(\hat{p}(y \mid X\hat{\beta}(y), \hat{\sigma}^2(y))\right)$$

$$+2(d+1)$$