本日の講義内容

第4章 誤特定された線型モデル

4.1 オラクル不等式 (Oracle Inequalities)

4.1.1. LSEのOI's

4.1.2. BIC推定量のスパースOI's

4.1.3. LASSOのスパースOI's

4.1.4 Maureyの論法

$$\underline{Y} = \underline{X}\,\underline{\beta}^* + \underline{\varepsilon}$$
$$= f(\underline{X}) + \underline{\varepsilon}$$

fは線型と仮定して, 回帰係数の推定量の精度を評価した.

線型性の仮定が誤りであった場合を考える.

次のモデルを考える. $f : \mathbb{R}^d \longrightarrow \mathbb{R}$

$$Y_j = f(\underline{X}_j) + \varepsilon_j \quad (j = 1, 2, \ldots, n), \ \underline{X}_j \in \mathbb{R}^d$$

ただし. $\underline{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T \sim \mathrm{subG}_n(\sigma^2)$

$\underline{f} = (f(x_1), f(x_2), \dots, f(x_n))^T \in \mathbb{R}^n$

とおく

$\hat{\underline{f}} = (\hat{f}(\underline{x}_1), \dots, \hat{f}(x_n))^T$

f の推定量を $\hat{f}$ と書くと 固定デザインの

$\hat{\underline{f}} = \underline{X}\,\hat{\underline{\beta}}^{LS}$

場合

$$\mathrm{MSE}(\hat{\underline{f}}) = \frac{1}{n}\left|\underline{f} - \hat{\underline{f}}\right|_{2,n}^2$$

$$= \frac{1}{n}\sum_{j=1}^{n}\{f(\underline{x}_j) - \hat{f}(\underline{x}_j)\}^2$$

となる。真のモデルが線型性をみたしてないと？

$\hat{\beta}^{LS}$, $\hat{\beta}^{BIC}$, $\hat{\beta}^{L}$ の精度を評価する)

4.1 オラクルの不等式

モデル $Y_j = f(\underline{x}_j) + \varepsilon_j \ (j = 1, 2, \dots, n)$ において、

線型性を仮定することは、f の推定量として

$$\mathbb{R}^d \ni \underline{x} \longmapsto \underline{x}^T \beta \in \mathbb{R}$$

に制限して推定量を考えていることになる。

オラクル推定量 $\hat{f}$ は f に最も近い線型関数のことをいう。

一般に、関数（既定の関数）

$$\varphi_j : \mathbb{R}^d \longrightarrow \mathbb{R} \quad (j = 1, 2, \ldots, M)$$

の集まり（辞書＝dictionary）$\{\varphi_1, \ldots, \varphi_M\}$

を考える.

$\{\varphi_1, \ldots, \varphi_M\}$ の線型結合

$$\varphi_\beta = \sum_{j=1}^{M} \beta_j \varphi_j \quad (\beta_j \in \mathbb{R})$$

を用いて、f の推定を考える.

辞書 $\{\varphi_1, \dots, \varphi_M\}$ を用いても LSE は制約付き

LSEの考え方を変わらない

I. LSE

$$\underline{\hat{\beta}}^{LS} \in \arg\min_{\underline{\beta} \in \mathbb{R}^M} \frac{1}{n} \sum_{j=1}^{n} \left\{ Y_j - \varphi_{\underline{\beta}}(\underline{X}_j) \right\}^2$$

ただし $\displaystyle\sum_{j=1}^{n} \left\{ Y_j - \varphi_{\underline{\beta}}(\underline{X}_j) \right\}^2 = \sum_{j=1}^{n} \left\{ Y_j - \sum_{k=1}^{M} \beta_k \varphi_k(\underline{X}_j) \right\}^2$

$$= | \underline{Y} - \underline{\beta}^T \underline{\varphi} |_{2,n}^2$$

$\underline{Y} = (Y_1, \dots, Y_n)^T$, $\underline{\beta} = (\beta_1, \dots, \beta_M)^T$, $\underline{\varphi} = (\varphi(\underline{X}_1), \dots, \varphi(\underline{X}_n))^T$

2 $K \subset \mathbb{R}^m$ に制約して LSE

$$\hat{\beta}_K^{LS} \in \arg\min_{\beta \in K} \frac{1}{n}\sum_{j=1}^{n}\{Y_j - \varphi_\beta(\underline{X}_j)\}^2$$

3. BIC の推定量 $(\tau > 0)$

$$\hat{\beta}^{BIC} \in \arg\min_{\beta \in \mathbb{R}^m}\left[\frac{1}{n}\sum_{j=1}^{n}\{Y_j - \varphi_\beta(\underline{X}_j)\}^2 + \tau^2 |\beta|_{0,m}\right.$$

$$\frac{1}{n}\sum_{j=1}^{n}\{Y_j - \varphi_\beta(\underline{X}_j)\}^2 = \frac{1}{n}\,|\underline{Y} - \beta^T \underline{\Phi}|_{2,n}$$

$$|\beta|_{0,m} = \#\{j \in \{1,\dots,m\};\ \beta_j \neq 0\}.$$

4. LASSO 推定量 $(\tau > 0)$

$$\hat{\beta}^{L} \in \underset{\beta \in \mathbb{R}^m}{\arg\min} \left[\frac{1}{n} \sum_{j=1}^{n} \{ Y_j - \varphi_{\beta}(\underline{x}_j) \}^2 + 2\tau |\underline{\beta}|_{1,m} \right.$$

ただし、$|\beta|_{1,m} = \sum_{j=1}^{m} |\beta_j|$.

定義 4.2 $M \in \mathbb{N}$ とし、$R(\cdot)$ をリスク関数とする。
$\mathcal{H} = \{\varphi_1, \varphi_2, \ldots, \varphi_M\}$ を $\mathbb{R}^d$ 上で定義された $\mathbb{R}$ 値の辞書とする。$K \subset \mathbb{R}^M$ を部分集合とする。
リスク関数 R に関する K 上のオラクル $\varphi_{\bar{\beta}}$ ($\bar{\beta} \in \mathbb{R}^M$) を

$$R(\varphi_{\bar{\beta}}) \leq R(\varphi_{\beta}) \quad (\beta \in \mathbb{R}^M)$$

をみたすものと定義する。さらに

$$R_K := R(\varphi_{\bar{\beta}})$$

を K 上のオラクルリスクという。推定量 $\hat{f}$ は剰余項 ϕ の K 上のオラクル不等式をみたすとは、次のいずれかが成立することをいう

(1) ある定数 $C \geq 1$ が存在して

$$E[R(\hat{f})] \leq C \inf_{\beta \in K} R(\varphi_\beta) + \phi_{n,m}(K).$$

(2) $\forall \delta > 0$ に対して

$$\Pr\left(R(\hat{f}) \leq C \inf_{\beta \in K} R(\varphi_\beta) + \phi_{n,m}(K) \right) \geq 1 - \delta.$$

4.1.1. LSEのオラクル不等式

$n > M$ とする.

定理4.3 回帰モデル

$$Y_j = f(\underline{x}_j) + \varepsilon_j \quad (j = 1, \dots, n)$$

$$\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T \sim \mathrm{subG}_n(\sigma^2)$$

を考える. このとき, LSE $\hat{\beta}^{LS}$ は次をみたす.

ある $C \geq 1$ が存在して, ${}^\forall \delta > 0$ に対して

$$\Pr\Bigg(\mathrm{MSE}(\varphi_{\hat{\underline{\beta}}^{LS}}) = \frac{1}{n} \sum_{j=1}^{n} \{ \underline{x}_j^T \hat{\underline{\beta}}^{LS} - f(\underline{x}_j) \}^2$$

$$\le \inf_{\bar{\beta}\in K} \frac{1}{n}\sum_{j=1}^{n}\{\varphi_{\bar{\beta}}(x_j)-f(x_j)\}^2 + C\sigma^2\frac{M_\tau \log(\frac{1}{\delta})}{n}\Big) \ge 1-\delta.$$

証明　$\bar{\beta}\in K$ は

$$R(\varphi_{\bar{\beta}}) \le R(\varphi_{\beta}) \quad (\forall \beta\in K)$$

であった。一方，LSEの定め方より

$$\|\underline{Y}-\underline{\varphi}_{\hat{\beta}^{LS}}\|_{2,n}^2 \le \|\underline{Y}-\underline{\varphi}_{\bar{\beta}}\|_{2,n}^2$$

である。ただし

$$\underline{Y}=(Y_1,\dots,Y_n)^T;\ \underline{\varphi}_{\hat{\beta}^{LS}}=(\underline{x}_1^T\hat{\beta}^{LS},\dots,\underline{x}_n^T\hat{\beta}^{LS})^T$$

$$\underline{\varphi}_{\bar{\beta}} = (\varphi_{\bar{\beta}}(\underline{x}_1), \dots, \varphi_{\bar{\beta}}(\underline{x}_n))^T.$$

$$\sum_{j=1}^{n} \{ f(x_j) - \varphi_{\bar{\beta}}(x_j) \} \varphi_{\bar{\beta}}(x_j) = 0$$

(すなわち, $\varphi_{\bar{\beta}}$ は f の $\mathrm{span}\{\varphi_1, \dots, \varphi_m\}$ への直交射影)

$$\underline{Y} = \underline{f} + \underline{\varepsilon} \ ; \ \underline{f} = (f(\underline{x}_1), \dots, f(\underline{x}_n))^T$$

なので

$$\| \underline{f} - \underline{\varphi}_{\hat{\beta}^{LS}} \|_{2,n}^2 \qquad <$$

$$= \| \underline{Y} - \underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varepsilon} \|_{2,n}^2 .$$

$$= |\underline{Y} - \underline{\Psi}\,\hat{\beta}^{LS}|_{2,n}^2 - 2\,\underline{\varepsilon}^T(\underline{Y} - \underline{\Psi}\,\hat{\beta}^{LS}) + |\underline{\varepsilon}|_{2,n}^2$$

$$\leq |\underline{Y} - \underline{\Psi}\,\bar{\beta}|_{2,n}^2 - 2\,\underline{\varepsilon}^T(\underline{Y} - \underline{\Psi}\,\hat{\beta}^{LS}) + |\underline{\varepsilon}|_{2,n}^2$$

$$= |\underline{f} + \underline{\varepsilon} - \underline{\Psi}\,\bar{\beta}|_{2,n}^2 - 2\,\underline{\varepsilon}^T(\underline{Y} - \underline{\Psi}\,\hat{\beta}^{LS}) + |\underline{\varepsilon}|_{2,n}^2$$

$$= |\underline{f} - \underline{\Psi}\,\bar{\beta}|_{2,n}^2 + 2\,\underline{\varepsilon}^T(\underline{f} - \underline{\Psi}\,\bar{\beta}) + |\underline{\varepsilon}|_{2,n}^2$$

$$- 2\,\underline{\varepsilon}^T(\underline{Y} - \underline{\Psi}\,\hat{\beta}^{LS}) + |\underline{\varepsilon}|_{2,n}^2$$

$$= |\underline{f} - \underline{\Psi}\,\bar{\beta}|_{2,n}^2 + 2\,\underline{\varepsilon}^T\{\cancel{\underline{f}} - \underline{\Psi}\,\bar{\beta} + \cancel{\underline{\varepsilon}} - \cancel{\underline{Y}} + \underline{\Psi}\,\hat{\beta}^{LS}\}$$

$$= |\underline{f} - \underline{\Psi}\,\bar{\beta}|_{2,n}^2 + 2\,\underline{\varepsilon}^T(\underline{\Psi}\,\hat{\beta}^{LS} - \underline{\Psi}\,\bar{\beta}).$$

よって

$$\|\underline{f} - \underline{\varphi}_{\hat{\beta}^{LS}}\|_{2,n}^2 \leq \|\underline{f} - \underline{\varphi}_{\bar{\beta}}\|_{2,n}^2 + 2\underline{\varepsilon}^T(\underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varphi}_{\bar{\beta}}) \quad (*)$$

を得る. さらに, Pythagorasの定理より

$$\|\underline{f} - \underline{\varphi}_{\hat{\beta}^{LS}}\|_{2,n}^2 = \|\underline{f} - \underline{\varphi}_{\bar{\beta}} + \varphi_{\bar{\beta}} - \varphi_{\hat{\beta}^{LS}}\|_{2,n}^2$$

$$= \|\underline{f} - \underline{\varphi}_{\bar{\beta}}\|_{2,n}^2 + 2(\underline{f} - \varphi_{\bar{\beta}})(\varphi_{\bar{\beta}} - \varphi_{\hat{\beta}^{LS}})$$

直交　　$\varphi_{\bar{\beta}} - \varphi_{\hat{\beta}^{LS}} \in \mathrm{span}\{\varphi_1, \ldots, \varphi_M\}$

$$+ \|\underline{\varphi}_{\bar{\beta}} - \varphi_{\hat{\beta}^{LS}}\|_{2,n}^2$$

$$= \|\underline{f} - \underline{\varphi}_{\bar{\beta}}\|_{2,n}^2 + \|\underline{\varphi}_{\bar{\beta}} - \underline{\varphi}_{\hat{\beta}^{LS}}\|_{2,n}^2$$

よって

$$\|\underline{f} - \underline{\varphi}_{\hat{\beta}^{LS}}\|_{2,n}^2 - \|\underline{f} - \underline{\varphi}_{\bar{\beta}}\|_{2,n}^2 = \|\underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varphi}_{\bar{\beta}}\|_{2,n}^2 \tag{4.5}$$

を得る. (※)と(4.5)を合わせると

$$\|\underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varphi}_{\bar{\beta}}\|_{2,n}^2 \leq 2\underline{\varepsilon}^T(\underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varphi}_{\bar{\beta}})$$

$$= 2\|\underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varphi}_{\bar{\beta}}\|_{2,n} \cdot \frac{\underline{\varepsilon}^T(\underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varphi}_{\bar{\beta}})}{\|\underline{\varphi}_{\hat{\beta}^{LS}} - \underline{\varphi}_{\bar{\beta}}\|_{2,n}}$$

を得る.

いま, $\underline{\Phi} \in \mathbb{R}^{n\times M}$ を

$$\begin{bmatrix} \varphi_1(\underline{x}_1) \\ \varphi_1(\underline{x}_2) \\ \vdots \\ \varphi_1(\underline{x}_n) \end{bmatrix}, \begin{bmatrix} \varphi_2(\underline{x}_1) \\ \varphi_2(\underline{x}_2) \\ \vdots \\ \varphi_2(\underline{x}_n) \end{bmatrix}, \cdots, \begin{bmatrix} \varphi_M(\underline{x}_1) \\ \varphi_M(\underline{x}_2) \\ \vdots \\ \varphi_M(\underline{x}_n) \end{bmatrix}$$

の張る線型部分空間の正規直交基底とする.

$\underline{\nu} \in \mathbb{R}^M$ が存在して

$$\underline{\psi}_{\hat{\beta}^{LS}} - \underline{\psi}_{\hat{\beta}} = \underline{\Phi}\,\underline{\nu}$$

と書ける. よって

$$\frac{\underline{\varepsilon}^T\left(\underline{\varphi}_{\hat{\beta}^{LS}}-\underline{\varphi}_{\bar{\beta}}\right)}{\left|\underline{\varphi}_{\hat{\beta}^{LS}}-\underline{\varphi}_{\bar{\beta}}\right|_{2,n}}=\frac{\underline{\varepsilon}^T\underline{\underline{\Phi}}\,\underline{\nu}}{\left|\underline{\underline{\Phi}}\,\underline{\nu}\right|_{2,n}}$$

$$=\frac{\underline{\varepsilon}^T\underline{\underline{\Phi}}\,\underline{\nu}}{\left|\underline{\nu}\right|_{2,m}}=\frac{\underline{\tilde{\varepsilon}}^T\underline{\nu}}{\left|\underline{\nu}\right|_{2,m}}\quad\left(\underline{\tilde{\varepsilon}}=\underline{\underline{\Phi}}^T\underline{\varepsilon}\in\mathbb{R}^m\right)$$

$$\le\sup_{|\underline{\mu}|_{2,m}=1}\underline{\tilde{\varepsilon}}^T\underline{\mu}$$

となる.

1になって

$$\|\underline{\varphi}_{\hat{\underline{\beta}}^{LS}} - \underline{\varphi}_{\bar{\underline{\beta}}}\|_{2,n}^2 \le \underbrace{4 \sup_{|\mu|_{2,M}=1} (\tilde{\underline{\xi}}^T \mu)^2}_{(a)} \quad (4.6)$$

を得る。

次に，$\tilde{\underline{\xi}} \sim \mathrm{subG}_M(\sigma^2)$ を確認する。

そのために，$\underline{v} \in S^{M-1}$ を得る。

$$|\underline{\Phi}\,\underline{v}|_{2,n}^2 = \underline{v}^T \underbrace{\underline{\Phi}^T \underline{\Phi}}_{= I_M} \underline{v} = \underline{v}^T \underline{v} = 1$$

となる。

よって, $\forall \lambda \in \mathbb{R}$ に対して

$$\mathbb{E}\left[e^{\lambda \tilde{\boldsymbol{\varepsilon}}^T \boldsymbol{v}}\right] = \mathbb{E}\left[e^{\lambda (\boldsymbol{\Phi}^T \boldsymbol{\varepsilon})^T \boldsymbol{v}}\right] \quad = \boldsymbol{\varepsilon}^T \boldsymbol{\Phi} \boldsymbol{v}$$

$$= \mathbb{E}\left[\exp\left\{\lambda (\boldsymbol{\Phi} \boldsymbol{v})^T \boldsymbol{\varepsilon}\right\}\right]$$

$$\leq \exp\left\{\frac{\lambda^2}{2}\sigma^2\right\}. \quad (\because)\ \boldsymbol{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$$

よって, $\tilde{\boldsymbol{\varepsilon}} \sim \mathrm{subG}_n(\sigma^2)$ がわかった.

補題 2.10 と 2.11 を用いる.

補題 2.10 より $X \sim \mathrm{subG}(\sigma^2) \Rightarrow \Pr(X > t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$

補題2.4より $E[|X|^2] \le 2\sigma^2 \cdot 2 \cdot \Gamma(1) = 4\sigma^2$

よって

$$4E\left[\sup_{|\underline{\mu}|_{2,M}=1} (\underline{\tilde{\varepsilon}}^T \underline{\mu})^2\right]$$

$$= 4\sum_{j=1}^{M} E[\tilde{\varepsilon}_j^2] \le 16\sigma^2 M.$$

となる。ただし、$\underline{\tilde{\varepsilon}} = (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_M)^T$

である。

さらに, $\underline{W} \sim \mathrm{subG}_d(\sigma^2)$ のとき

$$\Pr\left(\sup_{|\underline{\theta}|_{2,d}=1} \underline{\theta}^\top \underline{W} > t \right) \leq 6^d \exp\left(-\frac{t^2}{8\sigma^2} \right)$$

より

$$\Pr\left(\sup_{|\underline{\mu}|_{2,n}=1} (\hat{\underline{\varepsilon}}^\top \underline{\mu})^2 > t^2 \right)$$

$$= \Pr\left(\sup_{|\underline{\mu}|_{2,n}=1} (\hat{\underline{\varepsilon}}^\top \underline{\mu}) > t \right)$$

$$\leq 6^M \exp\left(-\frac{t^2}{8\sigma^2}\right)$$

よって

$$1 - 6^M \exp\left(-\frac{t^2}{8\sigma^2}\right) \leq \Pr\left(\sup_{\|u\|_{2,M}=1} (\xi^T u)^2 \leq t^2\right)$$

すると

$$6^M \exp\left(-\frac{t^2}{8\sigma^2}\right) \leq \delta$$

$$\Longleftrightarrow t^2 \geq 8\sigma^2 M \log 6 + 8\sigma^2 \log\left(\frac{1}{\delta}\right)$$

$\sqrt{8\log 6} < 4$ なので.

$$t := 40\sqrt{M} + 2\sigma\sqrt{2\log\left(\frac{1}{\delta}\right)}$$

とおけば

$$t^2 \geq 8\overbrace{\sigma^2 \log}^{M} 6 + 8\sigma^2 \log\left(\frac{1}{\delta}\right)$$

となるので.

$$\Pr\left(\sup_{\|\mu\|_{2,m}=1} (\mu^T \xi)^2 > 8\sigma^2 \log(6) + 8\sigma^2 \log\left(\frac{1}{\delta}\right)\right)$$

$$\leq \Pr\left(\sup_{\|\underline{\mu}\|_{2,n}=1} \underline{\mu}^T \underline{\xi} > t \right)$$

$$\leq 6^M \cdot \exp\left(- \frac{t^2}{8\sigma^2} \right)$$

$$= \exp\left(- \frac{t^2}{8\sigma^2} + M \log 6 \right)$$

$$\leq \delta$$

$\delta > 2$

$$\Pr\left(\sup_{\|\underline{\mu}\|_{2,n}=1} (\underline{\mu}^T \underline{\xi})^2 \leq 8\sigma^2 M \log 6 + 8\sigma^2 \log\left(\frac{1}{\delta}\right) \right) \geq 1-\delta$$

a

$$\sup_{\|\mu\|_{2,n}=1} (\mu^T \tilde{\xi})^2 \geq \frac{1}{4} \|\Psi \hat{\beta}^{LS} - \Psi \bar{\beta}\|_{2,n}^2$$

$$= \frac{1}{4} \left\{ \|f - \Psi \hat{\beta}^{LS}\|_{2,n}^2 - \|f - \Psi \bar{\beta}\|_{2,n}^2 \right\}$$

より

$$\Pr\left(\|f - \Psi \hat{\beta}^{LS}\|_{2,n}^2 \leq \|f - \Psi \bar{\beta}\|_{2,n}^2 + 8Q\sigma^2 \left(M + \log\left(\frac{1}{\delta}\right)\right) \right) \geq 1 - \delta$$

No.27

よって

$$\Pr\left(\frac{1}{n} \left| \underline{f} - \underline{\varphi}\, \hat{\beta}^{LS} \right|_{2,n}^{2} \leq \frac{1}{n} \left| \underline{f} - \underline{\varphi}\, \beta \right|_{2,n}^{2} + 32\sigma^2 \frac{M + \log\left(\frac{1}{\delta}\right)}{n} \right) \geq 1 - \delta$$

$\square$

4.1.2 BIC 推定量のスパース・オラクル不等式

定理 4.4 $n > M$ とし、回帰モデル

$$Y_j = f(\underline{X}_j) + \varepsilon_j \ ; \quad (j = 1, 2, \ldots, n)$$

$$\underline{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n) \sim \mathrm{subG}_n(\sigma^2) \quad (\sigma > 0)$$

を仮定する。さらに $\mathcal{H} = \{\varphi_1, \ldots, \varphi_M\}$ を辞書とし、

$$\varphi_{\underline{\beta}}(\underline{x}) = \sum_{j=1}^{M} \beta_j \varphi_j(\underline{x})$$

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

とする。さらに、調整パラメータ

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

をもつ BIC 推定量を $\hat{\beta}^{BIC}$ とする. すなわち,

$$\hat{\beta}^{BIC} \in \underset{\beta\in\mathbb{R}^M}{\arg\min}\left\{\frac{1}{n}\sum_{j=1}^{n}\{Y_j - \varphi_\beta(\underline{X}_j)\}^2 + \tau^2|\beta|_0\right\}$$

である. このとき, ${}^\forall\delta>0$ と ある $C>0$ に対して,

$$\Pr\left(\mathrm{MSE}(\varphi_{\hat{\beta}^{BIC}}) \leq \inf_{\beta\in\mathbb{R}^M}\left\{3\,\mathrm{MSE}(\varphi_\beta) + \frac{C\sigma^2}{n}|\beta|_{0,n}\log(eM)\right\} + \frac{C\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\right) \geq 1-\delta.$$

証明 BIC推定量の定義より

$$\frac{1}{n}\left|\underline{Y}-\underline{\varphi}_{\hat{\beta}^{BIC}}\right|_{2,n}^{2}+\tau^2\left|\hat{\beta}^{BIC}\right|_{0,M}$$

$$\leq \frac{1}{n}\left|\underline{Y}-\underline{\varphi}_{\beta}\right|_{2,n}^{2}+\tau^2\left|\beta\right|_{0,M} \qquad (a)$$

となることに注意する。ただし

$$\underline{\varphi}_{\beta}=\left(\varphi_\beta(\underline{x}_1),\ \dots\ ,\ \varphi_\beta(\underline{x}_n)\ \right)$$

$$\varphi_\beta(\underline{x})=\sum_{j=1}^{M}\beta_j\varphi_j(\underline{x}),\quad \beta=(\beta_1,\dots\ ,\beta_M)^T$$

$$|\underline{Y} - \underline{\Psi}\hat{\beta}^{BIC}|_{2,n}^2 = |\underline{f} + \underline{\varepsilon} - \underline{\Psi}\hat{\beta}^{BIC}|_{2,n}^2$$

$$= |f - \underline{\Psi}\hat{\beta}^{BIC}|_{2,n}^2 + 2\underline{\varepsilon}^T(\underline{f} - \underline{\Psi}\hat{\beta}^{BIC}) + |\underline{f}|_{2,n}^2$$

$$|\underline{Y} - \underline{\Psi}\beta|_{2,n}^2 = |f - \underline{\Psi}\beta|_{2,n}^2 + 2\underline{\varepsilon}^T(f - \underline{\Psi}\beta) + |f|_{2,n}^2$$

を(a)に代入して整理をすると

$$|\underline{f} - \underline{\Psi}\hat{\beta}^{BIC}|_{2,n}^2 + n\tau^2|\hat{\beta}^{BIC}|_{0,n} \qquad (4.9)$$

$$\leq |\underline{f} - \underline{\Psi}\beta|_{2,n}^2 + 2\underline{\varepsilon}^T(\underline{\Psi}\hat{\beta}^{BIC} - \underline{\Psi}\beta) + n\tau^2|\beta|_{0,n}$$

を得る. $\hat{\beta}^{BIC} = \beta$ のときは定理の主張は自明なので,

$\hat{\beta}^{BIC} \neq \beta$ と仮定する. このとき, $\alpha > 0$ に対して

$$2\underline{\varepsilon}^T(\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta})$$

$$= 2\underline{\varepsilon}^T \left(\frac{\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}}{|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}|_{2,n}} \right) |\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}|_{2,n}$$

$$\leq \frac{2}{\alpha} \left[\underline{\varepsilon}^T \left(\frac{\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}}{|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}|_{2,n}} \right) \right] + \frac{\alpha}{2} |\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}|_{2,n} \tag{4.11}$$

となる.

$\because)\quad 2ab \le \frac{2}{\alpha}a^2 + \frac{\alpha}{2}b^2 \quad (a, b \in \mathbb{R},\ \alpha > 0)$

次に、$(a+b)^2 \le 2a^2 + 2b^2 \quad (a, b \in \mathbb{R})$ から

$$\frac{1}{4}\left|\underline{\varphi}_{\hat{\beta}^{DRC}} - \underline{\varphi}_{\beta}\right|_{2,n}^2 = \frac{1}{4}\left|(\underline{\varphi}_{\hat{\beta}^{DRC}} - \underline{f}) - (\underline{\varphi}_{\beta} - \underline{f})\right|_{2,n}^2$$

$$\le \frac{1}{2}\left|\underline{\varphi}_{\hat{\beta}^{DRC}} - \underline{\varphi}_{\beta}\right|_{2,n}^2 + \frac{1}{2}\left|\underline{\varphi}_{\beta} - f\right|_{2,n}^2 \qquad (4.12)$$

となる。(4.9) に (4.11), (4.12) を代入すると

$$\left|\underline{\varphi}_{\hat{\beta}^{DRC}} - \underline{f}\right|_{2,n}^2 \le \underline{\left|\underline{\varphi}_{\beta} - \underline{f}\right|_{2,n}^2 + n\tau^2|\beta|_{0,n} - n\tau^2|\beta^{DRC}|_{0,n}}$$

$$+ 2\varepsilon^T\left(\frac{\underline{\varphi}_{\hat{\beta}^{DRC}} - \underline{\varphi}_{\beta}}{\left|\underline{\varphi}_{\hat{\beta}^{DRC}} - \underline{\varphi}_{\beta}\right|_{2,n}}\right)\left|\underline{\varphi}_{\hat{\beta}^{DRC}} - \underline{\varphi}_{\beta}\right|_{2,n}$$

$$\leq \qquad + 4\left[\underline{\varepsilon}^T\left(\frac{\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_\beta}{|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_\beta|_{2,n}}\right)\right]^2 + \frac{1}{4}|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_\beta|_{2,n} \geq \mathrm{Noise}$$

$$\leq \qquad + \qquad + \frac{1}{2}|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{f}|_{2,n}^2 + \frac{1}{2}|\underline{\varphi}_\beta - \underline{f}|_{2,n}^2$$

$$= \frac{3}{2}|\underline{\varphi}_\beta - \underline{f}|_{2,n}^2 + n\tau^2|\beta|_{0,n} - n\tau^2|\hat{\beta}^{BIC}|_{0,n}$$

$$+ 4\left[\underline{\varepsilon}^T\frac{(\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_\beta)}{|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_\beta|_{2,n}}\right]^2 + \frac{1}{2}|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{f}|_{2,n}^2$$

となる。これを整理すると

$$|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{f}|_{2,n}^2 \le 3|\underline{\varphi}_{\beta} - f|_{2,n}^2 + 2n\tau^2|\beta|_{0,M} - 2n\tau^2|\hat{\beta}^{BIC}|_{0,M}$$

$$+ \delta\left[\underline{\varepsilon}^T \mathcal{U}(\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta})\right]^2 \qquad (4.17)$$

を得る. ただし

$$\mathcal{U}(\underline{z}) = \frac{\underline{z}}{|\underline{z}|_{2,n}} \qquad (\underline{z} \in \mathbb{R}^n)$$

である. ここで, $S \subset \{1, 2, \ldots, M\}$ に対して

$$\underline{\Phi}_{r_{S,*}} = [\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_{r_{S,*}}]$$

を $\{\underline{\varphi}_j \,;\ j \in S \cup \mathrm{supp}(\beta)\}$ の張る

線型部分空間の正規直交基底とする S.

ただし

$$\underline{\varphi}_j = (\varphi_j(\underline{x}_1), \varphi_j(\underline{x}_2), \dots, \varphi_j(\underline{x}_n))^T$$

$$M = \{\varphi_1, \varphi_2, \dots, \varphi_M\} \text{ は辞書}$$

である. すると

$$r_{S,*} \leq \#(S) + |\beta|_{0,M}$$

である.

このとき

$$4\left[\underline{\varepsilon}^{T}\,\mathcal{U}\left(\underline{\psi}_{\hat{\beta}^{BIC}}-\underline{\psi}_{\beta}\right)\right]^{2}-n\tau^{2}\left|\hat{\beta}^{BIC}\right|_{0,M}$$

$$=\max_{1\le k\le M}\left\{\max_{\#(S)=k}\left\{\sup_{\mathrm{supp}(\tilde{\beta})=S}\left\{4\left[\underline{\varepsilon}^{T}\,\mathcal{U}\left(\underline{\psi}_{\tilde{\beta}}-\underline{\psi}_{\beta}\right)\right]^{2}-n\tau^{2}k\right\}\right\}\right\}$$

と書き直せる。よって、$\forall t>0$ に対して

$$\boxed{\Pr\left(\underline{\qquad\qquad\qquad}>\frac{t}{2}\right)}\longrightarrow \text{スライド No.44で利用}$$

$$\le\sum_{k=1}^{M}\sum_{\#(S)=k}\Pr\left(\sup_{\underline{u}\in B_{r_{S,n}}}\left[\underline{\varepsilon}^{T}\,\Phi_{r_{S,n}}\underline{u}\right]^{2}>\frac{t}{8}+\frac{n}{4}\tau^{2}k\right)$$

となる. ただし. $B_{r_{s,*}} = \{ \underline{u} \in \mathbb{R}^{r_{s,*}} ; \ |\underline{u}|_{2, r_{s,*}} = 1 \}$.

N を $B_{r_{s,*}}$ の $(1/2)$ 網とする. すなわち,

$N = \{ \underline{u}_1, \dots, \underline{u}_{\#(N)} \}$ で $^\forall \underline{u} \in B_{r_{s,*}}$ に対して

$$\inf_{\underline{v} \in N} |\underline{v} - \underline{u}|_{2, r_{s,*}} \leq \frac{1}{2}$$

となる $\#(N)$ が最小の集合. すると, 補題 2.29 より

$$\#(N) \leq 6^{r_{s,*}}$$

No.38

となる．$\forall \underline{u} \in B_{r_{s,*}}$ に対して，$\exists \underline{v} \in N$ と $\underline{x} \in \mathbb{R}^{r_{s,*}}$

があって

$$\underline{u} = \underline{v} + \underline{x} , \quad \|\underline{x}\|_{2,r_{s,*}} \leq \frac{1}{2}$$

と書ける．したがって

$$\max_{\underline{u} \in B_{r_{s,*}}} \underline{u}^T \underline{\Phi}_{r_{s,*}}^T \underline{\varepsilon} = \max_{\underline{v} \in N} \underline{v}^T \underline{\Phi}_{r_{s,*}}^T \underline{\varepsilon}$$

$$+ \max_{\underline{x} \in \frac{1}{2} B_{r_{s,*}}} \underline{x}^T \underline{\Phi}_{r_{s,*}}^T \underline{\varepsilon}$$

しかし

$$\max_{\underline{x} \in \frac{1}{2} B_{r_{s,r}}} \underline{x}^T \bar{\underline{\Phi}}_{r_{s,r}}^T \underline{\varepsilon} = \frac{1}{2} \max_{\underline{u} \in B_{r_{s,r}}} \underline{u}^T \bar{\underline{\Phi}}_{r_{s,r}}^T \underline{\varepsilon}$$

すので

$$\max_{\underline{u} \in B_{r_{s,r}}} \underline{u}^T \bar{\underline{\Phi}}_{r_{s,r}}^T \underline{\varepsilon} \leq 2 \max_{\underline{v} \in N} \underline{v}^T \bar{\underline{\Phi}}_{r_{s,r}} \underline{\varepsilon}$$

となる. よって

$$\Pr\left(\sup_{\underline{u} \in B_{r_{s,r}}} \left[\underline{\varepsilon}^T \bar{\Phi}_{r_{s,r}} \underline{u} \right]^2 > \frac{t}{8} + \frac{n}{4} \tau^2 P_k \right)$$

$$\leq \Pr\left(\sup_{\underline{v} \in N} \left[\underline{\varepsilon}^T \bar{\Phi}_{r_{s,r}} \underline{v} \right]^2 > \frac{t}{32} + \frac{n}{16} \tau^2 P_k \right)$$

補題 2.30 より, $\Phi_{r_{s,*}}^{T} \boldsymbol{\varepsilon}^{T} \sim \mathrm{subG}_{r_{s,*}}(\sigma^2)$

$$\Pr\left(\max_{\boldsymbol{v} \in N} |\boldsymbol{v}^T \tilde{\boldsymbol{\varepsilon}}| > \frac{t}{2} \right) \le 2\,\#(N) \exp\left(- \frac{t^2}{8\sigma^2} \right)$$

$$\le 2 \cdot 6^{r_{s,*}} \exp\left(- \frac{t^2}{8\sigma^2} \right)$$

$t = \sqrt{\frac{t}{8} + \frac{n}{4} \tau^2 R}$ とおけばよい.

$$\le 2 \cdot 6^{r_{s,*}} \exp\left(- \frac{\frac{t}{8} + \frac{n}{4} \tau^2 R}{8\sigma^2} \right)$$

$$\le 2 \exp\left(- \frac{t}{64\sigma^2} - \frac{n\tau^2 R}{32\sigma^2} + (R + |\beta|_{0,H}) \log 6 \right)$$

No. 42

$$\because)\quad r_{S,*} \leq \#(S) + |\beta|_{0,m}$$

$$= 2\exp\left(-\frac{t}{64\sigma^2} - \frac{nk}{32\sigma^2}\left(32\log(6)\frac{\sigma^2}{n} + 32\log(eM)\frac{\sigma^2}{n}\right) + (k + |\beta|_{0,m})\log 6 \right)$$

$$= 2\exp\left(-\frac{t}{64\sigma^2} - \cancel{k\log(6)} - k\log(eM) + (\cancel{k} + |\beta|_{0,m})\log(6) \right)$$

$$= \exp\left(\log 2 - \frac{t}{64\sigma^2} - k\log(eM) + |\beta|_{0,m}\log(6) \right)$$

$$\boxed{\leq \exp\left(-\frac{t}{64\sigma^2} - k\log(eM) + |\beta|_{0,m}\log(12) \right)}$$

がわかる．(4、13)に注意すると

$$|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{f}|_{2,n}^2 > 3|\underline{\varphi}_{\beta} - \underline{f}|_{2,n}^2 + 2n\tau^2|\beta|_{0,n} + \tau$$

が言えれば

$$\cancel{3|\underline{\varphi}_{\beta} - \underline{f}|_{2,n}^2} + \cancel{2n\tau^2|\beta|_{0,n}} - 2n\tau^2|\hat{\beta}^{BIC}|_{0,n}$$

$$+ 8[\underline{\varepsilon}^T \mathcal{U}(\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta})]^2$$

$$\geq |\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}|_{2,n}^2 > \cancel{3|\underline{\varphi}_{\beta} - \underline{f}|_{2,n}^2} + \cancel{2n\tau^2|\beta|_{0,n}^2} + t$$

$$\Longrightarrow 4[\underline{\varepsilon}^T \mathcal{U}(\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta})]^2 - n\tau^2|\hat{\beta}^{BIC}|_{0,n} > \frac{t}{2}$$

となる. よって

$$\|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{f}\|_{2,n}^2 > 3\|\underline{\varphi}_{\beta} - \underline{f}\|_{2,n}^2 + 2n\tau^2 |\beta|_{0,M} + t$$

$$\Longrightarrow 4[\underline{\varepsilon}^T U(\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta})]^2 - n\tau^2 |\hat{\beta}^{BIC}|_{0,M} > \frac{t}{2}$$

となるので

$$\Pr\left(\|\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{f}\|_{2,n}^2 > 3\|\underline{\varphi}_{\beta} - \underline{f}\|_{2,n}^2 + 2n\tau^2 |\beta|_{0,M} + t\right)$$

$$\leq \boxed{\Pr\left(4[\underline{\varepsilon}^T U(\underline{\varphi}_{\hat{\beta}^{BIC}} - \underline{\varphi}_{\beta}) - n\tau^2 |\hat{\beta}^{BIC}|_{0,M} > \frac{t}{2}\right)}$$

$$\leq \sum_{k=1}^{M} \sum_{|S|=k} \exp\left(-\frac{t}{64\sigma^2} - k\log(eM) + |\beta|_{0,n}\log(12)\right)^{N_{w.ks}}$$

$$\leq \sum_{k=1}^{M} \left(\frac{eM}{k}\right)^{k}$$

(∵ 補題3.10の議論を参照)

$$\leq \exp\left(-\frac{t}{64\sigma^2} + |\beta|_{0,n}\log(12)\right)$$

$$\because) \quad \left(\frac{eM}{k}\right)^{k} = \exp\left(\log\left(\frac{eM}{k}\right)^{k}\right)$$

$$= \exp\{k\log(eM) - k\log k\}$$

$$\leq \exp\{k\log(eM)\}$$

を得る.

よって

$$\Pr\left(\|\underline{Y}_{\hat{\beta}^{\mathrm{BIC}}} - \underline{f}\|_{2,n}^2 > 3\|\underline{Y}_{\beta} - \underline{f}\|_{2,n}^2 + 2n\tau^2 |\beta|_{0,n} + t \right)$$

$$\leq \exp\left(\underline{-\frac{t}{64\sigma^2} + |\beta|_{0,n} \log(12)} \right)$$

を得る. ここで

$$t := 64\sigma^2 |\beta|_{0,n} \log(12) + 64\sigma^2 \log\left(\frac{1}{\delta}\right)$$

とおけば

$$\left(\underline{\qquad\qquad} = \exp\left(-\log\left(\frac{1}{\delta}\right)\right) = \delta \right.$$

$$\Pr\Big(\|\underline{\varphi}_{\hat{\beta}^{\mathrm{ac}}} - \underline{f}\|_{2,n}^2 > 3\|\underline{\varphi}_{\beta} - \underline{f}\|_{2,n}^2 + 2n\tau^2|\beta|_{0,n}$$

$$+ 64\sigma^2|\beta|_{0,n}\log(12) + 64\sigma^2\log\left(\frac{1}{\delta}\right)\Big)$$

$$= \Pr\Big(\|\underline{\varphi}_{\hat{\beta}^{\mathrm{ac}}} - \underline{f}\|_{2,n}^2 > 3\|\underline{\varphi}_{\beta} - \underline{f}\|_{2,n}^2$$

$$+ 2n|\beta|_{0,n}\left(32\log(12)\frac{\sigma^2}{n} + 32\frac{\sigma^2}{n}\log(eM)\right)$$

$$+ 64\sigma^2|\beta|_{0,n}\log(12) + 64\sigma^2\log\left(\frac{1}{\delta}\right)\Big)$$

$$= \Pr\Big(\|\underline{\varphi}_{\hat{\beta}^{\mathrm{ac}}} - \underline{f}\|_{2,n}^2 > 3\|\underline{\varphi}_{\beta} - \underline{f}\|_{2,n}^2 + 128\sigma^2|\beta|_{0,n}\log(12)$$

$$+ 64\sigma^2 |\beta|_{0,M} \log(eM) + 64\sigma^2 \log\left(\frac{1}{\delta}\right)\Big)$$

$$\leq \delta$$

$$\begin{array}{r} 128 \\ 64 \\ \hline 192 \end{array}$$

さらに、　　$\leq 192\sigma^2 |\beta|_{0,M} \log(eM)$

に注意すれば

$$\Pr\Big(\|\Psi_{\hat\beta^{BIC}} - f\|_{2,n}^2 > 3\|\Psi_\beta - f\|_{2,n}^2 + 192\sigma^2 |\beta|_{0,M} \log(eM)$$

$$+ 64 \log\left(\frac{1}{\delta}\right)\Big)$$

$$\leq \Pr\Big(\|\Psi_{\hat\beta^{BIC}} - f\|_{2,n}^2 > 3\|\Psi_\beta - f\|_{2,n}^2 + \qquad + 64\sigma^2 \log\left(\frac{1}{\delta}\right)\Big)$$

$\le \delta$

を得る. よって

$$\Pr\left(\frac{1}{n}\left|\underline{\varphi}_{\hat{\beta}^{\mathrm{BIC}}}-\underline{f}\right|_{2,n}^{2} \le \frac{3}{n}\left|\underline{\varphi}_{\beta}-\underline{f}\right|_{2,n}^{2}+192\sigma^{2}|\beta|_{0,n}\log(eM)+64\sigma^{2}\log\left(\frac{1}{\delta}\right)\right)$$

$$\ge 1-\delta$$

を得る すなわち

$$\Pr\left(\mathrm{MSE}(\underline{\varphi}_{\hat{\beta}^{\mathrm{BIC}}}) \le 3\,\mathrm{MSE}(\underline{\varphi}_{\beta})+192\frac{\sigma^{2}}{n}\log(eM)\right)$$

$$+ 64\sigma^2 \log\left(\frac{1}{\delta}\right)\Bigg)\Bigg) \geq 1 - \delta$$

を得る.

$$\Pr\Bigg(\mathrm{MSE}(\mathbb{X}\hat{\beta}^{BIC}) \leq \inf_{\beta \in \mathbb{R}^M} \left\{ 3\,\mathrm{MSE}(\mathbb{X}\beta) + 192\frac{\sigma^2}{n}|\beta|_0 \, n^{\log(eM)} \right\}$$

$$+ 64\sigma^2 \log\left(\frac{1}{\delta}\right)\Bigg) \geq 1 - \delta$$

を得る.

$\square$

4.1.3 LASSOに対するスパースオラクル不等式

LASSO推定量に対するオラクル不等式を証明するためには,

incoherence条件のようなデザイン行列に対する条件が

必要となる. $n \times M$ のデザイン行列 $\underline{\Phi}$ の (i, j) 要素

を以下のように定める.

$$\Phi_{ij} = \varphi_j(X_i) \; ; \; i = 1, \ldots, n, \; j = 1, \ldots, M$$

$$\underline{\varphi}_j = (\varphi_j(X_1), \varphi_2(X_2), \ldots, \varphi_j(X_n))^T$$

$$\underline{\Phi} = (\underline{\varphi}_1, \underline{\varphi}_2, \ldots, \underline{\varphi}_M).$$

定理 4.6 回帰モデル

$$Y_j = f(X_j) + \varepsilon_j \; ; \; j = 1, \dots, n$$

$$\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T \sim \mathrm{subG}_n(\sigma^2), \quad \sigma > 0$$

を考える。さらに、ある正の整数 k があって、デザイン行列 $\underline{\Phi}$ は INC(k) 条件をみたすとする。

$$\left| \frac{\underline{\Phi}^T \underline{\Phi}}{n} - I_m \right|_\infty \leq \frac{1}{32k}$$

$$\Leftrightarrow \left| \frac{|\varphi_j|_{2,n}}{n} - 1 \right| \leq \frac{1}{32k} \; ; \; \left| \frac{\varphi_j^T \varphi_l}{n} \right| \leq \frac{1}{32k}.$$

・$\delta > 0$ に対して，調整パラメータ τ を

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

と定める．このとき

$$\hat{\beta}^{\lambda} \in \mathop{\arg\min}_{\beta\in\mathbb{R}^{n}} \left\{ \left\| \underline{Y} - \underline{\varphi}_{\hat{\beta}^{\lambda}} \right\|_{2,n}^{2} + 2\tau |\beta|_{1,M} \right\}$$

は以下をみたす．

$$\Pr\left(\mathrm{MSE}(\underline{\varphi}_{\hat{\beta}^{\lambda}}) \leq \inf_{\beta} \left\{ \mathrm{MSE}(\underline{\varphi}_{\beta}) + C\frac{\sigma^{2}}{n}|\beta|_{0,M}\log\left(\frac{2M}{\delta}\right) \right\} \right)$$

$$\geq 1 - \delta .$$

ただし

$$\mathrm{MSE}(\boldsymbol{\varphi}_{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{j=1}^{n} \left\{ Y_j - \varphi_{\boldsymbol{\beta}}(\boldsymbol{x}_j) \right\}^2$$

$$= \frac{1}{n} \sum_{j=1}^{n} \left\{ Y_j - \sum_{k=1}^{M} \beta_k \varphi_k(\boldsymbol{x}_j) \right\}^2$$

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_M)^T$$

$$\boldsymbol{\varphi}_j = (\varphi_1(\boldsymbol{x}_j), \varphi_2(\boldsymbol{x}_j), \dots, \varphi_M(\boldsymbol{x}_j))^T .$$

証明　LASSO $\hat{\beta}^L$ の定義より

$$\frac{1}{n}\left|\underline{Y} - \underline{\Psi}_{\hat{\beta}^L}\right|_{2,n}^2 \leq \frac{1}{n}\left|\underline{Y} - \underline{\Psi}_{\beta}\right|_{2,n}^2 + 2\tau|\beta|_{1,n} - 2\tau|\hat{\beta}^L|_{2,n} \tag{4.18}$$

が成立する。すると

$$\left|\underline{Y} - \underline{\Psi}_{\hat{\beta}^L}\right|_{2,n}^2 = \left|\underbrace{\underline{Y} - \underline{f}}_{=\underline{\varepsilon}}\right|_{2,n}^2 + \left|\underline{f} - \underline{\Psi}_{\hat{\beta}^L}\right|_{2,n}^2 - 2\underline{\varepsilon}^T\left(\underline{\Psi}_{\hat{\beta}^L} - \underline{f}\right)$$

$$\left|\underline{Y} - \underline{\Psi}_{\beta}\right|_{2,n}^2 = \left|\underline{Y} - \underline{f}\right|_{2,n}^2 + \left|\underline{f} - \underline{\Psi}_{\beta}\right|_{2,n}^2 - 2\underline{\varepsilon}^T\left(\underline{\Psi}_{\beta} - \underline{f}\right).$$

No.56

これらを(4.14)に代入して、両辺に $\tau|\hat{\beta}^\lambda-\beta|_{1,n}$ を加えて、n 倍すると

$$|\underline{\varphi}_{\hat{\beta}^\lambda}-\underline{f}|_{2,n}^2-|\underline{\varphi}_\beta-\underline{f}|_{2,n}^2+n\tau|\hat{\beta}^\lambda-\beta|_{1,n}$$

$$\leq 2\underline{\varepsilon}^T(\underline{\varphi}_{\hat{\beta}^\lambda}-\underline{\varphi}_\beta)+n\tau|\hat{\beta}^\lambda-\beta|_{1,n}$$

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

となる。INC(k)条件より

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

No.58

$$\Leftrightarrow n - \frac{n}{32k} \leq |\varphi_j|_{2,n}^2 \leq n + \frac{n}{32k}$$

$$\Rightarrow \quad |\varphi_j|_{2,n}^2 \leq 2n \leq 4n.$$

が成立する. よって, $\varphi_j^T \varepsilon \sim \mathrm{subG}(4n\sigma^2)$ となる.

次に, $\varphi_{\hat{\beta}^L} = \Phi^T \hat{\beta}^L$ に Hölder の不等式を

用いると

$$|\varepsilon^T(\varphi_{\hat{\beta}^L} - \varphi_\beta)| = |(\Phi^T \varepsilon)^T (\hat{\beta}^L - \beta)|$$

$$\leq |\Phi^T \varepsilon|_\infty \cdot |\hat{\beta}^L - \beta|_{1,n}$$

となる.

命題 2.24 $W_j \sim \mathrm{subG}(\sigma^2)$ $(j=1,2,\dots,M)$

$\forall \delta > 0$ に対して

$$\Pr\left(\max_{j=1,\dots,M}|W_j| > t\right) \le 2M\exp\left(-\frac{t^2}{2\sigma^2}\right)$$

これより　　　$\sigma^2 = 4n\sigma^2$

$$\Pr\left(\max_{j=1,\dots,M}|\underline{\psi}_j^T \underline{\varepsilon}| > t\right) \le 2M\exp\left(-\frac{t^2}{8n\sigma^2}\right)$$

を得る. よって

$$\Pr\left(2\underline{\varepsilon}^T(\underline{\psi}_{\hat{\beta}^\lambda} - \underline{\psi}_{\beta}) > \frac{n\tau}{2}|\hat{\beta}^\lambda - \beta|_{1,\psi}\right)$$

No. 59

$$\leq \Pr\left(2\left|\varepsilon^{T}(\Psi_{\hat{\beta}^{\lambda}} - \Psi_{\beta})\right| > \frac{n\tau}{2}\|\hat{\beta}^{\lambda} - \beta\|_{1,n}\right)$$

$$\leq \Pr\left(2|\Psi^{T}\varepsilon|_{\infty}\cdot\|\hat{\beta}^{\lambda} - \beta\|_{1,n} > \frac{n\tau}{2}\|\hat{\beta}^{\lambda} - \beta\|_{1,n}\right)$$

$$= \Pr\left(|\Psi^{T}\varepsilon|_{\infty} > \frac{n\tau}{4}\right)$$

$$= \Pr\left(\max_{1\leq j\leq M}|\Psi_j^{T}\varepsilon|_{\infty} > \frac{n\tau}{4}\right)$$

$$\leq 2M\exp\left(-\frac{n\tau^2}{64a^2}\right) \qquad (a)$$

を得る.

ここで, $S = \mathrm{supp}(\beta)$ とおくと

$$2\varepsilon^T(\Psi_{\hat{\beta}^\lambda} - \Psi_\beta) \le \frac{n\tau}{2}\|\hat{\beta}^\lambda - \beta\|_{1,n}$$

のとき, (4.15)の不等式の上限として以下のようなものを得る.

$$2\varepsilon^T(\Psi_{\hat{\beta}^\lambda} - \Psi_\beta) + n\tau\|\hat{\beta}^\lambda - \beta\|_{1,n} + 2n\tau\|\beta\|_{1,n} - 2n\tau\|\hat{\beta}^\lambda\|_{1,n}$$

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

$$= 2n\tau\|\hat{\beta}^\lambda_S - \beta_S\|_{1,n} + 2n\tau\|\cancel{\hat{\beta}^\lambda_{S^c}} - \underbrace{\beta_{S^c}}_{=0}\|_{1,n} + 2n\tau\|\beta_S\|_{1,n} + 2n\tau\|\underbrace{\beta_{S^c}}_{=0}\|_{1,n}$$

$$- 2n\tau\|\hat{\beta}^\lambda_S\| - 2n\|\cancel{\hat{\beta}_{S^c}}\|$$

$$= 2n\tau|\hat{\beta}_S^L - \beta_S|_{1,n} + 2n\tau|\beta_S|_{1,n} - 2n\tau|\hat{\beta}_S|_{1,n}$$ No.61

$$\leq 4n\tau|\hat{\beta}_S^L - \beta_S|_{1,n}$$

このことより

$$2\varepsilon^T(\varphi_{\hat{\beta}^L} - \varphi_\beta) \leq \frac{n\tau}{2}|\hat{\beta}^L - \beta|_{1,n}$$

$$\Rightarrow \text{(4.15)の右辺} \leq 4n\tau|\hat{\beta}^L - \beta|_{1,n}$$

$$\rightarrow |\varphi_{\hat{\beta}^L} - f|_{2,n}^2 - |\varphi_\beta - f|_{2,n}^2 + n\tau|\hat{\beta}^L - \beta|_{1,n}$$

$$\leq 4n\tau|\hat{\beta}_S^L - \beta_S|_{1,n}$$

がわかる.

(a) $\Pr\left(2\xi^{T}(\psi_{\hat{\beta}^{L}}-\psi_{\beta})>\frac{n\tau}{2}|\hat{\beta}^{L}-\beta|_{1,n}\right)\le 2M\exp\left(-\frac{n\tau^2}{8\,t\,\sigma^2}\right)$

No.62

に注意すると

$$\Pr\left(|\psi_{\hat{\beta}^{L}}-f|_{2,n}^{2}-|\psi_{\beta}-f|_{2,n}^{2}+n\tau|\hat{\beta}^{L}-\beta|_{1,n}\le 4n\tau|\hat{\beta}^{L}_{S}-\beta_{S}|_{1,n}\right)$$

$$\ge \Pr\left(2|\xi^{T}(\psi_{\hat{\beta}^{L}}-\psi_{\beta})|\le\frac{n\tau}{2}\right)$$

$$\ge \Pr\left(\max_{1\le j\le M}|\psi_{j}^{T}\xi|\le\frac{n\tau}{4}\right)$$

$$\ge 1-2M\exp\left(-\frac{n\tau^2}{6\,t\,\sigma^2}\right).$$

によって

$$\|\underline{\psi}_{\hat{\beta}^{\lambda}} - \underline{f}\|_{2,n}^2 - \|\underline{\psi}_{\beta} - \underline{f}\|_{2,n}^2 + n\tau|\hat{\beta}^{\lambda} - \beta|_{1,n} \leq 4n\tau|\hat{\beta}^{\lambda}_S - \beta_S|_{1,n} \tag{4.19}$$

が起きているとする。さらに

$$|\hat{\beta}^{\lambda}_{S^c} - \beta_{S^c}|_{1,n} \geq 3|\hat{\beta}^{\lambda}_S - \beta_S|_{1,n}$$

が起こると

$$4n\tau|\hat{\beta}^{\lambda}_S - \beta_S|_{1,n} - n\tau|\hat{\beta}^{\lambda} - \beta|_{1,n}$$

$$= 4n\tau|\hat{\beta}_S - \beta_S|_{1,n} - n\tau|\hat{\beta}_S - \beta_S|_{1,n} - n\tau|\hat{\beta}_{S^c} - \hat{\beta}_S|_{1,n}$$

$$\leq 3n\tau|\hat{\beta}_S^{\alpha}-\beta_S|_{1,n}-3|\hat{\beta}_S-\beta_S|_{1,n}=0$$

となるので

$$\mathrm{MSE}(\underline{y}_{\hat{\beta}})-\mathrm{MSE}(\underline{y}_{\beta})\leq 4n\tau|\hat{\beta}_S^{\alpha}-\beta_S|_{1,n}|\hat{\beta}^{\alpha}-\beta|_{1,n}$$
$$\leq 0.$$

以上から、(4.19) が起こること

$$\mathrm{MSE}(\underline{y}_{\hat{\beta}^{\alpha}})\leq \mathrm{MSE}(\underline{y}_{\beta})$$

または

$$|\hat{\beta}_{S^c}^{\alpha}-\beta_{S^c}|_{1,n}<3|\hat{\beta}_S^{\alpha}-\beta|_{1,n} \quad (4.21)$$

のいずれかが起こる.

(4.21) は 凸条件

$$|\tilde{\beta}_{S^c}|_{1,n} \le 3|\hat{\beta}^{\alpha}_S|_{1,n}$$

(b)

←スライドNo.66の下に記載する.

を $\tilde{\beta} = \hat{\beta}^{\alpha} - \beta$ に対してみたしている.

補題3.24　INC(R) 条件のもとで

$$(b) \Rightarrow |\tilde{\beta}|^2_{2,n} \le 2\,\frac{|\Psi^T\beta|^2_{2,n}}{n}$$

$$\left(\Psi^T(\hat{\beta}^{\alpha} - \beta) = \Psi\hat{\beta}^{\alpha} - \Psi_{\beta} \right)$$

したがって, (4.19)が起こると

(i) $\mathrm{MSE}(\varphi_{\hat\beta^L}) \leq \mathrm{MSE}(\varphi_{\beta})$

または

(ii) $|\hat\beta^L_{S^c} - \beta_{S^c}| < 3|\hat\beta^L_S - \beta_S|$

が起こる. よって, (i)の場合は

$$\mathrm{MSE}(\varphi_{\hat\beta^L}) \leq \inf\left\{\mathrm{MSE}(\varphi_{\beta}) + \frac{C\sigma^2}{n}|\beta|_0 \log\left(\frac{en}{\delta}\right)\right\}$$

が起ってことになるので

(ii) ⇒ となることを示せばよい.

以下では、(ii)(凸条件式之)を仮定して議論を進める.

さて

$$
\begin{aligned}
4n\tau \left| \hat{\beta}^{\alpha}_{S} - \beta_{S} \right|_{1,n}
&\leq 4n\tau \sqrt{\#(S)} \left| \hat{\beta}^{\alpha}_{S} - \beta_{S} \right|_{2,n} \\
&\qquad (\text{Cauchy-Schwarz}) \\
&\leq 4n\tau \sqrt{\#(S)} \left| \hat{\beta}^{\alpha} - \beta \right|_{2,n}
\end{aligned}
$$

∵ 項数をふやしたので

No.68

$$\leq 4n\tau\sqrt{\#(S)}\sqrt{\frac{2}{n}\|\Psi^{\tau}(\hat{\beta}^{s}-\beta)\|_{2,n}^{2}}$$

補題3.2より

$$\leq 4\tau\sqrt{2n\#(S)}\,\|\Psi_{\hat{\beta}^{s}}-\Psi_{\beta}\|_{2,n}$$

$$= 4\tau\sqrt{2n|\beta|_{0,n}}\,\|\Psi_{\hat{\beta}^{s}}-\Psi_{\beta}\|_{2,n}$$

$$= 2\underbrace{\left(2\tau\sqrt{2n|\beta|_{0,n}}\right)}_{=a}\underbrace{\|\Psi_{\hat{\beta}^{s}}-\Psi_{\beta}\|_{2,n}}_{=b}$$

$$\le \frac{2}{\alpha}\left(2\tau\sqrt{2n\,|\beta|_{0,n}}\right)^2 + \frac{\alpha}{2}\left|\underline{\Psi}\hat{\beta}^{s} - \underline{\Psi}\beta\right|_{2,n}^2$$

No. 69

$$2ab \le \frac{2}{\alpha}a^2 + \frac{\alpha}{2}b^2 \quad (a, b \in \mathbb{R},\ \alpha > 0)$$

$$= \frac{16\tau^2 n\,|\beta|_{0,n}}{\alpha} + \frac{\alpha}{2}\left|\underline{\Psi}\hat{\beta}^{s} - \underline{\Psi}\beta\right|_{2,n}^2$$

$$a = \underline{\Psi}\hat{\beta}^{s} - \underline{f} + \underline{f} - \underline{\Psi}\beta = b$$

$$\le \frac{16\tau^2 n\,|\beta|_{0,n}}{\alpha} + \alpha\left|\underline{\Psi}\hat{\beta}^{s} - \underline{f}\right|_{2,n}^2$$

$$+ \alpha\left|\underline{\Psi}\beta - \underline{f}\right|_{2,n}^2 \quad (※)$$

と得る.

$\because)\ |a+b| \le 2a^2 + 2b^2$

この式と(4.15), (4.19)を合わせると

$$|\underline{\Psi}_{\hat{\beta}^L} - \underline{f}|_{2,n}^2 + |\underline{\Psi}_{\beta} - \underline{f}|_{2,n}^2$$

$$\le |\underline{\Psi}_{\hat{\beta}^L} - \underline{f}|_{2,n}^2 + |\underline{\Psi}_{\beta} - \underline{f}|_{2,n}^2 + n\tau|\hat{\beta}^L - \beta|_{1,n}$$

$$\le 4n\tau|\hat{\beta}^L_S - \beta_S|_{1,n} \quad \text{by (4.15)}$$

$$\le \frac{16\tau^2 n|\beta|_{0,n}}{\alpha} + \alpha|\underline{\Psi}_{\hat{\beta}^L} - \underline{f}|_{2,n}^2 + \alpha|\underline{\Psi}_{\beta} - \underline{f}|_{2,n}^2 \quad \text{by (※)}$$

これより

$$(1-\alpha)\,\mathrm{MSE}(\underline{\varphi}_{\hat{\underline{\beta}}^S}) \leq (1+\alpha)\,\mathrm{MSE}(\underline{\varphi}_{\underline{\beta}}) + \frac{16\tau^2 n|\underline{\beta}|_{0,n}}{\alpha}$$

を得る．この両辺を$(1-\alpha)$でわり，$\alpha=\frac{1}{2}$とすると $\frac{1+\alpha}{1-\alpha}=3$

$$\mathrm{MSE}(\underline{\varphi}_{\hat{\underline{\beta}}^S}) \leq 3\,\mathrm{MSE}(\underline{\varphi}_{\underline{\beta}}) + 32\tau^2 n|\underline{\beta}|_{0,n}$$

以上の議論より

$$1 - 2M\exp\left(-\frac{n\tau^2}{64\sigma^2}\right)$$

$$\le \Pr\Big(\|\underline{\varphi}_{\hat{\beta}^\lambda} - \underline{f}\|_{2,n}^2 - \|\underline{\varphi}_{\beta} - \underline{f}\|_{2,n}^2 + n\tau\|\hat{\beta}^\lambda - \beta\|_{1,n}$$

$$\le 4n\tau\|\hat{\beta}^\lambda_S - \beta_S\|_{1,n}\Big)$$

(両辺をnでわる)

$$\le \Pr\left(\mathrm{MSE}(\underline{\varphi}_{\hat{\beta}^\lambda}) \le 3\,\mathrm{MSE}(\underline{\varphi}_{\beta}) + 32\tau^2|\beta|_{0,n}\right)$$

$$= \Pr\left(\mathrm{MSE}(\underline{\varphi}_{\hat{\beta}^\lambda}) \le 3\,\mathrm{MSE}(\varphi_{\beta}) + 128\frac{\sigma^2}{n}|\beta|_{0,n}\log\left(\frac{2M}{\delta}\right)\right)$$

$\because)\ 32\tau^2|\beta|_{0,n} = 8(2\tau)^2|\beta|_{0,n}$

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

$$\le 128\frac{\sigma^2}{n}\left\{2\log(2K) + 2\log\left(\frac{1}{\delta}\right)\right\}$$

$$= \cancel{128}\ 256\,\frac{\sigma^2}{n}\log\left(\frac{2K}{\delta}\right)$$

とする。 ーる

$$\frac{n\tau^2}{64\sigma^2} = \frac{n(2\tau)^2}{256\sigma^2} = \frac{n}{256\sigma^2}\left\{8\sigma\sqrt{\frac{2\log(2K)}{n}} + 8\sigma\sqrt{\frac{2\log(\frac{1}{\delta})}{n}}\right\}^2$$

$$\le \frac{n}{256\sigma^2}\left\{256\frac{\sigma^2}{n}\log(2K) + 256\frac{\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\right\}$$

No.27

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

$$8^2 \times 2 \times 2 = 256$$

$$= \log(2M) + \log\left(\frac{1}{\delta}\right)$$

とaて.

$$1 - 2M \exp\left(-\frac{n\varepsilon^2}{64\sigma^2}\right)$$

$$= 1 - \exp\left\{\log(2M) - \frac{n\varepsilon^2}{64\sigma^2}\right\}$$

$$\geq 1 - \exp\left\{\log\left(\frac{1}{\delta}\right)\right\} = 1 - \delta$$

No. 25

よって

$$\Pr\left(\mathrm{MSE}(\psi_{\hat{\beta}^s}) \geq \inf_{\beta} \left\{ \mathrm{MSE}(\psi_\beta) + C \frac{\sigma^2}{n} |\beta|_{0,n} \log \frac{2M}{\delta} \right\} \right)$$

$$\geq 1 - \delta$$