第3章　高次元線型回帰モデル

この章では, 次の回帰モデルを考える. $d, n \in \mathbb{N}$ とし[1]

$$Y_j = f(\boldsymbol{X}_j) + \epsilon_j \qquad (j = 1, 2, \ldots, n)$$

とする. ただし, f は $\mathbb{R}^d$ 上の実数値関数で, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\top$ は代行分散 $\sigma^2\,(\sigma > 0)$ の劣 Gauss 型確率ベクトルで, $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ である. 目標は, 線型性の仮定のもとで, f の推定を行うことである. すなわち, 未知のベクトル $\boldsymbol{\beta}^* \in \mathbb{R}^d$ があって

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta}^*$$

を仮定することである. したがって

$$Y_j = \boldsymbol{x}_j^\top \boldsymbol{\beta}^* + \epsilon_j \qquad (j = 1, 2, \ldots, n) \tag{3.1}$$

を考える.

3.1　ランダムと固定デザインの線型回帰モデル

3.1.1　ランダムデザインの場合

線型回帰モデル (3.1) において, ランダムデザインの場合は, 統計的機械学習理論の枠組みに対応する. $n \in \mathbb{N}$ とし

$$(\boldsymbol{X}_1, Y_1), (\boldsymbol{X}_2, Y_2), \ldots, (\boldsymbol{X}_{n+1}, Y_{n+1})$$

をランダムな組とする. 観測 $(\boldsymbol{X}_1, Y_1), (\boldsymbol{X}_2, Y_2), \ldots, (\boldsymbol{X}_n, Y_n)$ に基づき, f の推定量 $\widehat{f}_n$ をうまく構成して, $\widehat{f}_n(\boldsymbol{X}_{n+1})$ が Y_{n+1} の「よい」予測量とすることが目標である. $\widehat{f}_n$ を構成するときには, $\boldsymbol{X}_{n+1}$ は未知の観測 (未来の観測) であり, $\boldsymbol{X}_{n+1}$ のどんな値にたいしても対応しなければならない.

---

[1] d と n の大小関係は仮定していないことに注意せよ.

予測量の自然な評価尺度は L_2 リスクである. $\mathsf{E}[\cdot]$ を $(\boldsymbol{X}_{n+1}, Y_{n+1})$ の分布に関する期待値の記号とする.

$$\begin{aligned}\mathsf{R}(\widehat{f}_n) &= \mathsf{E}[\{Y_{n+1} - \widehat{f}_n(\boldsymbol{X}_{n+1})\}^2] \\ &= \mathsf{E}[\{Y_{n+1} - f(\boldsymbol{X}_{n+1}) + f(\boldsymbol{X}_{n+1}) - \widehat{f}_n(\boldsymbol{X}_{n+1})\}^2] \\ &= \mathsf{E}[\{Y_{n+1} - f(\boldsymbol{X}_{n+1})\}^2] + \mathsf{E}[\{f(\boldsymbol{X}_{n+1}) - \widehat{f}_n(\boldsymbol{X}_{n+1})\}^2] \\ &\quad +2\mathsf{E}[\{Y_{n+1} - f(\boldsymbol{X}_{n+1})\}\{f(\boldsymbol{X}_{n+1}) - \widehat{f}_n(\boldsymbol{X}_{n+1})\}] \\ &= \mathsf{E}[\{Y_{n+1} - f(\boldsymbol{X}_{n+1})\}^2] + \mathsf{E}[\{f(\boldsymbol{X}_{n+1}) - \widehat{f}_n(\boldsymbol{X}_{n+1})\}^2] \\ &= \mathsf{E}[\{Y_{n+1} - f(\boldsymbol{X}_{n+1})\}^2] + \|\widehat{f}_n - f\|^2_{L^2(P^{\boldsymbol{X}})}\end{aligned}$$

となる. なぜならば, $\widehat{f}_n$ は $(\boldsymbol{X}_1, Y_2), \ldots, (\boldsymbol{X}_n, Y_n)$ にのみ依存し, ϵ_{n+1} と $\boldsymbol{X}_{n+1}$ の独立性より

$$\begin{aligned}&\mathsf{E}[\{Y_{n+1} - f(\boldsymbol{X}_{n+1})\}\{f(\boldsymbol{X}_{n+1}) - \widehat{f}_n(\boldsymbol{X}_{n+1})\}] \\ &\qquad = \mathsf{E}[\epsilon_{n+1}\{f(\boldsymbol{X}_{n+1}) - \widehat{f}_n(\boldsymbol{X}_{n+1})\}] \\ &\qquad = \mathsf{E}[\epsilon_{n+1}]\,\mathsf{E}[f(\boldsymbol{X}_{n+1}) - \widehat{f}_n(\boldsymbol{X}_{n+1})] = 0\end{aligned}$$

となることからわかる. また, $\boldsymbol{X}_{n+1}$ の分布を $P^{\boldsymbol{X}}$ と書いたとき

$$\|\widehat{f}_n - f\|^2_{L^2(P^{\boldsymbol{X}})} = \int_{\mathbb{R}^d} \{\widehat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\}^2 \, dP^{\boldsymbol{X}}(\boldsymbol{x})$$

で定義している. Y_{n+1} の予測量としての「よさ」は $\boldsymbol{X}_{n+1}$ の分布に関して平均をとる. ϵ_{n+1} の分散を $\sigma^2\,(\sigma > 0)$ としたとき

$$\mathsf{R}(\widehat{f}_n) = \sigma^2 + \|\widehat{f}_n - f\|^2_{L^2(P^{\boldsymbol{X}})}$$

となる. したがって, ランダムデザインに対して, L_2 ノルム $\|\widehat{f}_n - f\|^2_{L^2(P^{\boldsymbol{X}})}$ が精度の尺度となる. $\boldsymbol{X}_{n+1}$ の実現の平均をとって, 未知の f に $\widehat{f}_n$ が近いかどうかを測っている.

3.1.2 固定デザインの場合

この場合には, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ はランダムではない. このことを強調するために, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ と記すことにする. もちろん, $\boldsymbol{x}_j\,(j = 1, 2, \ldots, n)$ を $\boldsymbol{X}_j$ の実現値とみなすこともできる. しかし, 固定デザインとランダムデザインの相違は大きく, 予測量の尺度が異なってくる. ランダムデザインでは, $\boldsymbol{X}_{n+1}$ の周辺分布に関して平均をとって, 予測量の精度を調べた. 固定デザインの場合には, そのようなこと ($\boldsymbol{X}_{n+1}$ の周辺分布に関して平均ととる) はできない. デザイン $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ は固定と考えるので, 目標は f の推定のみになる. 問題の目標は

$$(Y_1, \boldsymbol{x}_1), (Y_2, \boldsymbol{x}_2), \ldots, (Y_n, \boldsymbol{x}_n)$$

の観測値に基づき, $f(\boldsymbol{x}_1)$, $f(\boldsymbol{x}_2)$, ..., $f(\boldsymbol{x}_n)$ を回復することになるので, 「ノイズ除去問題」ともよばれている.

多くの場合, 固定デザインには構造がある. 典型的な例は, 閉区間 $[0, 1]$ 上の均一デザインである. すなわち, $d = 1$ で $x_j = j/n\,(j = 1, 2, \ldots, n)$ である. この間の補間はなんらかのやり方が可能であろう.

固定デザインにおいて, $\boldsymbol{\mu} + \boldsymbol{\epsilon}$ を観測する. ただし

$$\boldsymbol{\mu} = (f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots\ f(\boldsymbol{x}_n))^\top \in \mathbb{R}^n$$

で, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\top \in \mathbb{R}^n$ は代行分散 $\sigma^2\,(\sigma > 0)$ の劣 Gauss 型分布に従う. 関数 f の推定問題とみなすよりも, $\mathbb{R}^n$ のベクトルを推定する問題と考えた方が議論は容易になる. 固定デザインにおいて, 平均 2 乗誤差 (the mean squred error, MSE) を精度の尺度とする. これを

$$\mathsf{MSE}(\widehat{f}_n) = \frac{1}{n}\sum_{j=1}^{n}\{\widehat{f}_n(\boldsymbol{x}_j) - f(\boldsymbol{x}_j)\}^2$$

で定義する. ベクトルの推定問題とみなし

$$\mathsf{MSE}(\widehat{\boldsymbol{\mu}}) = \frac{1}{n}\sum_{j=1}^{n}\{\widehat{u}_j - \mu_j\}^2 = \frac{1}{n}|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}|_{2,n}^2$$

と定義してもよい. ただし, $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_n)$, $|\boldsymbol{x}|_{2,n} = \sqrt{\boldsymbol{x}^\top\boldsymbol{x}}\,(\boldsymbol{x} \in \mathbb{R}^n)$ である.

デザイン $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ を用いて, $n \times d$ のデザイン行列を

$$\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$$

とする. この記号を用いれば, $f(\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{\beta}^*$ より

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}, \qquad \boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top, \qquad \boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\top \tag{3.2}$$

を得る. さらに

$$\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}) = \frac{1}{n}|\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2 = \frac{1}{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top\mathbf{X}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \tag{3.3}$$

となることがわかる.

3.2 最小 2 乗推定量の再訪問

ここでは, $\boldsymbol{\beta}^*$ を推定するかわりに, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}^*$ の推定問題を扱う. $\mathbf{X}^\top\mathbf{X}$ が正則でない場合も考えたいので, 一般化逆行列を導入する.

定義 3.1. 任意の $\mathbf{A} \in \mathrm{Mat}(m,\, n;\, \mathbb{R})$ に対して, 次の条件をみたす行列 $\mathbf{G} \in \mathrm{Mat}(n,\, m;\, \mathbb{R})$ を $\mathbf{A}$ の Moore-Penrose の一般化逆行列といい, これを $\mathbf{A}^\dagger$ と記す.
(1) $\mathbf{AGA} = \mathbf{A}$.
(2) $\mathbf{GAG} = \mathbf{G}$.
(3) $(\mathbf{AG})^\top = \mathbf{AG}$.
(4) $(\mathbf{GA})^\top = \mathbf{GA}$.

注意 3.2. 条件 (1) ∼ (4) をみたす行列 $\mathbf{G}$ は一意的に存在することが知られている. ハーベィル (2012, 下巻の p.191) を参照のこと. □

3.2.1 制約なしの最小 2 乗推定量

命題 3.3. $\mathbf{X}\boldsymbol{\beta}^*$ の最小 2 乗推定量 $\widehat{\boldsymbol{\mu}}^{\mathrm{LS}} = \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ は方程式

$$\mathbf{X}^\top \widehat{\boldsymbol{\mu}}^{\mathrm{LS}} = \mathbf{X}^\top Y$$

をみたす. さらに

$$\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} = \mathbf{X}^\dagger \boldsymbol{Y}$$

と書ける.

Proof. 関数

$$\boldsymbol{\beta} \mapsto |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,\,n}^2$$

は凸なので, 最小値は

$$\nabla_{\boldsymbol{\beta}} |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,\,d}^2 = \mathbf{0}$$

をみたす. ただし

$$\nabla_{\boldsymbol{\beta}} = \left(\frac{\partial}{\partial \beta_1},\, \frac{\partial}{\partial \beta_2},\, \ldots,\, \frac{\partial}{\partial \beta_d} \right)^\top; \qquad \boldsymbol{\beta} = (\beta_1,\, \beta_2,\, \ldots,\, \beta_d)^\top$$

である. $k = 1,\, 2,\, \ldots,\, d$ に対して

$$\begin{aligned} \frac{\partial}{\partial \beta_k} |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_2^2 &= \frac{\partial}{\partial \beta_k} \sum_{j=1}^n \{Y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta}\}^2 \\ &= -2 \sum_{j=1}^n \{Y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta}\} x_{jk} \end{aligned}$$

となる. ただし, $\boldsymbol{x}_j = (x_{j1},\, x_{j2},\, \ldots,\, x_{jd})^\top$ である. よって

$$\nabla_{\boldsymbol{\beta}} |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,\,n}^2 = -2 \sum_{j=1}^n \bigl(Y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta}\bigr) \boldsymbol{x}_j = -2\bigl(\mathbf{X}^\top \boldsymbol{Y} - \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}\bigr)$$

となる．したがって, $\nabla_{\boldsymbol{\beta}}|\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta}|_{2,n}^2=\mathbf{0}$ より

$$\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}=\mathbf{X}^\top\boldsymbol{Y}$$

を得る．2 番目の主張は Moore-Penrose の一般化逆行列の定義より

$$\begin{aligned}\mathbf{X}^\top\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}&=\mathbf{X}^\top\mathbf{X}\mathbf{X}^\dagger\boldsymbol{Y}=\mathbf{X}^\top\big(\mathbf{X}\mathbf{X}^\dagger\big)\boldsymbol{Y}=\mathbf{X}^\top\big(\mathbf{X}\mathbf{X}^\dagger\big)^\top\boldsymbol{Y}=\big(\mathbf{X}\mathbf{X}^\dagger\mathbf{X}\big)^\top\boldsymbol{Y}\\&=\mathbf{X}^\top\boldsymbol{Y}\end{aligned}$$

となることからわかる． □

命題 3.4. 線型モデル (3.2) を仮定し, $\boldsymbol{\epsilon}\sim\mathsf{subG}_n(\sigma^2)\,(\sigma>0)$ とする．このとき, 最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ は方程式

$$\mathsf{E}\big[\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}})\big]=\frac{1}{n}\mathsf{E}\big[|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2\big]\lesssim\sigma^2\frac{r}{n}$$

をみたす．ただし, $r=\mathrm{rank}(\mathbf{X}^\top\mathbf{X})$ である．さらに, $\forall\delta>0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}})\lesssim\frac{\sigma^2\{r+\log(1/\delta)\}}{n}\right)\geq1-\delta$$

となる．

Proof. 最小 2 乗推定量の定義より

$$|\boldsymbol{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}|_{2,n}^2\leq|\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2=|\boldsymbol{\epsilon}|_{2,n}^2\tag{3.4}$$

である．さらに

$$\begin{aligned}|\boldsymbol{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}|_{2,n}^2&=|\mathbf{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}-\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}|_{2,n}^2\\&=|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2-2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\boldsymbol{\beta}^*)+|\boldsymbol{\epsilon}|_{2,n}^2\end{aligned}$$

となることがわかる．この式を (3.4) の右辺に代入すれば

$$\begin{aligned}|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2&\leq2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\boldsymbol{\beta}^*)\\&=2|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*|_{2,n}\frac{\boldsymbol{\epsilon}^\top(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*)}{|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*|_{2,n}}\end{aligned}\tag{3.5}$$

となることがわかる．しかし, (3.5) の最右辺の項 $\boldsymbol{\epsilon}^\top(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*)/|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*|_2$ を直接評価するのは難しい．なぜならば, この量は $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ に依存しており, その構造は複雑である． 古典的なやり方は, 上限 (sup) をとることにより $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ を取り除くことである．ただし, すこしだけ注意が必要である． $\boldsymbol{\Psi}=[\boldsymbol{\psi}_1,\boldsymbol{\psi}_2,\ldots,\boldsymbol{\psi}_r]\in\mathrm{Mat}(n,r;\mathbb{R})$ は $\mathbf{X}$ の列で張られ

る $\mathbb{R}^n$ の部分空間の正規直交基底とする．すると, $\boldsymbol{\nu} \in \mathbb{R}^r$ が存在して, $\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \boldsymbol{\beta}^*) = \boldsymbol{\Psi}\boldsymbol{\nu}$ と書ける．このことより

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \boldsymbol{\beta}^*)}{|\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \boldsymbol{\beta}^*)|_{2,n}} = \frac{\boldsymbol{\epsilon}^\top \boldsymbol{\Psi}\boldsymbol{\nu}}{|\boldsymbol{\Psi}\boldsymbol{\nu}|_{2,n}} = \frac{\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{\nu}}{|\boldsymbol{\nu}|_{2,r}} \le \sup_{\boldsymbol{u} \in \mathbb{B}_2} \tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u}$$

となることがわかる．ただし, B_2 は $\mathbb{R}^r$ の単位球であり, $\tilde{\boldsymbol{\epsilon}} = \boldsymbol{\Psi}^\top \boldsymbol{\epsilon}$ である．したがって

$$|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2 \le 2|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n} \sup_{\boldsymbol{u} \in B_2} \tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u}$$

より

$$|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n} \le 2 \sup_{\boldsymbol{u} \in B_2} \tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u}$$

となる．これより

$$|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2 \le 4 \sup_{\boldsymbol{u} \in B_2} (\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u})^2 \tag{3.6}$$

がわかる．次に, $\boldsymbol{v} \in \mathbb{S}^{r-1}$ としたとき

$$|\boldsymbol{\Psi}\boldsymbol{v}|_{2,n}^2 = \boldsymbol{v}^\top \boldsymbol{\Psi}^\top \boldsymbol{\Psi}\boldsymbol{v} = \boldsymbol{v}^\top \boldsymbol{v} = 1$$

であることに注意する．これより $\forall s \in \mathbb{R}$ に対して

$$\mathsf{E}[\exp\{s\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{v}\}] = \mathsf{E}\big[\exp\{s\boldsymbol{\epsilon}^\top \boldsymbol{\Psi}\boldsymbol{v}\}\big] \le e^{s^2\sigma^2/2}$$

となることがわかる．したがって, $\tilde{\boldsymbol{\epsilon}} \sim \mathsf{subG}_r(\sigma^2)$ となる．期待値の上限を得るために, 補題 2.11 を用いると

$$4\mathsf{E}\Big[\sup_{\boldsymbol{v} \in B_2} (\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{v})^2\Big] = 4\sum_{j=1}^r \mathsf{E}[\tilde{\epsilon}_j^2] \le 16\sigma^2 r \tag{3.7}$$

であることがわかる．よって, (3.6) と (3.7) を合わせると

$$\frac{1}{n}\mathsf{E}\big[|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2\big] \le \frac{4}{n}\mathsf{E}\Big[\sup_{\boldsymbol{u} \in B_2} \big(\boldsymbol{u}^\top \widetilde{\boldsymbol{\epsilon}}\big)^2\Big] \le \frac{16r\sigma^2}{n}$$

を得る．よって, 命題の 1 番目の主張が証明された．

つぎに, 2 番目の主張を証明するために, $B_2 \subset \mathbb{R}^r$ の (1/2) 網 N を考える．すると補題 2.16 から

$$\#(N) \le 6^r$$

となる．$\forall \boldsymbol{u} \in B_2$ に対して, ある $\boldsymbol{w} \in N$ と $\boldsymbol{v} \in (1/2)B_2$ が存在して

$$\boldsymbol{u} = \boldsymbol{w} + \boldsymbol{v}$$

と書ける．よって

$$\sup_{\boldsymbol{u}\in B_2}\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}\leq\max_{\boldsymbol{w}\in N}\boldsymbol{w}^\top\widetilde{\boldsymbol{\epsilon}}+\sup_{\boldsymbol{v}\in(1/2)B_2}\boldsymbol{v}^\top\widetilde{\boldsymbol{\epsilon}}$$

と

$$\sup_{\boldsymbol{v}\in(1/2)B_2}\boldsymbol{v}^\top\widetilde{\boldsymbol{\epsilon}}=\frac{1}{2}\sup_{\boldsymbol{v}\in B_2}\boldsymbol{v}^\top\widetilde{\boldsymbol{\epsilon}}$$

となるので

$$\sup_{\boldsymbol{u}\in B_2}\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}\leq 2\max_{\boldsymbol{u}\in N}\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}$$

がわかる．ここで命題 2.24 を用いると, $\forall t>0$ に対して

$$\begin{aligned}\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in B_2}\left(\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}\right)^2\geq t^2\right)&=\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in B_2}\left(\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}\right)\geq t\right)\leq\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in N}\left(\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}\right)\geq\frac{t}{2}\right)\\&\leq\#(N)\exp\left[-\frac{t^2}{8\sigma^2}\right]\leq 6^r\exp\left[-\frac{t^2}{8\sigma^2}\right]\end{aligned}$$

を得る．ここで

$$\delta:=6^r\exp\left[-\frac{t^2}{8\sigma^2}\right]\Leftrightarrow t^2=8\sigma^2\left\{r\log 6+\log(1/\delta)\right\}$$

となるので

$$\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in B_2}\left(\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}\right)^2\geq 8\sigma^2\left\{r\log 6+\log(1/\delta)\right\}\right)\leq\delta$$

を得る．(3.6) を上式に代入すると

$$\begin{aligned}&\mathsf{Pr}\left(\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2>2\sigma^2\left\{r\log 6+\log(1/\delta)\right\}\right)\\&\quad\leq\mathsf{Pr}\left(\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2\geq 2\sigma^2\left\{r\log 6+\log(1/\delta)\right\}\right)\\&\quad\leq\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in B_2}\left(\boldsymbol{u}^\top\widetilde{\boldsymbol{\epsilon}}\right)^2\geq 8\sigma^2\left\{r\log 6+\log(1/\delta)\right\}\right)\\&\quad\leq\delta\end{aligned}$$

を得る．最右辺の事象の補事象を取ると

$$\mathsf{Pr}\left(\frac{1}{n}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}-\mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2\leq\frac{2\sigma^2\left\{r\log 6+\log(1/\delta)\right\}}{n}\right)\geq 1-\delta$$

を得る． □

注意 3.5. $d \le n$ とする.

$$\mathbf{B} := \frac{1}{n}\mathbf{X}^\top \mathbf{X}$$

がランク d ならば,

$$|\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \boldsymbol{\beta}^*|_{2,d}^2 \lesssim \frac{\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}})}{\lambda_{\max}(\mathbf{B})}$$

である. ただし, $\lambda_{\max}(\mathbf{B})$ は行列 $\mathbf{B}$ の最小の固有値である. したがって, 命題 3.4 を用いて, $|\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} - \boldsymbol{\beta}^*|_{2,d}^2$ の上限を直接評価できる. しかし, 高次元データの設定の場合には, 構造上の仮定がより多く必要になってくる. □

3.2.2 制約付きの最小 2 乗推定量

$K \subset \mathbb{R}^d$ を対称な凸集合とする. すなわち, $-K := \{-\boldsymbol{x};\, \boldsymbol{x} \in K\}$ としたとき, $K = -K$ となることである.

事前情報 $\boldsymbol{\beta}^* \in K$ があると仮定する. すると, 制約付き最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}$ の方がよりよいであろう.

$$\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} \in \mathrm{argmin}_{\boldsymbol{\beta} \in K} |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2.$$

ただし, $\mathbb{R}^d$ 乗の実数値関数 g に対して

$$\{\boldsymbol{v} \in K;\, g(\boldsymbol{v}) \le g(\boldsymbol{u})\, (\forall \boldsymbol{u} \in K)\} \subset K$$

である. 基本方程式 (3.4) は常に成立しているので, $\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}})$ の上限はより小さくなる. 実際, (3.5) は次のように変形できる.

$$|\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2 \le 2\boldsymbol{\epsilon}^\top \mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*) \le 2 \sup_{\boldsymbol{u} \in K - K} (\boldsymbol{\epsilon}^\top \mathbf{X}\boldsymbol{u}).$$

ただし, $K - K := \{\boldsymbol{x} - \boldsymbol{y};\, \boldsymbol{x}, \boldsymbol{y} \in K\}$ である. K は対称かつ凸ならば, $K - K = 2K$. よって

$$2 \sup_{\boldsymbol{u} \in K - K} (\boldsymbol{\epsilon}^\top \mathbf{X}\boldsymbol{u}) = 4 \sup_{\boldsymbol{v} \in \mathbf{X}K} (\boldsymbol{\epsilon}^\top \boldsymbol{v}); \qquad \mathbf{X}K := \{\mathbf{X}\boldsymbol{u};\, \boldsymbol{u} \in K\} \subset \mathbb{R}^n$$

となる.

ℓ_1 制約付き最小 2 乗推定量

B_1 を $\mathbb{R}^d$ の ℓ_1 単位球とする.

$$B_1 := \{\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top \in \mathbb{R}^d;\, \sum_{j=1}^d |x_j| \le 1\}.$$

B_1 はちょうど $2d$ 個の点から成る頂点集合

$$\mathcal{V}(B_1) := \{\boldsymbol{e}_1,\, -\boldsymbol{e}_1,\, \boldsymbol{e}_2,\, -\boldsymbol{e}_2,\, \ldots,\, \boldsymbol{e}_d,\, -\boldsymbol{e}_d\}$$

を持つ．ただし，

$$\boldsymbol{e}_j = (\underbrace{0,\, \ldots,\, 0}_{j-1},\, 1,\, 0,\, \ldots,\, 0)^\top; \quad (j = 1,\, 2,\, \ldots,\, d)$$

である．$K = B_1$ と書くことにする．部分集合 $\mathbf{X}K := \{\mathbf{X}\boldsymbol{\beta}^*;\, \boldsymbol{\beta}^* \in K\} \subset \mathbb{R}^n$ は最大 $2d$ 個の頂点をもつ凸多面体で

$$\mathbf{X}\mathcal{V} := \{\boldsymbol{y}_1,\, -\boldsymbol{y}_1,\, \boldsymbol{y}_2,\, -\boldsymbol{y}_2,\, \ldots,\, \boldsymbol{y}_d,\, -\boldsymbol{y}_d\}; \qquad \mathbf{X} = (\boldsymbol{y}_1,\, \boldsymbol{y}_2,\, \ldots,\, \boldsymbol{y}_d)$$

に含まれる．実際，$d \le n$ のとき，凸多面体 K をリスケールすることにより，$\mathbf{X}K$ を得ることができる．行列 $\mathbf{X}$ のある列は $\mathbf{X}K$ の頂点ではないかもしれない．したがって，$\mathbf{X}\mathcal{V}$ は $\mathbf{X}K$ の頂点集合を含む集合となる．これは，$\mathbf{X}K$ のサイズを測っていることになる．$\boldsymbol{\epsilon} \sim \mathsf{N}_n(\mathbf{0}_n,\, \mathbf{I}_n)$ のとき，上式の上限の期待値は $\mathbf{X}K$ の Gauss 幅となる．$\boldsymbol{\epsilon}$ は正規分布ではなく，劣 Gauss 的な分布に従うときも同じような性質が成立する．

命題 3.6. *$d \ge 2$ とし，$K = B_1$ を $\mathbb{R}^d$ の ℓ_1 単位球とする．$\boldsymbol{\beta}^* \in B_1$ と仮定する．$\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)\,(\sigma > 0)$ とし，$\mathbf{X}$ の列は規準化されていて，*

$$\max_{1 \le j \le d} |\mathbf{X}_j|_{2,\,d} \le \sqrt{n}; \qquad X_j \text{ は } \mathbf{X} \text{ の第 } j \text{ 列}$$

とする．このとき，制約付き最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}$ は

$$\mathsf{E}[\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}})] = \frac{1}{n}\mathsf{E}\left[|\mathbf{X}\widehat{\boldsymbol{\beta}}_{B_1}^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,\,n}^2\right] \precsim \sigma\sqrt{\frac{\log d}{n}}$$

をみたす．さらに，$\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}) \precsim \sigma\sqrt{\frac{\log(d/\delta)}{n}}\right) \ge 1 - \delta$$

が成り立つ．

Proof. この命題の前の議論より

$$|\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_{2,\,n}^2 \le 4 \sup_{\boldsymbol{v} \in \mathbf{X}K} (\boldsymbol{v}^\top \boldsymbol{\epsilon})$$

となることがわかる．ここで，$\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)$ であり，$\mathbf{X}$ のすべての列は $|X_j|_{2,\,d} \le \sqrt{n}\,(j = 1,\, 2,\, \ldots,\, d)$ だったので，$\boldsymbol{\epsilon}^\top X_j \sim \mathsf{subG}(n\sigma^2)$ となることに注意する．したがって

$$\mathsf{Pr}\Big(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}\big) \ge t)\Big) \le \mathsf{Pr}\left(\sup_{\boldsymbol{v} \in \mathbf{X}K} (\boldsymbol{\epsilon}^\top \boldsymbol{v}) \ge \frac{t}{4}\right) \le 2d \exp\left(-\frac{nt^2}{32\sigma^2}\right)$$

となることがわかる．最後に

$$
\begin{aligned}
2d\exp\left(-\frac{nt^2}{32\sigma^2}\right) \le \delta \;&\Leftrightarrow\; \log(2d) - \frac{nt^2}{32\sigma^2} \le \log\delta \\
&\Leftrightarrow\; \log(2d) - \log\delta \le \frac{nt^2}{32\sigma^2} \\
&\Leftrightarrow\; t^2 \ge 32\sigma^2\left(\frac{\log(2d)}{n} + \frac{\log(1/\delta)}{n}\right) \\
&\Leftarrow\; t \ge C\sigma\sqrt{\frac{\log d}{n}}; \qquad C \text{ はある正の定数}
\end{aligned}
$$

より

$$
\mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}\big) \lesssim \sigma\sqrt{\frac{\log d}{n}}\right) \ge 1-\delta
$$

がわかる． □

注意 3.7. 命題 2.30 の議論を $\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}$ に適用することができるので,

$$
\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}\big) \lesssim \sigma^2\frac{d}{n}
$$

ともなる．したがって

$$
\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}_{B_1}^{\mathrm{LS}}) \lesssim \min\left(\sigma\sqrt{\frac{\log d}{n}},\, \sigma^2\frac{d}{n}\right)
$$

となる．これを肘作用とよび, 肘は

$$
d \approx \sqrt{n}
$$

付近になることがわかる． □

ℓ_0 制約付き最小 2 乗推定量

記号を乱用して, $\boldsymbol{\beta} \in \mathbb{R}^d$ の零でない要素の個数をその ℓ_0 ノルムということにする．すなわち,

$$
|\boldsymbol{\beta}|_{0,d} := \sum_{j=1}^{d} \mathbb{1}\{\beta_j \ne 0\}; \qquad \boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)^\top.
$$

ℓ_0 ノルムが小さい $\boldsymbol{\beta}^*$ をスパースなベクトルとよぶことにする．正確には, $|\boldsymbol{\beta}|_{0,d} \le k$ のとき, k スパースベクトルという．ベクトル $\boldsymbol{\beta}$ の台を $\mathrm{supp}(\boldsymbol{\beta})$ と書き,

$$
\mathrm{supp}(\boldsymbol{\beta}) := \{j \in \{1, 2, \ldots, n\};\, \beta_j \ne 0\}
$$

で定義する．すると

$$|\boldsymbol{\beta}|_{0,\,d} = \#(\mathrm{supp}(\boldsymbol{\beta}))$$

となる．ただし, 集合 A に対して, $\#(A)$ は A の要素の個数を表す.

注意 3.8. ℓ_0 ノルムという用語は下記から来ている．$\boldsymbol{\beta} = (\beta_1,\, \beta_2,\, \ldots,\, \beta_d)^\top \in \mathbb{R}^d$ の対して

$$\lim_{q \to 0+} \sum_{j=1}^{d} |\beta_j|^q = |\boldsymbol{\beta}|_{0,\,d}$$

となる． □

$1 \le k \le d$ に対し, $B_0(k)$ を $\mathbb{R}^d$ の ℓ_0 球とする．すなわち

$$B_0(k) := \{\boldsymbol{\beta} \in \mathbb{R}^d;\, |\boldsymbol{\beta}|_{0,\,d} \le k\}.$$

この節の目標は, $K = B_0(k)$ のとき, 制約付き最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)}$ の MSE を評価することである．$\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)}$ を計算するために, $\begin{pmatrix} d \\ k \end{pmatrix}$ 回最小 2 乗推定量を計算する必要がある．この計算量は k の指数オーダーである．実用上, これは計算困難であるが, $\begin{pmatrix} d \\ k \end{pmatrix}$ の統計的性質を調べることは意味がある.

命題 3.9. 正の整数 $k \le d/2$ を固定する．$K = B_0(k)$ を $\mathbb{R}^d$ の k スパースベクトルの集合とし, $\boldsymbol{\beta}^* \in B_0(k)$ と仮定する．$\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)\,(\sigma > 0)$ とする．このとき, $\forall \delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)}) \le \frac{\sigma^2}{n}\log\begin{pmatrix} d \\ 2k \end{pmatrix} + \frac{\sigma^2 k}{n} + \frac{\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\right) \ge 1 - \delta$$

が成立する.

Proof. $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)}$ を $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K}$ と書く．命題 3.4 の証明中の議論より

$$\begin{aligned}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K} - \mathbf{X}\boldsymbol{\beta}^*\right|^2_{2,\,n} &\le 2\boldsymbol{\epsilon}^\top(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K} - \mathbf{X}\boldsymbol{\beta}^*) \\ &= 2\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K} - \mathbf{X}\boldsymbol{\beta}\right|_{2,\,n} \frac{\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K} - \boldsymbol{\beta}^*)}{\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K} - \boldsymbol{\beta}^*)}\end{aligned}$$

となる．$\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K}$ と $\boldsymbol{\beta}^*$ は $B_0(k)$ の中にあるので, $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{K} - \boldsymbol{\beta}^* \in B_0(2k)$ となる．任意の $S \subset \{1,\, 2,\, \ldots,\, d\}$ に対して, $\mathbf{X}_S$ を $\mathbf{X}$ の $n \times \#(S)$ の部分行列で, $\mathbf{X}$ の第 j 列 $(j \in S)$ によって構成されているものとする．$\mathrm{rank}(\mathbf{X}_S) = r_S$ とおくと, $r_S \le \#(S)$ となる．$\boldsymbol{\Psi}_S = [\boldsymbol{\psi}_1,\, \boldsymbol{\psi}_2,\, \ldots,\, \boldsymbol{\psi}_{r_S}] \in \mathrm{Mat}(n,\, r_S;\, \mathbb{R})$

を $\mathbf{X}_S$ の列が張る部分空間の正規直交基底とする．さらに, $\boldsymbol{\beta} \in \mathbb{R}^d$ に対して, $\boldsymbol{\beta}(S) \in \mathbb{R}^{\#(S)}$ を $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_d)^\top$ の $\beta_j\ (j \in S)$ を成分ともつベクトルとする．いま

$$\widehat{S} = \mathrm{supp}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)$$

とおけば, $\#(\widehat{S}) \leq 2k$ となる．よって $\exists \boldsymbol{\nu} \in \mathbb{R}^{r_{\widehat{S}}}$ があって

$$\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*) = \mathbf{X}_{\widehat{S}}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}(\widehat{S}) - \boldsymbol{\beta}^*(\widehat{S})) = \boldsymbol{\Psi}_{\widehat{S}}\boldsymbol{\nu}$$

となる．したがって

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)}{|\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)|_{2,n}} = \frac{\boldsymbol{\epsilon}^\top \boldsymbol{\Psi}_{\widehat{S}}\boldsymbol{\nu}}{|\boldsymbol{\nu}|_{2, r_{\widehat{S}}}} \leq \max_{\#(S)=2k}\left(\sup_{\boldsymbol{u} \in B_2^{r_S}} \boldsymbol{\epsilon}^\top \boldsymbol{\Psi}\boldsymbol{u}\right)$$

となる．ただし, $B_2^{r_S}$ は $\mathbb{R}^{r_S}$ の単位球である．これより

$$\left|\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)\right|_{2,n}^2 \leq 2\left|\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)\right|_{2,n} \max_{\#(S)=2k}\left(\sup_{\boldsymbol{u} \in B_2^{r_S}} \boldsymbol{\epsilon}^\top \boldsymbol{\Psi}_S\boldsymbol{u}\right)$$

となる．よって

$$\left|\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)\right|_{2,n}^2 \leq 4 \max_{\#(S)=2k}\left(\sup_{\boldsymbol{u} \in B_2^{r_S}} (\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u})^2\right); \qquad \tilde{\boldsymbol{\epsilon}} = \boldsymbol{\Psi}_S^\top \boldsymbol{\epsilon} \sim \mathsf{subG}_{r_S}(\sigma^2)$$

となる．和集合上限を用いれば, $\forall t > 0$ に対して

$$\mathsf{Pr}\left(\max_{\#(S)=2k}\left\{\sup_{\boldsymbol{u} \in B_2^{r_S}} (\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u})^2\right\} \geq t\right) \leq \sum_{\#(S)=2k} \mathsf{Pr}\left(\sup_{\boldsymbol{u} \in B_2^{r_S}} (\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u})^2 \geq t\right)$$

を得る．命題 2.30 の証明中の (2.5) と $\#(S) \leq 2k$ より

$$\mathsf{Pr}\left(\sup_{\boldsymbol{u} \in B_2^{r_S}} (\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{u})^2 \geq t\right) \leq 6^{\#(S)} \exp\left(-\frac{t}{8\sigma^2}\right) \leq 6^{2k} \exp\left(-\frac{t}{8\sigma^2}\right)$$

となる．これらの式をあわせれば

$$\mathsf{Pr}\left(\left|\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)\right)\right|_{2,n}^2 \geq 4t\right) \leq \binom{d}{2k} 6^{2k} \exp\left(-\frac{t}{8\sigma^2}\right) \tag{3.8}$$

となることがわかる．ここで

$$\delta := \binom{d}{2k} 6^{2k} \exp\left(-\frac{t}{8\sigma^2}\right)$$

とおけば

$$\log\delta = \log\begin{pmatrix} d \\ 2k \end{pmatrix} + 2k\log 6 - \frac{t}{8\sigma^2}$$

を得る. よって

$$t = 8\sigma^2\left\{\log\begin{pmatrix} d \\ 2k \end{pmatrix} + 2k\log 6 + \log\frac{1}{\delta}\right\}$$

となることがわかる. これより

$$\begin{aligned}\mathsf{Pr}\Big(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}\big) \geq t\Big) &= \mathsf{Pr}\Big(\big|\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}} - \boldsymbol{\beta}^*)\big|_{2,\,d}^2 \geq nt\Big) \\ &\leq \begin{pmatrix} d \\ 2k \end{pmatrix} 6^{2k}\exp\left(-\frac{nt}{8\sigma^2}\right)\end{aligned}$$

なので

$$\mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\mathrm{LS}}\big) \precsim \frac{\sigma^2}{n}\log\begin{pmatrix} d \\ 2k \end{pmatrix} + \frac{\sigma^2 k}{n} + \frac{\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\right) \geq 1-\delta$$

を得る. □

補題 3.10. 整数 $1 \leq k \leq n$ に対して,

$$\begin{pmatrix} n \\ k \end{pmatrix} \leq \left(\frac{en}{k}\right)^k$$

が成り立つ.

Proof. 帰納法で示す. $k = 1$ のとき,

$$\begin{pmatrix} n \\ 1 \end{pmatrix} = n \leq en = \left(\frac{en}{1}\right)^1.$$

次に, ある $k \leq n-1$ で

$$\begin{pmatrix} n \\ k \end{pmatrix} \leq \begin{pmatrix} en \\ k \end{pmatrix}^k$$

が成立していると仮定する.

$$\begin{aligned}\begin{pmatrix} n \\ k+1 \end{pmatrix} &= \begin{pmatrix} n \\ k \end{pmatrix}\frac{n-k}{k+1} \leq \left(\frac{en}{k}\right)^k\frac{n-4}{k+1} \leq \frac{e^k n^{k+1}}{(k+1)^{k+1}}\left(1+\frac{1}{k}\right)^k \\ &\leq \frac{e^k n^{k+1}}{(k+1)^{k+1}}e = \left(\frac{en}{k+1}\right)^{k+1}\end{aligned}$$

よりわかる. □

系 3.11. 命題 3.29 と同じ仮定のもと, $\forall\delta > 0$ に対して,

$$\mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)}\big) \precsim \frac{\sigma^2}{n}\left\{k\log\left(\frac{ed}{2k}\right) + k\log 6 + \log\left(\frac{1}{\delta}\right)\right\}\right) \geq 1-\delta$$

が成り立つ.

Proof. 命題 3.9 と補題 3.10 より直ちにわかる. □

系 3.12. 命題 3.9 と同じ仮定のもとで

$$\mathsf{E}\left[\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)}\big)\right] \precsim \frac{\sigma^2 k}{n}\log\left(\frac{ed}{k}\right)$$

が成り立つ.

Proof. (3.8) から, $\forall H \geq 0$ に対して,

$$\begin{aligned}
\mathsf{E}\left[\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}_{B_0(k)}\big)\right] &= \int_0^\infty \mathsf{Pr}\left(\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)} - \mathbf{X}\boldsymbol{\beta}^*\right|_2^2 \geq nu\right) du \\
&\leq H + \int_0^\infty \mathsf{Pr}\left(\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{B_0(k)} - \mathbf{X}\boldsymbol{\beta}^*\right|_2^2 \geq n(u+H)\right) du \\
&\leq H + \binom{d}{2k} 6^{2k} \int_0^\infty \exp\left(-\frac{n(u+H)}{32\sigma^2}\right) du \\
&\leq H + \left(\frac{en}{2k}\right)^{2k} 6^k \left[-\frac{32\sigma^2}{n}\exp\left(-\frac{n(u+H)}{32\sigma^2}\right)\right]_0^\infty \\
&\leq H + \left(\frac{3de}{k}\right)^{2k} 6^{2k} \exp\left(-\frac{nH}{32\sigma^2}\right)
\end{aligned}$$

となる. ここで,

$$\left(\frac{3de}{k}\right)^{2k} 6^{2k} \exp\left(-\frac{nH}{32\sigma^2}\right) = 1$$

となるように H をとれば,

$$H = \frac{32\sigma^2 k}{n}\log\left(\frac{3de}{k}\right)$$

となるので,

$$H \precsim \frac{\sigma^2 k}{n}\log\left(\frac{de}{k}\right)$$

となる. □

3.3 Gauss 確率変数列モデル

Gauss 確率変数列モデルは以下のように書ける.

$$Y_j = \beta_j + \epsilon_j \qquad (j = 1, 2, \ldots, n) \tag{3.9}$$

ただし, $\epsilon_1, \epsilon_2, \ldots, \epsilon_n \sim N(0, \sigma^2)\,(\sigma > 0)$ である.

3.3.1 劣 Gauss 型確率変数列モデル

モデル (3.9) は固定デザインの回帰モデル (3.2) の特別な場合だといえる. $n = d$, $\boldsymbol{x}_j = \boldsymbol{e}_j\,(j = 1, 2, \ldots, d)$ とすればよい.

ORT 仮定 デザイン行列は以下をみたす.

$$\frac{\mathbf{X}^\top \mathbf{X}}{n} = I_d.$$

ORT 仮定は $d \le n$ の場合にのみ有効であり, 高次元モデル $d > n$ のときはみたされない.

この仮定のもとで, 回帰モデル (3.2) を以下のように変形できる.

$$\boldsymbol{y} := \frac{1}{n}\mathbf{X}^\top \boldsymbol{Y} = \frac{1}{n}\mathbf{X}^\top(\mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}) =: \boldsymbol{\beta}^* + \boldsymbol{\xi}.$$

ただし, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_d)^\top \sim \mathsf{subG}_n(\sigma^2/n)$. ORT 仮定のもとで, $\boldsymbol{\xi}$ が正規分布に従うとき, 回帰モデル (3.2) は正規確率変数列モデル (3.9) と (データ $\boldsymbol{y}$ をスケーリングすれば) 同じである. さらに, $\forall \boldsymbol{\beta}^* \in \mathbb{R}^d$ に対して

$$\begin{aligned}
|\boldsymbol{y} - \boldsymbol{\beta}^*|_{2,d}^2 &= \left|\frac{1}{n}\mathbf{X}^\top \boldsymbol{Y} - \boldsymbol{\beta}^*\right|_{2,d}^2 \\
&= |\boldsymbol{\beta}^*|_{2,d}^2 - \frac{2}{n}\boldsymbol{\beta}^{*\top}\mathbf{X}^\top \boldsymbol{Y} + \frac{1}{n^2}\boldsymbol{Y}^\top \mathbf{X}\mathbf{X}^\top \boldsymbol{Y} \\
&= \frac{1}{n}|\boldsymbol{\beta}^*\mathbf{X}|_{2,d}^2 - \frac{2}{n}(\mathbf{X}\boldsymbol{\beta}^*)^\top \boldsymbol{Y} + \frac{1}{n}|\boldsymbol{Y}|_{2,n}^2 + Q \\
&= \frac{1}{n}|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*|_{2,n}^2 + Q
\end{aligned}$$

となる. ただし,

$$Q = \frac{1}{n^2}\boldsymbol{Y}^\top \mathbf{X}\mathbf{X}^\top \boldsymbol{Y} - \frac{1}{n}|\boldsymbol{Y}|_2^2$$

である. よって, 最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ は $\boldsymbol{y}$ とある.

次に, もうすこし一般的なモデルである劣 Gauss 型確率変数列モデルを導入する.

$$\boldsymbol{y} = \boldsymbol{\beta}^* + \boldsymbol{\xi}; \qquad \xi \sim \mathsf{subG}_d(\sigma^2/n). \tag{3.10}$$

3.3.2 スパース適応的しきい値推定量

もし, $\boldsymbol{\beta}^* \in \mathbb{R}^d$ が k スパース（$1 \le k < d$）であることを知っているならば, 系 3.12 より, $\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{\mathbb{B}_0(k)}) \le C_\delta \frac{\sigma^2 k}{n} \log\left(\frac{ed}{2k}\right)\right) \ge 1-\delta$$

となることがわかる. ORT 条件のもと, スパース度 k を知っているという仮定をゆるめることができ, その値を適応的に求めることができる. ここで,「適応的」とは, k の真の情報はなくとも, $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}_{\mathbb{B}_0(k)}$ の MSE と乗数倍しか異ならない MSE をもつような推定量を構成できることである.

まず, 直観的な理解を得るために, 発見法的な議論をする.

(3.10) の劣 Gauss 的確率変数列モデルを仮定する. $\boldsymbol{\beta}^*$ の事前情報が全くないとき, 最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} = \boldsymbol{y}$ によって, $\boldsymbol{\beta}^*$ を推定するのが妥当であろう. この場合には, $\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}) = |\boldsymbol{y} - \boldsymbol{\beta}^*|_{2,\,d}^2 = |\boldsymbol{\xi}|_{2,\,d}^2 \le C_\delta \frac{\sigma^2 d}{n}\right) \ge 1-\delta$$

となる. ただし, C_δ は δ のみに依存する非負値定数である. ある $c \le 1$ があって, $k = cd$ のとき, 探しているものがこれである. k が d より十分小さいという事実を用いずに, $\boldsymbol{\beta}^*$ の要素の多くが零のとき, どれが零であるかを予想したい. $j \in \{1,\, 2,\, \ldots,\, d\}$ に対して, $\beta_j^* = 0$ のとき, $y_j = \xi_j$ は代行分散 $\sigma^2/n\,(\sigma > 0)$ の劣 Gauss 分布的な分布に従う. 特に補題 2.10 より $\forall t \in \mathbb{R}$ に対して

$$\mathsf{Pr}\left(|\xi_j| \ge t\right) \le 2\exp\left(-\frac{nt^2}{2\sigma^2}\right)$$

となる. ここで

$$\delta := 2\exp\left(-\frac{nt^2}{2\sigma^2}\right) \Longleftrightarrow t = \pm\sigma\sqrt{\frac{2\log(2/\delta)}{n}}$$

となるので, $\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(|\xi_j| \le \sigma\sqrt{\frac{2\log(2/\delta)}{n}} =: \tau\right) \ge 1-\delta \tag{3.11}$$

を得る. この不等式の結果は興味深い. (3.11) から, $j = 1,\, 2, \ldots,\, d$ に対して,

$$|y_j| \gg \tau \Longrightarrow \beta_j^* \ne 0$$

である. 他方,

$$|y_j| \ll \tau \Longrightarrow \beta_j^* \text{ は大きくない.}$$

したがって $|y_j| \le 2\tau$ ならば, $\beta_j^* = 0$ と考える.

定義 3.13. しきい値 2τ の硬しきい値推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}} = (\widehat{\beta}_1^{\mathrm{HRD}}, \widehat{\beta}_1^{\mathrm{HRD}}, \ldots, \widehat{\beta}_d^{\mathrm{HRD}})^\top$ は次で与えれれる．$j = 1, 2, \ldots, d$ に対して

$$\widehat{\beta}_j^{\mathrm{HRD}} = \begin{cases} y_j & (|y_j| > 2\tau), \\ 0 & (|y_j| \le 2\tau) \end{cases} .$$

簡単に

$$\widehat{\beta}_j^{\mathrm{HRD}} = y_j \mathbb{1}\{|y_j| > 2\tau\}$$

と書ける．上の考察から，(3.11) を考慮に入れれば，しきい値を τ としたくなるであろう．しかし，これでは十分に大きくとは言えない．なぜならば，τ は $\forall j \in \{1, 2, \ldots, d\}$ に対して，同時に

$$|y_1| \le \tau \iff \max_{1 \le j \le d} |y_j| \le \tau$$

とすべきである．これは最大不等式によって評価される．命題 2.24 より，$\forall t \in \mathbb{R}$ に対して

$$\mathsf{Pr}\left(\max_{1 \le j \le d} |\xi_j| > t \right) \le 2 \exp\left(-\frac{nt^2}{2\sigma^2} \right)$$

となるので

$$\delta := 2d \exp\left(-\frac{nt^2}{2\sigma^2} \right) \iff t = \pm\sigma \sqrt{\frac{2\log(2d/\delta)}{n}}$$

となる．よって

$$\mathsf{Pr}\left(\max_{1 \le j \le d} |\xi_j| \le \sigma \sqrt{\frac{2\log(2d/\delta)}{n}} \right) \ge 1 - \delta$$

を得る．

定理 3.14. 線型回帰モデル (3.2) と ORT 条件を仮定する．すなわち，劣 Gauss 的確率変数列モデル (3.10) を仮定する．$\delta > 0$ とする．このとき，しきい値

$$2\tau = 2\sigma \sqrt{\frac{2\log(2d/\delta)}{n}} \tag{3.12}$$

の硬しきい値推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}$ は以下の性質をみたす．

(1) $|\boldsymbol{\beta}^*|_{0,d} = k$ のとき，$\forall \delta > 0$ に対して，

$$\mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) = \frac{1}{n}|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 \precsim \sigma^2 \frac{k\log(2d/\delta)}{n} \right) \ge 1 - \delta.$$

(2) $\min_{j \in \mathrm{supp}(\boldsymbol{\beta}^*)} |\beta_j| > 3\tau$ のとき，$\forall \delta > 0$ に対して

$$\mathsf{Pr}\left(\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) = \mathrm{supp}(\boldsymbol{\beta}^*) \right) \ge 1 - \delta.$$

Proof. 事象 A を

$$A := \left\{ \max_{1 \le j \le d} |\xi_j| \le \tau \right\}$$

で定義する．命題 2.24 より

$$\mathsf{Pr}(A) \ge 1 - \delta$$

となる．事象 A 上では, $j \in \{1, 2, \ldots, d\}$ に対して, 以下のことが成立する．

まず, 次のことに注意する．

$$|y_j| > 2\tau \Longrightarrow |\beta_j^*| \ge |y_j| - |\xi_j| > \tau, \tag{3.13}$$

$$|y_j| \le 2\tau \Longrightarrow |\beta_j^*| \le |y_j| + |\xi_j| \le 3\tau. \tag{3.14}$$

これらより

$$\begin{aligned}
\left|\beta_j^{\mathrm{HRD}} - \beta_j^*\right| &= \left|\beta_j^{\mathrm{HRD}} - \beta_j^*\right| \mathbb{1}\{|y_j| > 2\tau\} + \left|\beta_j^{\mathrm{HRD}} - \beta_j^*\right| \mathbb{1}\{|y_j| \le 2\tau\} \\
&= \left|y_j - \beta_j^*\right| \mathbb{1}\{|y_j| > 2\tau\} + \left|\beta_j^*\right| \mathbb{1}\{|y_j| \le 2\tau\} \\
&\le |\xi_j| \mathbb{1}\{|\beta_j^*| > \tau\} + |\beta_j| \mathbb{1}\{|\beta_j^*| \le 3\tau\} \\
&\qquad (\because A \text{ 上では}, (3.13) \text{ と } (3.14)) \\
&\le \tau \mathbb{1}\{|\beta_j^*| > \tau\} + |\beta_j| \mathbb{1}\{|\beta_j^*| \le 3\tau\} \\
&\qquad (\because A \text{ 上では}, |\xi_j| \le \tau) \\
&\le 4 \min\{|\beta_j^*|, \tau\}
\end{aligned}$$

となることがわかる．この式より

$$\begin{aligned}
\left|\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}} - \boldsymbol{\beta}^*\right|_{2,d}^2 &= \sum_{j=1}^d \left|\beta_j^{\mathrm{HRD}} - \beta_j^*\right|^2 \\
&= 16 \sum_{j=1}^d \min\{|\beta_j^*|^2, \tau^2\} \\
&\le 16 |\boldsymbol{\beta}^*|_0 \tau^2 \\
&= 32 k \sigma^2 \frac{\log(2d/\delta)}{n} \\
&\precsim \sigma^2 \frac{k \log(2d/\delta)}{n}
\end{aligned}$$

を得る．よって

$$\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) \precsim \sigma^2 \frac{k \log(2d/\delta)}{n}$$

なので

$$\mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) \precsim \sigma^2 \frac{k \log(2d/\delta)}{n} \right) \ge \mathsf{Pr}(A) \ge 1 - \delta$$

となることがわかる．よって, (1) は示された．

(2) を示すために, $\beta_j^* \neq 0$ ならば, 命題の仮定より, $|\beta_j^*| > 3\tau$ に注意する．A 上では,

$$|y_j| = |\beta_j^* + \xi_j| > 3\tau - \tau = 2\tau$$

である．したがって $\widehat{\beta}_j^{\mathrm{HRD}} \neq 0$ となる．よって

$$\mathrm{supp}(\boldsymbol{\beta}^*) \subset \mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big)$$

である．

次に, $\widehat{\beta}_j^{\mathrm{HRD}} \neq 0$ のとき

$$\big|\widehat{\beta}_j^{\mathrm{HRD}}\big| = |y_j| > 2\tau$$

となることがわかる．これより, A 上では

$$|\beta_j^*| = |y_j - \xi_j| \geq |y_j| - |\xi_j| \geq 2\tau - \tau = \tau$$

である．したがって, $\beta_j^* \neq 0$ となるので

$$\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) \subset \mathrm{supp}(\boldsymbol{\beta}^*)$$

である．よって, A 上では

$$\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) = \mathrm{supp}(\boldsymbol{\beta}^*)$$

となる．このことより

$$\text{「}A\text{ が起きる」} \Longrightarrow \text{「}\Big\{\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) = \mathrm{supp}(\boldsymbol{\beta}^*)\Big\}\text{ が起きる」}$$

となるので

$$\mathsf{Pr}\left(\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}}\big) = \mathrm{supp}(\boldsymbol{\beta}^*)\right) \geq \mathsf{Pr}(A) \geq 1 - \delta$$

がわかる． □

次に, 軟しきい値推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}$ を

$$\widehat{\beta}_j^{\mathrm{SFT}} = \begin{cases} y_j - 2\tau & (y_j > 2\tau) \\ y_1 + 2\tau & (y_j < -2\tau) \\ 0 & (|y_j| \leq 2\tau) \end{cases} ; \qquad (j = 1, 2, \ldots, d)$$

を考える．これは

$$\widehat{\beta}_j^{\mathrm{SFT}} = \left(1 - \frac{2\tau}{|y_j|}\right)_+ y_j$$

とかける．ただし, $a \in \mathbb{R}$ に対して, $(a)_+ = \max(a, 0)$ である．

定理 3.15. 定理 3.14 と同様の仮定をおく．$\delta > 0$ とする．このとき，しきい値 (3.12) をもつ軟しきい値推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}$ は以下の性質をみたす．
(1) $|\boldsymbol{\beta}^*|_{0,d} = k$ のとき，$\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}) = \frac{1}{n}|\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}} - \boldsymbol{\beta}^*|_{2,d}^2 \lesssim \sigma^2 \frac{k\log(2d/\delta)}{n}\right) \geq 1 - \delta$$

となる．
(2) $\min_{j\in\mathrm{supp}(\boldsymbol{\beta}^*)} |\beta_j^*| > 3\tau$ のとき，$\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathrm{supp}(\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}) = \mathrm{supp}(\boldsymbol{\beta}^*)\right) \geq 1 - \delta$$

となる．

Proof. 事象 A を

$$A := \left\{\max_{1\leq j\leq d} |\xi_j| \leq \tau\right\}$$

で定める．(3.13) と (3.14) より，$j \in \{1, 2, \ldots, d\}$ に対して，A 上では，

$$\begin{aligned}
\left|\widehat{\beta}_j^{\mathrm{SFT}} - \beta_j^*\right| &= |y_j - 2\tau - \beta_j^*|\mathbb{1}\{y_j > 2\tau\} + |y_j + 2\tau - \beta_j^*|\mathbb{1}\{y_j < -2\tau\} \\
&\quad + |\beta_j^*|\mathbb{1}\{|y_j| \leq 2\tau\} \\
&= |\xi_j - 2\tau|\mathbb{1}\{y_j > 2\tau\} + |\xi + 2\tau|\mathbb{1}\{y_j < -2\tau\} \\
&\quad + |\beta_j^*|\mathbb{1}\{|y_j| \leq 2\tau\} \\
&\leq 3\tau\mathbb{1}\{|y_j| > 2\tau\} + |\beta_j^*|\mathbb{1}\{|y_j| \geq 2\tau\} \\
&\qquad (\because A \text{ 上では}, |\xi_j| \leq \tau \text{ なので}) \\
&\leq 3\tau\mathbb{1}\{|\beta_j^*| > \tau\} + |\beta_j|\mathbb{1}\{|\beta_j^*| \leq 2\tau\} \\
&= 5\min\{|\beta_j^*|, \tau\}
\end{aligned}$$

となる．この不等式と τ の定義に注意すれば，

$$\begin{aligned}
\left|\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}} - \boldsymbol{\beta}^*\right|_2^2 &= \sum_{j=1}^d \left|\widehat{\beta}_j^{\mathrm{SFT}} - \beta_j^*\right|^2 \leq 25\sum_{j=1}^d \min\{|\beta_j^*|^2, \tau^2\} \leq 25|\boldsymbol{\beta}^*|_{0,d}\tau^2 \\
&= 50k\sigma^2\frac{\log(2d/\delta)}{n} \lesssim \sigma^2\frac{k\log(2d/\delta)}{n}
\end{aligned}$$

を得る．これで (1) は示された．
(2) を示すために，$\beta_j^* \neq 0$ ならば，定理の仮定より，$|\beta_j^*| > 3\tau$ なので，

$$|y_j \pm 2\tau| = |\beta j^* + \xi_j \pm 2\tau| > |\beta_j^*| - 2\tau - |\xi_j| \geq |\beta_j^*| - 2\tau - \tau > 0.$$

したがって，$\widehat{\beta}_j^{\mathrm{SFT}} \neq 0$ となる．よって

$$\mathrm{supp}(\boldsymbol{\beta}^*) \subset \mathrm{supp}(\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}).$$

次に, $\widehat{\beta}_j^{\mathrm{SFT}} \neq 0$ のとき

$$|y_j| > 2\tau$$

となる. これより, A 上では

$$|\beta_j^*| = |y_j - \xi_j| \geq |y_j| - |\xi_j| \geq 2\tau - \tau = \tau$$

である. したがって $\beta_j^* \neq 0$ なので

$$\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}\big) \subset \mathrm{supp}(\boldsymbol{\beta}^*)$$

がわかる. よって, A 上では

$$\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}\big) = \mathrm{supp}(\boldsymbol{\beta}^*)$$

である. これより

$$\mathsf{Pr}\left(\mathrm{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}}\big) = \mathrm{supp}(\boldsymbol{\beta}^*)\right) \geq \mathsf{Pr}(A) \geq 1 - \delta.$$

□

3.4 高次元線型回帰モデル

3.4.1 BIC 推定量と LASSO 推定量

硬閾値推定量と軟ちきい値推定量は, 以下の罰則項付きの経験リスク最小化問題の解である.

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}} &\in \mathrm{argmax}_{\boldsymbol{\beta}\in\mathbb{R}^d} \left\{|\boldsymbol{y} - \boldsymbol{\beta}|_{2,d}^2 + 4\tau^2|\boldsymbol{\beta}|_{0,d}\right\}, \\
\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}} &\in \mathrm{argmax}_{\boldsymbol{\beta}\in\mathbb{R}^d} \left\{|\boldsymbol{y} - \boldsymbol{\beta}|_{2,d}^2 + 4\tau|\boldsymbol{\beta}|_{1,d}\right\}
\end{aligned}$$

(3.9) から, ORT 条件のもと, 上の変分問題は, 下記のように書きかえることができる.

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{\mathrm{HRD}} &\in \mathrm{argmax}_{\boldsymbol{\beta}\in\mathbb{R}^d} \left\{|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,n}^2 + 4\tau^2|\boldsymbol{\beta}|_{0,d}\right\}, \\
\widehat{\boldsymbol{\beta}}^{\mathrm{SFT}} &\in \mathrm{argmax}_{\boldsymbol{\beta}\in\mathbb{R}^d} \left\{|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,n}^2 + 4\tau|\boldsymbol{\beta}|_{1,d}\right\}.
\end{aligned}$$

ORT 条件をみたさないとき, これらはしきい値推定量に対応しないが, 上の問題自体は定義される.

定義 3.16. $\tau > 0$ を固定し, 線型回帰モデル (3.2) を仮定する. BIC 推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}$ を次で定義する.

$$\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} \in \mathrm{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^d} \left\{\frac{1}{n}|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,n}^2 + \tau^2|\boldsymbol{\beta}|_{0,d}\right\}.$$

さらに, $\widehat{\boldsymbol{\beta}}$ の LASSO 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ を

$$\widehat{\boldsymbol{\beta}}^{\mathcal{L}} \in \mathrm{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^d} \left\{ \frac{1}{n} |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,n}^2 + 2\tau |\boldsymbol{\beta}|_{1,d} \right\}.$$

注意 3.17. しきい値パラメータの乗数倍を変更した. この方が後の議論で都合がよいからである. □

この仮定のもとで, 回帰モデル (3.2) を以下のように変形できる.

注意 3.18. BIC 推定量の計算は NP 困難であり, 野蛮な計算より効率的なものは知られていない. 2^d のスパースパターンを計算することになる. 実際,

$$\begin{aligned} &\min_{\boldsymbol{\beta}\in\mathbb{R}^d} \left\{ \frac{1}{n} |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,n}^2 + \tau^2 |\boldsymbol{\beta}|_{0,d} \right\} \\ &\qquad = \min_{1\le k\le d} \left\{ \min_{\boldsymbol{\beta}\in\mathbb{R}^d:\, |\boldsymbol{\beta}|_{0,d}=k} \left(\frac{1}{n} |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}|_{2,n}^2 + \tau^2 k \right) \right\} \end{aligned}$$

である. $\min_{\boldsymbol{\beta}\in\mathbb{R}^d:\, |\boldsymbol{\beta}|_0=k} (1/n)|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*|_2^2$ を計算するために, $\begin{pmatrix} d \\ k \end{pmatrix}$ 回の最小 2 乗推定値の計算が必要になる. ひとつの最小 2 乗推定値の計算は $O(k^3)$ 回である. したがって, 計算コストは

$$C \sum_{k=0}^{d} \begin{pmatrix} d \\ k \end{pmatrix} k^3 = C d^3 2^d$$

となる. ただし, C は k, d に依存しない正の乗数である.

BIC 推定量と比較すると LASSO 推定値を求めるための効率的なアルゴリズムはたくさん提案されている. 代表的なものとして, 以下のものがある.

(1) 座標降下法. R のパッケージは glmet.

(2) LARS (least angle regression).

(3) FISTA アルゴリズム.

(4) 交互方向降下法.

川野 (2018, pp. 12–20) を参照のこと. □

3.4.2 BIC 推定量の精度評価

BIC 推定量は計算困難である. しかし, スパース推定の「よさ」のベンチマークを与える. 精度は硬しきい値推定量と似ている. 評価のために, ORT 条件を必要としない.

命題 3.19. 線型回帰モデル (3.2) を仮定し, $\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)\,(\sigma > 0)$ とし, $|\boldsymbol{\beta}^*|_{0,\,d} \geq 1$ とする. このとき, 罰則

$$\tau^2 = 16\log(12)\frac{\sigma^2}{n} + 32\frac{\sigma^2\log(ed)}{n} \tag{3.15}$$

をもつ BIC 推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}$ は以下の性質をみたす. $\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}) = \frac{1}{n}\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 \lesssim |\boldsymbol{\beta}^*|_{0,\,d}\sigma^2\frac{\log(ed/\delta)}{n}\right) \geq 1-\delta$$

が成り立つ.

Proof. まず, 推定量の定義より

$$\frac{1}{n}\big|\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{2,\,n}^2 + \tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,d} \leq \frac{1}{n}\big|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 + \tau^2\big|\boldsymbol{\beta}^*\big|_{0,\,d}$$

となることに注意せよ. これと簡単な計算により

$$\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 \leq n\tau^2|\boldsymbol{\beta}^*|_{0,\,d} + 2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \boldsymbol{\beta}^*) - n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,d}$$

を得る. 次に

$$\begin{aligned} 2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \boldsymbol{\beta}^*) &= 2\boldsymbol{\epsilon}^\top\left(\frac{\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*}{\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}}\right)\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n} \\ &\leq 2\left[\boldsymbol{\epsilon}^\top\left(\frac{\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*}{\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}}\right)\right]^2 + \frac{1}{2}\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 \end{aligned}$$

を得る. これは, $a,\,b \in \mathbb{R}$ に対して, $2ab \leq 2a^2 + (1/2)b^2$ からわかる. これらを合わせれば

$$\begin{aligned} \big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 \leq 2n\tau^2|\boldsymbol{\beta}^*|_{0,\,d} &+ 4\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\big)\big]^2 \\ &- 2n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,d} \end{aligned} \tag{3.16}$$

を得る. ただし, $\boldsymbol{z} \in \mathbb{R}^d$ に対して

$$\mathcal{U}(\boldsymbol{z}) := \frac{\boldsymbol{z}}{|\boldsymbol{z}|_{2,\,d}}$$

とおく. 次に, $\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}$ を sup-out する必要がある. このために, sup を

$$\sup_{\boldsymbol{\beta}\in\mathbb{R}^d} = \max_{1\leq k\leq d}\max_{\#(S)=k}\sup_{\mathrm{sup}(\boldsymbol{\beta})=S}$$

のように分解する．これを上の不等式に用いれば

$$
\begin{aligned}
&4\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}-\mathbf{X}\boldsymbol{\beta}^*\big)\big]^2-2n\tau^2\big|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\big|_{0,\,d}\\
&\qquad\le\max_{1\le k\le d}\left\{\max_{\#(S)=k}\sup_{\mathrm{supp}(\tilde{\boldsymbol{\beta}}=S)}\left\{4\big[\boldsymbol{\epsilon}^\top\mathcal{U}\big(\mathbf{X}\tilde{\boldsymbol{\beta}}^{\mathrm{BIC}}-\mathbf{X}\boldsymbol{\beta}^*\big)\big]^2-2n\tau^2k\right\}\right\}\\
&\qquad\le\max_{1\le k\le d}\left\{\max_{\#(S)=k}\sup_{S\subset\mathbb{R}^d;\,\#(S)=k}\sup_{\boldsymbol{u}\in B_2^{r_{S+}}}4[\boldsymbol{\epsilon}^\top\boldsymbol{\Psi}_{S+}\boldsymbol{u}]^2-2n\tau^2k\right\}
\end{aligned}
$$

となる．ただし，$j\in\{1,\,2,\,\ldots,\,d\}$ とし，$\boldsymbol{c}_j$ を $\mathbf{X}$ の第 j 列としたとき

$$
\boldsymbol{\Psi}_{S+}=[\boldsymbol{\psi}_1,\,\boldsymbol{\psi}_2,\,\ldots,\,\boldsymbol{\psi}_{S+}]
$$

は $\mathrm{span}\{\boldsymbol{c}_j;\,j\in S\cup\mathrm{supp}(\boldsymbol{\beta}^*)\}$ の正規直交基底であり，$r_{S+}\le\#(S)+|\boldsymbol{\beta}^*|_0$ はこの部分空間の次元である．和集合の上限より，$\forall t>0$ に対して

$$
\begin{aligned}
&\mathsf{Pr}\left(\max_{1\le k\le d}\left\{\max_{\#(S)=k}\sup_{S\subset\mathbb{R}^d;\,\#(S)=k}\sup_{\boldsymbol{u}\in B_2^{r_{S+}}}4[\boldsymbol{\epsilon}^\top\boldsymbol{\Psi}_{S+}\boldsymbol{u}]^2-2n\tau^2k\right\}\ge t\right)\\
&\qquad\le\sum_{k=1}^{d}\sum_{S\subset\mathbb{R}^d;\,\#(S)=k}\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in B_2^{r_{S+}}}\big[\boldsymbol{\epsilon}^\top\boldsymbol{\Psi}_{S+}\boldsymbol{u}\big]^2\ge\frac{t}{4}+\frac{1}{2}n\tau^2k\right)
\end{aligned}
$$

となる．上の式と命題 2.30 の ϵ 網の議論より $\#(S)=k$ に対して

$$
\begin{aligned}
\mathsf{Pr}\left(\sup_{\boldsymbol{u}\in B_2^{r_{S+}}}\big[\boldsymbol{\epsilon}^\top\boldsymbol{\Psi}_{S+}\boldsymbol{u}\big]^2\ge\frac{t}{4}+\frac{1}{2}n\tau^2k\right)&\le2\cdot6^{r_{S+}}\exp\left(-\frac{\frac{t}{4}+\frac{1}{2}n\tau^2k}{8\sigma^2}\right)\\
&\le2\exp\left(-\frac{t}{32\sigma^2}-\frac{n\tau^2k}{16\sigma^2}+(k+|\boldsymbol{\beta}^*|_{0,\,d})\log6\right)\\
&\le\exp\left(-\frac{t}{32\sigma^2}-2k\log(ed)+|\boldsymbol{\beta}^*|_{0,\,d}\log(12)\right)
\end{aligned}\tag{3.17}
$$

となる．よって

$$
\begin{aligned}
&\mathsf{Pr}\left(\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 \le 2n\tau^2|\boldsymbol{\beta}^*|_{0,d} + t\right) \\
&\quad \le \mathsf{Pr}\left(2n\tau^2|\boldsymbol{\beta}^*|_{0,d} + 4\left[\boldsymbol{\epsilon}^\top \mathcal{U}\left(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\right)\right]^2 - 2n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,d} \ge 2n\tau^2|\boldsymbol{\beta}^*|_{0,d} + t\right) \\
&\qquad (\because (3.18)) \\
&\quad \le \mathsf{Pr}\left(4\left[\boldsymbol{\epsilon}^\top \mathcal{U}\left(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\right)\right]^2 - 2n\tau^2\left|\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}}\right|_{0,d} \ge t\right) \\
&\quad \le \sum_{k=1}^{d} \sum_{S\subset\mathbb{R}^d:\,\#(S)=d} \mathsf{Pr}\left(\sup_{\boldsymbol{u}\in B_2^{r_{S+}}} \left[\boldsymbol{\epsilon}^\top \boldsymbol{\Psi}_{S+}\boldsymbol{u}\right]^2 \ge \frac{t}{4} + \frac{n\tau^2 k}{2}\right) \\
&\quad \le \sum_{k=1}^{n} \sum_{S\subset\mathbb{R}^d:\,\#(S)=d} \exp\left(-\frac{t}{32\sigma^2} - 2k\log(ed) + \left|\boldsymbol{\beta}^*\right|_{0,d}\log(12)\right) \qquad (\because (3.17)) \\
&\quad \le \sum_{k=1}^{d} \begin{pmatrix} k \\ d \end{pmatrix} \exp\left(-\frac{t}{32\sigma^2} - 2k\log(ed) + \left|\boldsymbol{\beta}^*\right|_{0,d}\log(12)\right) \\
&\quad \le \sum_{k=1}^{d} \exp\left(-\frac{t}{32\sigma^2} - k\log(ed) + \left|\boldsymbol{\beta}^*\right|_{0,d}\log(12)\right) \\
&\qquad \Bigg(\because \text{補題 3.10 より} \begin{pmatrix} d \\ k \end{pmatrix} \le \left(\frac{ed}{k}\right)^k. \\
&\qquad\quad \text{よって，} \begin{pmatrix} d \\ k \end{pmatrix} \le \exp\left\{k\log(ed) - k\log k\right\} \le \exp\left\{k\log(ed)\right\}\Bigg) \\
&\quad \le \sum_{k=1}^{d} \left(\frac{1}{ed}\right)^k \exp\left(-\frac{t}{32\sigma^2} + \left|\boldsymbol{\beta}^*\right|_{0,d}\log(12)\right) \\
&\quad \le \sum_{k=1}^{d} \left(\frac{1}{ed}\right)^k \exp\left(-\frac{t}{32\sigma^2} + \left|\boldsymbol{\beta}^*\right|_{0,d}\log(12)\right) \\
&\qquad \left(\because \sum_{k=1}^{d}\left(\frac{1}{ed}\right)^k = \frac{1}{ed}\sum_{k=1}^{d}\left(\frac{1}{ed}\right)^{k-1} = \frac{1}{ed}\frac{1-\left(\dfrac{1}{ed}\right)^k}{1-\dfrac{1}{ed}} \le \frac{1}{ed}\frac{ed}{ed-1}\right. \\
&\qquad\quad \left. = \frac{1}{ed-1} \le 1\right)
\end{aligned}
$$

がわかる. いま

$$
\delta = \exp\left(-\frac{t}{32\sigma^2} + \left|\boldsymbol{\beta}^*\right|_{0,\,d}\log(12)\right) \Leftrightarrow \log\delta = -\frac{t}{32\sigma^2} + \left|\boldsymbol{\beta}^*\right|_{0,\,d}\log(12)
$$

$$
\Leftrightarrow \frac{t}{32\sigma^2} = \left|\boldsymbol{\beta}^*\right|_{0,\,d}\log(12) + \log\left(\frac{1}{\delta}\right)
$$

$$
\Leftrightarrow t = 32\sigma^2\left|\boldsymbol{\beta}^*\right|_0\log(12) + 32\sigma^2\ \log\left(\frac{1}{\delta}\right)
$$

とおく. すると

$$
\begin{aligned}
\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{BIC}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,\,n}^2 &\leq 2n\tau^2 + t \\
&= 32\sigma^2\log(12)\left|\boldsymbol{\beta}^*\right|_{0,\,d} + 64\sigma^2\log(ed)\left|\boldsymbol{\beta}^*\right|_{0,\,d} + 32\sigma^2\log(12)\left|\boldsymbol{\beta}^*\right|_{0,\,d} \\
&\qquad + 32\sigma^2\log\left(\frac{1}{\delta}\right) \\
&= 64\sigma^2\log(ed)\left|\boldsymbol{\beta}^*\right|_{0,\,d} + 64\sigma^2\log(12)\left|\boldsymbol{\beta}^*\right|_{0,\,d} + 32\sigma^2\log\left(\frac{1}{\delta}\right) \\
&\leq 256\sigma^2\log(ed)\left|\boldsymbol{\beta}^*\right|_{0,\,d} + 32\sigma^2\log\left(\frac{1}{\delta}\right) \\
&\qquad\qquad \left(\because 64\log(12) \leq 64\log(2^3) \leq 192\log(ed)\right) \\
&\leq 256\sigma^2\log(ed)\left|\boldsymbol{\beta}^*\right|_{0,\,d} + 256\sigma^2\log\left(\frac{1}{\delta}\right)\left|\boldsymbol{\beta}^*\right|_{0,\,d} \\
&\qquad\qquad (\because \left|\boldsymbol{\beta}^*\right|_{0,\,d} \geq 1) \\
&\leq \left|\boldsymbol{\beta}^*\right|_{0,\,d} 256\sigma^2\log\left(\frac{ed}{\delta}\right)
\end{aligned}
$$

となり, 定理は証明された. □

3.4.3 LASSO 推定量の精度評価

定理 3.20. 線型回帰モデル (3.2) を仮定し, $\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)\,(\sigma > 0)$ とする. さらに, デザイン行列 $\mathbf{X}$ の列は規準化されているとする.

$$
\max_{1\leq j\leq d}|X_j|_2 \leq \sqrt{n}; \qquad \mathbf{X} = (X_1,\, X_2,\, \ldots,\, X_d); \qquad X_i \in \mathbb{R}^n\,(j = 1,\, 2,\, \ldots,\, d).
$$

このとき, 調整パラメータ

$$
2\tau := 2\sigma\sqrt{\frac{2\log(2d)}{n}} + 2\sigma\sqrt{\frac{2\log(1/\delta)}{n}} \tag{3.18}
$$

をもつ LASSO 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ は以下をみたす. $\forall\delta > 0$ に対して

$$
\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}) \leq 4|\boldsymbol{\beta}^*|_1\sigma\sqrt{\frac{2\log(2d)}{n}} + 4|\boldsymbol{\beta}^*|_1\sqrt{\frac{\log(1/\delta)}{n}}\right) \geq 1 - \delta.
$$

ただし

$$\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}) = \frac{1}{n}\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2$$

である.

Proof. 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ の定義より

$$\frac{1}{n}\big|\boldsymbol{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{2,\,n}^2+2\tau\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,d} \leq \frac{1}{n}|\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta}^*|_{2,\,n}^2+2\tau|\boldsymbol{\beta}^*|_{1,\,d} = \frac{1}{n}\big|\boldsymbol{\epsilon}\big|_{2,\,d}^2+2\tau|\boldsymbol{\beta}^*|_{1,\,d}$$

がわかる. Hölder の不等式 (命題 **??**) と上の不等式より

$$\begin{aligned}
\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 &= \big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{Y} + \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 \\
&= \big|\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{2,\,n}^2 + |\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*|_{2,\,n}^2 + 2(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*)^\top\big(\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}^* - \boldsymbol{\epsilon}\big) \\
&\leq 2\boldsymbol{\epsilon}^\top\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\big) - 2|\boldsymbol{\epsilon}|_{2,\,n}^2 + 2|\boldsymbol{\epsilon}|_{2,\,n}^2 + 2\tau n|\boldsymbol{\beta}^*|_{1,\,d} - 2\tau n|\boldsymbol{\beta}^*|_{1,\,d} \\
&= 2\boldsymbol{\epsilon}^\top\mathbf{X}\big(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\big) + 2\tau n\big(|\boldsymbol{\beta}^*|_{1,\,d} - \big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,d}\big) \\
&\leq 2\big|\mathbf{X}^\top\boldsymbol{\epsilon}\big|_\infty\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,d} + 2|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty|\boldsymbol{\beta}^*|_{1,\,d} + 2\tau n\big(|\boldsymbol{\beta}^*|_{1,\,d} - \big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,d}\big) \\
&= 2\big(|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty - n\tau\big)\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,d} + 2\big(|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty + n\tau\big)|\boldsymbol{\beta}^*|_{1,\,d}
\end{aligned}$$

となることがわかる. $j = 1,\,2,\,\ldots,\,d$ に対して, $|\boldsymbol{X}_j|_{2,n} \leq \sqrt{n}$ なので, $\boldsymbol{X}_j^\top\boldsymbol{\epsilon} \sim \mathsf{subG}(n\sigma^2)$ となる. このことに注意して, 命題 2.24 を用いると, $\forall t > 0$ に対して

$$\mathsf{Pr}\left(|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty \geq t\right) = \mathsf{Pr}\left(\max_{1\leq j\leq d}|\boldsymbol{X}_j^\top\boldsymbol{\epsilon}| \geq t\right) \leq 2d\exp\left(-\frac{t^2}{2n\sigma^2}\right) =: \delta \tag{3.19}$$

となることがわかる. ここで,

$$t := \sigma\sqrt{2n\log(2d)} + \sigma\sqrt{2n\log\left(\frac{1}{\delta}\right)} = n\tau$$

とる. すると $|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty \leq t$ 上では

$$\begin{aligned}
\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 &\leq 2\underbrace{(t - n\tau)}_{=0}\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_{1,\,d} + 2(t + n\tau)|\boldsymbol{\beta}^*|_{1,\,d} \\
&= 4n\tau|\boldsymbol{\beta}^*|_{1,\,d}
\end{aligned}$$

となるので

$$|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty \leq t \implies \big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,\,n}^2 \leq 4n\tau|\boldsymbol{\beta}^*|_{1,\,d}$$

である．よって

$$\begin{aligned}\mathsf{Pr}\left(\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}-\mathbf{X}\boldsymbol{\beta}^*\right|_{2,\,n}^2\le 4n\tau|\boldsymbol{\beta}^*|_{1,\,d}\right) &\ge \mathsf{Pr}\left(|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty\le t\right)\\ &= 1-\mathsf{Pr}\left(|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty> t\right)\\ &\ge 1-\delta \qquad (\because (3.19))\end{aligned}$$

を得る． □

注意 3.21. BIC 推定量では，調整パラメータは信頼レベル δ に依存しなかったが，LASSO は依存していることに注意せよ．また，定理 3.20 で与えられた収束レートは $\sqrt{\log(d/n)}$ である．BIC 推定量の $\log(d/n)$ と比較すると遅い． □

次に，デザイン行列 $\mathbf{X}$ により強い仮定をおくことで，$\sqrt{\log d/n}$ より速いレートをもつ LASSO を導出する．

Incoherence 仮定

仮定 INC(k) デザイン行列 $\mathbf{X}$ が次をみたすとき，incoherence k であるという．ある $k\in\mathbb{N}$ が存在して，

$$\left|\frac{\mathbf{X}^\top\mathbf{X}}{n}-I_d\right|_\infty\le\frac{1}{32k}.$$

ただし，正方行列 $\mathbf{A}$ に対して，$|\mathbf{A}|_\infty$ で固有値の絶対値が最大ものを表す．

INC(k) 条件と同値の条件は以下である．$\mathbf{X}$ の第 j 列 ($j=1,\,2,\,\ldots,\,d$)) を $\mathbf{X}_j$ とする．

(1) $\forall j\in\{1,\,2,\,\ldots,\,d\}$ に対して，

$$\left|\frac{|\boldsymbol{X}_j|_{2,\,n}^2}{n}-1\right|\le\frac{1}{32k}.$$

(2) $1\le\forall j,\,\ell\le d,\,j\ne\ell$ に対して，

$$\frac{|\boldsymbol{X}_j^\top\boldsymbol{X}_\ell|}{n}\le\frac{1}{32k}.$$

注意 3.22. ORT 条件は，$k\to\infty$ のとき，INC(k) 条件の極限と考えることができる．しかし，ORT 条件は $d\le n$ を要求しているが，INC(k) 条件はそれを求めていない．実際，次の命題は $d\gg n$ の場合を考えている． □

命題 3.23. $\mathbf{X} \in \mathrm{Mat}(n, d; \mathbb{R})$ はランダム行列とし, $\mathbf{X}$ の (i, j) 成分 $X_{ij}\,(i = 1, 2, \ldots, n; j = 1, 2, \ldots, d)$ は独立な Rademacher 確率変数列とする．すなわち

$$\mathsf{Pr}(X_{ij} = 1) = \mathsf{Pr}(X_{ij} = -1) = \frac{1}{2}$$

である．このとき, $\forall\delta > 0$ に対して

$$n \geq 2^{11}k^2 \log\left(\frac{2}{\delta}\right) + 2^{12}k^2 \log d$$

のとき

$$\mathsf{Pr}\Bigg(\mathbf{X} \text{ は } \mathsf{INC}(k) \text{ 条件をみたす}\Bigg) \geq 1 - \delta$$

が成り立つ．したがって, ある定数 C に対して,

$$n \geq Ck^2 \log d$$

ならば,

$$\mathsf{Pr}\Bigg(\mathbf{X} \text{ は } \mathsf{INC}(k) \text{ 条件をみたす}\Bigg) \geq 1 - \delta$$

となる．

Proof. $\epsilon_{ij}\,(i = 1, 2, \ldots, n; j = 1, 2, \ldots, d)$ は $\{-1, 1\}$ に値をとる Rademacher 確率変数列とする．すると, $\mathbf{X}^\top\mathbf{X}/n$ の第 j 番目 $(j = 1, 2, \ldots, d))$ の対角要素は

$$\frac{1}{n}\sum_{i=1}^{n} \epsilon_{ij}^2 = 1$$

となることに注意する．さらに, $\ell_1 \neq \ell_2\,(\ell_1, \ell_2 = 1, 2, \ldots, d)$ としたとき, $\mathbf{X}^\top\mathbf{X}/n$ の (ℓ_1, ℓ_2) 要素は

$$\frac{1}{n}\sum_{j=1}^{n} \epsilon_{j\ell_1}\epsilon_{j\ell_2} = \frac{1}{n}\sum_{j=1}^{n} \xi_j^{(\ell_1, \ell_2)}$$

と書ける．ここで,

$$\xi_j^{(\ell_1, \ell_2)} = \epsilon_{j\ell_1}\epsilon_{j\ell_2}$$

とした．$\xi_1^{(\ell_1, \ell_2)}, \xi_2^{(\ell_1, \ell_2)}, \ldots, \xi_n^{(\ell_1, \ell_2)}$ は i.i.d. の Rademacher 確率変数列で

あることに注意せよ．したがって, $\forall t > 0$ に対して

$$\begin{aligned}
\mathsf{Pr}\left(\left|\frac{\mathbf{X}^\top\mathbf{X}}{n} - I_d\right|_\infty \geq t\right) &= \mathsf{Pr}\left(\max_{k\neq\ell}\left|\frac{1}{n}\sum_{j=1}^n \xi_j^{(\ell_1,\,\ell_2)}\right|\right) \\
&\leq \sum_{\ell_1\neq\ell_2}\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^n \xi_j^{(\ell_1,\,\ell_2)}\right| \geq t\right) \\
&\leq \sum_{\ell_1\neq\ell_2} 2\exp\left(-\frac{nt^2}{2}\right) \\
&\leq 2d^2\exp\left(-\frac{nt^2}{n}\right)
\end{aligned}$$

となる．ここで,

$$t = \frac{1}{32k}$$

とする．すると

$$n \geq 2^{11}k^2\log\left(\frac{2}{\delta}\right) + 2^{12}k^2\log d$$

のとき

$$\begin{aligned}
\frac{n}{2^{11}k^2} \geq \log\left(\frac{1}{\delta}\right) + 4\log(d) &\iff \delta \leq \exp\left(4\log d - \frac{n}{2^{11}k^2}\right) \\
&\iff \delta \geq 2d^2\exp\left(-\frac{n}{2^{11}k^2}\right)
\end{aligned}$$

となる．よって

$$\mathsf{Pr}\left(\left|\frac{\mathbf{X}^\top\mathbf{X}}{n} - I_d\right|_\infty \geq \frac{1}{32k}\right) \leq 2d^2\exp\left(-\frac{n}{2^{11}k^2}\right) \leq \delta$$

となる． □

$\boldsymbol{\beta}^* \in \mathbb{R}^d$ と $S \subset \{1,\, 2,\, \ldots,\, d\}$ に対して

$$\boldsymbol{\beta}_S^* = (\beta_{S,\,1},\, \beta_{S,\,2},\, \ldots,\, \beta_{S,\,d})^\top \in \mathbb{R}^d$$

を次で与えらるベクトルとする．

$$\beta_{S,\,j} = \begin{cases} \beta_j & (j \in S), \\ 0 & (\text{otherwise}) \end{cases}$$

である．特に

$$|\boldsymbol{\beta}^*|_{1,\,d} = |\boldsymbol{\beta}_S^*|_{1,\,d} + |\boldsymbol{\beta}_{S^c}^*|_{1,\,d}$$

に注意する．

補題 3.24. 整数 $1 \le k \le d$ を固定し, デザイン行列 $\mathbf{X}$ は INC(k) 条件をみたすと仮定する. このとき, 任意の $S \subset \{1, 2, \dots, d\}$ で $\#(S) \le k$ なるものと $\forall \boldsymbol{\beta}^* \in \mathbb{R}^d$ に対して, 錐条件

$$|\boldsymbol{\beta}^*_{S^c}|_{1,d} \le 3|\boldsymbol{\beta}^*_S|_{1,d} \tag{3.20}$$

をみたしているとする. このとき

$$|\boldsymbol{\beta}^*|^2_{2,d} \le 2\frac{|\mathbf{X}\boldsymbol{\beta}^*|^2_{2,n}}{n}$$

が成立する.

Proof. まず,

$$\frac{|\mathbf{X}\boldsymbol{\beta}^*|^2_{2,n}}{n} = \frac{|\mathbf{X}\boldsymbol{\beta}^*_S|^2_{2,n}}{n} + \frac{|\mathbf{X}\boldsymbol{\beta}^*_{S^C}|^2_{2,n}}{n} + \frac{2{\boldsymbol{\beta}^*}^\top_S\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}^*_{S^C}}{n} \tag{3.21}$$

に注意する. 各項を別々に評価する.
(1) INC(k) 条件より

$$\begin{aligned}\frac{|\mathbf{X}\boldsymbol{\beta}^*_S|^2_{2,n}}{n} &= {\boldsymbol{\beta}^*}^\top_S\frac{\mathbf{X}^\top\mathbf{X}}{n}{\boldsymbol{\beta}^*}_S = |{\boldsymbol{\beta}^*}_S|^2_{2,d} + {\boldsymbol{\beta}^*}^\top_S\left(\frac{\mathbf{X}^\top\mathbf{X}}{n} - I_d\right){\boldsymbol{\beta}^*}_S \\ &\le |{\boldsymbol{\beta}^*}_S|^2_{2,d} - \frac{|{\boldsymbol{\beta}^*}_S|^2_{2,d}}{32k} \le |{\boldsymbol{\beta}^*}_S|^2_{2,d} - \frac{|{\boldsymbol{\beta}^*}_S|^2_{1,d}}{32k}.\end{aligned} \tag{3.22}$$

(2) 同様に

$$\frac{|\mathbf{X}\boldsymbol{\beta}^*_{S^C}|^2_{2,n}}{n} \ge |\boldsymbol{\beta}^*_{S^C}|^2_{2,d} - \frac{|\boldsymbol{\beta}^*_{S^C}|^2_{2,d}}{n} \ge |\boldsymbol{\beta}^*_{S^C}|^2_{2,d} - \frac{9|\boldsymbol{\beta}^*_{S^C}|^2_{1,d}}{n}. \tag{3.23}$$

最後の不等号は, 錐条件 (3.20) を用いた.
(3) 最後に

$$2\left|{\boldsymbol{\beta}^*}^\top_S\frac{\mathbf{X}^\top\mathbf{X}}{n}\boldsymbol{\beta}^*_{S^C}\right| \le \frac{2}{32k}|\boldsymbol{\beta}^*_S|_{1,d}|\boldsymbol{\beta}^*_{S^C}|_{1,d} \le \frac{6}{32k}|\boldsymbol{\beta}^*_S|^2_{1,d}. \tag{3.24}$$

最後の不等号は, 錐条件 (3.20) を用いた. Cauchy-Schwarz の不等式より

$$|\boldsymbol{\beta}^*_S|^2_{1,d} \le \#(S)|\boldsymbol{\beta}^*_S|^2_{2,d}$$

となることに注意する. したがって, (3.22) – (3.24) を (3.21) に代入すれば

$$\begin{aligned}\frac{|\mathbf{X}\boldsymbol{\beta}^*|^2_{2,n}}{n} &\ge |\boldsymbol{\beta}^*_S|^2_{2,d} + |\boldsymbol{\beta}^*_{S^C}|^2_{2,d} - \frac{16}{32k}|\boldsymbol{\beta}^*_S|^2_{1,d} \\ &\ge |\boldsymbol{\beta}^*_S|^2_{2,d} + |\boldsymbol{\beta}^*_{S^C}|^2_{2,d} - \frac{16\#(S)}{32k}|\boldsymbol{\beta}^*_S|^2_{2,d} \\ &\ge \frac{1}{2}|\boldsymbol{\beta}^*_S|^2_{2,d}.\end{aligned}$$

最後から 2 番目の不等号は, Cauchy-Schwarz の不等式より

$$|\boldsymbol{\beta}^*_S|^2_{1,d} \le \#(S)|\boldsymbol{\beta}^*_S|^2_{2,d}$$

となることと $\#(S) \le k$ であることからわかる. □

LASSO 推定量の早い収束レート

定理 3.25. $n \geq 2$ を固定する．線型回帰モデル (3.2) と $\boldsymbol{\epsilon} \sim \mathsf{subG}_n(\sigma^2)$ ($\sigma > 0$) を仮定する．さらに，ある $k \in \mathbb{N}$ があって，$|\boldsymbol{\beta}^*|_0 \leq k$ とし，デザイン行列 $\mathbf{X}$ は INC(k) 条件をみたすとする．$\delta > 0$ とする．このとき，調整パラメータ

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

をもつ LASSO 推定量 $\boldsymbol{\beta}^{*\mathcal{L}}$ は以下をみたす．$\forall\delta > 0$ に対して

$$\mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{L}}\big) = \frac{1}{n}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 \lesssim k\sigma^2\frac{\log(2d/\delta)}{n}\right) \geq 1 - \delta$$

と

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

が成り立つ．

Proof. $\widehat{\boldsymbol{\beta}}^{\mathcal{L}}$ の定義より

$$\frac{1}{n}\left|\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\right|_{2,n}^2 \leq \frac{1}{n}\left|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 + 2\tau|\boldsymbol{\beta}^*|_{1,d} - 2\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\right|_{1,d}$$

が成立する．上の式の両辺に $\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_1$ を加えて，n 倍し

$$\left|\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\right|_{2,n}^2 = \left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 + \left|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 - 2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*)$$

に注意すれば

$$\begin{aligned}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 + n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{2,n} \quad &\leq \quad 2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*) + n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{1,d} \\ &\quad + 2n\tau|\boldsymbol{\beta}^*|_{1,d} - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\right|_{1,d} \qquad (3.25)\end{aligned}$$

を得る．Hölder の不等式を用いて，定理 3.20 の証明と同じ議論をする．まず

$$\left|\boldsymbol{\epsilon}^\top\mathbf{X}\big(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\big)\right| \leq \left|\boldsymbol{\epsilon}^\top\mathbf{X}\right|_\infty \cdot \left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{1,d}$$

である．INC(k) 条件より，

$$\left|\boldsymbol{X}_j\right|_{2,n}^2 \leq n + \frac{1}{32k} \leq 2n$$

となるので，$\forall s \in \mathbb{R}$ に対して，

$$\begin{aligned}\mathsf{E}\big[\exp\big(s\boldsymbol{X}_j^\top\boldsymbol{\epsilon}\big)\big] \quad &\leq \quad \exp\left(s|\boldsymbol{X}_j|_{2,n}\frac{\boldsymbol{X}_j^\top}{|\boldsymbol{X}_j|_{2,n}}\boldsymbol{\epsilon}\right) \\ &\leq \quad \exp\left(s\sqrt{2n}\frac{\boldsymbol{X}_j^\top}{|\boldsymbol{X}_j|_{2,n}}\boldsymbol{\epsilon}\right) \\ &\leq \quad \exp\left(-\frac{\sigma^2}{2}(\sqrt{2n}s)^2\right) = \exp\left(-n\sigma^2s^2\right)\end{aligned}$$

となるので

$$\boldsymbol{X}_j^\top \boldsymbol{\epsilon} \sim \mathsf{subG}_n(2n\sigma^2)$$

がわかる．よって

$$\mathsf{Pr}\left(\max_{1\le j\le n} |\boldsymbol{X}_j^\top \boldsymbol{\epsilon}| > t\right) \le 2d\exp\left(-\frac{t^2}{4n\sigma^2}\right)$$

となる．ここで

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

とすれば, 上の式より

$$\mathsf{Pr}\left(|\mathbf{X}^T\boldsymbol{\epsilon}|_\infty > \frac{nt}{2}\right) \le 2d\exp\left(-\frac{n\tau^2}{16\sigma^2}\right)$$

を得る．

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

より

$$\begin{aligned}
\exp\left(-\frac{n\tau^2}{16\sigma^2}\right) &= \exp\left(-\frac{n}{64\sigma^2}(2\tau)^2\right) \\
&\le \exp\left(-\frac{n}{64\sigma^2} \times 64\sigma^2\left(\frac{\log(1/\delta)}{n}\right)\right) \\
&= \exp\left(-\log(2d) + \log(1/\delta)\right) \\
&= \frac{\delta}{2d}
\end{aligned}$$

となる．これより

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

を得る．よって

$$\mathsf{Pr}\left(|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty \ge \frac{n\tau}{2}\right) \le \delta$$

を得る．ただし, $\boldsymbol{X}_j\,(j = 1, 2, \ldots, d)$ はデザイン行列 $\mathbf{X}$ の第 j 列ベクトルである．$|\mathbf{X}^\top\boldsymbol{\epsilon}|_\infty \le \frac{n\tau}{2}$ 上では,

$$\left|\boldsymbol{\epsilon}^\top\mathbf{X}\big(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\big)\right| \le |\boldsymbol{\epsilon}\mathbf{X}^\top|_\infty \big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\big|_{1,d} \le \frac{n\tau}{2}\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\big|_{1,d}$$

となる．ここで,

$$S := \mathsf{supp}\big(\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big)$$

とおく. $|\mathbf{X}\boldsymbol{\epsilon}|_\infty \le (n\tau/2)$ 上では (3.25) より

$$
\begin{aligned}
&\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 + n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{1,d} \\
&\quad \le 2\left|\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*)\right| + n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{1,d} + 2n\tau|\boldsymbol{\beta}^*|_{1,d} - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\right|_{1,d} \\
&\quad \le 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*\right|_{1,d} 2n\tau|\boldsymbol{\beta}^*|_{1,d} - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S\right|_{1,d} \\
&\quad \le 4n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*\right|_{1,d} \qquad (3.26)
\end{aligned}
$$

となる. 特に

$$4\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{1,d} \le \left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*\right|_{1,d}$$

である. $S = \mathrm{supp}(\boldsymbol{\beta}^*)$ に注意すれば

$$\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*_S\right|_{1,d} + \left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C} - \boldsymbol{\beta}^*_{S^C}\right|_{1,d} = \left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{1,d} \le 4\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*\right|_{1,d} \le 4\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*_S\right|_{1,d}$$

なので

$$\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_{S^C} - \boldsymbol{\beta}^*_{S^C}\right|_{1,d} \le \left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*_S\right|_{1,d}$$

を得る. よって

$$\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*$$

としたとき, $\tilde{\boldsymbol{\beta}}$ は条件 (3.20) をみたすので, これに対して, Cauchy-Schwarz の不等式と補題 3.24 を用いる. $\#(S) \le k$ と $S = \mathrm{supp}(S)$ なので,

$$
\begin{aligned}
\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*\right|_{1,d} &\le \sqrt{\#(S)}\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*\right|_{2,d} \\
&\le \sqrt{\#(S)}\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{2,d} \\
&\quad (\because \beta_j = 0\,(j \notin S)) \\
&\le \sqrt{\frac{2k}{n}}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}
\end{aligned}
$$

がわかる. この式と (3.26) を合わせれば

$$
\begin{aligned}
\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n} &\le \left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n} + n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\right|_{1,d} \le 4n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{L}}_S - \boldsymbol{\beta}^*\right|_{1,d} \\
&\le \tau\sqrt{32kn}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}
\end{aligned}
$$

を得る. これより

$$
\begin{aligned}
\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n} &\le 32nk\tau^2 = 8nk(2\tau)^2 = 8nk\left\{8\sigma\sqrt{\frac{\log(2d)}{n}} + 8\sigma\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{n}}\right\}^2 \\
&\le \frac{2^{11}k\sigma^2}{n}\log\left(\frac{2d}{\delta}\right)
\end{aligned}
$$

がわかる．よって

$$\begin{aligned}
&\mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big) = \frac{1}{n}\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,n}^2 \le 2^{11}\cdot\frac{k\sigma^2}{n}\log\left(\frac{2d}{\delta}\right)\right)\\
&\ge \mathsf{Pr}\left(\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big|_\infty \le \frac{n\tau}{2}\right) \ge 1-\delta
\end{aligned}$$

を得る．再度, 補題 3.24 を用いれば

$$\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\big|_{2,d}^2 \le \frac{2}{n}\big|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*\big|_{2,n}^2 = 2\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big)$$

より

$$\begin{aligned}
&\mathsf{Pr}\left(\big|\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*\big|_{2,d}^2 \le 2^{10}\cdot\frac{k\sigma^2}{n}\log\left(\frac{2d}{\delta}\right)\right)\\
&\ge \mathsf{Pr}\left(\mathsf{MSE}\big(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}\big) \le 2^{11}\cdot\frac{k\sigma^2}{n}\log\left(\frac{2d}{\delta}\right)\right)\\
&\ge 1-\delta
\end{aligned}$$

を得る． □

注意 3.26. 証明のために必要なことは, incoherence 条件自体ではなく

$$\inf_{\#(S)\le k}\inf_{\beta\in\mathcal{C}_S}\frac{|\mathbf{X}\boldsymbol{\beta}^*|_2^2}{n|\boldsymbol{\beta}^*|_2^2} \ge \frac{1}{2} \tag{3.27}$$

である．ただし,

$$\mathcal{C}_S := \{\boldsymbol{\beta}^* \in \mathbb{R}^d;\ |\boldsymbol{\beta}^*_{S^C}|_1 \le 3|\boldsymbol{\beta}^*_S|_1\}$$

である．条件 (3.27) を制約固有値条件 (restricted eigenvalu (RE) condition) という．この名前は以下のことに由来する．すべての k スパースベクトル $\boldsymbol{\beta}^*$ は $\#(S)\le k$ とした錐 $\mathcal{C}_S$ に含まれる．したがって, $rm\ RE$ 条件は $\mathbf{X}_S$ の最小固有値が $\lambda_{\min}(\mathbf{X}_S) \ge \frac{n}{2}$ ($\forall S \subset \{1, 2, \dots, d\}$) で $\#(S)\le k$) をみたす．明らかに, RE 条件は INC(k) 条件よりも弱い．そして, デザイン行列 $\mathbf{X}$ が Rademacher 確率変数列ならば, $n \ge Ck\log d$ (C は正の定数) ならば, 高い確率で, RE 条件をみたす． □

3.4.4 SLOPE 推定量

BIC 推定量は, $k = |\boldsymbol{\beta}^*|_0$ のとき, 収束レートが

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

となる．計算が効率的できる推定量でこのオーダーに達する推定量があるかどうかは興味ある疑問である．LASSO を少しだけ修正することにより, このオーダーに達する推定量を構成することができる．

定義 3.27. $\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_d)$ を非増加の正の数列とする．すなわち，

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

$\boldsymbol{\beta}^* = (\beta_1, \beta_2, \dots, \beta_d)^\top \in \mathbb{R}^d$ に対して，

$$(\beta_1^*, \beta_2^*, \dots, \beta_d^*)^\top \in \mathbb{R}^d$$

は $|\beta_1|, |\beta_2|, \dots, |\beta_d|$ を非増加数列に並び変えたものとする．$\boldsymbol{\beta}^*$ の並び替え ℓ_1 ノルム (sorted ℓ_1 norum) を次で定義する．

$$|\boldsymbol{\beta}^*|_* := \sum_{j=1}^{d} \lambda_j \beta_j^*.$$

または，

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

である．ただし，$\mathfrak{S}_d$ は $\{1, 2, \dots, d\}$ の置換群である．調整パラメータ $\boldsymbol{\lambda}$ と $t > 0$ に対する SLOPE 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{S}}$ を

$$\widehat{\boldsymbol{\beta}}^{\mathcal{S}} \in \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^d} \left\{ \frac{1}{n} \left| \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^* \right|_{2,n}^2 + 2\tau |\boldsymbol{\beta}^*|_* \right\} \tag{3.28}$$

で定義する．

SLOPE 推定量は “Sorted L- One Penalized estimation” の略である．
以下では

$$\lambda_j = \sqrt{\log\left(\frac{2d}{j}\right)} \qquad (j = 1, 2, \dots, d) \tag{3.29}$$

とする．

補題 3.28. $g_1, g_2, \dots, g_d$ は平均 0, 分散が σ^2 $(\sigma > 0)$ 以下の正規分布に独立同一に従う確率変数列とする．このとき，$k \leq d$ と $\forall t > 0$ に対して，

$$\mathsf{Pr}\left(\frac{1}{k\sigma^2} \sum_{j=1}^{k} (g_j^*)^2 \geq t \log\left(\frac{2d}{k}\right) \right) \leq \left(\frac{2d}{k}\right)^{1-\frac{3t}{8}} \tag{3.30}$$

である．

Proof. Jensen の不等式を用いる．

$$\begin{aligned}
\mathsf{E}\left[\exp\left(\frac{3}{8k\sigma^2} \sum_{j=1}^{k} (g_j^*)^2 \right) \right] &\leq \frac{1}{k} \sum_{j=1}^{k} \left[\exp\left(\frac{3(g_j^*)^2}{8\sigma^2} \right) \right] \\
&\quad (\because \text{Jensen の不等式}) \\
&\leq \frac{1}{k} \sum_{j=1}^{d} \mathsf{E}\left[\exp\left(\frac{3g_j^2}{8\sigma^2} \right) \right] = \frac{2d}{k}.
\end{aligned}$$

最後の等号は, $g \sim N(0,1)$ ならば, $g^2 \sim \chi_1^2$ なので,

$$\mathsf{E}\left[\exp\left(\frac{3}{8}g^2\right)\right] = 2$$

よりわかる. 次に, Chernoff 限界を用いる.

$$\begin{aligned}
&\mathsf{Pr}\left(\frac{1}{k\sigma^2}\sum_{j=1}^{k}(g_j^*)^2 \geq t\log\left(\frac{2d}{k}\right)\right) \\
&\quad\leq \exp\left(-\frac{3t}{8}\log\left(\frac{2d}{k}\right)\right)\times\mathsf{E}\left[\exp\left(\frac{3}{8k\sigma^2}\sum_{j=1}^{k}(g_j^*)^2\right)\right] \\
&\quad\leq \exp\left(-\frac{3t}{8}\log\left(\frac{2d}{k}\right)\right)\frac{2d}{k} \\
&\quad= \left(\frac{2d}{k}\right)^{1-(3t)/8}.
\end{aligned}$$

□

補題 3.29. $[d] := \{1, 2, \ldots, d\}$ とする. 補題 3.28 と同じ仮定をおく. このとき, $1/2 > \forall\delta > 0$ に対して,

$$\mathsf{Pr}\left(\sup_{k\in[d]}\frac{(g_k^*)^2}{\sigma\lambda_k} \leq 4\sqrt{\log\left(\frac{1}{\delta}\right)}\right) \geq 1-\delta$$

が成り立つ.

Proof. まず, $g_1^* \geq g_2^* \geq \cdots \geq g_k^*$ なので

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

であることに注意する. 次に, 補題 3.28 を用いると

$$\mathsf{Pr}\left(\frac{(g_k^*)^2}{\sigma^2\lambda_k^2} \geq t\right) \leq \mathsf{Pr}\left(\frac{1}{k\sigma^2}\sum_{j=1}^{k}(g_j^*)^2 \geq t\log\left(\frac{2d}{k}\right)\right) \leq \left(\frac{2d}{8}\right)^{1-(3t)/8}$$

を得る. さらに

$$\lambda_k = \sqrt{\log\left(\frac{2d}{k}\right)}$$

に注意する. $t > 8$ に対して

$$\begin{aligned}
\mathsf{Pr}\left(\sup_{k\in[d]}\frac{(g_k^*)^2}{\sigma^2\lambda_k^2} \geq t\right) &\leq \sum_{k=1}^{d}\mathsf{Pr}\left(\frac{1}{k\sigma^2}\sum_{j=1}^{k}(g_j^*)^2 > t\log\left(\frac{2d}{j}\right)\right) \\
&\leq \sum_{j=1}^{d}\left(\frac{2d}{j}\right)^{1-(3t)/8} \leq (2d)^{1-(3t)/8}\sum_{j=1}^{d}\frac{1}{j^2} \\
&\leq 4\cdot 2^{-3t/8} = 2^{2-(3t)/8}
\end{aligned}$$

を得る．$\delta > 1/2$ なので, $16/3 > (4/3)\log(1/\delta)$ となることに注意すれば

$$\begin{aligned}\delta := 4\cdot 2^{-(3t)/8} \quad &\Rightarrow \quad 2-\frac{3t}{8} = \log\delta \Rightarrow -\frac{3}{8}t = \log\delta - 2 \Rightarrow \frac{3}{8}t = \log\left(\frac{1}{\delta}\right)+2 \\ &\Rightarrow \quad t = \frac{8}{3}\log\left(\frac{1}{\delta}\right) + \frac{16}{3} \geq 4\log\left(\frac{1}{\delta}\right).\end{aligned}$$

に注意する．よって,

$$\begin{aligned}\mathsf{Pr}\left(\sup_{k\in[k]}\frac{(g_k^*)^2}{\sigma\lambda_k} \leq 2\sqrt{\log\left(\frac{1}{\delta}\right)}\right) \quad &\geq \quad \mathsf{Pr}\left(\sup_{k\in[d]}\frac{g_k^2}{\sigma^2\lambda_k^2} \leq \frac{3}{8}\log\left(\frac{1}{\delta}\right)+\frac{16}{3}\right) \\ &\geq \quad 1-\delta\end{aligned}$$

がわかる． □

定理 3.30. $n \geq 2$ とする．線型回帰モデル (3.2) と $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma a^2 I_n)$ ($\sigma > 0$) を仮定する．さらに, $|\boldsymbol{\beta}^*|_0 \leq k$ とし,

$$k' \geq 4k\log\frac{2de}{k}$$

に対して, デザイン行列 $\mathbf{X}$ は $\mathsf{INC}(k')$ 条件をみたすとする．$\delta > 0$ に対して, 調整パラメータ τ を

$$\tau := 8\sqrt{2}\sigma\sqrt{\frac{\log(1/\delta)}{n}} \tag{3.31}$$

で定める．このとき, SLOPE 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{S}}$ は次の性質をみたす．$\forall\delta > 0$ に対して

$$\begin{aligned}&\mathsf{Pr}\left(\mathsf{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}}) = \frac{1}{n}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 \precsim \frac{k\sigma^2}{n}\log\left(\frac{2d}{k}\right)\log\left(\frac{1}{\delta}\right)\right) \\ &\quad \geq 1-\delta\end{aligned} \tag{3.32}$$

と

$$\mathsf{Pr}\left(\left|\widehat{\boldsymbol{\beta}}^{\mathcal{S}} - \boldsymbol{\beta}^*\right|_{2,d}^2 \precsim \frac{k\sigma^2}{n}\log\left(\frac{2d}{k}\right)\log\left(\frac{1}{\delta}\right)\right) \geq 1-\delta. \tag{3.33}$$

Proof. SLOPE 推定量 $\widehat{\boldsymbol{\beta}}^{\mathcal{S}}$ の定義より,

$$\left|\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}}\right|_{2,n}^2 + \left|\widehat{\boldsymbol{\beta}}^{\mathcal{S}}\right|_{1,d} \leq \left|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 + \left|\boldsymbol{\beta}^*\right|_{1,d}$$

である．さらに,

$$\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 = \left|\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}}\right|_{2,n}^2 + \left|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 - 2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathcal{S}} - \boldsymbol{\beta}^*)$$

である. これらより

$$\begin{aligned}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}}-\mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2+n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{S}}-\boldsymbol{\beta}^*\right|_* &\le 2\boldsymbol{\epsilon}^\top\mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathcal{S}}-\boldsymbol{\beta}^*)+n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{S}}-\boldsymbol{\beta}^*\right|_* \\ &\quad +2n\tau|\boldsymbol{\beta}^*|_*-2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{S}}\right|_* \qquad (3.34)\end{aligned}$$

となる.

$$\boldsymbol{u}=(u_1,\,u_2,\,\dots,\,u_d)^\top:=\widehat{\boldsymbol{\beta}}^{\mathcal{S}}-\boldsymbol{\beta}^*;\quad g_j:=(\mathbf{X}^\top\boldsymbol{\epsilon})_j\,(j=1,\,2,\,\dots,\,d)$$

と定める. 補題 3.29 (σ^2 を $2n\sigma^2$ と読み替える) より

$$\begin{aligned}\boldsymbol{\epsilon}^\top\mathbf{X}\boldsymbol{u} &= \sum_{j=1}^d(\mathbf{X}^\top\boldsymbol{\epsilon})_j\boldsymbol{u}_j\le\sum_{j=1}^d g_j^*u_j^*=\sup_{j\in[d]}\frac{g_j^*}{\lambda_j}(\lambda_ju_j^*)\le\sum_{j\in[k]}\left\{\frac{g_j^*}{\lambda_j}\right\}|\boldsymbol{u}|_* \\ &\le 4\sqrt{2}\sigma\sqrt{n\log\left(\frac{1}{\delta}\right)}|\boldsymbol{u}|_* \qquad (3.35)\end{aligned}$$

がわかる. なぜならば

$$|\boldsymbol{X}_j|_{2,n}^2\le 2n$$

なので,

$$g_j=X_j^\top\boldsymbol{\epsilon}\sim N(0,\,|\boldsymbol{X}_j|_{2,n}^2\sigma^2)$$

より, 分散は $2n\sigma^2$ より小さいことがわかるので, σ^2 を $2n\sigma^2$ と読み替えて補題 3.29 を適用する. 置換 $\underline{\sigma}\in\mathfrak{G}_d$ をうまく選んで

$$|\boldsymbol{\beta}^*|_*=\sum_{j=1}^k\lambda_j|\beta_{\underline{\sigma}(j)}|$$

かつ

$$|u_{\underline{\sigma}(k+1)}|\ge|u_{\underline{\sigma}(k+2)}|\ge\cdots\ge|u_{\underline{\sigma}(d)}|$$

とできる. $\{\lambda_j\}$ は非増加列だったので

$$\begin{aligned}|\boldsymbol{\beta}^*|_*-\left|\widehat{\boldsymbol{\beta}}^{\mathcal{S}}\right|_* &= \sum_{j=1}^k\lambda_j|\beta_{\underline{\sigma}(j)}|-\sum_{j=1}^d\lambda_j(\widehat{\boldsymbol{\beta}}^{\mathcal{S}})_j^* \\ &\le \sum_{j=1}^k\lambda_j\left\{\left|\beta_{\underline{\sigma}(j)}\right|-\left|(\widehat{\boldsymbol{\beta}}^{\mathcal{S}})_{\underline{\sigma}(j)}\right|\right\}-\sum_{j=k+1}^d\lambda_j\left|(\widehat{\boldsymbol{\beta}}^{\mathcal{S}})_{\underline{\sigma}(j)}\right| \\ &\le \sum_{j=1}^k\lambda_j|u_{\underline{\sigma}(j)}|-\sum_{j=k+1}^d\lambda_ju_j^*\end{aligned}$$

($\because j\ge k$ のとき, $\beta_j=0$)

$$\le \sum_{j=1}^k\lambda_ju_j^*-\sum_{j=k+1}^d\lambda_ju_j^*. \qquad (3.36)$$

$\tau = 8\sqrt{2}\sigma\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{n}}$, (3.34) − (3.37) より

$$\begin{aligned}\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 + n\tau|\boldsymbol{u}|_* &\leq 2n\tau|\boldsymbol{u}|_* + 2n\tau|\boldsymbol{\beta}^*|_* - 2n\tau\left|\widehat{\boldsymbol{\beta}}^{\mathcal{S}}\right|_* \\ &\leq 2n\tau|\boldsymbol{u}|_* + 2n\tau\sum_{j=1}^k \lambda_j u_j^* - 2n\tau\sum_{j=1}^d \lambda_j u_j^* \\ &\leq 4n\tau\sum_{j=1}^k \lambda_j u_j^*. \qquad (3.37)\end{aligned}$$

したがって

$$\left|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{S}} - \mathbf{X}\boldsymbol{\beta}^*\right|_{2,n}^2 \leq 4n\tau\sum_{j=1}^k \lambda_j u_j^* - n\tau|\boldsymbol{u}|_* \leq 3n\tau\sum_{j=1}^k \lambda_j u_j - \sum_{j=k+1}^d \lambda_j u_j$$

より

$$\sum_{j=k+1}^d \lambda_j u_j \leq 3\sum_{j=1}^k \lambda_j u_j \qquad (3.38)$$

がわかる．さらに，$k' \geq 4k\log(2de/k)$ に対して，INC(k') 条件をみたしていることと $\lambda_d \geq (1/2)$ であることに注意し，長い計算[2]をすれば，$\boldsymbol{u} = \widehat{\boldsymbol{\beta}}^{\mathcal{S}} - \boldsymbol{\beta}^*$ に対して

$$\frac{|\mathbf{X}\boldsymbol{u}|_{2,n}^2}{n} \geq |\boldsymbol{\beta}_S^*|_{2,d}^2 + |\boldsymbol{u}_{S^C}|_{2,d}^2 - \frac{36+12+1}{128}|\boldsymbol{u}_S|_{2,d}^2 \geq \frac{1}{2}|\boldsymbol{u}|_{2,d}^2 \qquad (3.39)$$

がわかる．ここで, Stirling の不等式

$$\left(\frac{k}{3}\right)^k \leq k!$$

を用いると

$$\begin{aligned}\sum_{j=1}^k \lambda_j &= \sum_{j=1}^k \log\left(\frac{2d}{j}\right) = \sum_{j=1}^k\{\log(2d) - \log(j)\} \\ &= \log(2d) - \log(1) + log(2d) - \log(2) + \cdots \\ &\quad + \log(2d) - \log(k) = k\log(2d) - \log k! \\ &\leq k\log(2d) - k\log\left(\frac{k}{e}\right) = k\log\left(\frac{2de}{k}\right) \qquad (3.40)\end{aligned}$$

[2]補遺を参照.

(3.37) と (3.38) を用いて, (3.36) を評価する.

$$
\begin{aligned}
|\mathbf{X}\boldsymbol{u}|_{2,n}^2 + n\tau|\boldsymbol{u}|_* &\leq 4n\tau\sum_{j=1}^k \lambda_j u_j \leq 4n\tau\sqrt{\sum_{j=1}^k \lambda_j^2}\sqrt{\sum_{j=1}^k u_j^2} \leq 4n\tau\sqrt{\sum_{j=1}^k \lambda_j^2}|\boldsymbol{u}|_2 \\
&\leq 4n\tau\sqrt{k\log\left(\frac{2de}{k}\right)}\sqrt{\frac{2|\mathbf{X}\boldsymbol{u}|_{2,d}^2}{n}} = 4\sqrt{2}\tau\sqrt{k\log\left(\frac{de}{k}\right)}|\mathbf{X}\boldsymbol{u}|_{2,n} \\
&= 64\sigma\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{n}}\sqrt{nk\log\left(\frac{2de}{k}\right)}|\mathbf{X}\boldsymbol{u}|_{2,n} \\
&= 2^6\sigma\sqrt{\log\left(\frac{1}{\delta}\right)}\sqrt{k\log\left(\frac{2de}{k}\right)}|\mathbf{X}\boldsymbol{u}|_{2,n}
\end{aligned}
$$

を得る. これと (3.38) より

$$
\begin{aligned}
\frac{1}{n}|\mathbf{X}\boldsymbol{u}|_{2,n}^2 &\leq 2^{12}\sigma^2\frac{k}{n}\log\left(\frac{2de}{k}\right)\log\left(\frac{1}{\delta}\right), \\
|\boldsymbol{u}|_{2,d}^2 &\leq 2^{13}\sigma^2\frac{k}{n}\log\left(\frac{2de}{k}\right)\log\left(\frac{1}{\delta}\right)
\end{aligned}
$$

を得る. よって, $\forall\delta > 0$ に対して,

$$
\sup_{j\in[d]}\left\{\frac{g_j^*}{\lambda_j}\right\} \leq 4\sqrt{\log\left(\frac{1}{\delta}\right)}
$$

ならば

$$
\frac{1}{n}|\mathbf{X}\boldsymbol{u}|_2^2 \leq 2^{13}\sigma^2\frac{k}{n}\log\left(\frac{2de}{k}\right)\log\left(\frac{1}{\delta}\right)
$$

となる. よって, $\forall\delta > 0$ に対して

$$
\begin{aligned}
&\mathsf{Pr}\left(\frac{1}{n}|\mathbf{X}\boldsymbol{u}|_{2,n}^2 \leq 2^{12}\sigma^2\frac{k}{n}\log\left(\frac{2de}{k}\right)\log\left(\frac{1}{\delta}\right)\right) \\
&\leq \mathsf{Pr}\left(\sup_{1\leq j\leq d}\left\{\frac{g_j^*}{\lambda_j}\right\} \leq 4\sqrt{\log\left(\frac{1}{\delta}\right)}\right) \geq 1-\delta
\end{aligned}
$$

を得る. 同様に, $\forall\delta > 0$ に対して,

$$
\sup_{1\leq j\leq d}\left\{\frac{g_j^*}{\lambda_j}\right\} \leq 4\sqrt{\log\left(\frac{1}{\delta}\right)}
$$

ならば,

$$
|\boldsymbol{u}|_{2,d}^2 \leq 2^{13}\sigma^2\frac{k}{n}\log\left(\frac{2ed}{k}\right)\log\left(\frac{1}{\delta}\right)
$$

なので

$$\Pr\left(|\boldsymbol{u}|_{2,d}^2 \le 2^{13}\sigma^2 \frac{k}{n}\log\left(\frac{2de}{k}\right)\log\left(\frac{1}{\delta}\right)\right)$$
$$\ge \Pr\left(\sup_{1\le j\le d}\left\{\frac{g_j^*}{\lambda_j}\right\} \le 4\sqrt{\log\left(\frac{1}{\delta}\right)}\right) \ge 1-\delta$$

がわかる. □

3.5 補遺

3.5.1 (3.39) の計算

まず,

$$\frac{|\mathbf{X}\boldsymbol{u}|_{2,n}^2}{n} = \frac{|\mathbf{X}\boldsymbol{u}_S|_{2,n}^2}{n} + \frac{|\mathbf{X}\boldsymbol{u}_{S^C}|_{2,n}^2}{n} + 2\boldsymbol{u}_S^\top \frac{\mathbf{X}^\top\mathbf{X}}{n}\boldsymbol{u}_{S^C} \tag{3.41}$$

に注意をする. (3.41) の右辺の各項を別々に評価している.

(1)

$$\begin{aligned}\frac{|\mathbf{X}\boldsymbol{u}_S|_{2,n}^2}{n} &= \boldsymbol{u}_S^\top \frac{\mathbf{X}^\top\mathbf{X}}{n}\boldsymbol{u}_S = |\boldsymbol{u}_S|_{2,d}^2 + \boldsymbol{u}_S^\top\left(\frac{\mathbf{X}^\top\mathbf{X}}{n} - I_d\right)\boldsymbol{u}_S \ge |\boldsymbol{u}_S|_{2,d}^2 - \frac{|\boldsymbol{u}_S|_{1,d}^2}{32k'} \\ &\ge |\boldsymbol{u}_S|_{2,d}^2 - \frac{k}{32k'}|\boldsymbol{u}_S|_{2,d}^2. \end{aligned} \tag{3.42}$$

(2)

$$\begin{aligned} 2\left|\boldsymbol{u}_S^\top\frac{\mathbf{X}^\top\mathbf{X}}{n}\boldsymbol{u}_{S^C}\right| &\le \frac{2}{32k'}\sum_{j=1}^k u_j^* \sum_{j=k+1}^d u_j^* \\ &\le \frac{2}{32k'}\sum_{j=1}^k u_j^* \sum_{j=k+1}^d \frac{\lambda_j}{\lambda_d} u_j^* \\ &\le \frac{6}{32k'\lambda_d}\sum_{j=1}^k u_j^* \sum_{j=1}^k \lambda_j u_j^* \\ &\quad (\because (3.38)\text{ より}) \\ &\le \frac{6\sqrt{k}}{32k'\lambda_d}\sqrt{\sum_{j=1}^k \lambda_j^2} \\ &\left(\because \sum_{j=1}^k u_j^* \le \sqrt{k}\sqrt{\sum_{j=1}^k (u_j^*)^2};\ \sum_{j=1}^k \lambda_j u_j^* \le \sqrt{\sum_{j=1}^k \lambda_j^2}\sqrt{\sum_{j=1}^k (u_j^*)^2} = \sqrt{\sum_{j=1}^k \lambda_j^2}|\boldsymbol{u}_S|_{2,d}\right) \\ &\le \frac{6k}{32k'\lambda_d}\sqrt{\log\left(\frac{2de}{k}\right)}|\boldsymbol{u}_S|_{2,d}^2. \\ &\quad (\because (3.40)\text{ より}) \end{aligned} \tag{3.43}$$

(3)

$$
\begin{aligned}
\frac{|\mathbf{X}\boldsymbol{u}_{S^C}|_{2,n}^2}{n} &= \boldsymbol{u}_{S^C}^\top \frac{\mathbf{X}^\top\mathbf{X}}{n}\boldsymbol{u}_{S^C} \geq |\boldsymbol{u}_{S^C}|_{2,d}^2 - \frac{1}{32k'}|\boldsymbol{u}_{S^C}|_{1,d}^2 \\
&= |\boldsymbol{u}_{S^C}|_{2,d}^2 - \frac{1}{32k'}\left(\sum_{j=k+1}^d u_j^*\right)^2 \\
&\geq |\boldsymbol{u}_{S^C}|_{2,d}^2 - \frac{1}{32k'\lambda_d^2}\left(\sum_{j=k+1}^d \lambda_j u_j^*\right)^2 \\
&\geq |\boldsymbol{u}_{S^C}|_{2,d}^2 - \frac{9}{32k'\lambda_d^2}\left(\sum_{j=1}^k \lambda_j u_j^*\right)^2 \\
&\geq |\boldsymbol{u}_{S^C}|_{2,d}^2 - \frac{9}{32k'\lambda_d^2}\sum_{j=1}^k \lambda_j^2 \sum_{j=1}^k (u_j^*)^2 \\
&\geq |\boldsymbol{u}_{S^C}|_{2,d}^2 - \frac{9k}{32k'\lambda_d^2}\log\left(\frac{2de}{k}\right)|\boldsymbol{u}_S|_{2,d}^2 \qquad (3.44)\\
&\quad \because (3.40) \text{ より} \qquad (3.45)
\end{aligned}
$$

(3.40) – (3.42) を (3.41) に代入すれば

$$
\begin{aligned}
\frac{|\mathbf{X}\boldsymbol{u}|_{2,n}^2}{n} \geq\ & |\boldsymbol{u}_S|_{2,d}^2 - \frac{k}{32k'}|\boldsymbol{u}_S|_{2,d}^2 - \frac{6k}{32k'\lambda_d}\sqrt{\log\left(\frac{2de}{k}\right)}|\boldsymbol{u}_S|_{2,d}^2 + |\boldsymbol{u}_{S^C}|_{2,d}^2 \\
& -\frac{9k}{32k'\lambda_d^2}\log\left(\frac{2de}{k}\right)|\boldsymbol{u}_S|_{2,d}^s \\
\geq\ & |\boldsymbol{u}_S|_{2,d}^2 - \frac{1}{128e}|\boldsymbol{u}_S|_{2,d}^2 - \frac{12}{128}|\boldsymbol{u}_S|_{2,d}^2 + |\boldsymbol{u}_{S^C}|_{2,d}^2 \\
& -\frac{36}{128}|\boldsymbol{u}_S|_{2,d}^s
\end{aligned}
$$

となる．これは

$$
\lambda_d \geq \frac{1}{2}; \quad \frac{k}{k'} \leq \frac{1}{4\log\left(\frac{2de}{k}\right)}
$$

に注意すれば

$$
\begin{aligned}
\frac{k}{32k'\lambda_d} &\geq \frac{1}{128}, \\
\frac{6k}{32k'\lambda_d}\sqrt{\log\left(\frac{2de}{k}\right)} &\geq \frac{12}{128}\frac{1}{\log\left(\frac{2de}{k}\right)}\sqrt{\log\left(\frac{2de}{k}\right)} \\
&= \frac{12}{128}\sqrt{\log\left(\frac{2de}{k}\right)} \geq \frac{12}{128}\sqrt{\log(e)} = \frac{12}{128}, \\
\frac{9k}{32k'\lambda_2}\log\frac{2de}{k} &\geq \frac{36}{128}
\end{aligned}
$$

であることよりわかる． □

3.6 章末注釈と参考文献

この章は [3] を参考にした.

3.7 演習問題

演習問題 3.1.

$\boxed{\text{演習問題 3.1 の解答}}$ □