講義内容

Least Absolute Shrinkage and Selection Operator

3.4.3 LASSO推定量の精度評価

- Incoherence仮定のもとて

Sorted L-One Penalized Estimation

3.4.4 SLOPE推定量

回帰モデル

$$\underset{n\times 1}{\underline{Y}} = \underset{n\times d}{\underline{X}}\ \underset{d\times 1}{\underline{\beta}^{*}} + \underset{n\times 1}{\underline{\varepsilon}} \quad ; \quad \underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$$

BIC推定量

$$\underline{\hat{\beta}}^{BIC} \in |\underline{Y} - \underline{X}\beta|_{2,n}^{2} + \tau^2 |\underline{\beta}|_{0,d}$$

LASSO推定量

$$\hat{\beta}^{L} \in |\underline{Y} - \underline{X}\underline{\beta}|_{2,n}^{2} + 2\tau |\underline{\beta}|_{1,d}$$

$$|\underline{\beta}|_{1,d} = \sum_{j=1}^{d} |\beta_j|$$

$$\underline{\beta} = (\beta_1, \ldots, \beta_d)^{T}$$

ただし $|\cdot|_{2,n}$: $\mathbb{R}^d$ の Euclid norm，$|\cdot|_{0,d}$: $\cdot$ の非零成分の個数

命題3.19 $|\beta^*|_{0,d} \geq 1$ とする. $\forall \delta > 0$ に対して

$$\tau^2 = 16\log(12)\frac{\sigma^2}{n} + 32\sigma^2\frac{\log(ed)}{n}$$

とおいたとき,

$$\Pr\left(\frac{1}{n}|\underset{\sim}{X}\hat{\beta}^{BIC} - \underset{\sim}{X}\beta^*|_{2,n}^2 \lesssim |\beta^*|_{0,d}\,\sigma^2\frac{\log(ed/\delta)}{n}\right) \geq 1-\delta.$$

$|\beta^*|_{0,d} = k$ のとすると

$\boxed{\frac{k}{n}\log(d)}$ のスピード.

定理3.20　$\underline{X}=(X_1,\dots,X_n)$ に対して

$$\max_{1\le j\le d} |\underline{X}_j|_{2,n} \le \sqrt{n}$$

とする。　$^\forall\delta>0$ に対して

$$2\tau = 2\sigma\sqrt{\frac{2\log(2d)}{n}} + 2\sigma\sqrt{\frac{2\log(1/\delta)}{n}}$$

とおいたとき

$$\Pr\left(\frac{1}{n}|\underline{X}\hat{\beta}^{L} - \underline{X}\beta^*|_{2,n}^2 \le 4|\beta^*|_{1,d}\,\sigma\sqrt{\frac{2\log(2d)}{n}} + 4|\beta^*|\,\sigma\sqrt{\frac{\log(1/\delta)}{n}}\right)$$

$$\ge 1-\delta$$

$$\boxed{|\beta^*|_{1,d}\sqrt{\frac{\log d}{n}}}$$

$$\hat{\beta}^{LS} \in \underset{\beta \in \mathbb{R}^d}{\arg\min} \; |\underline{Y} - \underline{X}\beta|_{2,n}^2$$

$$\hat{\beta}^{LS, B_1} \in \underset{\beta \in \mathbb{R}^d,\, \beta \in B_1}{\arg\min} \; |\underline{Y} - \underline{X}\beta|_{2,n}^2$$

Tc Tc 1. $B_1 = \{ \underline{x} \in (x_1, \dots, x_d) : \sum_{j=1}^{d} |x_j| \leq 1 \}$

$$\hat{\beta}^{LS, B_0} \in \underset{\beta \in \mathbb{R}^d;\, |\beta|_{0,d} \leq R}{\arg\min} \; |\underline{Y} - \underline{X}\beta|_{2,n}^2$$

推定量	$\frac{1}{n}\|\underline{X}\hat{\beta}-\underline{X}\beta^*\|_{2,n}^2$	
$\hat{\beta}^{LS}$	$\sigma^2\frac{r}{n}$	$\mathrm{rank}\,\underline{X}=r$
$\hat{\beta}^{LS,B_1}$	$\sigma\sqrt{\frac{\log d}{n}}$	
$\hat{\beta}^{LS,B_0}$	$\sigma^2\frac{k}{n}$	
$\hat{\beta}^{BIC}$	$\underbrace{\|\beta^*\|_{0,d}}_{=k}\sigma^2\frac{\log d}{n}$	$\sigma^2 k\frac{\log d}{n}$
$\hat{\beta}^{\mathcal{L}}$	$\boxed{\|\beta^*\|_{1,d}}\,\sigma\sqrt{\frac{\log d}{n}}$	← 遅いレート.

直交デザイン

$\hat{\beta}^{HRD} \leftrightarrow \hat{\beta}^{BIC}$ $\quad \sigma^2 k \dfrac{\log d}{n}$ $\quad \hat{\beta}^{BIC}$とかわらず→

同じ

$\hat{\beta}^{SFT} \leftrightarrow \hat{\beta}^{L}$ $\quad \sigma^2 k \dfrac{\log d}{n}$ $\quad$ 大きくずれず

直交デザイン $\longrightarrow$ 直交でないデザイン

$\hat{\beta}^{L} \quad k \dfrac{\log d}{n} \longrightarrow \|\beta^*\|_1 \sqrt{\dfrac{\log d}{n}}$

$\hat{\beta}^{HRD} \quad k \cdot \dfrac{\log d}{n} \longrightarrow k \dfrac{\log d}{n}$

中間のデザインを考える

記号

$\beta = (\beta_1, \dots, \beta_d)^T \in \mathbb{R}^d$ と $S \subset \{1, \dots, d\}$ に対して

$$\beta_S = (\beta_{S,1}, \beta_{S,2}, \dots, \beta_{S,d})^T$$

$$\beta_{S,j} = \begin{cases} \beta_j & (j \in S) \\ 0 & (\text{otherwise}) \end{cases} ; \quad j = 1, 2, \dots d$$

特に

$$|\beta|_{1,d} = |\beta_S|_{1,d} + |\beta_{S^c}|_{1,d}$$

となる.

Nag

131 $d = 5$, $S = \{1, 3, 5\}$, $S^c = \{2, 4\}$

$\underline{\beta} = (1, 2, 3, 4, 5)^T$ az:

$\underline{\beta}_S = (1, 0, 3, 0, 5)^T$

$\underline{\beta}_{S^c} = (0, 2, 0, 4, 0)^T$

$|\underline{\beta}|_{1,d} = 1 + 2 + 3 + 4 + 5 = 15$

$|\underline{\beta}_S|_{1,d} = 1 + 3 + 5 = 9$

$|\underline{\beta}_{S^c}|_{1,d} = 2 + 4 = 6$

INC(k) 条件

$\underset{\sim}{Y} = \underset{\sim}{X} \beta^* + \underset{\sim}{\varepsilon}$ に対して

ある $k \in \mathbb{N}$ が存在して

$$\left| \frac{\underset{\sim}{X}^T \underset{\sim}{X}}{n} - \underset{\sim}{I}_d \right|_\infty \leq \frac{1}{32k}$$

$\Longleftrightarrow$

(a) $\left| \frac{|\underset{\sim}{x}_j|_{2,n}^2}{n} - 1 \right| \leq \frac{1}{32k}$; $\forall j \in \{1, 2, \ldots, d\}$

(b) $\left| \frac{\underset{\sim}{x}_j^T \underset{\sim}{x}_\ell}{n} \right| \leq \frac{1}{32k}$; $\forall j \neq \ell,\ j, \ell \in \{1, \ldots, d\}$

ただし、$\underset{\sim}{X} = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_d)$; $\underset{\sim}{x}_j \in \mathbb{R}^n$ $(j = 1, \ldots, d)$

補題 3.24　整数 k $(1 \leq k \leq d)$ を固定し、デザイン行列 X は INC(k) 条件をみたすとする。さらに、

$\exists S \subset \{1, \ldots, d\}$ で $\#(S) \leq k$ なるもので $\beta \in \mathbb{R}^d$

Sの要素の個数

に対して、錐条件

$$|\beta_{S^c}|_{1,d} \leq 3 |\beta_S|_{1,d}$$

をみたしているとき

$$|\beta|_{2,d}^2 \leq 2 \frac{|X\beta|_{2,n}^2}{n}$$

をみたす。

証明　まず

$$\|X\beta\|_{2,n}^2 = \|X(\beta_S + \beta_{S^c})\|_{2,n}^2$$

$$= \|X\beta_S\|_{2,n}^2 + \|X\beta_{S^c}\|_{2,n}^2 + 2\beta_S^T X^T X \beta_{S^c}$$

に注意する。最右辺の各項をINC(R)条件を用いて評価して行く

(1)　INC(R)より

$$\frac{\|X\beta_S\|_{2,n}^2}{n} = \frac{1}{n}\beta_S^T X^T X \beta_S$$

No.13

$$= \beta_S^T \left(\frac{X^T X}{n} - I_d + I_d \right) \beta_S$$

$$= |\beta_S|_{2,d}^2 + \beta_S^T \left(\frac{X^T X}{n} - I_d \right) \beta_S$$

$$\geq |\beta|_{2,d}^2 - \frac{1}{32k} |\beta_S|_{2,d}^2$$

$$\geq |\beta|_{2,d}^2 - \frac{1}{32k} |\beta|_{1,d}^2$$

$$\because) \quad a^2 + b^2 \leq \{ |a| + |b| \}^2$$

$$\therefore \frac{|X\beta|_{2,n}^2}{n} \geq |\beta_S|_{2,d}^2 - \frac{1}{32k} |\beta_S|_{1,d}^2 \qquad (3.22)$$

No. 14

(2)

$$\frac{\|X\beta_{S^c}\|_{2,n}}{n} = \frac{1}{n}\beta_{S^c}^T X^T X \beta_{S^c}$$

$$= \beta_{S^c}^T\left(\frac{X^T X}{n} - I_d\right)\beta_{S^c} + \beta_{S^c}^T\beta_{S^c}$$

$$\geq \|\beta_{S^c}\|_{2,d}^2 - \frac{1}{32k}\|\beta_{S^c}\|_{2,d}^2$$

$a^2+b^2 \leq (|a|+|b|)^2$

$$\geq \|\beta_{S^c}\|_{2,d}^2 - \frac{1}{32k}\|\beta_{S^c}\|_{1,d}^2$$

錐条件 (3.23)

$$\geq \|\beta_{S^c}\|_{2,d}^2 - \frac{9}{32k}\|\beta_{S}\|_{1,d}^2$$

(3)

$$2\left|\beta_S^T \frac{X^T X}{n} \beta_{S^c}\right| = 2\left|\sum_{j\neq \ell} \beta_{S,j}\, \mathbf{x}_j^T \mathbf{x}_\ell\, \beta_{S^c,\ell}\right|$$

INC(k) (b)

$a_i b_j \le (\sum|a_i|)(\sum|b_j|)$

$$\le \frac{2}{32k}\left|\sum_{j\neq \ell} \beta_{S,j}\beta_{S^c,\ell}\right|$$

$$\le \frac{2}{32k}\|\beta_S\|_{1,d}\|\beta_{S^c}\|_{1,d}$$

[illegible]

(3.2k)

$$\le \frac{6}{32k}\|\beta_S\|_{1,d}^2$$

$$\curvearrowright \quad 2\beta_S^T \frac{X^T X}{n}\beta_{S^c} \ge -\frac{6}{32k}\|\beta_S\|_{1,d}^2$$

さらに

$$|\beta_S|_{1,d} = (\underbrace{1\times|\beta_{S,1}| + \cdots + 1\times|\beta_{S,d}|}_{\text{非零の項は }k\text{ 個}})$$

$\#(S)=k$ とおいて $(\underbrace{1,\cdots,1}_{k},0,\cdots,0)$ と $(\beta_{S,1},\cdots,\beta_{S,k},0,\cdots,0)$ に Cauchy-Schwarz を適用

$$\leq \sqrt{k}\,|\beta_S|_{2,d}$$

よって

$$|\beta_S|_{1,d}^2 \leq \#(S)\,|\beta_S|_{2,d}^2$$

$$\|\beta_S\|_{1,d}^2 \geq -\#(S)\,\|\beta_S\|_{2,d}^2 \quad (d)$$

を得る.

よって, (3.22)-(3.2x) より

$$\frac{\|X\beta\|_{2,n}^2}{n} = \frac{\|X\beta_S\|_{2,n}^2}{n} + \frac{\|X\beta_{S^c}\|_{2,n}^2}{n} + \frac{2\beta_S^T X^T X \beta_{S^c}}{n}$$

$$\geq \|\beta_S\|_{2,d}^2 - \frac{\|\beta_S\|_{1,d}^2}{32k} + \|\beta_{S^c}\|_{2,d}^2 - \frac{9\|\beta_S\|_{1,d}^2}{32k}$$

$$- \frac{6 \|\beta_S\|_{1,d}^2}{32 k}$$

$$= \|\beta_S\|_{2,d}^2 + \|\beta_{S^c}\|_{2,d}^2 - \frac{16}{32k} \|\beta_S\|_{1,d}^2$$

$$\underset{(d)}{\geq} \|\beta_S\|_{2,d}^2 + \|\beta_{S^c}\|_{2,d}^2 - \frac{16 \#(S)}{32 k} \|\beta_S\|_{2,d}^2$$

$$\geq \|\beta_S\|_{2,d}^2 + \|\beta_{S^c}\|_{2,d}^2 - \frac{1}{2} \|\beta_S\|_{2,d}^2$$

$$\because) \ \#(S) \leq k$$

$$\geq \frac{1}{2} \|\beta_S\|_{2,d}^2 \Rightarrow \frac{\|X\beta\|_{2,n}^2}{n} \geq \frac{1}{2} \|\beta_S\|_{2,d}^2 \quad \boxed{\leq}$$

LASSO推定量の速い収束レート

定理3.25 $n \geq 2$ とし、

$$\underline{Y} = \underline{X}\beta^* + \underline{\varepsilon} \;;\; \underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$$

を仮定する。さらに、$\exists k \in \{1, 2, \ldots, d\}$ に対して

$$|\beta^*|_{0,d} \leq k$$

とし、デザイン行列 $\underline{X}$ は INC(k) 条件をみたす

とする。

このとき, $^\forall \delta > 0$ に対して

$$2\tau := 8\sigma\sqrt{\frac{\log(2d)}{n}} + 8\sigma\sqrt{\frac{\log(1/\delta)}{n}}$$

$$\hat{\beta}^L \in \arg\min\left\{ \|\underline{Y} - \underline{X}\beta\|_{2,n}^2 + 2\tau\|\beta\|_{1,d} \right\}$$

とおいたとき

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^L - \underline{X}\beta^*\|_{2,n}^2 \lesssim \sigma^2 k \frac{\log(2d/\delta)}{n} \right) \geq 1-\delta$$

と

$$\Pr\left(\|\hat{\beta}^L - \beta^*\|_{2,d} \lesssim \sigma^2 k \frac{\log(2d/\delta)}{n} \right) \geq 1-\delta$$

が示す

証明　$\hat{\beta}^L$の定義から

$$\frac{1}{n}\|\underline{Y}-\underline{X}\hat{\beta}^L\|_{2,n}^2 \leq \frac{1}{n}\|\underline{Y}-\underline{X}\beta^*\|_{2,n}^2+2\tau\|\beta^*\|_{1,d}-2\tau\|\hat{\beta}^L\|_{1,d} \quad (a)$$

となる. 次に上の式の左辺を書き直すと

$$\|\underline{Y}-\underline{X}\hat{\beta}^L\|_{2,n}^2 = \|\underbrace{\underline{Y}-\underline{X}\beta^*}_{=\underline{\varepsilon}}+\underline{X}\beta^*-\underline{X}\hat{\beta}^L\|_{2,n}^2$$

$$= \|\underline{X}\hat{\beta}^L-\underline{X}\beta^*\|_{2,n}^2 + \|\underline{Y}-\underline{X}\beta^*\|_{2,n}^2$$

$$-2\,\underline{\varepsilon}^T X(\hat{\beta}^L-\beta^*)$$

となることに注意する. 上の式を(a)の左辺に代入して,

その両辺に $\tau|\hat{\beta}^L - \beta^*|_{1,d}$ を加えて n 倍すると

$$|\underline{X}\hat{\beta}^L - \underline{X}\beta^*|_{2,n}^2 + n\tau|\hat{\beta}^L - \beta^*|_{1,d}$$

$$\le 2\underline{\varepsilon}^T\underline{X}(\hat{\beta}^L - \beta^*) + n\tau|\hat{\beta}^L - \beta^*|_{1,d}$$

$$+2n\tau|\beta^*|_{1,d} - 2n\tau|\hat{\beta}^L|_{1,d} \qquad (3.25)$$

を得る。 を定理3.20の証明中の議論と同じように評価する。まず，Hölder不等式から

$$|\underline{\varepsilon}^T\underline{X}(\hat{\beta}^L - \beta^*)| \le |\underline{\varepsilon}^T\underline{X}|_\infty \cdot |\hat{\beta} - \beta^*|_{1,d}$$

INC(k) 条件から

$$\left| \frac{|\underline{x}_j|_{2,n}^2}{n} - 1 \right| \leq \frac{1}{32k}$$

$$\Longleftrightarrow$$

$$-\frac{1}{32k} \leq \frac{|\underline{x}_j|_{2,n}^2}{n} - 1 \leq \frac{1}{32k}$$

$$\Longleftrightarrow \quad |\underline{x}_j|_{2,n}^2 \leq n + \frac{1}{32k} n \leq 2n$$

となることに注意する.

このことより

$$\underline{X}_j^T \underline{\varepsilon} \sim \mathrm{subG}(2n\sigma^2)$$

となる。

∵) $^\forall \lambda > 0$ に対して

$$E\left[\exp\left(\lambda \underline{X}_j^T \underline{\varepsilon}\right)\right] = \exp\left(\lambda |\underline{X}_j|_{2,n} \frac{\underline{X}_j^T}{|\underline{X}|_{2,n}} \underline{\varepsilon}\right)$$

$$\leq \exp\left(\lambda \sqrt{2n} \frac{\underline{X}_j^T}{|\underline{X}_j|_{2,n}} \underline{\varepsilon}\right)$$

$$| \quad |_{2,n} = 1$$

$$\leq \exp\left(-\frac{\sigma^2}{2}(\sqrt{2n}\,\Delta)^2\right)$$

$\because\ \underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$

$$= \exp\left(-\frac{1}{2}(2n\sigma^2)\,\Delta^2\right).$$

//

よって，命題2.24

$$W_j \sim \mathrm{subG}(\sigma^2) \Rightarrow \Pr\left(\max_{1\leq j\leq d}|W_j| > t\right) \leq 2d\exp\left(-\frac{t^2}{2\sigma^2}\right)$$

ただし $W_j \to \underline{X}_j^T\underline{\varepsilon}$ $\quad \sigma^2 \to 2n\sigma^2$

$$\Pr\left(\max_{1 \leq j \leq d} |\underline{X}_j^T \underline{\varepsilon}| > t \right) \leq 2d \exp\left(-\frac{t^2}{4n\sigma^2} \right)$$

とする. ここで

$$t = \frac{n\tau}{2}$$

とおくと

$$\Pr\left(|\underline{X}^T \underline{\varepsilon}|_\infty > \frac{n\tau}{2} \right) \leq 2d \exp\left(-\frac{n\tau^2}{16\sigma^2} \right) \quad (a)$$

と得る. $|\underline{X}\underline{\varepsilon}|_\infty = \max_{1 \leq j \leq n} |\underline{x}_j^T \underline{\varepsilon}|$

ここに

$$2\tau = 8\sigma\sqrt{\frac{\log(2d)}{n}} + 8\sigma\sqrt{\frac{\log(\frac{1}{\delta})}{n}}$$

より

$(\sqrt{a}+\sqrt{b})^2 \geq a+b$

$$\exp\left(-\frac{n\tau^2}{16\sigma^2}\right) = \exp\left(-\frac{n}{64\sigma^2}(2\tau)^2\right)$$

$$= \exp\left[-n\left\{\sqrt{\frac{\log(2d)}{n}} + \sqrt{\frac{\log(\frac{1}{\delta})}{n}}\right\}^2\right]$$

$$\leq \exp\left[-n\left\{\frac{\log(2d)}{n} + \frac{\log(\frac{1}{\delta})}{n}\right\}\right]$$

$$= \frac{\delta}{2d}$$

となる。これより

$$2d\exp\left(-\frac{n\tau^2}{16}\right) \leq \delta$$

を得る。(a)より

$$\Pr\left(|\underline{X}\,\varepsilon|_\infty > \frac{n\tau}{2}\right) \leq 2d\exp\left(-\frac{n\tau^2}{16\sigma^2}\right) \leq \delta$$

よって

$$\Pr\left(|\underline{X}^T \underline{\varepsilon}|_\infty \le \frac{n\tau}{2} \right) \ge 1 - \delta \qquad (b)$$

がわかる. No.29

$|\underline{X}^T\underline{\varepsilon}|_\infty \le \frac{n\tau}{2}$ 上では

$$|\underline{\varepsilon}^T \underline{X}(\hat{\beta}^L - \beta^*)| \le |\underline{X}^T \underline{\varepsilon}|_\infty |\hat{\beta}^L - \beta^*|_{1,d}$$

$$\le \frac{n\tau}{2} |\hat{\beta}^L - \beta^*|_{1,d} \qquad (c)$$

となる

ここで

$$S := \mathrm{supp}(\beta^*) \subset \{1, 2, \dots, d\}$$

とおく. すると $\|X^\top \varepsilon\|_\infty \leq \frac{n\tau}{2}$ 上では

(3.25)は以下のように評価できる:

$$\|X\hat{\beta}^\lambda - X\beta^*\|_{2,n}^2 + n\tau\|\hat{\beta} - \beta^*\|_{1,d}$$

$$\leq 2|\varepsilon^\top X(\hat{\beta} - \beta^*)| + n\tau\|\hat{\beta}^\lambda - \beta^*\|_{1,d}$$

$$+ 2n\tau\|\beta^*\|_{1,d} - 2n\tau\|\hat{\beta}^\lambda\|_{1,d}$$

$$\leq n\tau|\hat{\beta} - \beta^*|_{1,d} + n\tau|\hat{\beta}^c - \beta^*|_{1,d}$$

$$+ 2n\tau|\beta^*|_{1,d} - 2n\tau|\hat{\beta}^c|_{1,d}$$

$$= 2n\tau|\hat{\beta} - \beta^*|_{1,d} + 2n\tau|\beta^*|_{1,d} - 2n\tau|\hat{\beta}^c|_{1,d}$$

$$= 2n\tau\{|\hat{\beta}_S - \beta^*_S|_{1,d} + |\hat{\beta}_{S^c}|_{1,d}\} + 2n\tau|\beta^*_S|_{1,d}$$

$$- 2n\tau\{|\hat{\beta}^c_S|_{1,d} + |\hat{\beta}^c_{S^c}|_{1,d}\}$$

$$\because \; S = \mathrm{supp}(\beta^*) \Rightarrow \beta^* = \beta^*_S$$

$$= 2n\tau \| \hat{\beta}_S - \beta^*_S \|_{1,d} + 2n\tau \| \beta^*_S \|_{1,d} - 2n\tau \| \hat{\beta}^c_S \|_{1,d}$$

No.31

$$\leq 4n\tau \| \hat{\underline{\beta}}_S - \underline{\beta}^*_S \|_{1,d}$$

$$\because) \quad \| \hat{\beta}_S - \beta^*_S \|_{1,d} \geq \| \beta^*_S \|_{1,d} - \| \hat{\underline{\beta}}^c \|_{1,d}$$

ここで、$\Sigma > 2$

$$\| \hat{\underline{\beta}}_S - \underline{\beta}^*_S \|_{1,d} + \| \hat{\underline{\beta}}_{S^c} - \beta^*_{S^c} \|_{1,d}$$

$$\leq \frac{1}{n\tau} \| \underline{X} \hat{\underline{\beta}}^c - \underline{X} \underline{\beta}^* \|^2_{2,n} + \| \hat{\underline{\beta}}^c - \beta^* \|_{1,d}$$

$$\leq 4 \| \hat{\underline{\beta}}_S - \underline{\beta}^*_S \|_{1,d}$$

より

$$|\hat{\beta}_{S^c} - \beta^*_{S^c}|_{1,d} \leq 3|\hat{\beta}_S - \beta^*_S|_{1,d}$$

を得る. したがって

$$\beta = \hat{\beta}^\lambda - \beta^*$$

で錐条件をみたすので

$$|\hat{\beta}^\lambda - \beta^*|_{2,d}^2 \leq \frac{2}{n}|X\hat{\beta}^\lambda - X\beta^*|_{2,n}^2 \quad (\mathrm{d})$$

を得る.

$S = \mathrm{supp}(\beta^*)$ と $\#(S) \le k$ に注意して

Cauchy-Schwarz の不等式を用いると

$$\|\hat{\beta}^\lambda_S - \beta^*_S\|_{1,d} \le \sqrt{\#(S)} \cdot \|\hat{\beta}^\lambda_S - \beta^*_S\|_{2,d}$$

$$\le \sqrt{\#(S)} \cdot \underline{\|\hat{\beta}^\lambda - \beta^*\|_{2,d}}$$

$$\because)\ \|\hat{\beta}^\lambda - \beta^*\|^2_{2,d} = \|\hat{\beta}^\lambda_S - \beta^*_S\|^2_{2,d} + \|\hat{\beta}^\lambda_{S^c}\|^2_{2,d}$$

$$\ge \|\hat{\beta}^\lambda_S - \beta^*_S\|^2_{2,d}$$

$$\le \sqrt{k}\,\underline{\sqrt{\frac{2}{n}\|X\hat{\beta}^\lambda - X\beta^*\|^2_{2,n}}} \quad \because (d)$$

(3.26) と $|\hat{\beta}^{c}_{S} - \beta^{*}_{S}|_{1,d} \leq \sqrt{\frac{2k}{n}} \, |\underline{X}\hat{\beta}^{c} - \underline{X}\beta^{*}|_{2,n}$

を合わせると

$$|\underline{X}\hat{\beta} - \underline{X}\beta^{*}|_{2,n}^{2} \leq |\underline{X}\hat{\beta} - \underline{X}\beta^{*}|_{2,n}^{2} + n\tau|\hat{\beta} - \beta^{*}|_{1,d}$$

$$\leq 4n\tau|\hat{\beta}^{c}_{S} - \beta^{*}_{S}|_{1,d}$$

$$\leq \tau\sqrt{32kn}\,|\underline{X}\hat{\beta}^{c} - \underline{X}\beta^{*}|_{2,n}$$

$$\Rightarrow \quad |\underline{X}\hat{\beta}^{c} - \underline{X}\beta^{*}|_{2,n} \leq \tau\sqrt{32kn}$$

$$|\underline{X}\hat{\beta}^{\lambda} - \underline{X}\beta^*|_{2,n}^2 \le 32 k n \tau^2$$

$$= 8 k n (2\tau)^2$$

$$= 8 k n \left\{ 8\sigma \sqrt{\frac{\log(2d)}{n}} + 8\sigma \sqrt{\frac{\log(\frac{1}{\delta})}{n}} \right\}^2$$

$$\le 2^9 k n \sigma^2 \left\{ 2 \left(\frac{\log(2d)}{n} + \frac{\log(\frac{1}{\delta})}{n} \right) \right\}$$

$$= 2^{10} k \sigma^2 \log\left(\frac{2d}{\delta}\right)$$

よって

$$\|\underline{X}^T\underline{\varepsilon}\|_\infty \le \frac{n\tau}{2} \Rightarrow \frac{1}{n}\|\underline{X}\hat{\beta}^\tau - \underline{X}\beta^*\|_{2,n}^2 \lesssim \frac{k\sigma^2}{n}\log\left(\frac{2d}{\delta}\right)$$

より

$$1-\delta \le \Pr\left(\|\underline{X}^T\varepsilon\|_\infty \le \frac{n\tau}{2}\right)$$

$$\le \Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^\tau - \underline{X}\beta^*\|_{2,n}^2 \lesssim \frac{k\sigma^2}{n}\log\left(\frac{2d}{8}\right)\right)$$

$\hat{\beta} - \beta^*$ は

$$|\hat{\beta}_{S^c} - \beta^*_{S^c}|_{1,d} \leq 3|\hat{\beta}^c_S - \beta^*_S|_{1,d}$$

なので、補題3.2より

$$|\hat{\beta}^c - \beta^*|_{2,d}^2 \leq \frac{2}{n}|X\hat{\beta}^c - X\beta^*|_{2,n}^2$$

より

$$\Pr\left(|\hat{\beta}^c - \beta^*|_{2,d}^2 \leq 2^9 \frac{k\sigma^2}{n}\log\left(\frac{2d}{\delta}\right)\right)$$

$$\geq \Pr\left(\frac{1}{n} \left\| \underline{X} \hat{\beta}^{\lambda} - \underline{X} \beta^{*} \right\|_{2,n}^{2} \leq 2^{10} \frac{R\sigma^{2}}{n} \log\left(\frac{2d}{\delta} \right) \right)$$

$$\geq 1 - \delta$$

がわかる.

$\square$

証明の概略

① $\hat{\beta}^L$ の定義より

$$
\begin{aligned}
&\|\underline{X}\hat{\beta}^L - \underline{X}\beta^*\|_{2,n}^2 + n\tau\|\hat{\beta} - \beta^*\|_{1,d} \\
&\quad \leq 2\|\underline{X}^T\underline{\varepsilon}\|_\infty \cdot \|\hat{\beta}^L - \beta^*\|_{1,d} + n\tau\|\hat{\beta}^L - \beta^*\|_{1,d} \\
&\quad + 2n\tau\|\beta^*\|_{1,d} - 2n\tau\|\hat{\beta}^L\|_{1,d} \qquad (3.25)
\end{aligned}
$$

② 最大不等式より

$$\Pr\left(|\underline{X}^{\top}\varepsilon|_{\infty} > \frac{n\tau}{2} \right) \leq \delta$$

$$\Longleftrightarrow \Pr\left(|\underline{X}^{\top}\varepsilon|_{\infty} \leq \frac{n\tau}{2} \right) \geq 1-\delta$$

③ $|\underline{X}^{\top}\varepsilon|_{\infty} \leq \frac{n\tau}{2}$ 上で (3.25) を評価

すると

$$|\underline{X}\hat{\beta} - \underline{X}\beta^{*}|_{2,n}^{2} + n\tau|\hat{\beta}^{L} - \beta^{*}|_{1,d} \leq 4n\tau|\hat{\beta}^{L}_{S} - \beta^{*}_{S}|_{1,d}$$

$$\Rightarrow \quad |\hat{\beta}_{S^c} - \beta^*_{S^c}|_{1,d} \leq 3|\hat{\beta}^\lambda_S - \beta^*|_{1,d}$$

を得るので、$\hat{\beta}^\lambda - \beta^*$ は錐条件をみたすので、補題 3.2 より

$$|\hat{\beta}^\lambda - \beta^*|^2_{2,d} \leq \frac{2}{n}|\underline{X}\hat{\beta}^\lambda - \underline{X}\beta^*|^2_{2,n}$$

$$\Rightarrow |\hat{\beta}^\lambda_S - \beta^*_S|_{1,d} \leq \sqrt{\frac{2k}{n}}|\underline{X}\hat{\beta}^\lambda - \underline{X}\beta^*|_{2,n}$$

(※)

④ (3.26)と(※)より

$$|\underline{X}\hat{\beta}^{\lambda} - \underline{X}\beta^{*}|_{2,n}^{2} \le |\underline{X}\hat{\beta}^{\lambda} - \underline{X}\beta^{*}| + n\tau|\hat{\beta}^{\lambda} - \beta^{*}|_{1,d}$$

$$\le 4n\tau|\hat{\beta}^{\lambda}_{S} - \beta^{*}|_{1,d}$$

$$\le \tau\sqrt{32kn}\ |\underline{X}\hat{\beta}^{\lambda} - \underline{X}\beta^{*}|_{2,n}$$

$$\Rightarrow\ |\underline{X}\hat{\beta}^{\lambda} - \underline{X}\beta^{*}|_{2,n}^{2} \le 32kn\tau^{2}$$

$$\le 2^{10}k\sigma^{2}\log\left(\frac{2d}{\delta}\right)$$

⑤ $\| X^\top \varepsilon \|_\infty \leq \frac{n\tau}{2}$ よって

$$\| \underline{X} \hat{\beta}^{\lambda} - \underline{X} \beta^* \|_{2,n}^2 \leq 2^{10} k \sigma^2 \log\left(\frac{2d}{\delta}\right)$$

$$1-\delta \leq P_r\left(\| \underline{X}^\top \varepsilon \|_\infty \lesssim \frac{n\tau}{2} \right)$$

$$\leq P_r\left(\frac{1}{n} \| X \hat{\beta}^{\lambda} - X \beta^* \|^2 \lesssim \frac{k\sigma^2}{n} \log\left(\frac{2d}{\delta}\right) \right)$$

$\square$

精度の比較

$|\beta^*|_{0,d} = k$ とおく.

$\hat\beta^{BIC}$		$\sigma^2 \frac{k}{n}\log(d)$ ⟹ $\sigma^2 \frac{k}{n}\log\frac{d}{k}$ まで改良できる 注意3.37
$\hat\beta^{L}$	直交デザイン	$\sigma^2 \frac{k}{n}\log(d)$
$\hat\beta^{L}$	INC(k)	$\sigma^2 \frac{k}{n}\log(2d)$
$\hat\beta^{L}$	$\|X_j\|_{2,n} \le \sqrt{n}$	$\sigma^2 \|\beta^*\|_{1,d}\sqrt{\frac{\log(2d)}{n}}$

3.4.4 SLOPE 推定量

$k = |\beta^*|_{0,d}$ とする。BIC 推定量は

$$\frac{k}{n}\log\left(\frac{ed}{k}\right)$$

となる。一方、LASSO は

$$\frac{k}{n}\log(d)$$

INC(k) 条件

である。

定理 3.27 $\underline{\lambda} = (\lambda_1, \ldots, \lambda_d)$ を非増加の正の数列とする. すなわち

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0.$$

$\underline{\beta} = (\beta_1, \ldots, \beta_d)^T \in \mathbb{R}^d$ に対して

$$(\beta_{*1}, \ldots, \beta_{*d})$$

を $|\beta_1|, |\beta_2|, \ldots, |\beta_d|$ を非増加数列に並べ替えたものとする. すなわち

$$\beta_{*1} \geq \beta_{*2} \geq \cdots \geq \beta_{*d} > 0$$

である. β の並み替え ℓ_1 ノルムを次で定める.

$$|\beta|_{*,d} := \sum_{j=1}^{d} \lambda_j \beta_{*j}$$

または

$$|\beta|_{*,d} = \max_{\sigma \in \mathfrak{S}_d} \sum_{j=1}^{d} \lambda_j |\beta_{\sigma(j)}|$$

である. ただし, $\mathfrak{S}_d$ は $\{1, 2, \ldots, d\}$ の置換群である.

$d=2$ のとき，$\underline{G}_2 = \{\underline{\sigma}_1, \underline{\sigma}_2\}$ で

$$\underline{\sigma}_1(1)=1,\ \underline{\sigma}_1(2)=2;\ \underline{\sigma}_2(1)=2,\ \underline{\sigma}_2(2)=1.$$

調整パラメータ $\underline{\lambda}$ と $\tau>0$ に対する

SLOPE 推定量 $\hat{\beta}^S$ を

$$\hat{\beta}^S \in \arg\min_{\beta\in\mathbb{R}^d}\left\{\frac{1}{n}\|Y-\underline{X}\beta\|_{2,n}^2 + 2\tau\|\beta\|_*\right\}$$

で定義する.

以下では

$$\lambda_j = \sqrt{\log\left(\frac{2d}{j}\right)} \qquad (j = 1, \ldots, d)$$

とする.

補題 3.28 $g_1, g_2, \ldots, g_d \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$

$(0 < \sigma < \infty)$ とする。このとき、$k \leq d$ と $t > 0$ に対して

$$\Pr\left(\frac{1}{k\sigma^2} \sum_{j=1}^{k} (g_{*j})^2 > t \log\left(\frac{2d}{k}\right) \right) \leq \left(\frac{2d}{k}\right)^{1-\frac{3t}{8}}$$

となる。

No.51

証明 Jensenの不等式を用いる.

確率変数 W と凸関数 $g:\mathbb{R}\to\mathbb{R}$ で

$E[|g(W)|]<\infty$ かつ $|E[W]|<\infty$ をみたす.

このとき

$$E[g(W)] \geq g(E[W]).$$

$g(x)=e^{x}$ と $\Pr(W=w_j)=\dfrac{1}{k}$　$w_j=\dfrac{3(g_{*,j})^2}{8\sigma^2}$

$$E\left[\exp\left(\frac{3}{8k\sigma^2}\sum_{j=1}^{k}(g_{*j})^2\right)\right]$$

$$= E\left[\exp\left(E_w[w]\right)\right]$$

$$\leq E\left[E_w\left[\exp(w)\right]\right]$$

$$= E\left[\frac{1}{k}\sum_{j=1}^{k}\exp\left(\frac{3(g_{*j})^2}{8\sigma^2}\right)\right]$$

$$\leq E\left[\frac{1}{k}\sum_{j=1}^{d}\exp\left(\frac{3(g_{*j})^2}{8\sigma^2}\right)\right]$$

$$= \frac{1}{k}\sum_{j=1}^{d}\underbrace{E\left[\exp\left(\frac{3(g_{*j})^2}{8\sigma^2}\right)\right]}_{=2}$$

$$= \frac{2d}{R_2}. \qquad (a)$$

次に, Chernoff 限界を用いる.

$W \geq 0$ を確率変数列で $E[e^{\lambda W}] < \infty \ (\lambda > 0)$

とする. このとき

$$P_r(W > a) = P_r(e^{\lambda W} > e^{\lambda a})$$

$$\leq e^{-\lambda a} E[e^{\lambda a}]$$

$$\Pr\left(\frac{1}{k\sigma^2} \sum_{j=1}^{k} (g_{*j})^2 > t \log\left(\frac{2d}{k}\right)\right)$$

$$= \Pr\left(\frac{3}{8k\sigma^2} \sum_{j=1}^{k} (g_{*j})^2 > \frac{3\tau}{8} \log\left(\frac{2d}{k}\right)\right)$$

$$\leq \exp\left(-\frac{3\tau}{8}\log\left(\frac{2d}{k}\right)\right) E\left[\exp\left(\frac{3}{8k\sigma^2}\sum_{j=1}^{k}(g_{*j})^2\right)\right]$$

$$\leq \exp\left(-\frac{3\tau}{8}\log\left(\frac{2d}{k}\right)\right) \frac{2d}{k} \qquad \because (a)$$

$$= \left(\frac{2d}{k}\right)^{1-\frac{3\tau}{8}}.$$

補題3.29 $[d]=\{1,2,\dots,d\}$とする.

$g_1, g_2, \dots, g_d \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)\ (0<\sigma<\infty)$

とする. $0.5 > {}^\forall \delta > 0$に対して

$$\Pr\left(\sup_{k\in[d]} \frac{g_k}{\sigma\lambda_k} \leq 4\sqrt{\log\left(\frac{1}{\delta}\right)} \right) \geq 1-\delta.$$

となる.

証明　$g_{*1} \geq g_{*2} \geq \cdots \geq g_{*k}$ なので

$$(g_{*k})^2 \leq \frac{1}{k} \sum_{j=1}^{k} (g_{*j})^2$$

となることに注意する. 補題3.28を用いると

$$\Pr\left(\frac{1}{\sigma^2 \lambda_k^2} (g_{*k})^2 > t \right)$$

$$\leq \Pr\left(\frac{1}{k\sigma^2} \sum_{j=1}^{k} (g_{*j})^2 > \underbrace{\lambda_k^2}_{= \log\left(\frac{2d}{k}\right)} t \right)$$

∵定義

$$= \Pr\left(\frac{1}{k\sigma^2} \sum_{j=1}^{k} (g_{*j})^2 > t \log\left(\frac{2d}{k}\right) \right)$$ No.56

$$\leq \left(\frac{2d}{k}\right)^{1-\frac{3\tau}{8}}$$

となる. $t > 8$ に対して

$$\Pr\left(\sup_{k\in[d]} \frac{(g_{*k})^2}{\sigma^2 \lambda_k^2} > t \right)$$

$$= \Pr\left(\bigcup_{j=1}^{d} \left\{ \frac{(g_{*j})^2}{\sigma^2 \lambda_j^2} > \tau \right\} \right)$$

No.57

$$\leq \sum_{j=1}^{d} \Pr\left(\frac{(g_{*j})^2}{D^2 \lambda_j^2} > t \right)$$

$$\leq \sum_{j=1}^{d} \Pr\left(\frac{1}{j D^2} \sum_{k=1}^{j} (g_{*k})^2 > t \log\left(\frac{2d}{\delta}\right) \right)$$

$$\leq \sum_{j=1}^{d} \left(\frac{2d}{\delta}\right)^{1-\frac{3\tau}{8}}$$

$$\leq 4 \cdot 2^{-\frac{3\tau}{8}} = 2^{2-\frac{3\tau}{8}} \,〃$$

$\delta > \frac{1}{2}$ なので $\frac{16}{3} > \frac{t}{3} \log\left(\frac{1}{\delta}\right)$

に注意すると

$$\delta := 4 \cdot 2^{-\frac{3t}{8}} \Rightarrow t = \frac{8}{3}\log\left(\frac{1}{\delta}\right) + \frac{16}{3}$$

$$\geq 4\log\left(\frac{1}{\delta}\right)$$

となる。よって

$$\Pr\left(\sup_{k\in[d]} \frac{g * k}{\alpha \lambda_k} > 2\sqrt{\log\left(\frac{1}{\delta}\right)}\right)$$

$$\geq \Pr\left(\sup_{k\in[d]} \frac{(g * k)^2}{\alpha^2 \lambda_k^2} > \frac{8}{3}\log\left(\frac{1}{\delta}\right) + \frac{16}{3}\right)$$

No. 55

$$= \Pr\left(\sup_{k \in [d]} \frac{(g * k)^2}{\sigma \lambda_k^2} > t \right)$$

$$\geq \delta .$$

よって

$$\Pr\left(\sup_{k \in [d]} \frac{g * k}{\sigma \lambda_k} \leq 2\sqrt{\log\left(\frac{1}{\delta}\right)} \right) \geq 1 - \delta .$$

$\square$

定理 3.31 $n \geq 2$ とし, $|\beta^*|_{0,d} = k < d$,

$$\underline{Y} = \underline{X}\beta^* + \underline{\varepsilon}, \qquad \underline{\varepsilon} \sim N_n(\underline{0}_n, \sigma^2 I_n)$$

に対して, デザイン行列 $\underline{X}$ は

$$\mathrm{INC}(k'); \qquad k' \geq 4k \log\left(\frac{2de}{k}\right)$$

をみたすとする. $\forall \delta > 0$ に対して

$$\tau = 8\sqrt{2}\,\sigma \sqrt{\frac{\log \delta^{-1}}{n}}$$

とする. このとき

$$\hat{\beta}^s \in \arg\min \left\{ \frac{1}{n} |\underline{Y} - \underline{X}\underline{\beta}|_{2,n}^2 + 2\tau |\beta^*|_{*,1} \right\}$$

に対して

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^S - \underline{X}\beta^*\|_{2,n} \lesssim \sigma^2 \frac{k}{n}\log\left(\frac{2de}{k}\right)\log\left(\frac{1}{\delta}\right)\right) \geq 1-\delta$$

と

$$\Pr\left(\|\hat{\beta}^S - \beta^*\|_{2,d} \lesssim \sigma^2 \frac{k}{n}\log\left(\frac{2de}{k}\right)\log\left(\frac{1}{\delta}\right)\right) \geq 1-\delta$$

をみたす。

証明の方針　$\hat{\beta}^S$の定め方から

$$
\|\underline{X}\hat{\beta}^S - \underline{X}\beta^*\|_{2,n}^2 + n\tau\|\hat{\beta}^S - \beta^*\|_{2,d}
$$

$$
\leq 2\underline{\varepsilon}^T X(\hat{\beta}^S - \beta^*) + n\tau\|\hat{\beta}^S - \beta^*\|_{2,d}
$$

$$
+ 2n\tau\|\beta^*\|_{*,d} - 2n\tau\|\hat{\beta}^S\|_{*,d}
$$

がわかる.

$$
g_j = (\underline{X}^T\underline{\varepsilon})_j \;;\; j = 1, \cdots, d,\quad \underline{X} = (\underline{x}_1, \cdots, \underline{x}_d)
$$

とおくと　　　　$\underline{x}_j$は$\underline{X}$のj列

$$g_j \sim N(0, \|\underline{x}_j\|_{2,n}^2 \sigma^2)$$

となる. さらに, INC(R') 条件より

$$\|\underline{x}_j\|_{2,n}^2 \leq n + \frac{1}{32R'} n \leq 2n$$

となるので, 補題 3.30 ($\sigma^2 \longrightarrow 2n\sigma^2$) を用いると

$$\Pr\left(\max_{k \in [d]} \frac{g_{x,\sigma}}{\lambda n} \leq 4\sqrt{2}\,\sigma \sqrt{n \log\left(\frac{1}{\delta}\right)} \right) \geq 1-\delta$$

がわかる. さらに

$$\max_{k \in [d]} \frac{g_{*,k}}{\lambda k} \leq 4\sqrt{2}\,\sigma \sqrt{\log\left(\frac{1}{\delta}\right)}$$

$$\Rightarrow \quad \underline{\varepsilon}^{T} \underline{X} (\hat{\beta}^{S} - \beta^{*}) \leq 4\sqrt{2}\,\sigma \sqrt{\log\left(\frac{1}{\delta}\right)}$$

$$\Rightarrow \quad \|\underline{X}\hat{\beta}^{S} - \underline{X}\beta^{*}\|_{2,n}^{2} + n\tau \|\hat{\beta}^{S} - \beta^{*}\|_{*,d}$$

$$\lesssim \sigma \sqrt{\log\left(\frac{1}{\delta}\right)} \sqrt{k \log\left(\frac{2de}{k}\right)}\, \|\underline{X}\hat{\beta}^{S} - \underline{X}\beta^{*}\|_{2,n}$$

がわかる.

以上から

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^{S} - \underline{X}\beta^{*}\|_{2,n} \lesssim \sigma \frac{k}{n}\sqrt{\log\left(\frac{2de}{k}\right)}\sqrt{\log\left(\frac{1}{\delta}\right)} \right)$$

$$\geq \Pr\left(\max_{k\in[d]} \frac{g_{x,k}}{\lambda_k} \leq 4\sqrt{2}\,\sigma\sqrt{\log\left(\frac{1}{\delta}\right)} \right)$$

$$\geq 1-\delta$$

とわかる。 $\square$