

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,

$$Y_j = f(X_j) + \epsilon_j \quad (j = 1, 2, \dots, n).$$

ただし、 f は \mathbb{R}^d 上の実数値関数で、 $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$ は交代分散 σ^2 ($\sigma > 0$) の劣 Gauss 型確率ベクトルで、 $\mathbb{E}[\epsilon] = \mathbf{0}$ 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 $\beta^* \in \mathbb{R}^d$ があって、

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

を仮定することである。

1 固定デザインの線型回帰モデル

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、

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_{n+1}, Y_{n+1})$$

をランダムな組とする。観測 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ に基づき、 f の推定量 \hat{f}_n をうまく構成して、 $\hat{f}_n(X_{n+1})$ が Y_{n+1} の「よい」予測量とすることが目標である。 \hat{f}_n を構成するときには、 X_{n+1} は未知の観測（未来の観測）であり、 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。

予測量の自然な評価尺度は L_2 リスクである。 $\mathbb{E}[\cdot]$ を (X_{n+1}, Y_{n+1}) の分布

¹ d と n の大小関係は仮定していないことに注意せよ。

に関する期待値の記号とする.

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

となる. なぜならば, ϵ_{n+1} と X_{n+1} の独立性より,

$$\begin{aligned}
 &\mathbb{E}[\{Y_{n+1} - f(X_{n+1})\}\{f(X_{n+1}) - \widehat{f}_n(X_{n+1})\}] \\
 &= \mathbb{E}[\epsilon_{n+1}\{f(X_{n+1}) - \widehat{f}_n(X_{n+1})\}] \\
 &= \mathbb{E}[\epsilon_{n+1}]\mathbb{E}[f(X_{n+1}) - \widehat{f}_n(X_{n+1})] = 0
 \end{aligned}$$

となることからわかる. また, X_{n+1} の分布を P_X と書いたとき,

$$\|\widehat{f}_n - f\|_{L^2(P_X)}^2 = \int_{\mathbb{R}^d} \{\widehat{f}_n(\mathbf{x}) - f(\mathbf{x})\}^2 dP_X(\mathbf{x})$$

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

$$R(\widehat{f}_n) = \sigma^2 + \|\widehat{f}_n - f\|_{L^2(P_X)}^2$$

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 $\|\widehat{f}_n - f\|_{L^2(P_X)}^2$ が精度の尺度となる. 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.

1.1 固定デザインの場合

この場合には, X_1, X_2, \dots, X_n はランダムではない. このことを強調するために, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ と記すことにする. もちろん, \mathbf{x}_j ($j = 1, 2, \dots, n$) を 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, 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 (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 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 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, $(Y_1, \mathbf{x}_1), (Y_2, \mathbf{x}_2), \dots, (Y_n, \mathbf{x}_n)$ の観測値に基づき, $f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)$ 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 $x_j = j/n$ ($j = 1, 2, \dots, n$) である. この間の補間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, $\boldsymbol{\mu} = (f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n))^\top \in \mathbb{R}^n$ で, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top \in \mathbb{R}^n$ は代行分散 σ^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 squared error, MSE) を精度の尺度とする. これを

$$\text{MSE}(\hat{f}_n) = \frac{1}{n} \sum_{j=1}^n \{\hat{f}_n(\mathbf{x}_j) - f(\mathbf{x}_j)\}^2$$

で定義する. ベクトルの推定問題とみなし,

$$\text{MSE}(\hat{\boldsymbol{\mu}}) = \frac{1}{n} \sum_{j=1}^n \{\hat{\mu}_j - \mu_j\}^2 = \frac{1}{n} \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2$$

と定義してもよい. ただし, $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n)$, $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\top \mathbf{x}}$ ($\mathbf{x} \in \mathbb{R}^d$) である.

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

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

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

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}, \quad \mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top, \quad \boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top \quad (3.1)$$

を得る. さらに,

$$\text{MSE}(\mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{n} \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}^*\|_2^2 = \frac{1}{n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top \mathbf{X}^\top \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*). \quad (3.2)$$

2 最小 2 乗推定量の再訪問

ここでは、 β^* を推定するかわりに、 $\mu = X\beta^*$ の推定問題を扱う。 $X^\top X$ 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 $A \in \text{Mat}(m, n; \mathbb{R})$ に対して、次の条件をみたす行列 $G \in \text{Mat}(n, m; \mathbb{R})$ を A の Moore-Penrose の一般化逆行列といい、これを A^\dagger と記す。

- (1) $AGA = A$.
- (2) $GAG = G$.
- (3) $(AG)^\top = AG$.
- (4) $(GA)^\top = GA$.

注意 3.2 条件 (1) ~ (4) をみたす行列 G は一意的に存在することが知られている。ハーベイル (2012, 下巻の p.191) を参照のこと。 \square

2.1 制約なしの最小 2 乗推定量

命題 3.3 $X\beta^*$ の最小 2 乗推定量 $\hat{\mu}^{\text{LS}} = X\hat{\beta}^{\text{LS}}$ は方程式

$$X^\top \hat{\mu}^{\text{LS}} = X^\top Y$$

をみたす。さらに、

$$\hat{\beta}^{\text{LS}} = (X^\top X)^\dagger X^\top Y$$

と書ける。

証明 関数

$$\beta \mapsto |Y - X\beta|_2^2$$

は凸なので、最小値は

$$\nabla_\beta |Y - X\beta|_2^2 = 0$$

をみたす。ただし、

$$\nabla_\beta = \left(\frac{\partial}{\partial \beta_1}, \frac{\partial}{\partial \beta_2}, \dots, \frac{\partial}{\partial \beta_d} \right)^\top; \quad \beta^* = (\beta_1, \beta_2, \dots, \beta_d)^\top$$

である. $k = 1, 2, \dots, d$ に対して,

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

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

$$\nabla_{\boldsymbol{\beta}} |Y - \mathbf{X}\boldsymbol{\beta}^*|_2^2 = -2 \sum_{j=1}^n (Y - \boldsymbol{\beta}^{*\top} \boldsymbol{\beta}_j^*) \mathbf{x}_j = -2(Y^\top \mathbf{X} - \boldsymbol{\beta}^{*\top} \mathbf{X}^\top \mathbf{X})$$

となる. したがって, $\nabla_{\boldsymbol{\beta}} |Y - \mathbf{X}\boldsymbol{\beta}^*|_2^2 = \mathbf{0}$ より

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

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

命題 3.4 線型モデル (3.1) を仮定し, $\boldsymbol{\epsilon} \sim \text{subG}_n(\sigma^2)$ ($\sigma > 0$) とする. このとき, 最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\text{LS}}$ は方程式

$$\mathbb{E}[\text{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\text{LS}})] = \frac{1}{n} \mathbb{E}[|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2] \lesssim \sigma^2 \frac{r}{n}$$

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

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}^{\text{LS}}) \lesssim \frac{\sigma^2 \{r + \log(1/\delta)\}}{n}\right) > 1 - \delta$$

となる.

証明 最小 2 乗推定量の定義より

$$|Y - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\text{LS}}|_2^2 \leq |Y - \mathbf{X}\boldsymbol{\beta}^*|_2^2 = |\boldsymbol{\epsilon}|_2^2. \quad (3.3)$$

さらに,

$$|Y - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\text{LS}}|_2^2 = |\mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\text{LS}}|_2^2 = |\mathbf{X}\widehat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 - 2\boldsymbol{\epsilon}^\top \mathbf{X}(\widehat{\boldsymbol{\beta}}^{\text{LS}} - \boldsymbol{\beta}^*) + |\boldsymbol{\epsilon}|_2^2.$$

この式を (3.3) の右辺に代入すれば,

$$|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 \leq 2\boldsymbol{\epsilon}^\top \mathbf{X}(\hat{\boldsymbol{\beta}}^{\text{LS}} - \boldsymbol{\beta}^*) = 2|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2 \frac{\boldsymbol{\epsilon}^\top (\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*)}{|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2}. \quad (3.4)$$

項 $\boldsymbol{\epsilon}^\top (\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*) / |\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2$ を直接評価するのは難しい. なぜならば, この量は $\hat{\boldsymbol{\beta}}^{\text{LS}}$ 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 $\hat{\boldsymbol{\beta}}^{\text{LS}}$ 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. $\Psi = [\psi_1, \psi_2, \dots, \psi_r] \in \text{Mat}(n, r; \mathbb{R})$ 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, $\mathbf{X}(\hat{\boldsymbol{\beta}}^{\text{LS}} - \boldsymbol{\beta}^*) = \Psi\boldsymbol{\nu}$ と書ける. このことより,

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{X}(\hat{\boldsymbol{\beta}}^{\text{LS}} - \boldsymbol{\beta}^*)}{|\mathbf{X}(\hat{\boldsymbol{\beta}}^{\text{LS}} - \boldsymbol{\beta}^*)|_2} = \frac{\boldsymbol{\epsilon}^\top \Psi\boldsymbol{\nu}}{|\Psi\boldsymbol{\nu}|_2} = \frac{\tilde{\boldsymbol{\epsilon}}^\top \boldsymbol{\nu}}{|\boldsymbol{\nu}|_2} \leq \sup_{\mathbf{u} \in \mathbb{B}_2} \tilde{\boldsymbol{\epsilon}}^\top \mathbf{u}.$$

ただし, \mathbb{B}_2 は \mathbb{R}^r の単位球であり, $\tilde{\boldsymbol{\epsilon}} = \Psi^\top \boldsymbol{\epsilon}$ である. したがって,

$$|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 \leq 2|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2 \sup_{\mathbf{u} \in \mathbb{B}_2} \tilde{\boldsymbol{\epsilon}}^\top \mathbf{u}$$

より

$$|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2 \leq 2 \sup_{\mathbf{u} \in \mathbb{B}_2} \tilde{\boldsymbol{\epsilon}}^\top \mathbf{u}$$

となる. これより

$$|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{LS}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 \leq 4 \sup_{\mathbf{u} \in \mathbb{B}_2} (\tilde{\boldsymbol{\epsilon}}^\top \mathbf{u})^2.$$

次に, $\mathbf{v} \in \mathbb{S}^{r-1}$ としたとき,

$$|\Psi\mathbf{v}|_2^2 = \mathbf{v}^\top \Psi^\top \Psi \mathbf{v} = \mathbf{v}^\top \mathbf{v} = 1.$$

これより, $\forall s \in \mathbb{R}$ に対して,

$$\mathbb{E}[\exp\{s\tilde{\boldsymbol{\epsilon}}^\top \mathbf{v}\}] \leq \mathbb{E}[\exp\{s\boldsymbol{\epsilon}^\top \Psi \mathbf{v}\}] \leq e^{s^2 \sigma^2 / 2}.$$

したがって, $\tilde{\boldsymbol{\epsilon}} \sim \text{subG}_r(\sigma^2)$ となる. 期待値の上限を得るために, 補題 2.12 を用いる.

$$4\mathbb{E}\left[\sup_{\mathbf{v} \in \mathbb{B}_2} (\tilde{\boldsymbol{\epsilon}}^\top \mathbf{v})^2\right] = 4 \sum_{j=1}^r \mathbb{E}[\tilde{c}_j^2] \leq 16\sigma^2 r.$$

さらに, 命題 2.30 より

$$1 - 6^d \exp\left(-\frac{t^2}{2s\sigma^2}\right) < \mathbb{P}\left(\sup_{\mathbf{u} \in \mathbb{B}_2} (\mathbf{u}^\top \tilde{\epsilon}) \leq t\right) = \mathbb{P}\left(\sup_{\mathbf{u} \in \mathbb{B}_2} (\mathbf{u}^\top \tilde{\epsilon}) \leq t^2\right).$$

命題 2.30 の証明中の議論と同じようにすれば,

$$\mathbb{P}\left(\sup_{\mathbf{u} \in \mathbb{B}_2} (\mathbf{u}^\top \tilde{\epsilon}) \leq 8 \log(6) \sigma^2 r + 8 \sigma^2 \log\left(\frac{1}{\delta}\right)\right) > 1 - \delta$$

を得る. □

注意 3.5 $d \leq n$ とする.

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

がランク d ならば,

$$|\hat{\beta}^{\text{LS}} - \beta^*|_2^2 \lesssim \frac{\text{MSE}(\mathbf{X} \hat{\beta}^{\text{LS}})}{\lambda_{\max}(B)}$$

である. ただし, $\lambda_{\max}(B)$ は行列 B の最小の固有値である. したがって, 命題 3.4 を用いて, $|\hat{\beta}^{\text{LS}} - \beta^*|_2^2$ 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. □

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, $-K := \{-\mathbf{x}; \mathbf{x} \in K\}$ としたとき, $K = -K$ となることである.

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

$$\hat{\beta}_K^{\text{LS}} \in \operatorname{argmin}_{\beta \in K} |Y - \mathbf{X}\beta|_2^2.$$

ただし, \mathbb{R}^d 乗の実数値関数 g に対して,

$$\operatorname{argmin}(g) := \{\mathbf{v} \in K; g(\mathbf{v}) \leq g(\mathbf{u}) (\forall \mathbf{u} \in K)\} \subset K$$

である. 基本方程式 (3.3) は常に成立しているので, $\text{MSE}(\mathbf{X} \hat{\beta}_K^{\text{LS}})$ の上限はより小さくなる. 実際, (3.4) は次のように変形できる.

$$|\mathbf{X} \hat{\beta}_K^{\text{LS}} - \mathbf{X} \beta^*|_2^2 \leq 2\epsilon^\top \mathbf{X} (\hat{\beta}_K^{\text{LS}} - \beta^*) \leq 2 \sup_{\beta \in K-K} (\epsilon^\top \mathbf{X} \beta).$$

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

$$2 \sup_{\beta \in K-K} (\epsilon^\top X\beta^*) = 4 \sup_{v \in XK} (\epsilon^\top v); \quad XK := \{X\beta^*; \beta^* \in K\} \subset \mathbb{R}^n$$

となる.

2.2.1 ℓ_1 制約付き最小 2 乗推定量

\mathbb{B}_1 を \mathbb{R}^d の ℓ_1 単位球とする.

$$\mathbb{B}_1 := \{x = (x_1, x_2, \dots, x_d)^\top \in \mathbb{R}^d; \sum_{j=1}^d |x_j| \leq 1\}.$$

\mathbb{B}_1 はちょうど $2d$ 個の点から成る頂点集合

$$\mathcal{V}(\mathbb{B}_1) := \{e_1, -e_1, e_2, -e_2, \dots, e_d, -e_d\}$$

を持つ. ただし,

$$e_j = (\underbrace{0, \dots, 0}_{j-1}, 1, 0, \dots, 0)^\top; \quad (j = 1, 2, \dots, d)$$

である. $K = \mathbb{B}_1$ と書くことにする. 部分集合 $XK := \{X\beta^*; \beta^* \in K\} \subset \mathbb{R}^n$ は最大 $2d$ 個の頂点をもつ凸多面体で,

$$X\mathcal{V} := \{y_1, -y_1, y_2, -y_2, \dots, y_d, -y_d\}; \quad X = (y_1, y_2, \dots, y_d)$$

に含まれる. 実際, $d \leq n$ のとき, 凸多面体 K をリスケールすることにより, XK を得ることができる. 行列 X のある列は XK の頂点ではないかもしれない. したがって, $X\mathcal{V}$ は XK の頂点集合を含む集合となる. これは, XK のサイズを測っていることになる. $\epsilon \sim N_n(0, I_n)$ のとき, 上式の上限の期待値は XK の Gauss 幅となる. ϵ 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 $d \geq 2$ とし, $K = \mathbb{B}_1$ を \mathbb{R}^d の ℓ_1 単位球とする. $\beta^* \in \mathbb{B}_1$ と仮定する. $\epsilon \sim \text{subG}_n(\sigma^2)$ ($\sigma > 0$) とし, $r = \text{rank}(X^\top X)$ とし, X の列は規準化されていて,

$$\max_{1 \leq j \leq d} |X_j|_2 \leq \sqrt{n}; \quad X_j \text{ は } X \text{ の第 } j \text{ 列}$$

とする。このとき、制約付き最小 2 乗推定量 $\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}}$ は次をみたす。

$$\mathbb{E}[\text{MSE}(\mathbf{X}\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}})] = \frac{1}{n} \mathbb{E}[\|\mathbf{X}\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}} - \mathbf{X}\beta^*\|_2^2] \lesssim \sigma \sqrt{\frac{\log d}{n}}.$$

さらに、 $\forall \delta > 0$ に対して、

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}}) \lesssim \sigma \sqrt{\frac{\log(d/\delta)}{n}}\right) \geq 1 - \delta$$

が成り立つ。

証明 この命題の前の議論より

$$\|\mathbf{X}\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}} - \mathbf{X}\beta^*\|_2^2 \leq 4 \sup_{\mathbf{v} \in \mathbf{X}_K} (\epsilon^\top \mathbf{v}).$$

ここで、 $\epsilon \sim \text{subG}_n(\sigma^2)$ であり、 \mathbf{X} のすべての列は $|X_j|_2 \leq \sqrt{n}$ ($j = 1, 2, \dots, d$) だったので、 $\epsilon^\top X_j \sim \text{subG}(n\sigma^2)$ となることに注意する。したがって、

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}}) > t\right) \leq \mathbb{P}\left(\sup_{\mathbf{v} \in \mathbf{X}_K} (\epsilon^\top \mathbf{v}) > \frac{t}{4}\right) \leq 2d \exp\left(-\frac{nt^2}{32\sigma^2}\right).$$

最後に、

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

より

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}}) \lesssim \sigma \sqrt{\frac{\log d}{n}}\right) \geq 1 - \delta$$

がわかる。 □

注意 3.7 命題 2.30 の議論を $X\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}}$ に適用することができるので,

$$\text{MSE}(X\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}}) \lesssim \sigma^2 \frac{r}{n}$$

ともなる。したがって,

$$\text{MSE}(X\widehat{\beta}_{\mathbb{B}_1}^{\text{LS}}) \lesssim \min\left(\sigma\sqrt{\frac{\log d}{n}}, \sigma^2 \frac{r}{n}\right)$$

となる。これを肘作用とよび、肘は

$$r \approx \sqrt{n}$$

付近になることがわかる。□

2.2.2 ℓ_0 制約付き最小 2 乗推定量

記号を乱用して、 $\beta^* \in \mathbb{R}^d$ 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,

$$|\beta^*|_0 := \sum_{j=1}^d \mathbb{1}\{\beta_j \neq 0\}; \quad \beta^* = (\beta_1, \beta_2, \dots, \beta_d)^\top.$$

ℓ_0 ノルムが小さい β^* をスパースなベクトルとよぶことにする。正確には、 $|\beta^*|_0 \leq k$ のとき、 k スパースベクトルという。ベクトル β^* の台を $\text{supp}(\beta^*)$ と書き,

$$\text{suupp}(\beta^*) := \{j \in \{1, 2, \dots, n\}; \beta_j \neq 0\}$$

で定義する。すると

$$|\beta^*|_0 = \#(\text{supp}(\beta^*))$$

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。

$$\lim_{q \rightarrow 0^+} \sum_{j=1}^d |\beta_j|^q = |\beta^*|_0.$$

□

$1 \leq k \leq d$ に対し, $\mathbb{B}_0(k)$ を \mathbb{R}^d の ℓ_0 球とする. すなわち,

$$\mathbb{B}_0(k) := \{\beta^* \in \mathbb{R}^d; |\beta^*|_0 \leq k\}.$$

この節の目標は, $K = \mathbb{B}_0(k)$ のとき, 制約付き最小 2 乗推定量 $\hat{\beta}_{\mathbb{B}_0(k)}^{\text{LS}}$ の MSE を評価することである. $\hat{\beta}_{\mathbb{B}_0(k)}^{\text{LS}}$ を計算するために, $\binom{d}{k}$ 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, $\binom{d}{k}$ 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 $k \leq d/2$ を固定する. $K = \mathbb{B}_0(k)$ を \mathbb{R}^d の k スパースベクトルの集合とし, $\beta^* \in \mathbb{B}_0(k)$ と仮定する. $\epsilon \sim \text{subG}_n(\sigma^2)$ ($\sigma > 0$) とし, $r = \text{rank}(\mathbf{X}^\top \mathbf{X})$ とする. このとき, $\forall \delta > 0$ に対して,

$$\mathbb{P} \left(\text{MSE}(\mathbf{X} \hat{\beta}_{\mathbb{B}_0(k)}^{\text{LS}}) \leq \frac{\sigma^2}{n} \log \binom{d}{2k} + \frac{\sigma^2 k}{n} + \frac{\sigma^2}{n} \log \left(\frac{1}{\delta} \right) \right) > 1 - \delta$$

が成立する.

証明 $\hat{\beta}_{\mathbb{B}_0(k)}^{\text{LS}}$ を $\hat{\beta}_K^{\text{LS}}$ と書く. 命題 3.4 の証明中の議論より

$$\begin{aligned} |\mathbf{X} \hat{\beta}_K^{\text{LS}} - \mathbf{X} \beta^*| &\leq 2\epsilon^\top (\mathbf{X} \hat{\beta}_K^{\text{LS}} - \mathbf{X} \beta^*) \\ &= 2|\mathbf{X} \hat{\beta}_K^{\text{LS}} - \mathbf{X} \beta^*|_2 \frac{\epsilon^\top \mathbf{X} (\hat{\beta}_K^{\text{LS}} - \beta^*)}{\mathbf{X} (\hat{\beta}_K^{\text{LS}} - \beta^*)} \end{aligned}$$

となる. $\hat{\beta}_K^{\text{LS}}$ と β^* は $\mathbb{B}_0(k)$ の中にあるので, $\hat{\beta}_K^{\text{LS}} - \beta^* \in \mathbb{B}_0(2k)$ となる. 任意の $S \subset \{1, 2, \dots, d\}$ 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. $\text{rank}(\mathbf{X}_S) = r_S$ とおくと, $r_S \leq \#(S)$ となる. $\Psi_S = [\psi_1, \psi_2, \dots, \psi_{r_S}] \in \text{Mat}(n, r_S; \mathbb{R})$ 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, $\beta^* \in \mathbb{R}^d$ に対して, $\beta^*(S) \in \mathbb{R}^{\#(S)}$ を $\beta^* = (\beta_1, \beta_2, \dots, \beta_d)^\top$ の β_j ($j \in S$) を成分ともつベクトルとする.

$$\hat{S} = \text{supp}(\hat{\beta}_K^{\text{LS}} - \beta^*)$$

とおけば, $\#(\hat{S}) \leq 2k$ となり, $\exists \nu \in \mathbb{R}^{r_S}$ があって,

$$\mathbf{X} (\hat{\beta}_K^{\text{LS}} - \beta^*) = \mathbf{X}_{\hat{S}} (\hat{\beta}_K^{\text{LS}}(\hat{S}) - \beta^*(\hat{S})) = \Psi_{\hat{S}} \nu.$$

したがって,

$$\frac{\epsilon^\top \mathbf{X}(\widehat{\beta}_K^{\text{LS}} - \beta^*)}{|\mathbf{X}(\widehat{\beta}_K^{\text{LS}} - \beta^*)|_2} = \frac{\epsilon^\top \Psi_{\widehat{S}} \nu}{|\nu|_2} \leq \max_{\#(S)=2k} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_S}} \epsilon^\top \Psi_S \mathbf{u} \right)$$

となる. ただし, $\mathbb{B}_2^{r_S}$ は \mathbb{R}^{r_S} の単位球である. これより

$$|\mathbf{X}(\widehat{\beta}_K^{\text{LS}} - \beta^*)|_2^2 \leq 2|\mathbf{X}(\widehat{\beta}_K^{\text{LS}} - \beta^*)|_2 \max_{\#(S)=2k} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_S}} \epsilon^\top \Psi_S \mathbf{u} \right)$$

より

$$|\mathbf{X}(\widehat{\beta}_K^{\text{LS}} - \beta^*)|_2^2 \leq 4 \max_{\#(S)=2k} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_S}} (\tilde{\epsilon}^\top \mathbf{u})^2 \right); \quad \tilde{\epsilon} = \Psi_S^\top \epsilon \sim \text{subG}_{r_S}(\sigma^2)$$

和集合上限を用いれば, $\forall t > 0$ に対して

$$\mathbb{P} \left(\max_{\#(S)=2k} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_S}} (\tilde{\epsilon}^\top \mathbf{u})^2 \right) > t \right) \leq \sum_{\#(S)=2k} \mathbb{P} \left(\max_{\#(S)=2k} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_S}} (\tilde{\epsilon}^\top \mathbf{u})^2 \right) > t \right).$$

命題 2.30 と $\#(S) \leq 2k$ より

$$\mathbb{P} \left(\max_{\#(S)=2k} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_S}} (\tilde{\epsilon}^\top \mathbf{u})^2 \right) > t \right) \leq 6^{\#(S)} \exp \left(-\frac{t}{8\sigma^2} \right) \leq 6^{2k} \exp \left(-\frac{t}{8\sigma^2} \right)$$

となる. これらの式をあわせれば,

$$\mathbb{P} \left(|\mathbf{X}(\widehat{\beta}_K^{\text{LS}} - \beta^*)|_2^2 > 4t \right) \leq \binom{d}{2k} 6^{2k} \exp \left(-\frac{t}{8\sigma^2} \right). \quad (3.5)$$

ここで,

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

とおけば,

$$\log \delta = \log \binom{d}{2k} + 2k \log 6 - \frac{t}{8\sigma^2}.$$

よって,

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

これより

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\text{LS}}) > t\right) = \mathbb{P}\left(|\mathbf{X}(\widehat{\boldsymbol{\beta}}_K^{\text{LS}} - \boldsymbol{\beta}^*)|_2^2 > nt\right) \leq \binom{d}{2k} 6^{2k} \exp\left(-\frac{nt}{8\sigma^2}\right)$$

なので,

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\widehat{\boldsymbol{\beta}}_K^{\text{LS}}) \lesssim \frac{\sigma^2}{n} \log\left(\frac{d}{2k}\right) + \frac{\sigma^2 k}{n} + \frac{\sigma^2}{n} \log\left(\frac{1}{\delta}\right)\right) > 1 - \delta$$

を得る. □

補題 3.10 整数 $1 \leq k \leq n$ に対して,

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

が成り立つ.

証明 帰納法で示す. $k = 1$ のとき,

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

次に, ある $k \leq n - 1$ で

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

が成立していると仮定する.

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

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,

$$\mathbb{P} \left(\text{MSE}(\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbb{B}_0(k)}^{\text{LS}}) \lesssim \frac{\sigma^2}{n} \left\{ k \log \left(\frac{ed}{2k} \right) + k \log 6 + \log \left(\frac{1}{\delta} \right) \right\} \right) > 1 - \delta$$

が成り立つ.

証明 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,

$$\mathbb{E} \left[\text{MSE}(\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbb{B}_0(k)}^{\text{LS}}) \right] \lesssim \frac{\sigma^2 k}{n} \log \left(\frac{ed}{k} \right)$$

が成り立つ.

証明 (3.5) から, $\forall H \geq 0$ に対して,

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

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,

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

となる. □

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.

$$Y_j = \beta_j + \epsilon_j \quad (j = 1, 2, \dots, n) \quad (3.6)$$

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

3.1 劣 Gauss 型確率変数列モデル

モデル (3.6) は固定デザインの回帰モデル (3.1) の特別な場合だえる. $n = d$, $\mathbf{x}_j = \mathbf{e}_j$ ($j = 1, 2, \dots, d$) 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 $d \leq n$ の場合にのみ有効であり, 高次元モデル $d > n$ のときはみだされない.

この仮定のもとで, 回帰モデル (3.1) を以下のように変形できる.

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

ただし, $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_d)^\top \sim \text{subG}_n(\sigma^2/n)$. ORT 仮定のもとで, $\boldsymbol{\xi}$ が正規分布に従うとき, 回帰モデル (3.1) は正規確率変数列モデル (3.6) と (データ \mathbf{y} をスケールングすれば) 同じである. さらに, $\forall \boldsymbol{\beta}^* \in \mathbb{R}^d$ に対して,

$$\begin{aligned} \|\mathbf{y} - \boldsymbol{\beta}^*\|_2^2 &= \left\| \frac{1}{n} \mathbf{X}^\top \mathbf{Y} - \boldsymbol{\beta}^* \right\|_2^2 \\ &= \|\boldsymbol{\beta}^*\|_2^2 - \frac{2}{n} \boldsymbol{\beta}^{*\top} \mathbf{X}^\top \mathbf{Y} + \frac{1}{n^2} \mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y} \\ &= \frac{1}{n} \|\boldsymbol{\beta}^* \mathbf{X}\|_2^2 - \frac{2}{n} (\mathbf{X} \boldsymbol{\beta}^*)^\top \mathbf{Y} + \frac{1}{n^2} \|\mathbf{Y}\|_2^2 + Q \\ &= \frac{1}{n} \|\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}^*\|_2^2 + Q \end{aligned}$$

となる. ただし,

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

である。よって、最小 2 乗推定量 $\hat{\beta}^{\text{LS}}$ は \mathbf{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。

$$\mathbf{y} = \beta^* + \xi; \quad \xi \sim \text{subG}_d(\sigma^2/n). \quad (3.7)$$

3.2 スパース適応的しきい値推定量

もし、 $\beta^* \in \mathbb{R}^d$ が k スパース ($1 \leq d \leq k$) 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、

$$\mathbb{P} \left(\text{MSE}(\mathbf{X}\hat{\beta}_{\mathbb{B}_0(k)}^{\text{LS}}) \leq C_\delta \frac{\sigma^2 k}{n} \log \left(\frac{ed}{2k} \right) \right) > 1 - \delta$$

となるのがわかる。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、 $\hat{\beta}_{\mathbb{B}_0(k)}^{\text{LS}}$ 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。

(3.7) の劣 Gauss 的確率変数列モデルを仮定する。 β^* の事前情報が全くないとき、最小 2 乗推定量 $\hat{\beta}^{\text{LS}} = \mathbf{y}$ によって、 β^* 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、

$$\mathbb{P} \left(\text{MSE}(\mathbf{X}\hat{\beta}^{\text{LS}}) = \|\mathbf{y} - \beta^*\|_2^2 = \|\xi\|_2^2 \leq C_\delta \frac{\sigma^2 d}{n} \right) > 1 - \delta$$

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 $c \leq 1$ があって、 $k = cd$ のとき、探している者がこれである。 k が d より十分小さいという事実を用いずに、 β^* 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。 $j \in \{1, 2, \dots, d\}$ に対して、 $\beta_j = 0$ のとき、 $y_j = \xi_j$ は代行分散 σ^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、

$$\mathbb{P}(|\xi_j| > t) \leq 2 \exp \left(-\frac{nt^2}{2\sigma^2} \right)$$

となる。ここで、

$$\delta := 2 \exp \left(-\frac{nt^2}{2\sigma^2} \right) \iff t = \pm \sigma \sqrt{\frac{2 \log(1/\delta)}{n}}$$

となるので, $\forall \delta > 0$ に対して,

$$\mathbb{P} \left(|\xi_j| \leq \sigma \sqrt{\frac{2 \log(1/\delta)}{n}} =: \tau \right) > 1 - \delta \quad (3.8)$$

を得る. この不等式の結果は興味深い. (3.8) から, $j = 1, 2, \dots, d$ に対して,

$$|y_j| \gg \tau \implies \beta_j \neq 0$$

である. 他方,

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

したがって, $|y_j| \leq 2\tau$ 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 $\hat{\beta}^{\text{HRD}} = (\hat{\beta}_1^{\text{HRD}}, \hat{\beta}_1^{\text{HRD}}, \dots, \hat{\beta}_d^{\text{HRD}})^\top$ は次で与えられる. $j = 1, 2, \dots, d$ に対して,

$$\hat{\beta}_j^{\text{HRD}} = \begin{cases} y_j & (|y_j| > 2\tau), \\ 0 & (|y_j| \leq 2\tau) \end{cases}.$$

簡単に,

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

と書ける. 上の考察から, (3.8) を考慮に入れば, 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 $\forall j \in \{1, 2, \dots, d\}$ に対して, 同時に

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

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,

$$\mathbb{P} \left(\max_{1 \leq j \leq d} |\xi_j| > t \right) \leq 2 \exp \left(-\frac{t^2}{2\sigma^2} \right)$$

となるので,

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

となるので,

$$\mathbb{P}\left(\max_{1 \leq j \leq d} |\xi_j| \leq \sigma \sqrt{\frac{2 \log(2d/\delta)}{n}}\right) \geq 1 - \delta$$

を得る.

定理 3.14 線型回帰モデル (3.1) と ORT 条件を仮定する. すなわち, 劣 Gauss 的確率変数列モデル (3.7) を仮定する. $\delta > 0$ とする. このとき, しきい値

$$2\tau = 2\sigma \sqrt{\frac{2 \log(2d/\delta)}{n}} \quad (3.9)$$

の硬しきい値推定量 $\hat{\beta}^{\text{HRD}}$ は以下の性質をみtas.

(1) $|\beta^*|_0 = k$ のとき, $\forall \delta > 0$ に対して,

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\hat{\beta}^{\text{HRD}}) = \frac{1}{n} \|\mathbf{X}\hat{\beta}^{\text{HRD}} - \mathbf{X}\beta^*\|_2^2 \lesssim \sigma^2 \frac{k \log(2d/\delta)}{n}\right) > 1 - \delta.$$

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

$$\mathbb{P}\left(\text{supp}(\hat{\beta}^{\text{HRD}}) = \text{supp}(\beta^*)\right) > 1 - \delta.$$

証明 事象 A を

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

で定義する. 命題 2.24 より

$$\mathbb{P}(A) > 1 - \delta$$

となる. 事象 A 上では, $j \in \{1, 2, \dots, d\}$ 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.

$$|y_j| > 2\tau \implies |\beta_j| \geq |y_j| - |\xi_j| > \tau, \quad (3.10)$$

$$|y_j| \leq 2\tau \implies |\beta_j| \leq |y_j| + |\xi_j| \leq 3\tau. \quad (3.11)$$

これらより,

$$\begin{aligned}
|\beta_j^{\text{HRD}} - \beta_j| &= |\beta_j^{\text{HRD}} - \beta_j| \mathbb{1}\{|y_j| > 2\tau\} + |\beta_j^{\text{HRD}} - \beta_j| \mathbb{1}\{|y_j| \leq 2\tau\} \\
&= |y_j - \beta_j| \mathbb{1}\{|y_j| > 2\tau\} + |\beta_j| \mathbb{1}\{|y_j| \leq 2\tau\} \\
&\leq |\xi_j| \mathbb{1}\{|\beta_j| > \tau\} + |\beta_j| \mathbb{1}\{|\beta_j| \leq 3\tau\} \\
&\quad (\because A \text{ 上では, (3.10) と (3.11)}) \\
&\leq \tau \mathbb{1}\{|\beta_j| > \tau\} + |\beta_j| \mathbb{1}\{|\beta_j| \leq 3\tau\} \\
&\quad (\because A \text{ 上では, } |\xi_j| \leq \tau) \\
&\leq 4 \min\{|\beta_j|, \tau\}.
\end{aligned}$$

この式より

$$\begin{aligned}
|\hat{\beta}^{\text{HRD}} - \beta^*|_2^2 &= \sum_{j=1}^d |\beta_j^{\text{HRD}} - \beta_j|^2 \\
&= 16 \sum_{j=1}^d \min\{|\beta_j|^2, \tau^2\} \\
&\leq 16 |\beta^*|_0 \tau^2 \\
&= 32k\sigma^2 \frac{\log(2d/\delta)}{n} \\
&\lesssim \sigma^2 \frac{k \log(2d/\delta)}{n}
\end{aligned}$$

を得る. よって,

$$\text{MSE}(\mathbf{X}\hat{\beta}^{\text{HRD}}) \lesssim \sigma^2 \frac{k \log(2d/\delta)}{n}$$

なので,

$$\mathbb{P}\left(\text{MSE}(\mathbf{X}\hat{\beta}^{\text{HRD}}) \lesssim \sigma^2 \frac{k \log(2d/\delta)}{n}\right) \geq \mathbb{P}(A) > 1 - \delta.$$

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、 $\widehat{\beta}_j^{\text{HRD}} \neq 0$ 。よって、

$$\text{supp}(\beta^*) \subset \text{supp}(\widehat{\beta}^{\text{HRD}}).$$

次に、 $\widehat{\beta}_j^{\text{HRD}} \neq 0$ のとき、

$$|\widehat{\beta}_j^{\text{HRD}}| = |y_j| > 2\tau.$$

これより、 A 上では、

$$|\beta_j| = |y_j - \xi_j| \geq |y_j| - |\xi_j| \geq 2\tau - \tau = \tau.$$

したがって、 $\beta_j \neq 0$ となるので、

$$\text{supp}(\widehat{\beta}^{\text{HRD}}) \subset \text{supp}(\beta^*).$$

よって、 A 上では、

$$\text{supp}(\widehat{\beta}^{\text{HRD}}) = \text{supp}(\beta^*).$$

このことより、

$$\text{「} A \text{ が起きる」} \implies \text{「} \left\{ \text{supp}(\widehat{\beta}^{\text{HRD}}) = \text{supp}(\beta^*) \right\} \text{ が起きる」}$$

となるので、

$$\mathbb{P} \left(\text{supp}(\widehat{\beta}^{\text{HRD}}) = \text{supp}(\beta^*) \right) \geq \mathbb{P}(A) \geq 1 - \delta.$$

□

次に、軟しきい値推定量 $\widehat{\beta}^{\text{SFT}}$ を

$$\widehat{\beta}_j^{\text{SFT}} = \begin{cases} y_j - 2\tau & (y_j > 2\tau), \\ y_j + 2\tau & (y_j < -2\tau), \\ 0 & (|y_j| \leq 2\tau), \end{cases} \quad (j = 1, 2, \dots, d)$$

を考える。これは

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

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 (3.9) をもつ軟しきい値推定量 $\widehat{\beta}^{\text{SFT}}$ は以下の性質をみす.

(1) $|\beta^*|_0 = k$ のとき, $\forall \delta > 0$ に対して,

$$\mathbb{P} \left(\text{MSE}(\mathbf{X}\widehat{\beta}^{\text{SFT}}) = \frac{1}{n} |\widehat{\beta}^{\text{SFT}} - \beta^*|_2^2 \lesssim \sigma^2 \frac{k \log(2d/\delta)}{n} \right) > 1 - \delta.$$

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

$$\mathbb{P} \left(\text{supp}(\widehat{\beta}^{\text{SFT}}) = \text{supp}(\beta^*) \right) > 1 - \delta.$$

証明 事象 A を

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

で定める. (3.10) と (3.11) より, $j \in \{1, 2, \dots, d\}$ に対して, A 上では,

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

となる. この不等式と τ の定義に注意すれば,

$$\begin{aligned} |\widehat{\beta}^{\text{SFT}} - \beta^*|_2^2 &= \sum_{j=1}^d |\widehat{\beta}_j^{\text{SFT}} - \beta_j|^2 \leq 25 \sum_{j=1}^d \min\{|\beta_j|^2, \tau^2\} \leq 25 |\beta^*|_0 \tau^2 \\ &= 50k\sigma^2 \frac{\log(2d/\delta)}{n} \lesssim \sigma^2 \frac{k \log(2d/\delta)}{n} \end{aligned}$$

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、 $\widehat{\beta}_j^{\text{SFT}} \neq 0$. よって、

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

次に、 $|\widehat{\beta}_j^{\text{SFT}}| \neq 0$ のとき、

$$|y_j| > 2\tau.$$

これより、 A 上では、

$$|\beta_j| = |y_j - \xi_j| \geq |y_j| - |\xi_j| \geq 2\tau - \tau = \tau.$$

したがって、 $\beta_j \neq 0$ なので、

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

よって、 A 上では

$$\text{supp}(\widehat{\beta}^{\text{SFT}}) = \text{supp}(\beta^*).$$

これより、

$$\mathbb{P}\left(\text{supp}(\widehat{\beta}^{\text{SFT}}) = \text{supp}(\beta^*)\right) \geq \mathbb{P}(A) > 1 - \delta.$$

□

4 高次元線型回帰モデル

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。

$$\begin{aligned} \widehat{\beta}^{\text{HRD}} &= \operatorname{argmax}_{\beta \in \mathbb{R}^d} \{|\mathbf{y} - \beta^*|_2^2 + 4\tau^2|\beta^*|_0\}, \\ \widehat{\beta}^{\text{SFT}} &= \operatorname{argmax}_{\beta \in \mathbb{R}^d} \{|\mathbf{y} - \beta^*|_2^2 + 4\tau^2|\beta^*|_1\}. \end{aligned}$$

(3.6) 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。

$$\begin{aligned} \widehat{\beta}^{\text{HRD}} &= \operatorname{argmax}_{\beta \in \mathbb{R}^d} \{|\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + 4\tau^2|\beta^*|_0\}, \\ \widehat{\beta}^{\text{SFT}} &= \operatorname{argmax}_{\beta \in \mathbb{R}^d} \{|\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + 4\tau^2|\beta^*|_1\}. \end{aligned}$$

ORT 条件をみたまないとき、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 (3.1) を仮定する。BIC 推定量 $\hat{\beta}^{\text{BIC}}$ を次で定義する。

$$\hat{\beta}^{\text{BIC}} \in \operatorname{argmin}_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{n} |\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + \tau^2 |\beta^*|_0 \right\}.$$

さらに、 $\hat{\beta}$ の LASSO 推定量 $\hat{\beta}^{\mathcal{L}}$ を

$$\hat{\beta}^{\mathcal{L}} \in \operatorname{argmin}_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{n} |\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + 2\tau |\beta^*|_1 \right\}.$$

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 (3.1) 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、

$$\begin{aligned} & \min_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{n} |\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + \tau^2 |\beta^*|_0 \right\} \\ &= \min_{1 \leq k \leq d} \left\{ \min_{\beta \in \mathbb{R}^d: |\beta|_0=k} \left(\frac{1}{n} |\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + \tau^2 k \right) \right\}. \end{aligned}$$

$\min_{\beta \in \mathbb{R}^d: |\beta|_0=k} (1/n) |\mathbf{Y} - \mathbf{X}\beta^*|_2^2$ を計算するために、 $\binom{d}{k}$ 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

$$C \sum_{k=0}^d \binom{d}{k} k^3 = Cd^3 2^d$$

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 glmnet.

(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. □

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 (3.1) を仮定し, $\epsilon \sim \text{subG}_n(\sigma^2)$ ($\sigma > 0$) とし, $|\beta^*|_0 \geq 1$ とする. このとき, 罰則

$$\tau^2 = 16 \log(6) \frac{\sigma^2}{n} + 32 \frac{\sigma^2 \log(ed)}{n} \quad (3.12)$$

をもつ BIC 推定量 $\hat{\beta}^{\text{BIC}}$ は以下の性質をみたく. $\forall \delta > 0$ に対して,

$$\mathbb{P} \left(\text{MSE}(\mathbf{X}\hat{\beta}^{\text{BIC}}) = \frac{1}{n} |\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*|_2^2 \lesssim |\beta^*|_0 \sigma^2 \frac{\log(ed/\delta)}{n} \right) > 1 - \delta.$$

証明 まず, 推定量の定義より,

$$\frac{1}{n} |\mathbf{Y} - \mathbf{X}\hat{\beta}^{\text{BIC}}|_2^2 + \tau^2 |\hat{\beta}^{\text{BIC}}|_0 \leq \frac{1}{n} |\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + \tau^2 |\beta^*|_0$$

となることに注意せよ. これと簡単な計算により,

$$|\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*|_2^2 \leq n\tau^2 |\beta^*|_0 + 2\epsilon^\top \mathbf{X}(\hat{\beta}^{\text{BIC}} - \beta^*) - n\tau^2 |\hat{\beta}^{\text{BIC}}|_0$$

を得る. 次に,

$$\begin{aligned} 2\epsilon^\top \mathbf{X}(\hat{\beta}^{\text{BIC}} - \beta^*) &= 2\epsilon^\top \left(\frac{\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*}{|\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*|_2} \right) |\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*|_2 \\ &\leq 2 \left[\epsilon^\top \left(\frac{\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*}{|\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*|_2} \right) \right]^2 + \frac{1}{2} |\mathbf{X}\hat{\beta}^{\text{BIC}} - \mathbf{X}\beta^*|_2^2 \end{aligned}$$

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,

$$|\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{BIC}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 \leq 2n\tau^2|\boldsymbol{\beta}^*|_0 + 4[\boldsymbol{\epsilon}^\top \mathcal{U}(\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{BIC}} - \mathbf{X}\boldsymbol{\beta}^*)]^2 - 2n\tau^2|\hat{\boldsymbol{\beta}}^{\text{BIC}}|_0 \quad (3.13)$$

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

$$\mathcal{U}(\mathbf{z}) := \frac{\mathbf{z}}{|\mathbf{z}|_2}.$$

次に, $\hat{\boldsymbol{\beta}}^{\text{BIC}}$ を sup-out する必要がある. このために, sup を以下のように分解する.

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

これを上の不等式に用いれば,

$$\begin{aligned} & 4[\boldsymbol{\epsilon}^\top \mathcal{U}(\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{BIC}} - \mathbf{X}\boldsymbol{\beta}^*)]^2 - 2n\tau^2|\hat{\boldsymbol{\beta}}^{\text{BIC}}|_0 \\ & \leq \max_{1 \leq k \leq d} \left\{ \max_{\#(S)=k} \sup_{\text{supp}(\hat{\boldsymbol{\beta}})=S} \left\{ 4[\boldsymbol{\epsilon}^\top \mathcal{U}(\mathbf{X}\hat{\boldsymbol{\beta}}^{\text{BIC}} - \mathbf{X}\boldsymbol{\beta}^*)]^2 - 2n\tau^2 k \right\} \right\} \\ & \leq \max_{1 \leq k \leq d} \left\{ \max_{\#(S)=k} \sup_{S \subset \mathbb{R}^d; \#(S)=k} \sup_{\mathbf{u} \in \mathbb{B}_2^{r_{S+}}} 4[\boldsymbol{\epsilon}^\top \Psi_{S+} \mathbf{u}]^2 - 2n\tau^2 k \right\}. \end{aligned}$$

ただし, $j \in \{1, 2, \dots, d\}$ とし, \mathbf{c}_j を \mathbf{X} の第 j 列としたとき,

$$\Psi_{S+} = [\psi_1, \psi_2, \dots, \psi_{S+}]$$

は $\text{span}\{\mathbf{c}_j; j \in S \cup \text{supp}(\boldsymbol{\beta}^*)\}$ の正規直交基底であり, $r_{S+} \leq \#(S) + |\boldsymbol{\beta}^*|_0$ 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,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq k \leq d} \left\{ \max_{\#(S)=k} \sup_{S \subset \mathbb{R}^d; \#(S)=k} \sup_{\mathbf{u} \in \mathbb{B}_2^{r_{S+}}} 4[\boldsymbol{\epsilon}^\top \Psi_{S+} \mathbf{u}]^2 - 2n\tau^2 k \right\} \geq t \right) \\ & \leq \sum_{j=1}^d \sum_{S \subset \mathbb{R}^d; \#(S)=k} \mathbb{P} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_{S+}}} [\boldsymbol{\epsilon}^\top \Psi_{S+} \mathbf{u}]^2 \geq \frac{t}{4} + \frac{1}{2}n\tau^2 k \right) \end{aligned}$$

となるので、命題 2.30 の ϵ 網の議論より $\#(S) = k$ に対して、

$$\begin{aligned} \mathbb{P} \left(\sup_{\mathbf{u} \in \mathbb{B}_2^{r_{S^+}}} [\boldsymbol{\epsilon}^\top \Psi_{S^+} \mathbf{u}]^2 \geq \frac{t}{4} + \frac{1}{2} n \tau^2 k \right) &\leq 2 \cdot 6^{r_{S^+}} \exp \left(-\frac{\frac{t}{4} + \frac{1}{2} n \tau^2 k}{8\sigma^2} \right) \\ &\leq 2 \exp \left(-\frac{t}{32\sigma^2} - \frac{n\tau^2 k}{16\sigma^2} + (k + |\boldsymbol{\beta}^*|_0) \log 6 \right) \\ &\leq \exp \left(-\frac{t}{32\sigma^2} - 2k \log(ed) + |\boldsymbol{\beta}^*|_0 \log(12) \right). \end{aligned}$$

□

4.3 LASSO 推定量の精度評価

定理 3.20 線型回帰モデル (3.1) を仮定し、 $\boldsymbol{\epsilon} \sim \text{subG}_n(\sigma)$ ($\sigma > 0$) 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。

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

このとき、調整パラメータ

$$2\tau := 2\sigma \sqrt{\frac{2 \log(2d)}{n}} + 2\sigma \sqrt{\frac{2 \log(1/\delta)}{n}} \quad (3.14)$$

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、

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

ただし、

$$\text{MSE}(\mathbf{X} \widehat{\boldsymbol{\beta}}^\mathcal{L}) = \frac{1}{n} \|\mathbf{X} \widehat{\boldsymbol{\beta}}^\mathcal{L} - \mathbf{X} \boldsymbol{\beta}^*\|_2^2$$

である。

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

$$\frac{1}{n} \|\mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}^\mathcal{L}\|_2^2 + 2\tau |\widehat{\boldsymbol{\beta}}^\mathcal{L}|_1 \leq \frac{1}{n} \|\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}^*\|_2^2 + 2\tau |\boldsymbol{\beta}^*|_1 = \frac{1}{n} \|\boldsymbol{\epsilon}\|_2^2 + 2\tau |\boldsymbol{\beta}^*|_1.$$

Hölder の不等式 (命題 ??) と上の不等式より,

$$\begin{aligned}
 |\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 &= |\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{Y} + \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 \\
 &= |\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}}|_2^2 + |\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 + 2(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*)^\top (\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^* - \boldsymbol{\epsilon}) \\
 &\leq 2\boldsymbol{\epsilon}^\top (\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*) - 2|\boldsymbol{\epsilon}|_2^2 + 2|\boldsymbol{\epsilon}|_2^2 + 2\tau n|\boldsymbol{\beta}^*|_1 - 2\tau n|\boldsymbol{\beta}^*|_1 \\
 &= 2\boldsymbol{\epsilon}^\top \mathbf{X}(\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \boldsymbol{\beta}^*) + 2\tau n(|\boldsymbol{\beta}^*|_1 - |\widehat{\boldsymbol{\beta}}^{\mathcal{L}}|) \\
 &\leq 2|\mathbf{X}^\top \boldsymbol{\epsilon}|_\infty |\widehat{\boldsymbol{\beta}}^{\mathcal{L}}|_1 + 2|\mathbf{X}^\top \boldsymbol{\epsilon}|_\infty |\boldsymbol{\beta}^*|_1 + 2\tau n(|\boldsymbol{\beta}^*|_1 - |\widehat{\boldsymbol{\beta}}^{\mathcal{L}}|) \\
 &= 2(|\mathbf{X}^\top \boldsymbol{\epsilon}|_\infty - n\tau) |\widehat{\boldsymbol{\beta}}^{\mathcal{L}}|_1 + 2(|\mathbf{X}^\top \boldsymbol{\epsilon}|_\infty + n\tau) |\boldsymbol{\beta}^*|_1.
 \end{aligned}$$

$j = 1, 2, \dots, d$ に対して, $|X_j|_2 \leq \sqrt{n}$ なので, $X_j^\top \boldsymbol{\epsilon} \sim \text{subG}(n\sigma^2)$ 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,

$$\mathbb{P}\left(|\mathbf{X}^\top \boldsymbol{\epsilon}|_\infty > t\right) = \mathbb{P}\left(\max_{1 \leq j \leq d} |X_j^\top \boldsymbol{\epsilon}| > t\right) \leq 2d \exp\left(-\frac{t^2}{2n\sigma^2}\right) =: \delta$$

となることがわかる. ここで,

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

とすれば, $|\mathbf{X}^\top \boldsymbol{\epsilon}|_\infty \leq t$ 上では,

$$\begin{aligned}
 |\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 &\leq \underbrace{2(t - n\tau)}_{=0} |\widehat{\boldsymbol{\beta}}^{\mathcal{L}}|_1 + 2(t + n\tau) |\boldsymbol{\beta}^*|_1 \\
 &= 4n\tau |\boldsymbol{\beta}^*|_1
 \end{aligned}$$

なので,

$$\mathbb{P}\left(|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathcal{L}} - \mathbf{X}\boldsymbol{\beta}^*|_2^2 \leq 4n\tau |\boldsymbol{\beta}^*|_1\right) \geq \mathbb{P}\left(|\mathbf{X}^\top \boldsymbol{\epsilon}|_\infty \leq t\right) > 1 - \delta$$

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 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。

4.3.1 Incoherence 仮定

仮定 $\text{INC}(k)$ デザイン行列 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、

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

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

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

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

$$\left| \frac{|X_j|_2^2}{n} - 1 \right| \leq \frac{1}{32k}.$$

(2) $1 \leq \forall j, \ell \leq d, j \neq \ell$ に対して、

$$\frac{|X_j^\top X_\ell|}{n} \leq \frac{1}{32k}.$$

注意 3.22 ORT 条件は、 $k \rightarrow \infty$ のとき、 $\text{INC}(k)$ 条件の極限と考えることができる。しかし、 ORT 条件は $d \leq n$ を要求しているが、 $\text{INC}(k)$ 条件はそれを求めている。実際、次の命題は $d \gg n$ の場合を考えている。 \square

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

$$\mathbb{P}(X_{ij} = 1) = \mathbb{P}(X_{ij} = -1) = \frac{1}{2}.$$

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,

$$\mathbb{P}\left(\mathbf{X} \text{ は INC}(k) \text{ 条件をみたす}\right) > 1 - \delta$$

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,

$$\mathbb{P}\left(\mathbf{X} \text{ は INC}(k) \text{ 条件をみたす}\right) > 1 - \delta$$

となる.

証明 ϵ_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, d$) 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 ($j = 1, 2, \dots, d$) 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, $\ell_1 \neq \ell_2$ ($\ell_1, \ell_2 = 1, 2, \dots, d$) 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. $\xi_1^{(\ell_1, \ell_2)}, \xi_2^{(\ell_1, \ell_2)}, \dots, \xi_n^{(\ell_1, \ell_2)}$ 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,

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

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,

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

なので,

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

となる. □

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

$$\beta_S^* = (\beta_{S,1}, \beta_{S,2}, \dots, \beta_{S,d})^\top \in \mathbb{R}^d$$

を次で与えられるベクトルとする.

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

特に,

$$|\beta^*|_1 = |\beta_S^*|_1 + |\beta_{S^c}^*|_1$$

に注意する.

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

$$|\beta_{S^c}^*|_1 \leq 3|\beta_S^*|_1 \tag{3.15}$$

をみたしているとする。このとき、

$$|\boldsymbol{\beta}^*|_2^2 \leq 2 \frac{|\mathbf{X}\boldsymbol{\beta}^*|_2^2}{n}$$

が成立する。

証明 まず、

$$\frac{|\mathbf{X}\boldsymbol{\beta}^*|_2^2}{n} = \frac{|\mathbf{X}\boldsymbol{\beta}_S^*|_2^2}{n} + \frac{|\mathbf{X}\boldsymbol{\beta}_{S^c}^*|_2^2}{n} + \frac{2\boldsymbol{\beta}_S^{*\top} \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}_{S^c}^*}{n} \quad (3.16)$$

に注意する。各項を別々に評価する。

(1) INC(k) 条件より、

$$\begin{aligned} \frac{|\mathbf{X}\boldsymbol{\beta}_S^*|_2^2}{n} &= \boldsymbol{\beta}_S^{*\top} \frac{\mathbf{X}^\top \mathbf{X}}{n} \boldsymbol{\beta}_S^* = |\boldsymbol{\beta}_S^*|_2^2 + \boldsymbol{\beta}_S^{*\top} \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} - I_d \right) \boldsymbol{\beta}_S^* \\ &\leq |\boldsymbol{\beta}_S^*|_2^2 - \frac{|\boldsymbol{\beta}_S^*|_2^2}{32k} \leq |\boldsymbol{\beta}_S^*|_2^2 - \frac{|\boldsymbol{\beta}_S^*|_1^2}{32k}. \end{aligned} \quad (3.17)$$

(2) 同様に、

$$\frac{|\mathbf{X}\boldsymbol{\beta}_{S^c}^*|_2^2}{n} \geq |\boldsymbol{\beta}_{S^c}^*|_2^2 - \frac{|\boldsymbol{\beta}_{S^c}^*|_2^2}{n} \geq |\boldsymbol{\beta}_{S^c}^*|_2^2 - \frac{9|\boldsymbol{\beta}_{S^c}^*|_1^2}{n}. \quad (3.18)$$

最後の不等号は、錐条件 (3.15) を用いた。

(3) 最後に、

$$2 \left| \boldsymbol{\beta}_S^{*\top} \frac{\mathbf{X}^\top \mathbf{X}}{n} \boldsymbol{\beta}_{S^c}^* \right| \leq \frac{2}{32k} |\boldsymbol{\beta}_S^*|_1 |\boldsymbol{\beta}_{S^c}^*|_1 \leq \frac{6}{32k} |\boldsymbol{\beta}_S^*|_1^2. \quad (3.19)$$

最後の不等号は、錐条件 (3.15) を用いた。Cauchy-Schwarz の不等式より

$$|\boldsymbol{\beta}_S^*|_1^2 \leq \#(S) |\boldsymbol{\beta}_S^*|_2^2$$

となることに注意する。したがって、(3.17) – (3.19) を (3.16) に代入すれば、

$$\begin{aligned} \frac{|\mathbf{X}\boldsymbol{\beta}^*|_2^2}{n} &\geq |\boldsymbol{\beta}_S^*|_2^2 + |\boldsymbol{\beta}_{S^c}^*|_2^2 - \frac{16}{32k} |\boldsymbol{\beta}_S^*|_1^2 \\ &\geq |\boldsymbol{\beta}_S^*|_2^2 + |\boldsymbol{\beta}_{S^c}^*|_2^2 - \frac{16\#(S)}{32k} |\boldsymbol{\beta}_S^*|_2^2 \\ &\geq \frac{1}{2} |\boldsymbol{\beta}_S^*|_2^2. \end{aligned}$$

最後から 2 番目の不等号は, Cauchy-Schwarz の不等式より

$$|\beta_S^*|_1^2 \leq \#(S)|\beta_S^*|_2^2$$

となることと $\#(S) \leq k$ であることからわかる. \square

4.3.2 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 (3.1) と $\epsilon \sim \text{subG}_n(\sigma^2)$ ($\sigma > 0$) を仮定する. さらに, ある $k \in \mathbb{N}$ があって, $|\beta^*|_0 \leq k$ とし, デザイン行列 \mathbf{X} は $\text{INC}(k)$ 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

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

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

$$\mathbb{P} \left(\text{MSE}(\mathbf{X}\hat{\beta}^{\mathcal{L}}) = \frac{1}{n} \left| \mathbf{X}\hat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^* \right|_2^2 \lesssim k\sigma^2 \frac{\log(2d/\delta)}{n} \right) > 1 - \delta$$

と

$$\mathbb{P} \left(\left| \hat{\beta}^{\mathcal{L}} - \beta^* \right|_2^2 \lesssim k\sigma^2 \frac{\log(2d/\delta)}{n} \right) > 1 - \delta.$$

証明 $\hat{\beta}^{\mathcal{L}}$ の定義より,

$$\frac{1}{n} \left| \mathbf{Y} - \mathbf{X}\hat{\beta}^{\mathcal{L}} \right|_2^2 \leq \frac{1}{n} \left| \mathbf{Y} - \mathbf{X}\beta^* \right|_2^2 + 2\tau |\beta^*|_1 - 2\tau \left| \hat{\beta}^{\mathcal{L}} \right|_1$$

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

$$\left| \mathbf{Y} - \mathbf{X}\hat{\beta}^{\mathcal{L}} \right|_2^2 = \left| \mathbf{X}\hat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^* \right|_2^2 + \left| \mathbf{Y} - \mathbf{X}\beta^* \right|_2^2 - 2\epsilon^\top \mathbf{X}(\hat{\beta}^{\mathcal{L}} - \beta^*)$$

に注意すれば,

$$\begin{aligned} \left| \mathbf{X}\hat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^* \right|_2^2 + n\tau \left| \hat{\beta}^{\mathcal{L}} - \beta^* \right|_1 &\leq 2\epsilon^\top \mathbf{X}(\hat{\beta}^{\mathcal{L}} - \beta^*) + n\tau \left| \hat{\beta}^{\mathcal{L}} - \beta^* \right|_1 \\ &\quad + 2n\tau |\beta^*|_1 - 2n\tau \left| \hat{\beta}^{\mathcal{L}} \right|_1 \end{aligned} \quad (3.20)$$

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,

$$|\epsilon^\top \mathbf{X}(\widehat{\beta}^\mathcal{L} - \beta^*)| \leq |\epsilon^\top \mathbf{X}|_\infty \cdot |\widehat{\beta}^\mathcal{L} - \beta^*|$$

である. $\text{INC}(k)$ 条件より,

$$|X_j|_2^2 \leq n + \frac{1}{32k} \leq 2n$$

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

$$\begin{aligned} \mathbb{E}[\exp(sX_j^\top \epsilon)] &\leq \exp\left(s|X_j|_2 \frac{X_j^\top}{|X_j|_2} \epsilon\right) \\ &\leq \exp\left(s\sqrt{2n} \frac{X_j^\top}{|X_j|_2} \epsilon\right) \\ &\leq \exp\left(-\frac{\sigma^2}{2}(\sqrt{2ns})^2\right) = \exp(-n\sigma^2 s^2) \end{aligned}$$

となるので,

$$X_j^\top \epsilon \sim \text{subG}_n(2n\sigma^2)$$

がわかる. よって,

$$\mathbb{P}\left(\max_{1 \leq j \leq n} |X_j^\top \epsilon| > t\right) \leq 2d \exp\left(-\frac{t^2}{4n\sigma^2}\right).$$

ここで,

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

とすれば, 上の式より,

$$\mathbb{P}\left(|\mathbf{X}^T \epsilon|_\infty > \frac{nt}{2}\right) \leq 2d \exp\left(-\frac{n\tau^2}{16\sigma^2}\right)$$

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,

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

となる. これより

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

を得る. よって,

$$\mathbb{P}\left(|\mathbf{X}^\top \epsilon|_\infty > \frac{n\tau}{2}\right) \leq \delta$$

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

$$|\epsilon^\top \mathbf{X}(\hat{\beta}^\mathcal{L} - \beta^*)| \leq |\epsilon \mathbf{X}^\top|_\infty |\hat{\beta}^\mathcal{L} - \beta^*|_1 \leq \frac{n\tau}{2} |\hat{\beta}^\mathcal{L} - \beta^*|_1$$

となる. ここで,

$$S := \text{supp}(\hat{\beta}^\mathcal{L})$$

とおく. $|\mathbf{X}\epsilon|_\infty \leq (n\tau/2)$ 上では (3.20) より,

$$\begin{aligned}
&|\mathbf{X}\hat{\beta}^\mathcal{L} - \mathbf{X}\beta^*|_2^2 + n\tau |\hat{\beta}^\mathcal{L} - \beta^*|_1 \\
&\leq 2|\epsilon^\top \mathbf{X}(\hat{\beta}^\mathcal{L} - \beta^*)| + n\tau |\hat{\beta}^\mathcal{L} - \beta^*|_1 + 2n\tau |\beta^*|_1 - 2n\tau |\hat{\beta}^\mathcal{L}|_1 \\
&\leq 2n\tau |\hat{\beta}_S^\mathcal{L} - \beta^*|_1 + 2n\tau |\beta^*|_1 - 2n\tau |\hat{\beta}_S^\mathcal{L}|_1 \\
&\leq 4n\tau |\hat{\beta}_S^\mathcal{L} - \beta^*|_1 \tag{3.21}
\end{aligned}$$

となる. 特に,

$$4|\hat{\beta}^\mathcal{L} - \beta^*|_1 \leq |\hat{\beta}_S^\mathcal{L} - \beta^*|_1$$

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

$$|\hat{\beta}_S^\mathcal{L} - \beta_S^*|_1 + |\hat{\beta}_{S^c}^\mathcal{L} - \beta_{S^c}^*|_1 = |\hat{\beta}^\mathcal{L} - \beta^*|_1 \leq 4|\hat{\beta}_S^\mathcal{L} - \beta^*|_1 \leq 4|\hat{\beta}_S^\mathcal{L} - \beta_S^*|_1$$

なので,

$$|\widehat{\beta}_{Sc}^{\mathcal{L}} - \beta_{Sc}^*|_1 \leq |\widehat{\beta}_S^{\mathcal{L}} - \beta_S^*|_1$$

を得る. よって,

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

としたとき, $\tilde{\beta}$ は条件 (3.15) 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. $\#(S) \leq k$ と $S = \text{supp}(S)$ なので,

$$\begin{aligned} |\widehat{\beta}_S^{\mathcal{L}} - \beta_S^*|_1 &\leq \sqrt{\#(S)} |\widehat{\beta}_S^{\mathcal{L}} - \beta_S^*|_2 \\ &\leq \sqrt{\#(S)} |\widehat{\beta}^{\mathcal{L}} - \beta^*|_2 \\ &\quad (\because \beta_j = 0 (j \notin S)) \\ &\leq \sqrt{\frac{2k}{n}} |\mathbf{X}\widehat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^*|_2. \end{aligned}$$

この式と (3.21) を合わせれば,

$$|\mathbf{X}\widehat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^*|_2 \leq |\mathbf{X}\tilde{\beta}^{\mathcal{L}} - \mathbf{X}\beta^*|_2 + n\tau |\widehat{\beta}^{\mathcal{L}} - \beta^*|_1 \leq 4n\tau |\widehat{\beta}_S^{\mathcal{L}} - \beta_S^*|_1 \leq \tau\sqrt{32kn} |\mathbf{X}\widehat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^*|_2$$

を得る. これより,

$$\begin{aligned} |\mathbf{X}\widehat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^*|_2 &\leq 32nk\tau^2 = 8nk(2\tau)^2 = 8nk \left\{ 8\sigma\sqrt{\frac{\log(2d)}{n}} + 8\sigma\sqrt{\frac{\log(\frac{1}{\delta})}{n}} \right\}^2 \\ &\leq \frac{2^{11}k\sigma^2}{n} \log\left(\frac{2d}{\delta}\right). \end{aligned}$$

よって,

$$\begin{aligned} \mathbb{P}\left(\text{MSE}(\mathbf{X}\widehat{\beta}^{\mathcal{L}}) = \frac{1}{n} |\mathbf{X}\widehat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^*|_2^2 \leq 2^{11} \cdot \frac{k\sigma^2}{n} \log\left(\frac{2d}{\delta}\right)\right) \\ \geq \mathbb{P}\left(|\mathbf{X}\widehat{\beta}^{\mathcal{L}}|_{\infty} \leq \frac{n\tau}{2}\right) > 1 - \delta \end{aligned}$$

を得る. 再度, 補題 3.24 を用いれば,

$$|\widehat{\beta}^{\mathcal{L}} - \beta^*|_2^2 \leq \frac{2}{n} |\mathbf{X}\widehat{\beta}^{\mathcal{L}} - \mathbf{X}\beta^*|_2^2 = 2\text{MSE}(\mathbf{X}\widehat{\beta}^{\mathcal{L}})$$

より,

$$\begin{aligned} \mathbb{P} \left(|\widehat{\beta}^{\mathcal{L}} - \beta^*|_2^2 \leq 2^{10} \cdot \frac{k\sigma^2}{n} \log \left(\frac{2d}{\delta} \right) \right) \\ \geq \mathbb{P} \left(\text{MSE}(\mathbf{X}\widehat{\beta}^{\mathcal{L}}) \leq 2^{11} \cdot \frac{k\sigma^2}{n} \log \left(\frac{2d}{\delta} \right) \right) \\ > 1 - \delta \end{aligned}$$

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

$$\inf_{\#(S) \leq k} \inf_{\beta \in C_S} \frac{|\mathbf{X}\beta^*|_2^2}{n|\beta^*|_2^2} \geq \frac{1}{2} \quad (3.22)$$

である. ただし,

$$C_S := \{\beta^* \in \mathbb{R}^d; |\beta_{S^c}^*|_1 \leq 3|\beta_S^*|_1\}$$

である. 条件 (3.22) を制約固有値条件 (restricted eigenvalue (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 β^* は $\#(S) \leq k$ とした錐 C_S に含まれる. したがって, *rm RE* 条件は X_S の最小固有値が $\lambda_{\min}(X_S) \geq \frac{n}{2}$ ($\forall S \subset \{1, 2, \dots, d\}$ で $\#(S) \leq k$) 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 X が Rademacher 確率変数列ならば, $n \geq Ck \log d$ (C は正の定数) ならば, 高い確率で, RE 条件をみたす. □

4.4 SLOPE 推定量

BIC 推定量は, $k = |\beta^*|_0$ 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 $\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_d)$ を非増加の正の数数列とする. すなわち,

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

$\beta^* = (\beta_1, \beta_2, \dots, \beta_d)^\top \in \mathbb{R}^d$ 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. β^* の並び替え ℓ_1 ノルム (sorted ℓ_1 norm) を次で定義する.

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

または,

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

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

$$\hat{\beta}^S \in \operatorname{argmin}_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{n} \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + 2\tau |\beta^*|_* \right\} \quad (3.23)$$

で定義する.

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

以下では

$$\lambda_j = \sqrt{\log \left(\frac{2d}{j} \right)} \quad (j = 1, 2, \dots, d) \quad (3.24)$$

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数列とする. このとき, $k \leq d$ と $\forall t > 0$ に対して,

$$\mathbb{P} \left(\frac{1}{k\sigma^2} \sum_{j=1}^k (g_j^*)^2 > t \log \left(\frac{2d}{k} \right) \right) \leq \left(\frac{2d}{k} \right)^{1 - \frac{3t}{8}} \quad (3.25)$$

である.

証明 Jensen の不等式を用いる.

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

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

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

よりわかる. 次に, Chernoff 限界を用いる.

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

□

補題 3.29 $[d] := \{1, 2, \dots, d\}$ 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,

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

が成り立つ.

証明 まず, $g_1^* \geq g_2^* \geq \dots \geq g_k^*$ なので,

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

補題 3.28 を用いると

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

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

に注意すれば, $t > 8$ に対して,

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

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

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

に注意する. よって,

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

がわかる. □

定理 3.30 $n \geq 2$ とする. 線型回帰モデル (3.1) と $\epsilon \sim N_n(\mathbf{0}, \sigma a^2 I_n)$ ($\sigma > 0$) を仮定する. さらに, $|\beta^*|_0 \leq k$ とし,

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

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

$$\tau := 8\sqrt{2}\sigma\sqrt{\frac{\log(1/\delta)}{n}} \quad (3.26)$$

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

$$\mathbb{P}\left(\text{MSE}(X\widehat{\beta}^S) = \frac{1}{n}|\mathbf{X}\widehat{\beta}^S - \mathbf{X}\beta^*|_2^2 \lesssim \frac{k\sigma^2}{n} \log\left(\frac{2d}{k}\right) \log\left(\frac{1}{\delta}\right)\right) > 1 - \delta \quad (3.27)$$

と

$$\mathbb{P}\left(|\widehat{\beta}^S - \beta^*|_2^2 \lesssim \frac{k\sigma^2}{n} \log\left(\frac{2d}{k}\right) \log\left(\frac{1}{\delta}\right)\right) > 1 - \delta. \quad (3.28)$$

証明 SLOPE 推定量 $\widehat{\beta}^S$ の定義より、

$$|\mathbf{Y} - \mathbf{X}\widehat{\beta}^S|_2^2 + |\widehat{\beta}^S|_1 \leq |\mathbf{Y} - \mathbf{X}\beta^*|_2^2 + |\beta^*|_1$$

である。さらに、

$$|\mathbf{X}\widehat{\beta}^S - \mathbf{X}\beta^*|_2^2 = |\mathbf{Y} - \mathbf{X}\widehat{\beta}^S|_2^2 + |\mathbf{Y} - \mathbf{X}\beta^*|_2^2 - 2\epsilon^\top \mathbf{X}(\widehat{\beta}^S - \beta^*)$$

である。これらより、

$$\begin{aligned} |\mathbf{X}\widehat{\beta}^S - \mathbf{X}\beta^*|_2^2 + n\tau|\widehat{\beta}^S - \beta^*|_* &\leq 2\epsilon^\top \mathbf{X}(\widehat{\beta}^S - \beta^*) + n\tau|\widehat{\beta}^S - \beta^*|_* \\ &\quad + 2n\tau|\beta^*|_* - 2n\tau|\widehat{\beta}^S|_* \end{aligned} \quad (3.29)$$

となる。

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

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

$$\begin{aligned} \epsilon^\top \mathbf{X}\mathbf{u} &= \sum_{j=1}^d (\mathbf{X}^\top \epsilon)_j u_j \leq \sum_{j=1}^d g_j^* u_j^* = \sup_{j \in [d]} \frac{g_j^*}{\lambda_j} (\lambda_j u_j^*) \leq \sum_{j \in [k]} \left\{ \frac{g_j^*}{\lambda_j} \right\} |\mathbf{u}|_* \\ &\leq 4\sqrt{2}\sigma\sqrt{n \log\left(\frac{1}{\delta}\right)} |\mathbf{u}|_*. \end{aligned} \quad (3.30)$$

なぜならば,

$$|X_j|_2^2 \leq 2n$$

なので,

$$g_j = X_j^\top \epsilon \sim N(0, |X_j|_2^2 \sigma^2)$$

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,

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

かつ

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

とできる. $\{\lambda_j\}$ は非増加列だったので,

$$\begin{aligned} |\beta^*|_* - |\widehat{\beta}^S|_* &= \sum_{j=1}^k \lambda_j |\beta_{\underline{\sigma}(j)}| - \sum_{j=1}^d \lambda_j (\widehat{\beta}^S)_j^* \\ &\leq \sum_{j=1}^k \lambda_j \{|\beta_{\underline{\sigma}(j)}| - |(\widehat{\beta}^S)_{\underline{\sigma}(j)}|\} - \sum_{j=k+1}^d \lambda_j |(\widehat{\beta}^S)_{\underline{\sigma}(j)}| \\ &\leq \sum_{j=1}^k \lambda_j |u_{\underline{\sigma}(j)}| - \sum_{j=k+1}^d \lambda_j u_j^* \end{aligned}$$

($\because j \geq k$ のとき, $\beta_j = 0$)

$$\leq \sum_{j=1}^k \lambda_j u_j^* - \sum_{j=k+1}^d \lambda_j u_j^*. \quad (3.31)$$

$\tau = 8\sqrt{2}\sigma \sqrt{\frac{\log(\frac{1}{\delta})}{n}}$, (3.29) – (3.32) より

$$\begin{aligned} |\mathbf{X}\widehat{\beta}^S - \mathbf{X}\beta^*|_2^2 + n\tau|\mathbf{u}|_* &\leq 2n\tau|\mathbf{u}|_* + 2n\tau|\beta^*|_* - 2n\tau|\widehat{\beta}^S|_* \\ &\leq 2n\tau|\mathbf{u}|_* + 2n\tau \sum_{j=1}^k \lambda_j u_j^* - 2n\tau \sum_{j=1}^d \lambda_j u_j^* \\ &\leq 4n\tau \sum_{j=1}^k \lambda_j u_j^*. \quad (3.32) \end{aligned}$$

したがって,

$$\|\mathbf{X}\widehat{\boldsymbol{\beta}}^S - \mathbf{X}\boldsymbol{\beta}^*\|_2^2 \leq 4n\tau \sum_{j=1}^k \lambda_j u_j^* - n\tau \|\mathbf{u}\|_* \leq 3n\tau \sum_{j=1}^k \lambda_j u_j - \sum_{j=k+1}^d \lambda_j u_j$$

より,

$$\sum_{j=k+1}^d \lambda_j u_j \leq 3 \sum_{j=1}^k \lambda_j u_j. \quad (3.33)$$

$k' \geq 4k \log(2de/k)$ に対して, $\text{INC}(k')$ 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, $\mathbf{u} = \widehat{\boldsymbol{\beta}}^S - \boldsymbol{\beta}^*$ に対して,

$$\frac{\|\mathbf{X}\mathbf{u}\|_2^2}{n} \geq \|\boldsymbol{\beta}_S^*\|_2^2 + \|\mathbf{u}_{S^c}\|_2^2 - \frac{36 + 12 + 1}{128} \|\mathbf{u}_S\|_2^2 \geq \frac{1}{2} \|\mathbf{u}\|_2^2 \quad (3.34)$$

がわかる. さらに, Stirling の不等式

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

を用いると

$$\begin{aligned} \sum_{j=1}^k \lambda_j &= \sum_{j=1}^k \log\left(\frac{2d}{j}\right) = \sum_{j=1}^k \{\log(2d) - \log(j)\} \\ &= \log(2d) - \log(1) + \log(2d) - \log(2) + \cdots \\ &\quad + \log(2d) - \log(k) = k \log(2d) - \log k! \\ &\leq k \log(2d) - k \log\left(\frac{k}{e}\right) = k \log\left(\frac{2de}{k}\right) \end{aligned} \quad (3.35)$$

²補遺を参照.

(3.32) と (3.33) を用いて, (3.31) を評価する.

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

を得る. これと (3.33) より

$$\begin{aligned}
\frac{1}{n}|\mathbf{X}\mathbf{u}|_2^2 &\leq 2^{12}\sigma^2 \frac{k}{n} \log\left(\frac{2de}{k}\right) \log\left(\frac{1}{\delta}\right), \\
|\mathbf{u}|_2^2 &\leq 2^{13}\sigma^2 \frac{k}{n} \log\left(\frac{2de}{k}\right) \log\left(\frac{1}{\delta}\right)
\end{aligned}$$

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,

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

となる. よって, $\forall \delta > 0$ に対して,

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}|\mathbf{X}\mathbf{u}|_2^2 \leq 2^{12}\sigma^2 \frac{k}{n} \log\left(\frac{2de}{k}\right) \log\left(\frac{1}{\delta}\right)\right) \\
\leq \mathbb{P}\left(\sup_{1 \leq j \leq d} \left\{ \frac{g_j^*}{\lambda_j} \right\} \leq 4\sqrt{\log\left(\frac{1}{\delta}\right)}\right) > 1 - \delta
\end{aligned}$$

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,

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

なので,

$$\begin{aligned} & \mathbb{P}\left(|\mathbf{u}|_2^2 \leq 2^{13} \sigma^2 \frac{k}{n} \log\left(\frac{2de}{k}\right) \log\left(\frac{1}{\delta}\right)\right) \\ & \geq \mathbb{P}\left(\sup_{1 \leq j \leq d} \left\{ \frac{g_j^*}{\lambda_j} \right\} \leq 4 \sqrt{\log\left(\frac{1}{\delta}\right)}\right) > 1 - \delta \end{aligned}$$

がわかる. □

5 補遺

5.1 (3.34) の計算

まず,

$$\frac{|\mathbf{X}\mathbf{u}|_2^2}{n} = \frac{|\mathbf{X}\mathbf{u}_S|_2^2}{n} + \frac{|\mathbf{X}\mathbf{u}_{S^c}|_2^2}{n} + 2\mathbf{u}_S^\top \frac{\mathbf{X}^\top \mathbf{X}}{n} \mathbf{u}_{S^c} \quad (3.36)$$

に注意をする. (3.36) の右辺の各項を別々に評価している.

(1)

$$\begin{aligned} \frac{|\mathbf{X}\mathbf{u}_S|_2^2}{n} &= \mathbf{u}_S^\top \frac{\mathbf{X}^\top \mathbf{X}}{n} \mathbf{u}_S = |\mathbf{u}_S|_2^2 + \mathbf{u}_S^\top \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} - I_d \right) \mathbf{u}_S \geq |\mathbf{u}_S|_2^2 - \frac{|\mathbf{u}_S|_1^2}{32k'} \\ &\geq |\mathbf{u}_S|_2^2 - \frac{k}{32k'} |\mathbf{u}_S|_2^2. \end{aligned} \quad (3.37)$$

(2)

$$\begin{aligned}
2 \left| \mathbf{u}_S^\top \frac{\mathbf{X}^\top \mathbf{X}}{n} \mathbf{u}_{S^c} \right| &\leq \frac{2}{32k'} \sum_{j=1}^k u_j^* \sum_{j=k+1}^d u_j^* \\
&\leq \frac{2}{32k'} \sum_{j=1}^k u_j^* \sum_{j=k+1}^d \frac{\lambda_j}{\lambda_d} u_j^* \\
&\leq \frac{6}{32k' \lambda_d} \sum_{j=1}^k u_j^* \sum_{j=1}^k \lambda_j u_j^* \\
&\quad (\because (3.33) \text{ よリ}) \\
&\leq \frac{6\sqrt{k}}{32k' \lambda_d} \sqrt{\sum_{j=1}^k \lambda_j^2} \\
&\quad \left(\because \sum_{j=1}^k u_j^* \leq \sqrt{k} \sqrt{\sum_{j=1}^k (u_j^*)^2}; \sum_{j=1}^k \lambda_j u_j^* \leq \sqrt{\sum_{j=1}^k \lambda_j^2} \sqrt{\sum_{j=1}^k (u_j^*)^2} = \sqrt{\sum_{j=1}^k \lambda_j^2} \sqrt{\sum_{j=1}^k (u_j^*)^2} \right) \\
&\leq \frac{6k}{32k' \lambda_d} \sqrt{\log \left(\frac{2de}{k} \right)} |\mathbf{u}_S|_2^2. \\
&\quad (\because (3.35) \text{ よリ})
\end{aligned}$$

(3)

$$\begin{aligned}
\frac{|\mathbf{X}\mathbf{u}_{SC}|_2^2}{n} &= \mathbf{u}_{SC}^\top \frac{\mathbf{X}^\top \mathbf{X}}{n} \mathbf{u}_{SC} \geq |\mathbf{u}_{SC}|_2^2 - \frac{1}{32k'} |\mathbf{u}_{SC}|_1^2 \\
&= |\mathbf{u}_{SC}|_2^2 - \frac{1}{32k'} \left(\sum_{j=k+1}^d u_j^* \right)^2 \\
&\geq |\mathbf{u}_{SC}|_2^2 - \frac{1}{32k' \lambda_d^2} \left(\sum_{j=k+1}^d \lambda_j u_j^* \right)^2 \\
&\geq |\mathbf{u}_{SC}|_2^2 - \frac{9}{32k' \lambda_d^2} \left(\sum_{j=1}^k \lambda_j u_j^* \right)^2 \\
&\geq |\mathbf{u}_{SC}|_2^2 - \frac{9}{32k' \lambda_d^2} \sum_{j=1}^k \lambda_j^2 \sum_{j=1}^k (u_j^*)^2 \\
&\geq |\mathbf{u}_{SC}|_2^2 - \frac{9k}{32k' \lambda_d^2} \log \left(\frac{2de}{k} \right) |\mathbf{u}_S|_2^2 \tag{3.39}
\end{aligned}$$

$$\because (??) \text{ より} \tag{3.40}$$

(3.35) – – (??) を (3.36) に代入すれば,

$$\begin{aligned}
\frac{|\mathbf{X}\mathbf{u}|_2^2}{n} &\geq |\mathbf{u}_S|_2^2 - \frac{k}{32k'} |\mathbf{u}_S|_2^2 - \frac{6k}{32k' \lambda_d} \sqrt{\log \left(\frac{2de}{k} \right)} |\mathbf{u}_S|_2^2 + |\mathbf{u}_{SC}|_2^2 \\
&\quad - \frac{9k}{32k' \lambda_d^2} \log \left(\frac{2de}{k} \right) |\mathbf{u}_S|_2^2 \\
&\geq |\mathbf{u}_S|_2^2 - \frac{1}{128e} |\mathbf{u}_S|_2^2 - \frac{12}{128} |\mathbf{u}_S|_2^2 + |\mathbf{u}_{SC}|_2^2 \\
&\quad - \frac{36}{128} |\mathbf{u}_S|_2^2
\end{aligned}$$

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.

□