5月17日の回の講義内容 No.1

第3章 高次元線型回帰モデル

3.1 ランダムと固定デザインの線型回帰モデル

3.2 最小2乗推定量(LSE)

3.2.1 制約なしLSE

3.2.3. ℓ_0制約付きLSE

3.1 ランダムと固定デザインの線型回帰モデル

$\underline{x}_j \in \mathbb{R}^d$

3.1.1 ランダムデザインの場合

$n \in \mathbb{N}$ とし、データ

i.i.d.

観測　予測したい

$$\underbrace{(\underline{X}_1, Y_1), (\underline{X}_2, Y_2), \dots, (\underline{X}_n, Y_n)}_{\text{データ}}, \underbrace{(\underline{X}_{n+1}, Y_{n+1})}_{\text{将来のデータ}}$$

モデル

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

$\varepsilon_1, \dots, \varepsilon_{n+1}, \underline{X}_1, \dots, \underline{X}_{n+1}$ は互いに独立で

$E[\varepsilon_j] = 0 \ (j = 1, 2, \dots, n), \ E[\varepsilon_j^2] < \infty$

No.3

$\hat{f}_n$: $(\underline{X}_1, Y_1), \dots, (\underline{X}_n, Y_n)$ を基づく推定量

$\hat{f}_n(\underline{X}_{n+1})$ により Y_{n+1} をうまく予測していく

予測量の精度の尺度

$$R(\hat{f}_n) = E\left[\left\{Y_{n+1} - \hat{f}_n(\underline{X}_n)\right\}^2\right]$$

ただし、$E[\cdot]$ は $\{(Y_j, \underline{X}_j)\}_{j=1}^{n+1}$ の同時分布に関して取る。

$R(\hat{f}_n)$の書き直し

No.4

$$
\begin{aligned}
R(\hat{f}_n) &= E\left[\left\{Y_{n+1} - f(\underline{X}_{n+1}) + f(\underline{X}_{n+1}) - \hat{f}_n(\underline{X}_{n+1})\right\}^2\right] \\
&= E\left[\left\{Y_{n+1} - f(\underline{X}_{n+1})\right\}^2\right] \\
&\quad + E\left[\left\{f(\underline{X}_{n+1}) - \hat{f}_n(\underline{X}_{n+1})\right\}^2\right] \\
&\quad + 2E\left[\underbrace{\left\{Y_{n+1} - f(\underline{X}_{n+1})\right\}\left\{f(\underline{X}_{n+1}) - \hat{f}_n(\underline{X}_{n+1})\right\}}_{=:\,A}\right]
\end{aligned}
$$

No. 5

$$A = E\left[\left\{Y_{n+1} - f(\underline{X}_{n+1})\right\}\left\{f(\underline{X}_{n+1}) - \hat{f}_n(\underline{X}_{n+1})\right\}\right]$$

$$= E\left[\varepsilon_{n+1} \underline{\left\{ f(\underline{X}_{n+1}) - \hat{f}_n(\underline{X}_{n+1}) \right\}}\right]$$

は独立

$$= \underbrace{E[\varepsilon_{n+1}]}_{=0} \, E\left[f(\underline{X}_{n+1}) - \hat{f}_n(\underline{X}_{n+1})\right]$$

$$= 0$$

よって

$$R(\hat{f}_n) = E[\{Y_{n+1} - f(\underline{X}_{n+1})\}^2]$$

$$+ E[\{f(\underline{X}_{n+1}) - \hat{f}_n(\underline{X}_{n+1})\}^2]$$

$$= E[\varepsilon_{n+1}^2] + \|\hat{f}_n - f\|_{L^2(P_X)}$$

ただし

$$\|\hat{f}_n - f\|_{L^2(P_X)} = \int \{\hat{f}_n(\underline{x}) - f(\underline{x})\}^2 dP_{\underline{X}}(\underline{x})$$

また，$A \subset \mathbb{R}^d$ に対して

$$P_{\underline{x}}(A) = \Pr(\underline{X} \in A)$$

$\underline{X}$ は X_j の generic な変数

$E[\varepsilon_j^2] = \sigma^2$ のとき．

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

精度の尺度

3.1.2 固定デザインの場合

$X_1, X_2, \ldots, X_{n+1}$ はランダムではないと仮定

デザインを $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$

目標　観測

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

に基づき、$f(\underline{x}_1), f(\underline{x}_2), \ldots, f(\underline{x}_n)$

を回復すること

$\Rightarrow$「ノイズ除去問題」

$$\underline{Y_j} = f(\underline{x_j}) + \underline{\varepsilon_j}$$

観測　ノイズ

Y_jは $f(x_j)$とε_jの和から、ε_jをのぞく.

固定デザインの最も簡単な場合として以下で考える.

$d=1$ かつ $x_j = \dfrac{j}{n}$ $(j=1,2,\dots,n)$

No.16

$\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$

$\underline{\mu} = (f(x_1), f(x_2), \ldots, f(x_n))^T$

$\underline{\mu} + \underline{\varepsilon}$ を観測

$\underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$ $(0 < \sigma < \infty)$ と仮定

評価尺度

$$\mathrm{MSE}(\hat{f}_n) = \frac{1}{n}\sum_{j=1}^{n}\{\hat{f}_n(x_j) - f(x_j)\}^2$$

No. 11

$$\hat{\underline{\mu}} = (\hat{f}_n(x_1), \ldots, \hat{f}_n(x_1))^T =: (\hat{\mu}_1, \ldots, \hat{\mu}_n)$$

とおくと

$$\mathrm{MSE}(\hat{\underline{\mu}}) = \frac{1}{n}\sum_{j=1}^{n}\{\hat{\mu}_j - \mu_j\}^2$$

$$= \frac{1}{n}\|\hat{\underline{\mu}} - \underline{\mu}\|_{2,n}^2$$

ただし、 $\underline{x} = (x_1, \ldots, x_n)^T$

$$\|\underline{x}\|_{2,n} = \sqrt{\underline{x}^T \underline{x}} = \sqrt{\sum_{j=1}^{n} x_j^2};$$

No.12

次に，$\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n \in \mathbb{R}^d$

と仮定する．　　縦ベクトル

$$Y_j = f(\underline{x}_j) + \varepsilon_j \quad (j = 1, 2, \dots, n)$$

記号

$$\underline{Y} = (Y_1, Y_2, \dots, Y_n)^T$$

$$\underline{X} = \begin{pmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_n^T \end{pmatrix} ; n \times d ; \quad \underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$$

仮定　線型モデルの仮定

$$f(\underline{x}) = \underline{x}^T \beta^* \quad ; \quad \beta^* \in \mathbb{R}^d$$

とすると

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

と書ける

さらに，$\hat{\beta}$ を β^* の推定量と

すると f の推定量 $\hat{f}_n$ は

$$\hat{f}_n(\underline{x}) = \underline{x}^T \hat{\beta}_n$$

と書ける．よって

$$\mathrm{MSE}(\underline{X}\hat{\beta}_n)$$

$$= \frac{1}{n} \| \underline{X}\hat{\beta}_n - \underline{X}\beta^* \|_{2,n}^2$$

No.15

$$= \frac{1}{n}(\hat{\beta}-\beta^*)^T X^T X (\hat{\beta}-\beta^*)$$

となる。

注意 通常は

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

をMSEとよぶことがおおい。

3.2 最小2乗推定量の再訪問

β^*の推定問題のかわりに、

$\mu = X\beta^*$の推定問題を扱う。

$X^\top X$は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。

No.17

定義3.1 任意の $m\times n$ の実行列 $\underline{A}$ に対して、次の条件をみたす $n\times m$ の行列 $\underline{G}$ を $\underline{A}$ の Moore-Penrose の一般化逆行列という

(1) $\underline{A}\,\underset{m\times n}{\underline{G}}\,\underline{A}=\underline{A}$; (2) $\underline{G}\,\underset{n\times m}{\underline{A}}\,\underline{G}=\underline{G}$;

(3) $(\underset{m\times n}{\underline{A}}\,\underline{G})^T=\underline{A}\,\underline{G}$; (4) $(\underline{G}\,\underset{n\times n}{\underline{A}})^T=\underline{G}\,\underline{A}$

G を A^+ とかくことにする。

注 必ず存在し、それも一意的であることを証明できる。

3.2.1 制限なしの最小2乗推定量 (LSE)

命題3.3 $\underline{X}\beta^*$ の LSE $\hat{\underline{\mu}}^{LS} = \underline{X}\hat{\beta}^{LS}$ は方程式

$$\underline{X}^T \hat{\underline{\mu}}^{LS} = \underline{X}^T \underline{Y}$$

をみたす. さらに

$$\hat{\beta}^{LS} = \underline{X}^+ \underline{Y}$$

証明 関数

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

は凸なので、最小値は

$$\nabla_{\underline{\beta}} \| \underline{Y} - \underline{X}\,\underline{\beta} \|_{2,n}^2 = \underline{0}_n$$

をみたすβ

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

No.20

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

$\Longleftrightarrow$

$$\frac{\partial}{\partial \beta_k} \| \mathbf{Y} - \mathbf{X}\beta \|_{2,n}^2 = \frac{\partial}{\partial \beta_k} \sum_{j=1}^{n} \{ Y_j - \mathbf{x}_j^T \beta \}^2 = 0$$

$(k = 1, \dots, d)$

よって

$$\frac{\partial}{\partial \beta_k} \| \mathbf{Y} - \mathbf{X}\beta \|_{2,n}^2$$

$$= \frac{\partial}{\partial \beta_k} \sum_{j=1}^{n} \{ Y_j - \mathbf{x}_j^T \beta \}^2$$

No.21

$$= -2\sum_{j=1}^{n}\{Y_j - \underline{x}_j^T\beta\}x_{jk}.$$

ただし, $\underline{x}_j = (x_{j1}, \dots, x_{jd})^T$

よって

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

$$= -2(\underline{X}^T\underline{Y} - \underline{X}^T\underline{X}\beta)$$

である.

したがって

$$\nabla_{\beta} \|\underline{Y} - \underline{X}\beta\|_{2,n}^{2} = \underline{0}_n$$

$\iff$

$$\underbrace{\underbrace{\underline{X}^{T}}_{d\times n}\,\underbrace{\underline{X}}_{n\times d}\,\underbrace{\beta}_{d\times 1}}_{d\times 1} = \underbrace{\underbrace{\underline{X}^{T}}_{d\times n}\,\underbrace{\underline{Y}}_{n\times 1}}_{d\times 1} \quad \cdots \quad (※)$$

となる.

すると $\hat{\beta}^{LS} = \underline{X}^{+}\underline{Y}$ は(※)の一つ　No 23

の解である.

$$\underline{X}^{T}\underline{X}\hat{\beta}^{LS} = \underline{X}^{T}\underline{X}\,\underline{X}^{+}\underline{Y}$$

$$= \underline{X}^{T}(\underline{X}\,\underline{X}^{+})^{T}\underline{Y}$$

$$= (\underline{X}\,\underline{X}^{+}\underline{X})^{T}\underline{Y}$$

$$= \underline{X}^{T}\underline{Y}$$

さらに $\underline{\hat{\mu}}^{LS} = \underline{X}\,\underline{\hat{\beta}}^{LS} = \underline{X}\,\underline{X}^{+}\underline{Y}$ のこと ⇒

$$\underline{X}^{T}\underline{\hat{\mu}}^{LS} = \underline{X}^{T}\underline{X}\,\underline{X}^{+}\underline{Y} = \underline{X}^{T}(\underline{X}\,\underline{X}^{+})^{T}\underline{Y}$$

$$= (\underbrace{\underline{X}\,\underline{X}^{+}\underline{X}}_{=\underline{X}})^{T}\underline{Y} = \underline{X}^{T}\underline{Y}$$

よって、$\underline{\hat{\mu}}^{LS} = \underline{X}\,\underline{\hat{\beta}}^{LS}$ は

$$\underline{X}^{T}\underline{\hat{\mu}}^{LS} = \underline{X}^{T}\underline{Y}$$

をみたす。 $\square$

命題3.4

線型回帰モデル

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

$r = \mathrm{rank}(X^\top X)$

を考える。このとき、LSE $\hat{\beta}^{LS}$ は

$$\mathbb{E}[\mathrm{MSE}(\hat{\beta}^{LS})] = \frac{1}{n}\mathbb{E}[|\underline{X}\hat{\beta}^{LS} - X\beta^*|_{2,n}^2] \lesssim \frac{r\sigma^2}{n}$$

ただし、$a \lesssim b \Leftrightarrow {}^\exists C > 0$ s.t. $a \le Cb$

$\delta > 0$ に対して

$$\Pr\left(\mathrm{MSE}(\hat{\beta}^{LS}) \lesssim \frac{\sigma^2\{r + \log(1/\delta)\}}{n}\right) > 1-\delta$$

証明　LSEの定義より

$$|\underline{Y} - \underline{X}\hat{\beta}^{LS}|_{2,n}^2 \leq |\underline{Y} - \underline{X}\beta^*|_{2,n}^2 = |\underline{\varepsilon}|_{2,n}^2 . \quad (3.4)$$

さらに

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

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

この式を(3.4)に代入すると

No. 27

$$\|\underline{X}\hat{\beta}^{LS} - \underline{X}\beta^*\|_{2,n}^2 \le 2\,\underline{\varepsilon}^T \underline{X}(\hat{\beta}^{LS} - \beta^*)$$

$$= 2\,\|\underline{X}\hat{\beta}^{LS} - X\beta^*\|_{2,n} \boxed{\frac{\underline{\varepsilon}^T(\underline{X}\hat{\beta}^{LS} - X\beta^*)}{\|\underline{X}\hat{\beta}^{LS} - X\beta^*\|_{2,n}}}$$

↓

sup-out

$\underline{\Psi} = [\underline{\psi}_1, \ldots, \underline{\psi}_r] \in \mathrm{Mat}(n, r; \mathbb{R})$ は

$$\mathrm{span}(\underline{X}_1, \ldots, \underline{X}_d) = \mathrm{span}(\underline{\psi}_1, \ldots, \underline{\psi}_r)$$

Xの列ベクトル　正規直交基底

$$\psi_i^T \psi_j = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

$^\exists \underline{\nu} \in \mathbb{R}^r$ s.t.

$$\underline{X}(\hat{\beta}^{LS} - \beta^*) = \underline{\Phi}\,\underline{\gamma}$$

と書ける. よって

$$\frac{\underline{\varepsilon}^T \underline{X}(\hat{\beta}^{LS} - \beta^*)}{|\underline{X}(\hat{\beta}^{LS} - \beta^*)|_{2,n}} = \frac{\underline{\varepsilon}^T \underline{\Phi}\,\underline{\nu}}{|\underline{\Phi}\,\underline{\gamma}|_{2,p}} = \frac{\tilde{\varepsilon}^T \underline{\nu}}{|\underline{\nu}|_{2,r}}$$

||u||が1

$$\leq \sup_{|\underline{u}|_{2,p}=1} \tilde{\underline{\varepsilon}}^T \underline{u} \qquad \left(\underset{r\times 1}{\tilde{\underline{\varepsilon}}} = \underset{r\times n}{\underline{\Phi}}\ \underset{n\times 1}{\underline{\varepsilon}} \right)$$

よって

$$\|\underline{X}\hat{\beta}^{LS} - \underline{X}\beta^*\|_{2,n} \leq 2 \sup_{\|u\|_{2,p}=1} \underline{\tilde{\varepsilon}}^T \underline{u}$$

$$\Rightarrow \quad \|\underline{X}\hat{\beta}^{LS} - \underline{X}\beta^*\|_{2,n}^2 \leq 4 \sup_{\|u\|_{2,p}=1} (\underline{\tilde{\varepsilon}}^T \underline{u})^2 \quad (3.6)$$

次に, $\underline{v} \in S^{r-1}$ としたとき

$$\|\underline{\bar{\Psi}} \underline{v}\|_{2,n}^2 = \underline{v}^T \underbrace{\underline{\bar{\Psi}}^T \underline{\bar{\Psi}}}_{= I_r} \underline{v} = \underline{v}^T \underline{v} = 1$$

である.

$\forall s \in \mathbb{R}$ に対して

$$E[\exp\{s\, \underline{\tilde{\varepsilon}}^T \underline{v}\}] = E[\exp\{s\, \underline{\varepsilon}^T \underline{\Phi}\, \underline{v}\}]$$

$\|\Phi v\| \le 1$

$$\le \exp\left(\frac{s^2\sigma^2}{2}\right) \quad \because)\ \underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$$

よって、$\underline{\tilde{\varepsilon}} \sim \mathrm{subG}_p(\sigma^2)$

補題2.11

$$X \sim \mathrm{subG}(\sigma^2) \Rightarrow E[|X|^k] \le (2\sigma^2)^{k/2}\, k\, \Gamma\left(\frac{k}{2}\right)$$

$k=2$ のとき、 $E[X^2] \le 4\sigma^2$

No.31

$$4E\left[\sup_{\|\underline{u}\|_{2,p}=1}(\underline{\tilde{\varepsilon}}^T\underline{u})^2\right]=4\sum_{j=1}^{r}\tilde{\varepsilon}_j^2\le 16\sigma^2 r \quad (3.7)$$

ただし、$\underline{\tilde{\varepsilon}}=(\tilde{\varepsilon}_1,\dots,\tilde{\varepsilon}_r)^T$.

(3.6) と (3.7) より

$$\frac{1}{n}E\left[\|\underline{X}\,\underline{\hat{\beta}}^{LS}-\underline{X}\,\underline{\beta}^*\|_{2,n}^2\right]$$

$$\underset{(3.6)}{\le}\frac{4}{n}E\left[\sup_{\underline{u}\in B_2}(\underline{u}^T\underline{\tilde{\varepsilon}})^2\right]$$

$$\underset{(3.7)}{\le}\frac{16r}{n}\sigma^2.$$

よって、1番目の主張は

Okay!

2番目の主張の証明

$B_2 = \{\underline{u} \in \mathbb{R}^r ;\ |\underline{u}|_{2,r} = 1\} \subset \mathbb{R}^r$ の $(1/2)$ 網 N をとる.

補題 2.16 より

$$\#(N) \leq 6^r. \qquad (\divideontimes)$$

$\forall \underline{u} \in B_2$ に対して, $\exists \underline{w} \in N$ と $\underline{v} \in (1/2)B_2$ が存在して

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

と書ける.

よって

$$\sup_{\underline{u}\in B_2} \underline{u}^T\tilde{\varepsilon} \le \max_{\underline{w}\in N} \underline{w}^T\tilde{\varepsilon} + \boxed{\sup_{\underline{v}\in \frac{1}{2}B_2} \underline{v}^T\tilde{\varepsilon}},$$

$$\sup_{\underline{u}\in \frac{1}{2}B_2} \underline{u}^T\tilde{\varepsilon} = \boxed{\frac{1}{2}\sup_{\underline{u}'\in B_2} \underline{u}^T\tilde{\varepsilon}}.$$

より

$$\sup_{\underline{u}\in B_2} \underline{u}^T\tilde{\varepsilon} \le 2\max_{\underline{w}\in N} \underline{w}^T\tilde{\varepsilon} \qquad (*)$$

となる.

よって，$\forall t > 0$ に対して

$$\Pr\left(\sup_{\underline{u} \in B_2} (\underline{u}^T \underset{\sim}{\tilde{\varepsilon}})^2 \geq t^2 \right)$$

$$= \Pr\left(\sup_{\underline{u} \in B_2} (\underline{u}^T \underset{\sim}{\tilde{\varepsilon}}) \geq t \right)$$

$$\underset{(※)}{\leq} \Pr\left(\sup_{\underline{u} \in \textcolor{red}{N}} (\underline{u}^T \underset{\sim}{\tilde{\varepsilon}}) \geq \frac{t}{2} \right)$$

$$\underset{\text{命題2.2より}}{\leq} \#(N) \exp\left(-\frac{t^2}{8\sigma^2}\right) \underset{(※※)}{\leq} \overbrace{6^r \exp\left(-\frac{t^2}{8\sigma^2}\right)}^{\delta}$$

$$\delta = 6^r \exp\left(-\frac{t^2}{8a^2}\right) \Leftrightarrow t^2 = 8a^2\{r\log 6 + \log(\tfrac{1}{\delta})\}$$

よ')

$$\Pr\left(\sup_{\underline{u}\in B_2}(\underline{u}^T\underline{\tilde{\varepsilon}})^2 \geq 8a^2\{r\log 6 + \log(\tfrac{1}{\delta})\}\right) \leqq \delta \quad (a)$$

よって

$$\Pr\left(\underline{|\underline{X}\hat{\beta}^{LS} - \underline{X}\beta^*|_{2,n}^2 > 2a^2\{r\log 6 + \log(\tfrac{1}{\delta})\}}\right)$$

$$\leqq \Pr\left(|\underline{X}\hat{\beta}^{LS} - \underline{X}\beta^*|_{2,n}^2 \geq 2a^2\{r\log 6 + \log(\tfrac{1}{\delta})\}\right)$$

No.36

$$\underset{(3.6)}{\leq} \Pr\left(\sup_{\underline{u} \in B_2} (\underline{u}^T \tilde{\underline{\varepsilon}})^2 \geq 8\sigma^2 \left\{ r\log 6 + \log\left(\frac{1}{\delta}\right) \right\} \right)$$

$$\underset{(a)}{\leq} \delta$$

よって、＿＿＿ の補事象を取ると

$$1 - \Pr\left(\underline{\frac{1}{n} \| \underline{X} \hat{\beta} - \underline{X}^* \beta^* \|_{2,n}^2 \leq \frac{2\sigma^2 \left\{ r\log 6 + \log\left(\frac{1}{\delta}\right) \right\}}{n}} \right)$$

$$\leq \delta$$

$$\Leftrightarrow \Pr\left(\underline{\qquad\qquad} \right) \geq 1 - \delta.$$

$\square$

注意 3.5　$d \leq n$ とする.

$$\underset{d\times d}{\underline{B}} := \frac{1}{n} \underline{X}^{\top} \underline{X}$$

とおくと

$$\operatorname{rank} \underline{B} = r.$$

$$= \mathrm{MSE}(\underline{X}\hat{\beta}^{LS})$$

$$\boxed{\frac{1}{n} \left| \underline{X}\hat{\beta}^{LS} - \underline{X}\beta^* \right|_{2,n}^2} = \frac{1}{n} (\hat{\beta}^{LS} - \beta^*)^{\top} X^{\top} X (\hat{\beta}^{LS} - \beta^*)$$

$$= (\hat{\beta}^{LS} - \beta^*)^{\top} \boxed{\underline{B}} (\hat{\beta}^{LS} - \beta^*)$$

$$\leq (\hat{\underline{\beta}}^{LS} - \underline{\beta}^*)(\lambda_{max}(\underline{B})\, \underline{I}_r)(\hat{\underline{\beta}}^{LS} - \underline{\beta}^*)$$

$$= \lambda_{max}(\underline{B})\, |\hat{\underline{\beta}}^{LS} - \underline{\beta}^*|_{2,d}^2$$

より

$$|\hat{\underline{\beta}}^{LS} - \underline{\beta}^*|_{2,d} \leq \frac{MSE(\underline{X}\hat{\underline{\beta}}^{LS})}{\lambda_{max}(\underline{B})}$$

となる. よって

$$Pr\left(|\hat{\beta}^{LS} - \beta^*|_{2,d} \leq \lambda_{max}(\underline{B}) \frac{2\sigma^2\{r\log 6 + \log(\frac{1}{\delta})\}}{n}\right)$$

$$\geq 1 - \delta.$$

$\boxed{証}$

ℓ_0制約付きLSE

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

$$|\beta|_{0,d} := \sum_{j=1}^{d} \mathbb{I}\{\beta_j \neq 0\}$$

← βの要素で0でないものの個数

と定める. ただし

$$\mathbb{I}\{\ \_\_\_\_\ \} = \begin{cases} 1 & (\ \_\_\_\_\ は真の場合) \\ 0 & (\ \_\_\_\_\ は偽の場合) \end{cases}$$

である.

β^*のスパース仮定

$k < d$ に対して

$$|\beta^*|_{0,d} \leq k \iff k \text{ スパース}.$$

記号 $\beta = (\beta_1, \beta_2, \ldots, \beta_d)^T \in \mathbb{R}^d$ に対して

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

と定める.

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

例 $d=5$; $\beta=(-1,0,1,0,0)$ のとき

$$\mathrm{supp}(\beta)=\{1,3\}$$

注意 3.8 $\beta=(\beta_1,\ldots,\beta_d)\in\mathbb{R}^d$ に対して

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

$1 \leq k \leq d$ に対して

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

と定める.

$$\hat{\underline{\beta}}^{LS}_{B_0(k)} \in \underset{\underline{\beta} \in B_0(k)}{\arg\min} \ |\underline{Y} - \underline{X}\,\underline{\beta}|_{2,n}$$

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

命題3.9 $k \leq \frac{d}{2}$ を整数とし

$$\beta^* \in B_0(k)\ ;\ \varepsilon \sim \mathrm{subG}_n(\sigma^2)\ ;\ 0<\sigma<\infty$$

を仮定する. このとき, $^\forall \delta > 0$ に対して

$$\Pr\left(\mathrm{MSE}(X\hat{\beta}^{LS}) \leq \frac{\sigma^2}{n}\log\binom{d}{2k} + \frac{\sigma^2 k}{n} + \frac{\sigma^2}{n}\log\left(\frac{1}{\delta}\right)\right) \geq 1-\delta$$

が成立する. ただし

$$\binom{d}{2k} = \frac{d!}{(2k)!\,(d-2k)!}$$

である.

証明 $k := B_0(R)$ とし，$\hat{\beta}^{LS}_{B_0(R)}$ を $\hat{\beta}^{LS}_k$ とかく．

まず

$$\|\underline{Y} - \underline{X}\hat{\beta}^{LS}_k\|^2_{2,n} \leq \|\underline{Y} - \underline{X}\beta^*\|^2_{2,n} = \|\underline{\varepsilon}\|^2_{2,n}. \quad (a)$$

さらに

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

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

(b)

(a)と(b)を合せると

$$\|\underline{X}\hat{\beta}^{LS}_k - \underline{X}\beta^*\|_{2,n}^2 \le 2\underline{\varepsilon}^T(\underline{X}\hat{\beta}^{LS}_k - \underline{X}\beta^*)$$

$$= 2\|\underline{X}\hat{\beta}^{LS}_k - \underline{X}\beta^*\|_{2,n} \boxed{\frac{\underline{\varepsilon}^T(\underline{X}\hat{\beta}^{LS}_k - \underline{X}\beta^*)}{\|\underline{X}\hat{\beta}^{LS}_k - \underline{X}\beta^*\|_{2,n}}} \quad (c)$$

を以下で評価

を得る.

$\hat{\beta}^{LS}_k \in B_0(k)$ かつ $\beta^* \in B_0(k)$

$$\Rightarrow \hat{\beta}^{LS}_k - \beta^* \in B_0(2k)$$

記号 $S \subset \{1, 2, \ldots, d\}$ に対して

$$\underset{n \times \#(S)}{\underline{X}_S} = (x_j)_{j \in S} \quad (j = 1, 2, \ldots, d)$$

$$\underline{X} = (x_1, x_2, \ldots, x_d);\ x_j \in \mathbb{R}^n$$

$\mathrm{rank}(\underline{X}_S) =: r_S \leq \#(S)$ とかく.

$$\underline{\Psi}_S = (\underline{\psi}_1, \underline{\psi}_2, \ldots, \underline{\psi}_{r_S});\ \psi_j \in \mathbb{R}^n\ (j = 1, 2, \ldots, r_S)$$

とし,

$$\underline{\psi}_i^T \psi_j = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j). \end{cases}$$

$$\mathrm{span}(\underline{\Psi}_S) = \mathrm{span}(\underline{X}_S)$$

となるように取る.

さらに, $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_d)^T \in \mathbb{R}^d$ と $S \subset \{1, 2, \dots, d\}$ に対して

$$\underline{\beta}(S) = (\beta_j)_{j \in S}$$

と定める.

例1 $d=5$, $S=\{1,3,5\}$, $\beta=(\beta_1,\dots,\beta_5)$

$$\beta(S)=(\beta_1,\beta_3,\beta_5)^T$$

である.

いま

$$\hat{S}:=\mathrm{supp}(\hat{\beta}^{LS}_k-\beta^*)$$

とおく. すると

$$\#(\hat{S})\leq 2k.$$

また，$\underline{\nu} \in \mathbb{R}^{r_{\hat{S}}}$ があって

$$\underline{X}(\underline{\hat{\beta}}^{LS}_{k} - \underline{\beta}^{*}) = \underline{X}_{\hat{S}}(\underline{\hat{\beta}}^{LS}_{k}(\hat{S}) - \underline{\beta}^{*}(\hat{S}))$$

$$= \underline{\Psi}_{\hat{S}}\underline{\nu}$$

と書ける．

したがって

$$\frac{\underline{\varepsilon}^{T}\underline{X}(\underline{\hat{\beta}}^{LS}_{k} - \underline{\beta}^{*})}{|\underline{X}\underline{\hat{\beta}}^{LS}_{k} - \underline{X}\underline{\beta}^{*}|_{2,n}} = \frac{\underline{\varepsilon}^{T}\underline{\Psi}_{\hat{S}}\underline{\nu}}{|\underline{\nu}|_{2,r_{\hat{S}}}}$$

No. 50

$$\leq \max_{\#(S)=2k} \left(\sup_{\underline{u} \in B_2^{r_S}} \underline{\varepsilon}^T \Phi_S \underline{u} \right) \quad \text{(d)}$$

とする.

$$\tilde{\varepsilon} := \Phi_S^T \underline{\varepsilon} \sim \mathrm{subG}_{r_S}(\sigma^2)$$

となることに注意する.

よって，(c)と(d)より

$$\|\underline{X}\hat{\beta}^{LS}_k - \underline{X}\beta^*\|_{2,n} \underset{(c)}{\leq} 2\frac{\underline{\varepsilon}^T\underline{X}(\hat{\beta}^{LS}_k-\beta^*)}{\|\underline{X}\hat{\beta}^{LS}_k-\beta^*\|_{2,n}}$$

$$\underset{(d)}{\leq} 2\max_{\#(S)=2k}\left(\sup_{\underline{u}\in B_2^{rs}} \underline{\varepsilon}^T\underline{\underline{\Phi}}_S\underline{u}\right)$$

となる．ただし，B_2^{rs} は $\mathbb{R}^{rs}$ の単位球．

これより

No.52

$$\|\underline{X}\hat{\beta}_{k}^{LS} - \underline{X}\beta^*\|_{2,n}^2 \leq 4 \max_{\#(S)=2k} \left(\sup_{\underline{u} \in B_2^{rs}} (\tilde{\underline{\varepsilon}}^T \underline{u})^2 \right)$$

ただし、$\tilde{\underline{\varepsilon}} := \underline{\Psi}_s^T \underline{\varepsilon} \sim \mathrm{subG}_{r_s}(\sigma)$.

よって、$\forall t > 0$ に対して

$$\Pr\left(\max_{\#(S)=2k} \left(\sup_{\underline{u} \in B_2^{rs}} (\tilde{\underline{\varepsilon}}^T \underline{u})^2 \right) \geq t \right)$$

$$\leq \sum_{\#(S)=2k} \Pr\left(\sup_{\underline{u} \in B_2^{rs}} (\tilde{\underline{\varepsilon}}^T \underline{u})^2 \geq t \right)$$

$\sup_{\underline{u} \in B_2^{rs}} \tilde{\underline{\varepsilon}}^T \underline{u}) \geq \sqrt{t}$

$$\leq \sum_{\#(S)=2k} 6^{\#(S)} \exp\left(-\frac{t}{8a^2}\right)$$

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

$$\boxed{\Pr\left(\sup_{\underline{u}\in B_2} \underline{u}^T \underline{X} > t\right) \leq 6^d \exp\left(-\frac{t^2}{8a^2}\right) \qquad X \sim \mathrm{subG}_d(a^2)}$$

No.55

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

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

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

これより

$$\Pr\left(\mathrm{MSE}(\underline{X} \hat{\underline{\beta}}_k^{LS}) \geq t \right)$$

$$= \Pr\left(|\underline{X} \hat{\underline{\beta}}_k^{LS} - \underline{X} \underline{\beta}^{*}| \geq n\tau \right)$$

No.55

$$\leq \Pr\left(\max_{\#(S)=2k} \left(\sup_{\underline{u} \in B_2^{rs}} (\tilde{\underline{\varepsilon}}^T \underline{u})^2 \right) \geq nt \right)$$

$$\leq \sum_{\#(S)=2k} \Pr\left(\sup_{\underline{u} \in B_2^{rs}} (\tilde{\underline{\varepsilon}}^T \underline{u}) \geq \sqrt{n\tau} \right)$$

$$\leq \sum_{\#(S)=2k} 6^{rs} \exp\left(- \frac{n\tau}{8\sigma^2} \right)$$

$$\leq \binom{d}{2k} 6^{2k} \exp\left(- \frac{n\tau}{8\sigma^2} \right) \quad (\text{※※})$$

$$=: \delta$$

No.56

$$\Leftrightarrow \log \delta = \log\left(\frac{d}{2k}\right) + 2k\log 6 - \frac{n\tau}{8\sigma^2}$$

$$t = \frac{8\sigma^2}{n}\left\{ \log\left(\frac{d}{2k}\right) + k\,\underline{(2\log 6)} + \log\left(\frac{1}{\delta}\right) \right\}$$

$$\approx \frac{\sigma^2}{n}\log\left(\frac{d}{2e}\right) + \frac{\sigma^2 k}{\underline{n}} + \frac{\sigma^2}{n}\log\left(\frac{1}{\delta}\right)$$

$\frac{k}{n}$ のオーダーになっている

$$\Pr\left(\mathrm{MSE}(\underline{X}\hat{\beta}^{LS}_{k}) \lesssim \frac{\sigma^2}{n}\log\boxed{\left(\frac{d}{2e}\right)} + \frac{\sigma^2 k}{n} + \frac{\sigma^2}{n}\log\left(\frac{1}{\delta}\right) \right)$$

$$\geq 1-\delta.$$

次の補題で評価する

$\square$

補題 3.10 整数 $k \leq d$ に対して

$$\binom{d}{k} \leq \left(\frac{ed}{k}\right)^k.$$

証明 帰納法で示す.

- $k=1$ のとき

$$\binom{d}{1} = \frac{d!}{(d-k)!\cdot 1!} = d \leq de \leq \left(\frac{ed}{1}\right)^1$$

- ある $k\ (\leq d-1)$ で

$$\binom{d}{k} \leq \left(\frac{ed}{k}\right)^k$$ で成立と仮定

• $k+1$ の場合

$$\binom{d}{k+1} = \frac{d!}{(k+1)!\,(d-k-1)!} = \frac{d!}{k!\,(d-k)!}\cdot\frac{d-k}{k+1}$$

$$\le \left(\frac{ed}{k}\right)^{k}\frac{d-k}{k+1}$$

帰納法の仮定

$$= \frac{e^{k} d^{k}}{k^{k}}\,\frac{d-k}{k+1}$$

$d-k \le d$

$$\le \frac{e^{k} d^{k+1}}{(k+1)^{k+1}}\,\frac{(k+1)^{k}}{k^{k}}$$

$$= \frac{e^{k} d^{k+1}}{(k+1)^{k+1}}\left(1+\frac{1}{k}\right)^{k}$$

No.59

$$\leq \frac{e^{k} d^{k+1}}{(k+1)^{k+1}} e \qquad \because \left(1+\frac{k}{k}\right)^{k} \uparrow e \quad (k \to \infty)$$

$$= \left(\frac{ed}{k+1}\right)^{k+1}. \qquad \square$$

系3.11 $\forall \delta > 0$ に対して

$$\Pr\left(\mathrm{MSE}\left(\underline{X}\hat{\beta}^{LS}_{B_k(0)}\right) \lesssim \frac{\sigma^2}{n}\left\{k\log\left(\frac{ed}{2k}\right) + \frac{k}{n} + \log\left(\frac{1}{\delta}\right)\right\}\right) \geq 1 - \delta.$$

証明 補題3.10より

$$\binom{d}{2k} \leq \left(\frac{ed}{2k}\right)^{2k}$$

より

$$\underbrace{\frac{\sigma^2}{n}\left\{\log\binom{d}{2k} + \frac{k}{n} + \log\left(\frac{1}{\delta}\right)\right\}}_{=A}$$

No. 61

$$\le \frac{\sigma^2}{n}\left\{2k\log\left(\frac{ed}{2k}\right) + \frac{k}{n} + \log\left(\frac{1}{\delta}\right)\right\}. = B$$

よって

$$\left\{MSE(X\hat{\beta}^{LS}_{B_0(k)}) \lesssim A\right\}$$

$$\subset \left\{MSE(X\hat{\beta}^{LS}_{B_0(k)}) \lesssim B\right\}$$

より

$$1-\delta \le \Pr\left(MSE(X\hat{\beta}^{LS}_{B_0(k)}) \lesssim A\right)$$

$$\le \Pr\left(MSE(X\hat{\beta}_{B_0(k)}) \lesssim B\right). \quad \square$$

系3.12

No.62

$$E\left[\mathrm{MSE}(\underset{\sim}{X}\hat{\beta}^{LS}_{B_0(k)}\right] \lesssim \frac{\sigma^2 k}{n}\log\left(\frac{ed}{k}\right).$$

証明　まず，正値確率変数 W に対して

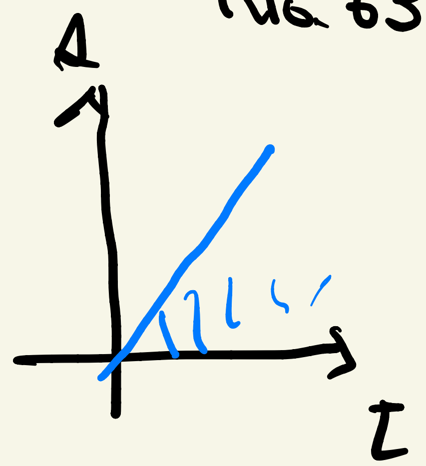
$$E[W] = \int_0^\infty \Pr(W \geq t)\,dt$$

なぜならば，W の p.d.f. を p^W とした とき

$$E[W] = \int_0^\infty t\, p^W(t)\, dt$$

Na 63

$$= \int_0^{\omega} \left\{ \int_0^{\tau} ds \right\} p^{\omega}(\tau)\, d\tau$$



$$= \int_0^{\omega} \left\{ \int_0^{\tau} p^{\omega}(\tau)\, ds \right\} d\tau$$

$$= \int_0^{\omega} \left\{ \int_s^{\omega} p^{\omega}(\tau)\, d\tau \right\} ds$$

$$= \int_0^{\omega} \Pr(\omega \geq s)\, ds .$$

$\square$

これより：$E[MSE(\underline{X}\hat{\beta}^{LS}_{B_0(R)})]$を評価する　No. 64

$\forall H>0$に対して

$$E[MSE(\underline{X}\hat{\beta}^{LS}_{B_0(R)})]$$

$$=\int_0^\infty P_r\left(MSE(\underline{X}\hat{\beta}^{LS}_{B_0(R)})\geq u\right)du$$

$$=\int_0^\infty P_r\left(|\underline{X}\hat{\beta}^{LS}_{B_0(R)}-\underline{X}\beta^*|_{2,n}^2\geq nu\right)du$$

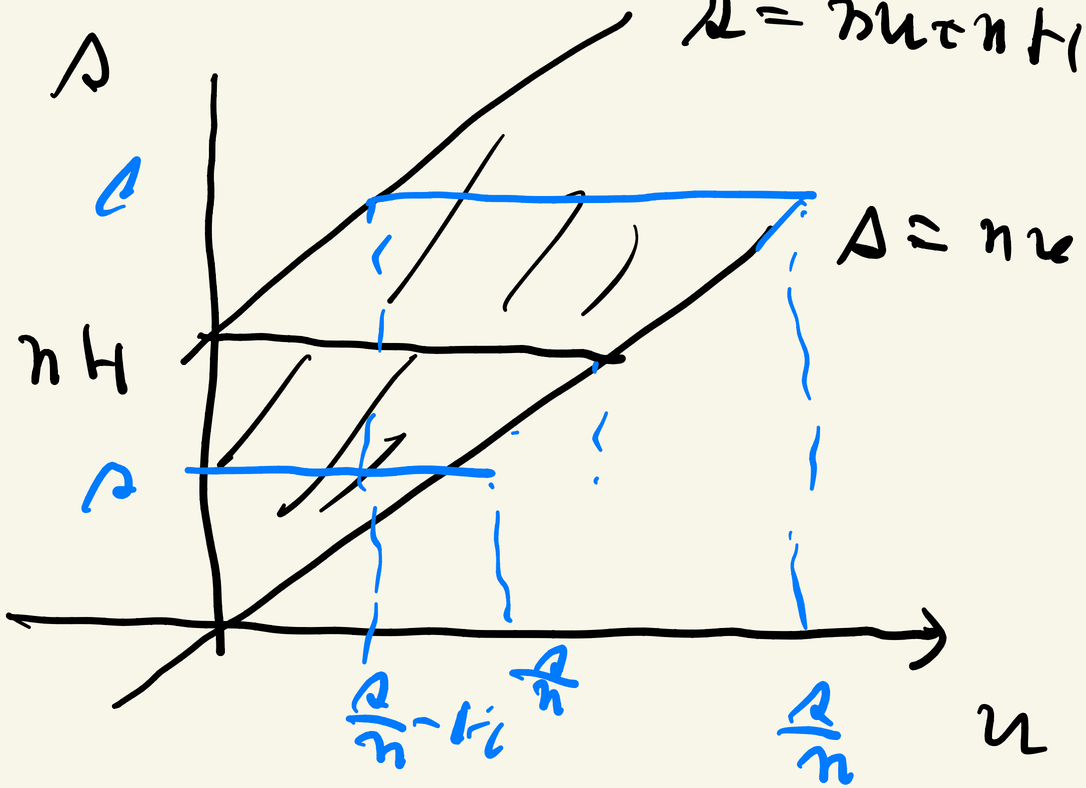
$=(※)$

正値確率変数 W とし、その p.d.f. を p^W とかく Nubs

すると
$$\int_0^\infty \Pr(W \ge nu)\,du \le \int_0^\infty \Pr(W \ge n(u+H))\,du + H$$
となる.

∵)
$$\int_0^\infty \Pr(W \ge nu)\,du$$
$$= \int_0^\infty \Big\{ \Pr(W \ge n(u+H)) \Big)\,du$$
$$+ \underbrace{\Pr(nH \le W < n(u+H))}_{=: A(u)}\,du$$

$$\int_0^\infty A(u)\,du = \int_0^\infty \left\{ \int_{nu}^{n(u+H)} p^w(s)\,ds \right\} du$$



$$= \int_0^{nH} \left\{ \int_0^{\frac{s}{n}} p^w(s)\,du \right\} ds$$

No. 69

$$+\int_{nH}^{\infty}\left\{\int_{\frac{s}{n}-H}^{\frac{s}{n}} p^{w}(s)\,du\right\} ds$$

$$=\int_{0}^{nH}\frac{s}{n}p^{w}(s)\,ds+\int_{nH}^{\infty}\left\{\frac{s}{n}-\left(\frac{s}{n}-H\right)\right\}p^{w}(s)\,ds$$

$$=\int_{0}^{nH}\underbrace{\boxed{\frac{s}{n}}}_{\leq H}p^{w}(s)\,ds+H\int_{nH}^{\infty}p^{w}(s)\,ds$$

$$\leq H\overbrace{\boxed{\int_{0}^{\infty}p^{w}(s)\,ds}}^{=1}=H$$

$\square$

$$(※) \leq H + \int_0^\infty \Pr\left(\| \underline{X} \hat{\beta}^{LS}_{B_0(k)} - \underline{X}\beta^* \|^2_{2,n} \geq n(u+H) \right) du$$

$$\leq H + \binom{d}{2k} 6^{2k} \int_0^\infty \exp\left(- \frac{n(u+H)}{32\sigma^2} \right) du$$

(※※)

$$= H + \left(\frac{ed}{2k}\right)^{2k} 6^{2k} \left[- \frac{32\sigma^2}{n} \exp\left(- \frac{n(u+H)}{32\sigma^2} \right) \right]_0^\infty$$

$$= H + \left(\frac{ed}{2k}\right)^{2k} 6^{2k} \frac{32}{n} \sigma^2 \exp\left[- \frac{nH}{32\sigma^2} \right]$$

$$= H + \left(\frac{3de}{k}\right)^{2k} \frac{32\sigma^2}{n} \exp\left[- \frac{nH}{32\sigma^2} \right]$$

$= I$ とする。

ここで

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

$$\Longleftrightarrow\quad \exp\left[-\frac{nH}{32\sigma^2}\right]=\left(\frac{3de}{k}\right)^{-2k}$$

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

$$\Longleftrightarrow\quad H=64\,\frac{k}{n}\,\sigma^2\log\left(\frac{3de}{k}\right)\quad\cdots$$

$$(\text{※}) \leq H + \frac{32}{n}\sigma^2$$

$$= 64\frac{k}{n}\sigma^2 \log\left(\frac{3de}{k}\right) + \frac{32}{n}\sigma^2$$

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