中 央 大 学 理 工 学 研 究 所 論 文 集　第 14 号　2008 年
Journal of the Institute of Science and Engineering. Chuo University

正準相関分析における変数選択に対するブートストラップ情報量規準

藤越 康祝*，櫻井 哲朗*，神田 真吾†，杉山 高一*

Bootstrap Information Criterion for Selection of Variables in Canonical Correlation Analysis

Yasunori Fujikoshi*, Tetsuro Sakurai*,
Shingo Kanda†, Takakazu Sugiyama*

abstract

In this paper, we consider the problem for selection of variables based on selection of redundancy models in canonical correlation analysis. Under normality, DIC and $CDIC$ were proposed by Fujikoshi [2] and Fujikoshi and Kurata [3]. Our purpose is to extend such approaches to nonnormality and to propose bootstrap information criterion for selection of redundancy models. By some simulations it is shown that our bootstrap information criterion is useful in nonnormal situations.

1　はじめに

本論文では，正準相関分析における変数選択問題において，有効な変数の組あるいは冗長な変数の組を選ぶ問題を考える．ここでの冗長性とは，判別分析において Rao [5] によって導入された考え方であって，それを正準相関分析に拡張したものが Siotani, Hayakawa and Fujikoshi [6] などに詳しく紹介されている．したがって，正準相関分析における変数選択問題へのアプローチとして冗長性モデルの選択問題とみなす方法がある．さらに，冗長性モデルはある種の共分散構造をもつモデルとして定式化できる．これより，正準相関分析での有効な変数の選択は，共分散行列がある種の構造をもつモデルの選択として考えることができる．

ここでは，特に非正規のもとでの変数選択に興味がある．そこで，本論文では尤度の代りに，共分散行列間の距離にもとづいたリスクを用いる (Fujikoshi and Kurata [3])．これは，母集団に正規分布を仮定した Kullback-Leibler 距離にもとづく AIC 型のリスクの自然な拡張であることが知られている．また，候補のモデルのもとでの共分散行列の推定量は，標本共分散行列との距離を最小にするものとして与えられる．これは，正規分布のもとでの最尤推定量と密接に関連している．このようなリスクに対して不偏推定量を構成し変数選択規準を導入する方法は，正規性の仮定のもとで，Fujikoshi [2], Fujikoshi and Kurata [3] によって検討されている．

本論文では，非正規性のもとでリスクの推定量を構成し変数選択規準を提案するが，その考え方はブートストラップ法の利用に基づいている．

本論文の構成は，次のようになっている．まず第 2 節で，共分散行列間の距離にもとづく共分散構造モデルの選択規準について説明する．第 3 節では，正準相関分析における冗長性と，その正規性のもとでの選択規準

*中央大学理工学研究科
†株式会社メディサイエンスプラニング

について簡単に紹介する．第 4 節において，非正規性のもとでの変数選択規準を提案する．そこでは，ブートストラップ変数選択規準の構成法を与える．最後に，数値実験を通して，今回提案したブートストラップ変数選択規準が広い範囲で有用であることを確認する．とくに，正規性のもとで提案された変数選択規準と比較し，ブートストラップ変数選択規準の有用性を明らかにする．

2　共分散構造の選択規準

いま，p 次元母集団からの大きさ $N=n+1$ の標本にもとづく標本共分散行列を S とする．このとき，母共分散行列 Σ がある種の構造をもつモデル

$$M_k:\ \Sigma \in \mathcal{S}_k$$

を考える．ここに，$\mathcal{S}_k$ は共分散行列のある集りで，添字 k は異なる共分散構造を示し，$k=1,2,\ldots,K$ とする．標本共分散行列 S と任意の共分散行列 Σ との距離を

$$D(S,\Sigma)=n\log|\Sigma S^{-1}|+n\mathrm{tr}\Sigma^{-1}(S-\Sigma)$$

と定義する．また，モデル M_k のもとでの推定量として

$$\min_{\Sigma\in\mathcal{S}_k} D(S,\Sigma)=D(S,\hat{\Sigma}_k)$$

となる最小距離推定量を考える．このような推定量 $\hat{\Sigma}_k=\hat{\Sigma}_k(S)$ は，正規分布のもとでの最尤推定量 $\tilde{\Sigma}_k(S)$ と密接に関連していて，次の関係がある．

$$\tilde{\Sigma}_k(S)=\hat{\Sigma}_k\left(\frac{n}{N}S\right).$$

モデル M_k の良さとして，次のようにリスクを定義する．

$$R_k=\mathrm{E}^*_{S_z}\mathrm{E}^*_S[D(S_z,\hat{\Sigma}_k)].$$

ここに，S_z は将来の標本共分散行列に相当するものであり，S_z は，S と同じ分布に従い，かつ，S とは独立であるとする．また，$\mathrm{E}^*[\ ]$ は真のモデルのもとでの期待値とする．このとき，R_k が小さければ，それだけよいモデルであることを意味する．ここで，R_k を推定する必要があるが，$D(S_z,\hat{\Sigma}_k)$ において S_z を S で置き換えたナイーブな推定量 $D(S,\hat{\Sigma}_k)$ を考え，これを改良することを考える．このとき，R_k を

$$R_k=\mathrm{E}^*_{S_z}\mathrm{E}^*_S[D(S,\hat{\Sigma}_k)]+B_k$$

と表し，B_k を推定することを考える．ここに，

$$B_k=\mathrm{E}^*_{S_z}\mathrm{E}^*_S[D(S_z,\hat{\Sigma}_k)-D(S,\hat{\Sigma}_k)]$$

である．偏りの項 B_k は，次のように簡単化される．

$$\begin{aligned}B_k&=\mathrm{E}^*_{S_z}\mathrm{E}^*_S[D(S_z,\hat{\Sigma}_k)-D(S,\hat{\Sigma}_k)]\\&=\mathrm{E}^*_{S_z}\mathrm{E}^*_S[n\log|\hat{\Sigma}_kS_z^{-1}|+n\mathrm{tr}\hat{\Sigma}_k^{-1}(S_z-\hat{\Sigma}_k)-\{n\log|\hat{\Sigma}_kS^{-1}|+n\mathrm{tr}\hat{\Sigma}_k^{-1}(S-\hat{\Sigma}_k)\}]\\&=\mathrm{E}^*_{S_z}\mathrm{E}^*_S[-n\log|S_z|+n\log|S|+n\mathrm{tr}\hat{\Sigma}_k^{-1}S_z-n\mathrm{tr}\hat{\Sigma}_k^{-1}S]\\&=\mathrm{E}^*_S[n\mathrm{tr}\hat{\Sigma}_k^{-1}(\Sigma-S)]=\mathrm{E}^*_S[n\mathrm{tr}\hat{\Sigma}_k^{-1}\Sigma]-np.\end{aligned}$$

共分散構造の選択規準は，B_k の推定量 $\hat{B}_k$ を考えることによって，次のように提案される．

$$DIC_k=n\log|\hat{\Sigma}_kS^{-1}|+n\mathrm{tr}\hat{\Sigma}_k^{-1}(S-\hat{\Sigma}_k)+\hat{B}_k.$$

3 冗長性モデル

ここでは，Fujikoshi [2], Fujikoshi and Kurata [3] で与えれている正準相関分析における冗長性モデルとそれを選択するための規準量を紹介する．

いま，$(p+q)$ 次元の母集団 G から，大きさ $N=n+1$ の標本が次のように与えられているとする．

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{y}_1 \end{pmatrix}, \ldots, \begin{pmatrix} \boldsymbol{x}_N \\ \boldsymbol{y}_N \end{pmatrix} \overset{i.i.d.}{\sim} G : \mathrm{E}^*\left[\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{pmatrix}\right] = \boldsymbol{\mu}_*, \quad \mathrm{Var}^*\left[\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{pmatrix}\right] = \Sigma_* = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}.$$

ここで，$\mathrm{E}^*[\ ]$, $\mathrm{Var}^*[\ ]$ は真のモデルのもとでの期待値と共分散行列とし，$\boldsymbol{x}_i : p\times 1$，$\boldsymbol{y}_i : q\times 1$，$i=1,\ldots,N$，$\Sigma_{xx} : p\times p$，$\Sigma_{yy} : q\times q$ である．これらの標本をまとめて

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1,\ldots,\boldsymbol{x}_N \\ \boldsymbol{y}_1,\ldots,\boldsymbol{y}_N \end{pmatrix}$$

とし，標本平均ベクトル，標本共分散行列を

$$\begin{pmatrix} \bar{\boldsymbol{x}} \\ \bar{\boldsymbol{y}} \end{pmatrix}, \quad S = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}$$

とする．ここに，

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i, \quad \bar{\boldsymbol{y}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{y}_i, \quad S_{xx} = \frac{1}{n}\sum_{i=1}^N (\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})', \quad S_{yy} = \frac{1}{n}\sum_{i=1}^N (\boldsymbol{y}_i-\bar{\boldsymbol{y}})(\boldsymbol{y}_i-\bar{\boldsymbol{y}})',$$

$$S_{xy} = \frac{1}{n}\sum_{i=1}^N (\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{y}_i-\bar{\boldsymbol{y}})', \quad S_{yx} = S_{xy}'.$$

ここでは，次の問題に関心がある．

(a) $\boldsymbol{x}$ と $\boldsymbol{y}$ の正準相関分析における $\boldsymbol{x}=(x_1,\ldots,x_p)'$，または $\boldsymbol{y}=(y_1,\ldots,y_q)$ に関する有効な変数の選択

(b) $\boldsymbol{x}$ と $\boldsymbol{y}$ の正準相関分析における $\boldsymbol{x}=(x_1,\ldots,x_p)'$ と $\boldsymbol{y}=(y_1,\ldots,y_q)$ に関する有効な変数の選択

3.1 1組の冗長性モデル

ここでは, 一般性を失うことなく $\boldsymbol{x}=(x_1,\ldots,x_p)'$ の変数選択問題として考える．変数 $\boldsymbol{x}$ を $\boldsymbol{x}=(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})'$ と分割して，$\boldsymbol{x}^{(2)}$ の冗長性モデルを考える．この冗長性モデルは

$$\mathrm{tr}\Sigma_{11}^{-1}\Sigma_{1y}\Sigma_{yy}^{-1}\Sigma_{y1} = \mathrm{tr}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \tag{3.1}$$

として定義される．ここに，$\Sigma_{11}; p_1\times p_1$，$\Sigma_{1y}; p_1\times q$ であり，これらは，$\boldsymbol{x}=(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})'$，$\boldsymbol{y}$ に対応した Σ の分割

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1y} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2y} \\ \Sigma_{y1} & \Sigma_{y2} & \Sigma_{yy} \end{pmatrix}, \qquad \begin{matrix} \Sigma_{11} : p_1\times p_1, \\ \Sigma_{22} : p_2\times p_2, \\ \Sigma_{yy} : q\times q \end{matrix}$$

における部分行列である．また，冗長性条件 (3.1) は，

$$M_1 :\ \Sigma_{2y\cdot 1} \equiv \Sigma_{2y} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{1y} = O \tag{3.2}$$

と同値である．ここに，$\Sigma_{ij\cdot k} = \Sigma_{ij} - \Sigma_{ik}\Sigma_{kk}^{-1}\Sigma_{kj}$ であり，他の行列についても同様な記号を用いる．

モデル M_1 のもとでの最小距離推定量や規準量の導出においては，次の補題を利用する．(例えば，Siotani, Hayakawa and Fujikoshi [6] を参照)

補題 3.1 行列 A，C は同じ次数の正則行列で，次のように分割する．

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}.$$

ここで，A_{11}，C_{11} は正方行列とする．このとき，

(1) $|A| = |A_{11}||A_{22\cdot 1}|$.

(2) $A^{-1} = \begin{pmatrix} A_{11}^{-1} & O \\ O & O \end{pmatrix} + \begin{pmatrix} -A_{11}^{-1}A_{12} \\ I \end{pmatrix} A_{22\cdot 1}^{-1} \begin{pmatrix} -A_{21}A_{11}^{-1} & I \end{pmatrix}$.

(3) $\mathrm{tr}A^{-1}C = \mathrm{tr}A_{11}^{-1}C_{11} + \mathrm{tr}A_{22\cdot 1}^{-1}\left\{C_{22\cdot 1} + (A_{21}A_{11}^{-1} - C_{21}C_{11}^{-1})C_{11}(A_{21}A_{11}^{-1} - C_{21}C_{11}^{-1})'\right\}$.

(4) 正定値行列 A の関数 $f(A) = \log|A| + \mathrm{tr}A^{-1}C$ を考える．
ここに, C はある定数正定値行列である．このとき, $f(A)$ は $A = C$ で最小となる．

上記の補題より，モデル M_1 のもとでの最小距離推定量は

$$\hat{\Sigma}_{22\cdot 1} = S_{22\cdot 1}, \quad \hat{\Sigma}_{yy\cdot 1} = S_{yy\cdot 1}, \quad \hat{\Sigma}_{2y\cdot 1} = O,$$
$$\hat{B}_{21} = S_{21}S_{11}^{-1}, \quad \hat{B}_{y1} = S_{y1}S_{11}^{-1}, \quad \hat{\Sigma}_{11} = S_{11}$$

となる．なぜなら，Σ と S を

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1y} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2y} \\ \Sigma_{y1} & \Sigma_{y2} & \Sigma_{yy} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{(2y)1} \\ \Sigma_{(2y)1} & \Sigma_{(2y)(2y)} \end{pmatrix}, \quad \Sigma_{(2y)(2y)} : (p_2+q)\times(p_2+q)$$

$$S = \begin{pmatrix} S_{11} & S_{12} & S_{1y} \\ S_{21} & S_{22} & S_{2y} \\ S_{y1} & S_{y2} & S_{yy} \end{pmatrix} = \begin{pmatrix} S_{11} & S_{(2y)1} \\ S_{(2y)1} & S_{(2y)(2y)} \end{pmatrix}, \quad S_{(2y)(2y)} : (p_2+q)\times(p_2+q)$$

と分割すると補題 3.1 (1)，(2)，(3) から

$$|\Sigma| = |\Sigma_{11}|\cdot|\Sigma_{(2y)(2y)\cdot 1}| = |\Sigma_{11}|\cdot|\Sigma_{22\cdot 1}|\cdot|\Sigma_{yy\cdot 1}|, \tag{3.3}$$

$$\begin{aligned} \mathrm{tr}\Sigma^{-1}S &= \mathrm{tr}\Sigma_{11}^{-1}S_{11} + \mathrm{tr}\Sigma_{(2y)(2y)\cdot 1}^{-1}\left\{S_{(2y)(2y)\cdot 1} + (B_{(2y)1} - S_{(2y)1}S_{11}^{-1})S_{11}(B_{(2y)1} - S_{(2y)1}S_{11}^{-1})'\right\} \\ &= \mathrm{tr}\Sigma_{11}^{-1}S_{11} \\ &\quad + \mathrm{tr}\Sigma_{22\cdot 1}^{-1}\left\{S_{22\cdot 1} + (B_{21} - S_{21}S_{11}^{-1})S_{11}(B_{21} - S_{21}S_{11}^{-1})'\right\} \\ &\quad + \mathrm{tr}\Sigma_{yy\cdot 1}^{-1}\left\{S_{yy\cdot 1} + (B_{y1} - S_{y1}S_{11}^{-1})S_{11}(B_{y1} - S_{y1}S_{11}^{-1})'\right\}, \end{aligned} \tag{3.4}$$

となる．ここに $B_{(2y)1} = \Sigma_{(2y)1}\Sigma_{11}^{-1}, B_{21} = \Sigma_{21}\Sigma_{11}^{-1}$，$B_{y1} = \Sigma_{y1}\Sigma_{11}^{-1}$ である．これらの表示と補題 3.1 (4) から結論を得る．

このとき，正規性と

$$\mathrm{A0}; \quad p, q : \mathrm{fix}, \quad n \to \infty$$

の仮定のもとで B_1 を漸近的に求め，次の規準が提案されている．

$$DIC = -n\log\left\{\frac{|S_{(2y)(2y)\cdot 1}|}{|S_{22\cdot 1}|\cdot|S_{yy\cdot 1}|}\right\} + B_0. \tag{3.5}$$

ここで,

$$B_0 = 2\left\{\frac{(p+q)(p+q+1)}{2} - p_2 q\right\}.$$

また，真のモデルが候補のモデルに含まれるという条件のもとで B_1 を求めることによって，次の修正規準 $CDIC$ が提案されている (Fujikoshi [2]，Fujikoshi and Kurata [3])．

$$CDIC = -n\log\left\{\frac{|S_{(2y)(2y)\cdot 1}|}{|S_{22\cdot 1}|\cdot|S_{yy\cdot 1}|}\right\} + B_c. \tag{3.6}$$

ここで，

$$B_c = n\left\{\frac{np}{n-p-1} + \frac{n(p_1+q)}{n-p_1-q-1} - \frac{np_1}{n-p_1-1} - (p+q)\right\}.$$

なお，B_c は

$$B_c = B_0 + \frac{p(p+1)^2}{n+p+1} + \frac{(p_1+q)(p_1+q+1)^2}{n+p_1+q+1} - \frac{p_1(p_1+1)^2}{n+p_1+1}.$$

と表すことができる．したがって，B_c は漸近的には B_0 に一致している．

3.2 2組の冗長性モデル

ここでは，$\boldsymbol{x} = (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})'$，$\boldsymbol{y} = (\boldsymbol{y}^{(3)}, \boldsymbol{y}^{(4)})'$ と分割して，$\boldsymbol{x}^{(2)}$ と $\boldsymbol{y}^{(4)}$ の冗長性モデルを考える．分割 $\boldsymbol{x} = (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})'$，$\boldsymbol{y} = (\boldsymbol{y}^{(3)}, \boldsymbol{y}^{(4)})'$ に対応して Σ を

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{pmatrix}, \qquad \begin{array}{c} \Sigma_{11} : p_1 \times p_1, \\ \Sigma_{22} : p_2 \times p_2, \\ \Sigma_{33} : q_3 \times q_3, \\ \Sigma_{44} : q_4 \times q_4 \end{array}$$

と分割する．このとき，$\boldsymbol{x}^{(2)}$ と $\boldsymbol{y}^{(4)}$ の冗長性モデルは

$$\mathrm{tr}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} = \mathrm{tr}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \tag{3.7}$$

として定義される．この条件は，

$$\begin{aligned} M_2 :\ & \Sigma_{24\cdot 13} = O, \quad \Sigma_{23\cdot 1} = O, \quad \Sigma_{14\cdot 3} = O. \\ \Leftrightarrow\ & \Sigma_{24\cdot 13} = O, \quad B_{23} = O, \quad B_{41} = O \end{aligned} \tag{3.8}$$

と表せる．ここに，$\Sigma_{24\cdot 13} = \Sigma_{24} - \Sigma_{2(13)}\Sigma_{(13)(13)}^{-1}\Sigma_{(13)4}$，

$$B_{(24)(13)} = \begin{pmatrix} B_{21} & B_{23} \\ B_{41} & B_{43} \end{pmatrix} = \begin{pmatrix} \Sigma_{21} & \Sigma_{23} \\ \Sigma_{41} & \Sigma_{43} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}^{-1},$$

$$\Sigma_{(13)(13)} = \begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}, \quad \Sigma_{2(13)} = \begin{pmatrix} \Sigma_{21} & \Sigma_{23} \end{pmatrix}, \quad \Sigma_{(13)4} = \begin{pmatrix} \Sigma_{14} \\ \Sigma_{34} \end{pmatrix}$$

である．さらに，補題 3.1 (1)，(2)，(3) から

$$\begin{aligned} |\Sigma| &= |\Sigma_{(13)(13)}|\cdot|\Sigma_{(24)(24)\cdot 13}| = |\Sigma_{(13)(13)}|\cdot|\Sigma_{22\cdot 13}|\cdot|\Sigma_{44\cdot 13}| \\ \mathrm{tr}\Sigma^{-1}S &= \mathrm{tr}\Sigma_{(13)(13)}^{-1}S_{(13)(13)} \\ &\quad + \mathrm{tr}\Sigma_{(24)(24)\cdot 13}^{-1}\Big\{S_{(24)(24)\cdot 13} \\ &\quad + (B_{(24)(13)} - S_{(24)(13)}S_{(13)(13)}^{-1})S_{(13)(13)}(B_{(24)(13)} - S_{(24)(13)}S_{(13)(13)}^{-1})'\Big\} \\ &= \mathrm{tr}\Sigma_{(13)(13)}^{-1}S_{(13)(13)} \\ &\quad + \mathrm{tr}\Sigma_{22\cdot 13}^{-1}\left\{S_{22\cdot 1} + (B_{21} - S_{21}S_{11}^{-1})S_{11}(B_{21} - S_{21}S_{11}^{-1})'\right\} \\ &\quad + \mathrm{tr}\Sigma_{33\cdot 1}^{-1}\left\{S_{44\cdot 3} + (B_{43} - S_{43}S_{33}^{-1})S_{33}(B_{43} - S_{43}S_{33}^{-1})'\right\} \end{aligned}$$

となる．ここに，S_{ij} は Σ_{ij} に対応する S の部分行列であり，

$$S_{(13)(13)} = \begin{pmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{pmatrix}, \quad S_{(24)(13)} = \begin{pmatrix} S_{21} & S_{23} \\ S_{41} & S_{43} \end{pmatrix}$$

である．1 組の場合と同様に，モデル M_2 のもとでの最小距離推定量は次のように与えられる．

$$\hat{\Sigma}_{22\cdot13} = S_{22\cdot1}, \quad \hat{\Sigma}_{44\cdot13} = S_{44\cdot3}, \quad \hat{\Sigma}_{24\cdot13} = O, \quad \hat{\Sigma}_{(13)(13)} = S_{(13)(13)},$$
$$\hat{B}_{21} = S_{21}S_{11}^{-1}, \quad \hat{B}_{24} = O, \quad \hat{B}_{41} = O, \quad \hat{B}_{43} = S_{43}S_{33}^{-1}.$$

このとき，正規性と

$$\mathrm{A0}; \quad p, q : \mathrm{fix}, \quad n \to \infty$$

の仮定のもとで B_2 を漸近的に求め，次の変数選択規準が提案されている．

$$DIC = -n \log \left\{ \frac{|S_{(24)(24)\cdot13}|}{|S_{22\cdot1}| \cdot |S_{44\cdot3}|} \right\} + B_0. \tag{3.9}$$

ここで，

$$B_0 = 2\left\{ \frac{(p+q)(p+q+1)}{2} - p_2q_4 - p_1q_4 - p_2q_3 \right\}.$$

また，真のモデルが候補のモデルに含まれるという条件のもとで B_2 を求めることによって，次の修正規準 $CDIC$ が提案されている (Fujikoshi and Kurata [3])．

$$CDIC = -n \log \left\{ \frac{|S_{(24)(24)\cdot13}|}{|S_{22\cdot1}| \cdot |S_{44\cdot3}|} \right\} + B_c. \tag{3.10}$$

ここで，

$$B_c = n\left\{ \frac{n(p_1+q_3)}{n-p_1-q_3-1} + \frac{np}{n-p-1} + \frac{nq}{n-q-1} - \frac{np_1}{n-p_1-1} - \frac{nq_3}{n-q_3-1} - (p+q) \right\}.$$

なお，B_c は

$$B_c = B_0 + \frac{(p_1+q_3)(p_1+q_3+1)^2}{n+p_1+q_3+1} + \frac{p(p+1)^2}{n+p+1} + \frac{q(q+1)^2}{n+q+1} - \frac{p_1(p_1+1)^2}{n+p_1+1} - \frac{q_3(q_3+1)^2}{n+q_3+1}$$

と表すことができる．したがって，B_c は漸近的には B_0 に一致している．

4　ブートストラップ変数選択規準

真のモデルが正規でない場合には，候補のモデルとして非正規分布を想定する必要がある．この場合には偏りの項 B_k の別な推定法が重要となる．そこで，本論文では偏りの項をブートストラップ法を適用し数値的に求めることを考え，これをブートストラップ変数選択規準とよぶ．

偏りの項は第 2 章で示しているように，一般に

$$B_k = \mathrm{E}_S^*[n\mathrm{tr}\hat{\Sigma}_k^{-1}\Sigma] - np$$

と表せる．この期待値をブートストラップ法で近似するには，Σ を標本共分散行列で置き換え，$\hat{\Sigma}_k$ をブートストラップ標本から求められる標本共分散行列で置き換えたものを，繰り返し計算しそれらの平均を考えればよい (Hall [4])．具体的には，ブートストラップ回数を B 回とし，次のようにして計算される．

4.1　1 組の冗長性モデル

1 組の冗長性モデルの場合，次のようにしてブートストラップ変数選択規準を求める．

(1) $j = 1$ とする.

(2) $j \leq B$ である間は (2)∼(5) を繰り返す.

(3) 標本 X, Y からのリサンプルであるブートストラップ標本

$$X^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_N^*), \quad Y^* = (\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_N^*)$$

を作る.

(4) 次に, ブートストラップ標本に基づく標本共分散行列を S^* として, S^* に基づく Σ の推定量を求める.

$$\hat{\Sigma}_{22\cdot1}^* = S_{22\cdot1}^*, \quad \hat{\Sigma}_{yy\cdot1}^* = S_{yy\cdot1}^*, \quad \hat{\Sigma}_{2y\cdot1}^* = O,$$
$$\hat{B}_{21}^* = S_{21}^* S_{11}^{*-1}, \quad \hat{B}_{y1}^* = S_{y1}^* S_{11}^{*-1}, \quad \hat{\Sigma}_{11}^* = S_{11}^*.$$

(5) 偏りの項

$$B_k^{(j)} = \frac{n^2}{N} \mathrm{tr} \hat{\Sigma}^{*-1} \hat{\Sigma} - n(p+q)$$

を計算し, $j = j + 1$ とする.

(6) ブートストラップ変数選択規準を計算する.

$$BDIC = -n \log \left\{ \frac{|S_{(2y)(2y)\cdot1}|}{|S_{22\cdot1}| \cdot |S_{yy\cdot1}|} \right\} + \frac{1}{B} \sum_{j=1}^{B} B_k^{(j)}.$$

なお, 偏りの項の計算においては,

$$\begin{aligned} B_k^{(j)} &= \frac{n^2}{N} \mathrm{tr} \hat{\Sigma}^{*-1} \hat{\Sigma} - n(p+q) \\ &= \frac{n^2}{N} \Big\{ \mathrm{tr} \hat{\Sigma}_{11}^{*-1} S_{11} + \mathrm{tr} \hat{\Sigma}_{22\cdot1}^{*-1} \left\{ S_{22\cdot1} + (\hat{B}_{21}^* - S_{21} S_{11}^{-1}) S_{11} (\hat{B}_{21}^* - S_{21} S_{11}^{-1})' \right\} \\ &\qquad + \mathrm{tr} \hat{\Sigma}_{yy\cdot1}^{*-1} \left\{ S_{yy\cdot1} + (\hat{B}_{y1}^* - S_{y1} S_{11}^{-1}) S_{11} (\hat{B}_{y1}^* - S_{y1} S_{11}^{-1})' \right\} \Big\} - n(p+q) \\ &= \frac{n^2}{N} \left\{ \mathrm{tr} S_{xx}^{*-1} S_{xx} + \mathrm{tr} \begin{pmatrix} S_{11}^* & S_{1y}^* \\ S_{y1}^* & S_{yy}^* \end{pmatrix}^{-1} \begin{pmatrix} S_{11} & S_{1y} \\ S_{y1} & S_{yy} \end{pmatrix} - \mathrm{tr} S_{11}^{*-1} S_{11} \right\} - n(p+q) \end{aligned}$$

を利用することができる.

4.2 2 組の冗長性モデル

2 組の冗長性モデルの場合, 次のようにしてブートストラップ変数選択規準を求める.

(1) $j = 1$ とする.

(2) $j \leq B$ である間は (2)∼(5) を繰り返す.

(3) 標本 X, Y からのリサンプルであるブートストラップ標本

$$X^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_N^*), \quad Y^* = (\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_N^*)$$

を作る.

(4) 次に，ブートストラップ標本に基づく標本共分散行列を S^* として，S^* に基づく Σ の推定量を求める．

$$\hat{\Sigma}^*_{22\cdot 13} = S^*_{22\cdot 1}, \quad \hat{\Sigma}^*_{44\cdot 13} = S^*_{44\cdot 3}, \quad \hat{\Sigma}^*_{24\cdot 13} = O, \quad \hat{\Sigma}^*_{(13)(13)} = S^*_{(13)(13)},$$
$$\hat{B}^*_{21} = S_{21}S^{*-1}_{11}, \quad \hat{B}^*_{24} = O, \quad \hat{B}^*_{41} = O, \quad \hat{B}^*_{43} = S_{43}S^{*-1}_{33}.$$

(5) 偏りの項

$$B_k^{(j)} = \frac{n^2}{N}\mathrm{tr}\hat{\Sigma}^{*-1}\hat{\Sigma} - n(p+q)$$

を計算し，$j = j+1$ とする．

(6) ブートストラップ変数選択規準を計算する．

$$BDIC = -n\log\left\{\frac{|S_{(24)(24)\cdot 13}|}{|S_{22\cdot 1}|\cdot|S_{44\cdot 3}|}\right\} + \frac{1}{B}\sum_{j=1}^{B} B_k^{(j)}.$$

なお，偏りの項の計算においては，

$$\begin{aligned}
B_k^{(j)} &= \frac{n^2}{N}\Bigg\{\mathrm{tr}\hat{\Sigma}^{*-1}_{(13)(13)}S_{(13)(13)} + \mathrm{tr}\hat{\Sigma}^{*-1}_{22\cdot 13}\left\{S_{22\cdot 1} + (\hat{B}^*_{21} - S_{21}S^{-1}_{11})S_{11}(\hat{B}^*_{21} - S_{21}S^{-1}_{11})'\right\} \\
&\qquad + \mathrm{tr}\hat{\Sigma}^{*-1}_{33\cdot 1}\left\{S_{44\cdot 3} + (\hat{B}^*_{43} - S_{43}S^{-1}_{33})S_{33}(\hat{B}^*_{43} - S_{43}S^{-1}_{33})'\right\} - n(p+q) \\
&= \frac{n^2}{N}\Bigg\{\mathrm{tr}\begin{pmatrix} S^*_{11} & S^*_{13} \\ S^*_{31} & S^*_{33}\end{pmatrix}^{-1}\begin{pmatrix} S_{11} & S_{13} \\ S_{31} & S_{33}\end{pmatrix} + \mathrm{tr}S^{*-1}_{xx}S_{xx} + \mathrm{tr}S^{*-1}_{yy}S_{yy} - \mathrm{tr}S^{*-1}_{11}S_{11} - \mathrm{tr}S^{*-1}_{33}S_{33}\Bigg\} \\
&\quad -n(p+q)
\end{aligned}$$

を利用することができる．

5 シミュレーション

ここでは，実際に $BDIC$ が R_k の漸近不偏推定量になっていることをシミュレーションを通して確認する．さらに，DIC，$CDIC$ と比較し，非正規のもとで $BDIC$ が R_k の推定量としてよいことを確認する．ここで，シミュレーション回数を 10^4 とし，ブートストラップ回数を 10^3 として，次の 2 通りの場合を考えた．何れの場合も，$p = q = 4$ とし，$p_1 = q_1 = 2$ で，x_3, x_4 と y_3, y_4 が冗長であるというモデルである．具体的には，$\boldsymbol{x}^{(1)}$，$\boldsymbol{y}^{(3)}$ から次のように $\boldsymbol{x}^{(2)}$，$\boldsymbol{y}^{(4)}$ を構成している．

$$\boldsymbol{x}^{(2)} = A\boldsymbol{x}^{(1)} + \boldsymbol{\varepsilon}_1, \quad \boldsymbol{y}^{(4)} = B\boldsymbol{y}^{(3)} + \boldsymbol{\varepsilon}_2.$$

ここで，A，B は定数行列で，$\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$ は $\boldsymbol{x}^{(1)}$，$\boldsymbol{y}^{(3)}$ とは独立な分布に従う確率変数である．定数行列 A，B および $\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$ の共分散行列を適当に与えることによって，シミュレーションで用いる共分散行列を定めた．また，Case 1 が共分散行列 Σ の多変量正規分布で，Case 2 が Case 1 と同等な相関行列をもつ自由度 N の多

変量 $\boldsymbol{t}$ 分布である．

$$\begin{array}{c} \text{Case 1} \\ N_8(\mathbf{0},\Sigma) \end{array} \quad \Sigma = \begin{pmatrix} 1.00 & 0.00 & 1.000 & 1.300 & 0.10 & 0.20 & 0.470 & 0.530 \\ 0.00 & 1.00 & 1.200 & 1.400 & 0.20 & 0.10 & 0.460 & 0.520 \\ 1.00 & 1.20 & 5.440 & 3.580 & 0.34 & 0.32 & 1.022 & 1.154 \\ 1.30 & 1.40 & 3.580 & 8.050 & 0.41 & 0.40 & 1.255 & 1.417 \\ 0.10 & 0.20 & 0.340 & 0.410 & 1.00 & 0.00 & 1.500 & 1.700 \\ 0.20 & 0.10 & 0.320 & 0.400 & 0.00 & 1.00 & 1.600 & 1.800 \\ 0.47 & 0.46 & 1.022 & 1.255 & 1.50 & 1.60 & 6.810 & 5.630 \\ 0.53 & 0.52 & 1.154 & 1.417 & 1.70 & 1.80 & 5.630 & 11.730 \end{pmatrix},$$

$$\begin{array}{c} \text{Case 2} \\ T_8(N,\Phi) \end{array} \quad \Phi = \begin{pmatrix} 1.00 & 0.00 & 0.43 & 0.46 & 0.10 & 0.20 & 0.18 & 0.15 \\ 0.00 & 1.00 & 0.51 & 0.49 & 0.20 & 0.10 & 0.18 & 0.15 \\ 0.43 & 0.51 & 1.00 & 0.54 & 0.15 & 0.14 & 0.17 & 0.14 \\ 0.46 & 0.49 & 0.54 & 1.00 & 0.14 & 0.14 & 0.17 & 0.15 \\ 0.10 & 0.20 & 0.15 & 0.14 & 1.00 & 0.00 & 0.57 & 0.50 \\ 0.20 & 0.10 & 0.14 & 0.14 & 0.00 & 1.00 & 0.61 & 0.53 \\ 0.18 & 0.18 & 0.17 & 0.17 & 0.57 & 0.61 & 1.00 & 0.63 \\ 0.15 & 0.15 & 0.14 & 0.15 & 0.50 & 0.53 & 0.63 & 1.00 \end{pmatrix}.$$

このとき，適当な $N = 30, 100, 300$ の場合に，リスク，リスクの推定量 (DIC，$CDIC$，$BDIC$) の平均および選択確率をシミュレーションによって求めた．ここで，Risk:R_k は，$(\boldsymbol{x}, \boldsymbol{y})'$ とは独立な $(\boldsymbol{x}^*, \boldsymbol{y}^*)'$ を発生させ数値的に求めた．

また，それぞれのグラフにおいて，□が真のリスク R_k，○が DIC，△が $CDIC$，×が $BDIC$ である．グラフにおける横軸は，p_1 の値を示している．

この結果より，正規性を仮定した場合において，$N = 30$ では $BDIC$ の当てはまりは良いとは言えない．しかし，$N = 100$ ぐらいの，ある程度の標本の大きさが得ることができれば，リスクの推定量として妥当な結果を与えることがわかった．また，傾向として，$N = 30$ では DIC は比較的，大きなモデルを選ぶ傾向があり，$CDIC$，$BDIC$ は真のモデルより小さいモデルを選ぶ傾向がある．しかし，$N = 100$ 以上になると DIC，$CDIC$，$BDIC$ に違いはなくなる．次に，非正規の場合においては，DIC や $CDIC$ は リスク の推定量として妥当ではないが，$BDIC$ は妥当な推定量であることがわかる．このとき，真のモデルを選ぶ確率は，$BDIC$ がもっとも高いことがわかる．

表 1. リスク・規準量の期待値

Case 1 ($N=30$)

R_k	$q_1=1$	2^*	3	4
$p_1=1$	71.79	72.21	74.12	76.36
2^*	72.28	74.17	78.37	83.35
3	74.21	78.33	85.17	93.36
4	76.46	83.21	93.15	105.21
DIC	$q_1=1$	2^*	3	4
$p_1=1$	62.72	62.42	63.27	64.08
2^*	62.36	62.61	64.26	65.83
3	63.21	64.28	66.72	69.00
4	64.03	65.87	69.07	72.00
$CDIC$	$q_1=1$	2^*	3	4
$p_1=1$	71.44	71.84	73.71	75.92
2^*	71.77	73.73	77.81	82.67
3	73.65	77.82	84.57	92.58
4	75.87	82.71	92.65	104.40
$BDIC$	$q_1=1$	2^*	3	4
$p_1=1$	72.45	73.02	75.31	78.35
2^*	72.95	75.51	80.87	88.04
3	75.27	80.88	90.41	103.29
4	78.32	88.09	103.35	124.45

Case 2 ($N=30$)

R_k	$q_1=1$	2^*	3	4
$p_1=1$	90.45	91.17	93.59	96.34
2^*	91.19	93.84	99.13	105.26
3	93.57	98.98	107.54	117.72
4	96.41	105.20	117.87	132.79
DIC	$q_1=1$	2^*	3	4
$p_1=1$	66.18	65.65	66.28	66.87
2^*	65.69	65.48	66.70	67.82
3	66.31	66.67	68.42	70.02
4	66.88	67.77	69.96	72.00
$CDIC$	$q_1=1$	2^*	3	4
$p_1=1$	74.91	75.07	76.71	78.71
2^*	75.10	76.60	80.25	84.65
3	76.74	80.22	86.28	93.60
4	78.73	84.60	93.54	104.40
$BDIC$	$q_1=1$	2^*	3	4
$p_1=1$	86.93	87.75	90.43	94.05
2^*	87.80	91.02	97.29	105.79
3	90.48	97.25	108.41	123.78
4	94.08	105.74	123.63	148.82

Case 1 ($N=100$)

R_k	$q_1=1$	2^*	3	4
$p_1=1$	69.35	66.36	67.55	68.81
2^*	66.38	63.48	65.99	68.63
3	67.61	65.98	69.84	73.87
4	68.85	68.53	73.77	79.25
DIC	$q_1=1$	2^*	3	4
$p_1=1$	67.00	63.95	64.92	65.86
2^*	63.94	60.65	62.56	64.43
3	64.90	62.56	65.41	68.22
4	65.85	64.46	68.27	72.00
$CDIC$	$q_1=1$	2^*	3	4
$p_1=1$	69.23	66.35	67.56	68.82
2^*	66.34	63.47	65.93	68.50
3	67.55	65.93	69.73	73.70
4	68.81	68.53	73.75	79.20
$BDIC$	$q_1=1$	2^*	3	4
$p_1=1$	68.62	65.72	66.92	68.21
2^*	65.71	62.82	65.29	67.91
3	66.92	65.30	69.13	73.21
4	68.20	67.94	73.25	78.89

Case 2 ($N=100$)

R_k	$q_1=1$	2^*	3	4
$p_1=1$	88.92	86.25	87.80	89.42
2^*	86.46	84.22	87.41	90.79
3	88.03	87.39	92.29	97.46
4	89.65	90.79	97.45	104.56
DIC	$q_1=1$	2^*	3	4
$p_1=1$	71.48	68.14	68.78	69.43
2^*	67.98	64.14	65.45	66.76
3	68.64	65.48	67.45	69.42
4	69.28	66.77	69.39	72.00
$CDIC$	$q_1=1$	2^*	3	4
$p_1=1$	73.71	70.55	71.43	72.39
2^*	70.38	66.96	68.82	70.83
3	71.28	68.85	71.76	74.90
4	72.24	70.84	74.87	79.20
$BDIC$	$q_1=1$	2^*	3	4
$p_1=1$	85.78	83.08	84.44	85.89
2^*	82.92	80.42	83.25	86.26
3	84.31	83.29	87.67	92.38
4	85.75	86.30	92.36	98.85

Case 1 ($N = 300$)

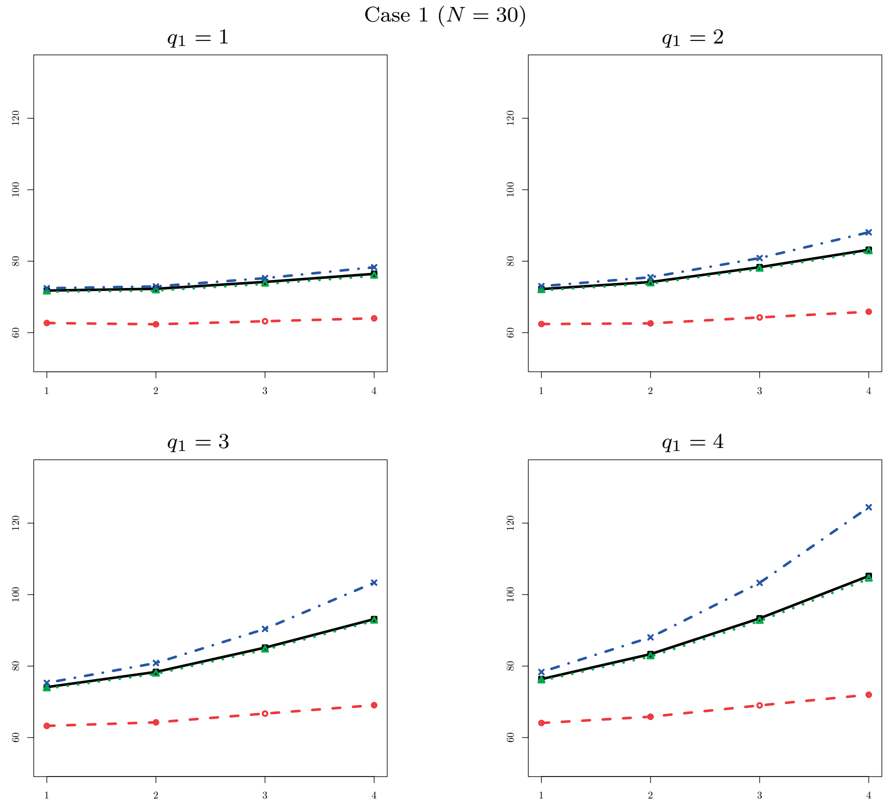
R_k	$q_1 = 1$	2^*	3	4
$p_1 = 1$	86.07	74.78	75.85	76.88
2^*	74.91	61.17	63.32	65.48
3	76.01	63.37	66.66	69.90
4	77.11	65.55	70.00	74.35
DIC	$q_1 = 1$	2^*	3	4
$p_1 = 1$	85.38	74.10	75.08	76.09
2^*	74.18	60.28	62.25	64.21
3	75.13	62.21	65.13	68.09
4	76.11	64.17	68.06	72.00
$CDIC$	$q_1 = 1$	2^*	3	4
$p_1 = 1$	86.10	74.87	75.92	77.03
2^*	74.95	61.18	63.32	65.49
3	75.98	63.28	66.49	69.80
4	77.06	65.46	69.78	74.23
$BDIC$	$q_1 = 1$	2^*	3	4
$p_1 = 1$	85.80	74.56	75.61	76.71
2^*	74.64	60.85	62.98	65.14
3	75.66	62.94	66.14	69.44
4	76.73	65.11	69.41	73.85

Case 2 ($N = 300$)

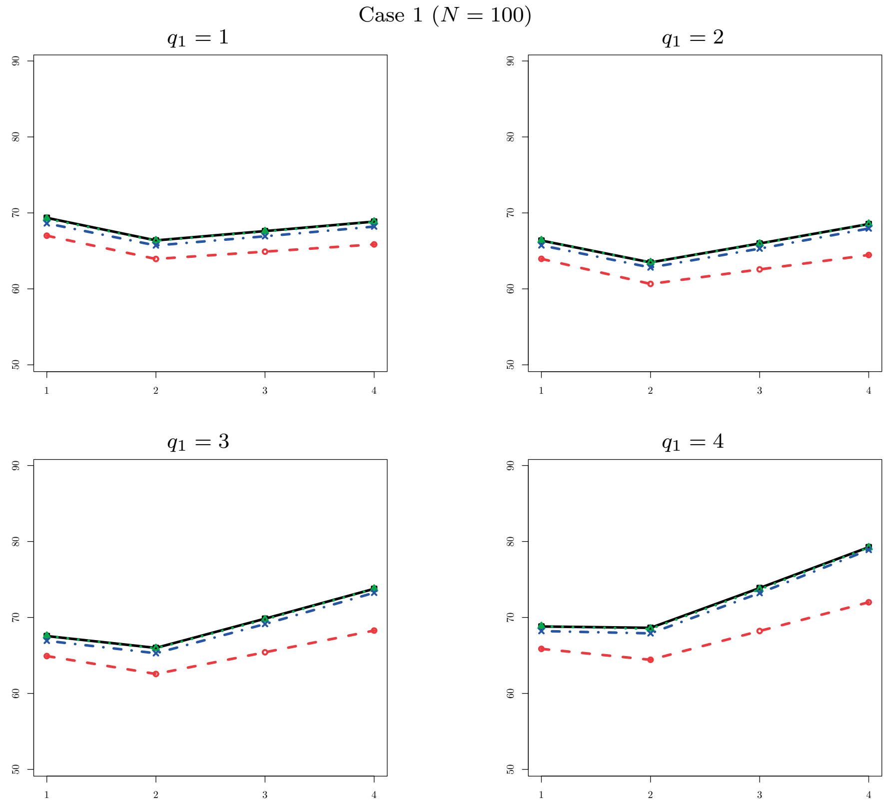
R_k	$q_1 = 1$	2^*	3	4
$p_1 = 1$	107.30	96.14	97.54	98.94
2^*	96.36	83.08	85.91	88.77
3	97.75	85.90	90.13	94.45
4	99.17	88.74	94.42	100.20
DIC	$q_1 = 1$	2^*	3	4
$p_1 = 1$	89.81	78.30	78.96	79.65
2^*	78.31	63.92	65.24	66.60
3	78.99	65.27	67.28	69.28
4	79.70	66.67	69.32	72.00
$CDIC$	$q_1 = 1$	2^*	3	4
$p_1 = 1$	90.53	79.07	79.80	80.59
2^*	79.08	64.81	66.31	67.88
3	79.83	66.35	68.64	71.00
4	80.64	67.96	71.04	74.23
$BDIC$	$q_1 = 1$	2^*	3	4
$p_1 = 1$	104.81	93.90	95.17	96.48
2^*	93.91	80.74	83.31	85.92
3	95.19	83.33	87.21	91.13
4	96.53	86.00	91.19	96.44

図 1. リスク・規準量の期待値のグラフ表示

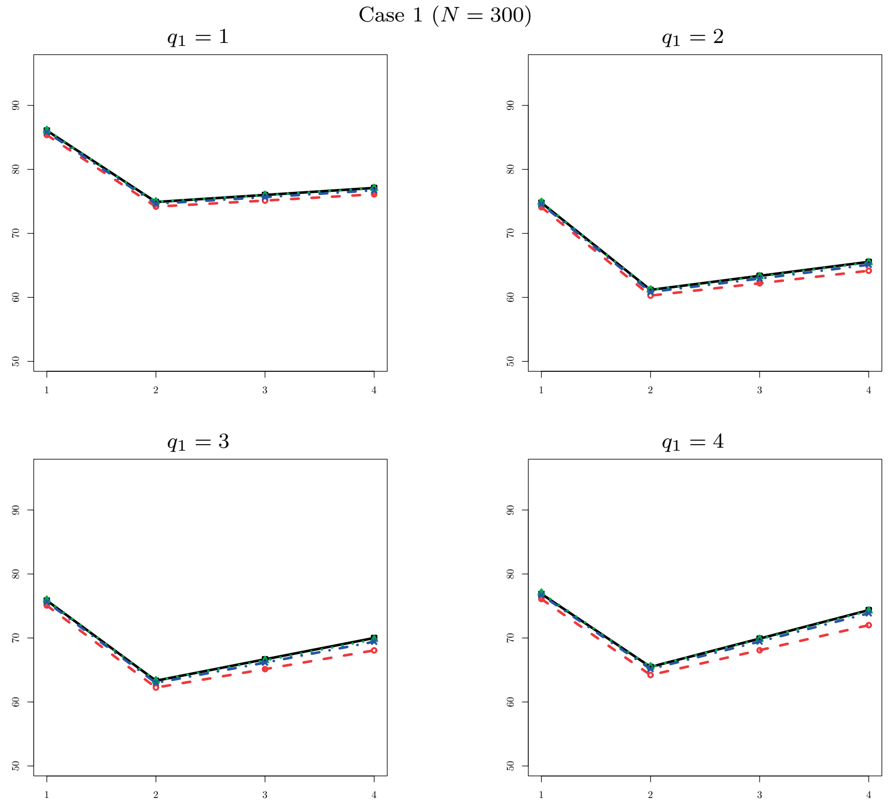
Case 1 ($N = 30$)



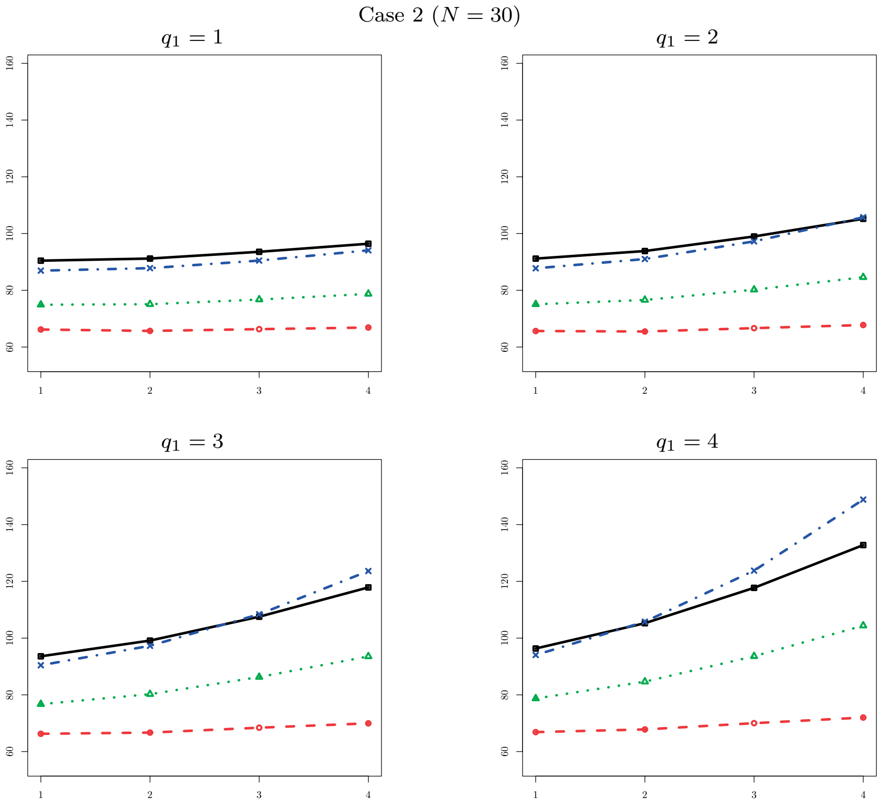
Case 1 ($N = 100$)



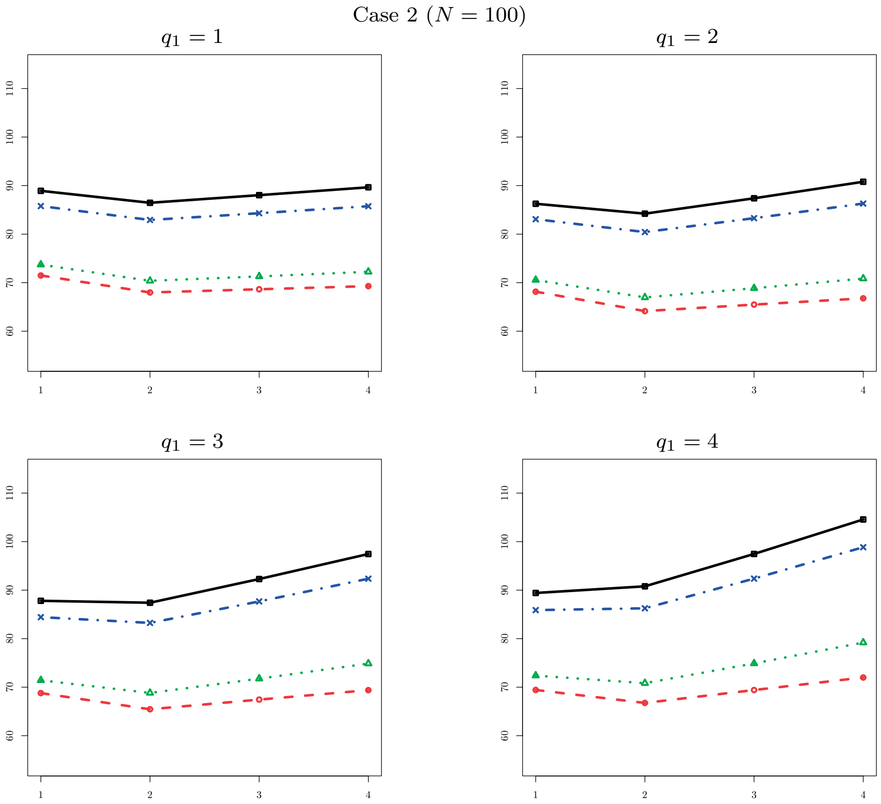
Case 1 ($N = 300$)



Case 2 ($N = 30$)



Case 2 ($N = 100$)



Case 2 ($N = 300$)

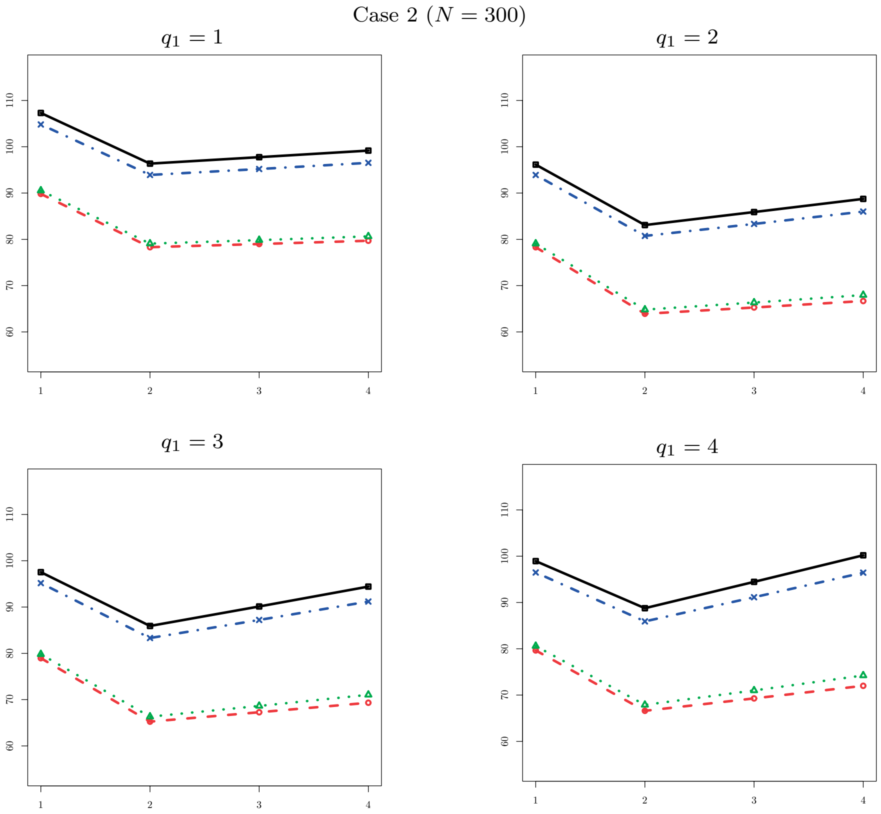


表 2. 選択確率

Case 1 ($N = 30$)

DIC	$q=1$	2^*	3	4
$p=1$	0.22	0.14	0.05	0.05
2^*	0.15	0.10	0.04	0.03
3	0.06	0.03	0.02	0.01
4	0.04	0.03	0.01	0.01
$CDIC$	$q=1$	2^*	3	4
$p=1$	0.39	0.19	0.04	0.02
2^*	0.19	0.07	0.01	0.00
3	0.04	0.01	0.00	0.00
4	0.02	0.00	0.00	0.00
$BDIC$	$q=1$	2^*	3	4
$p=1$	0.41	0.20	0.04	0.01
2^*	0.20	0.06	0.01	0.00
3	0.04	0.01	0.00	0.00
4	0.01	0.00	0.00	0.00

Case 2 ($N = 30$)

DIC	$q=1$	2^*	3	4
$p=1$	0.16	0.12	0.05	0.06
2^*	0.11	0.09	0.04	0.05
3	0.06	0.05	0.03	0.03
4	0.06	0.05	0.03	0.03
$CDIC$	$q=1$	2^*	3	4
$p=1$	0.34	0.18	0.05	0.03
2^*	0.17	0.08	0.02	0.01
3	0.05	0.02	0.01	0.00
4	0.03	0.01	0.00	0.00
$BDIC$	$q=1$	2^*	3	4
$p=1$	0.42	0.19	0.04	0.02
2^*	0.18	0.06	0.01	0.00
3	0.05	0.01	0.00	0.00
4	0.02	0.00	0.00	0.00

Case 1 ($N = 100$)

DIC	$q = 1$	2^*	3	4
$p = 1$	0.04	0.14	0.03	0.02
2^*	0.13	0.33	0.07	0.03
3	0.03	0.06	0.02	0.01
4	0.03	0.04	0.01	0.00
$CDIC$	$q = 1$	2^*	3	4
$p = 1$	0.05	0.16	0.04	0.02
2^*	0.16	0.34	0.05	0.02
3	0.04	0.05	0.01	0.00
4	0.02	0.02	0.00	0.00
$BDIC$	$q = 1$	2^*	3	4
$p = 1$	0.05	0.16	0.04	0.02
2^*	0.16	0.34	0.06	0.02
3	0.04	0.05	0.01	0.00
4	0.02	0.02	0.00	0.00

Case 2 ($N = 100$)

DIC	$q = 1$	2^*	3	4
$p = 1$	0.04	0.14	0.03	0.02
2^*	0.13	0.33	0.07	0.03
3	0.03	0.06	0.02	0.01
4	0.03	0.04	0.01	0.00
$CDIC$	$q = 1$	2^*	3	4
$p = 1$	0.05	0.16	0.04	0.02
2^*	0.16	0.34	0.05	0.02
3	0.04	0.05	0.01	0.00
4	0.02	0.02	0.00	0.00
$BDIC$	$q = 1$	2^*	3	4
$p = 1$	0.05	0.16	0.04	0.02
2^*	0.16	0.34	0.06	0.02
3	0.04	0.05	0.01	0.00
4	0.02	0.02	0.00	0.00

Case 1 ($N = 300$)

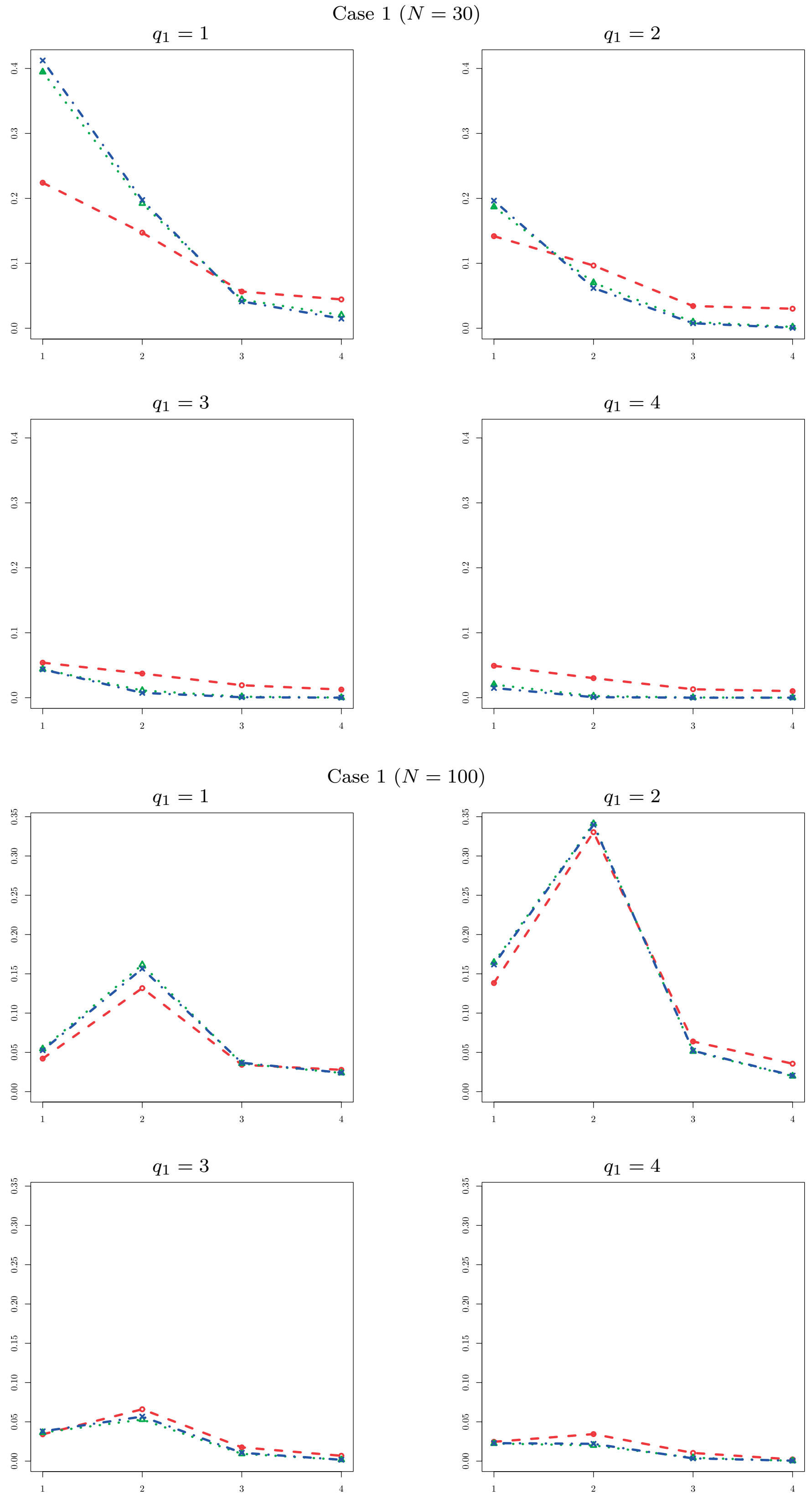
DIC	$q = 1$	2^*	3	4
$p = 1$	0.00	0.01	0.00	0.00
2^*	0.01	0.64	0.10	0.05
3	0.00	0.10	0.02	0.01
4	0.00	0.05	0.01	0.00
$CDIC$	$q = 1$	2^*	3	4
$p = 1$	0.00	0.01	0.00	0.00
2^*	0.01	0.66	0.09	0.04
3	0.00	0.09	0.02	0.01
4	0.00	0.05	0.01	0.00
$BDIC$	$q = 1$	2^*	3	4
$p = 1$	0.00	0.01	0.00	0.00
2^*	0.01	0.65	0.09	0.04
3	0.00	0.10	0.02	0.01
4	0.00	0.04	0.01	0.00

Case 2 ($N = 300$)

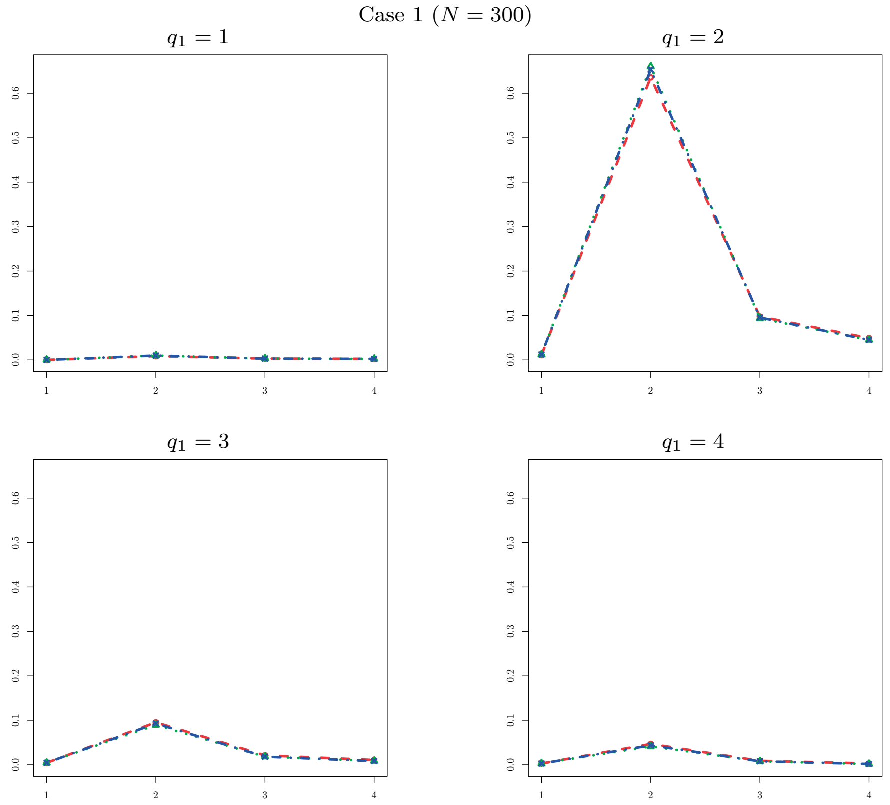
DIC	$q = 1$	2^*	3	4
$p = 1$	0.00	0.01	0.00	0.00
2^*	0.01	0.64	0.10	0.05
3	0.00	0.10	0.02	0.01
4	0.00	0.05	0.01	0.00
$CDIC$	$q = 1$	2^*	3	4
$p = 1$	0.00	0.01	0.00	0.00
2^*	0.01	0.66	0.09	0.04
3	0.00	0.09	0.02	0.01
4	0.00	0.05	0.01	0.00
$BDIC$	$q = 1$	2^*	3	4
$p = 1$	0.00	0.01	0.00	0.00
2^*	0.01	0.65	0.09	0.04
3	0.00	0.10	0.02	0.01
4	0.00	0.04	0.01	0.00

図 2. 選択確率のグラフ表示

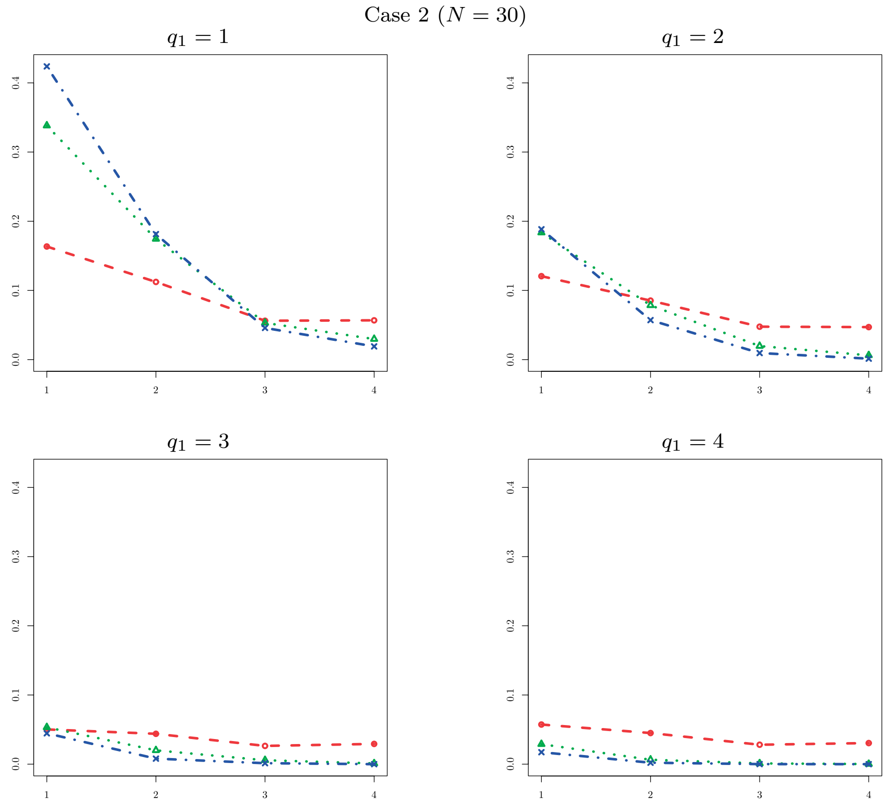
Case 1 ($N = 30$)



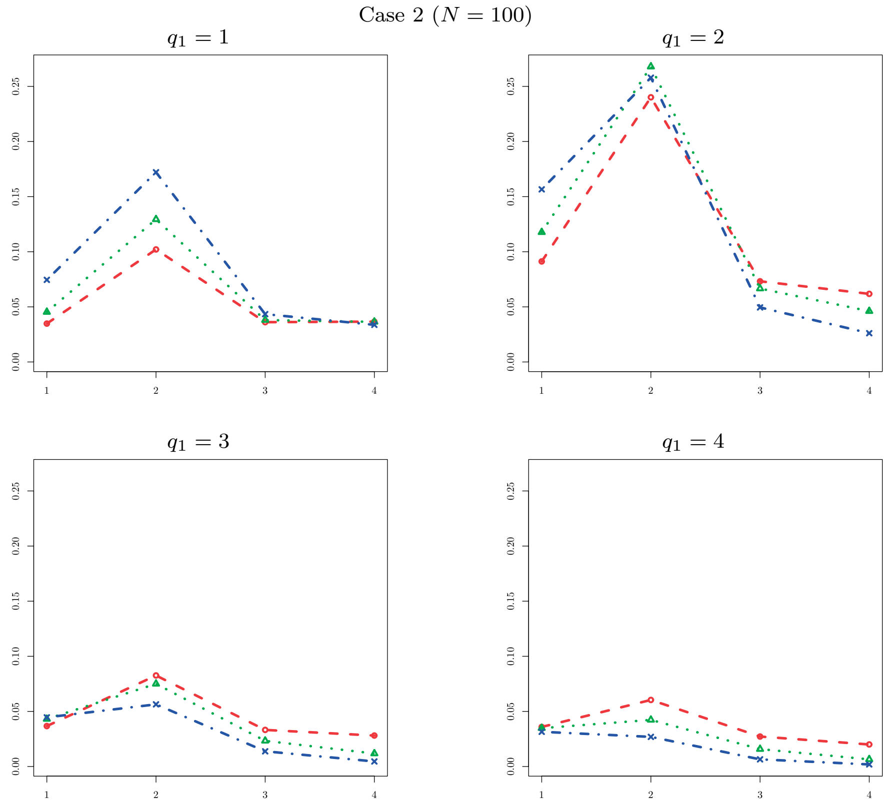
Case 1 ($N = 300$)



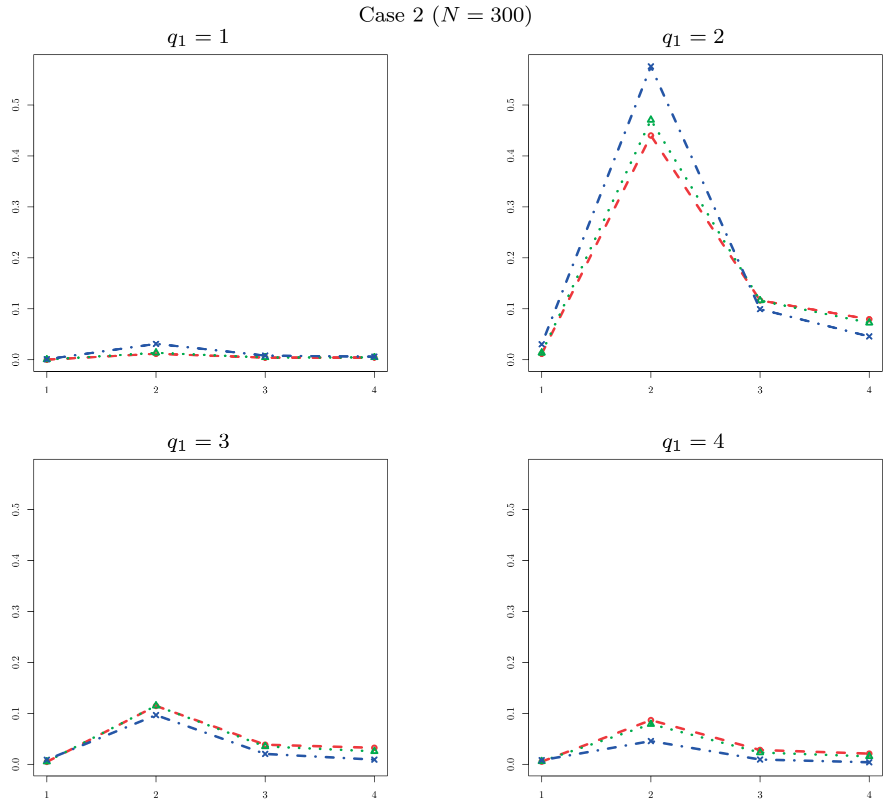
Case 2 ($N = 30$)



Case 2 ($N = 100$)



Case 2 ($N = 300$)



6 結論

本論文では，正準相関分析における変数選択問題を冗長性モデルの選択問題として考察した．とくに，第 4 節で提案した $BDIC$ の有効性についてシミュレーションにより調べた．この結果，非正規母集団の場合には，その次元や分布によって妥当な標本の大きさは異なるが，$N = 100$ ぐらいのある程度の標本の大きさが得られている場合には，$BDIC$ を用いる方が有用であることが指摘される．

謝辞

査読者の方には貴重なコメントを頂きました．ここに記して，お礼申し上げます．また本研究は，理工学研究所，共同研究第 2 類「多変量高次元データ解析の理論と応用」(2006 年度) から研究助成を受けております．

参考文献

[1] Akaike, H.(1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, Eds. B. N. Petrov and F. Cs*á*ki, 267–281, Budapest: Akad*é*mia Kiado.

[2] Fujikoshi, Y. (1985). Selection of variables in discriminant analysis and canonical correlation analysis. *Multivariate Analysis*-VI, Ed. P. R. Krishnaiah, 219–236, Elsevier Science Publishers B. V.

[3] Fujikoshi, Y. and Kurata H. (2008). Information Criterion for Some Conditional Independence Structures. To appear in New Trends in Pychometrics (K. Sigemasu et al., Ed.).

[4] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

[5] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley, New York.

[6] Siotani, M. , Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statical Analysis: A Graduate Course and Handbook*. American Sciences Press, Ohio.