数理解析研究所講究録
第2047巻 2017年 39-54

一般化推定方程式による解析における変数選択規準

広島大学・理学研究科数学専攻 佐藤 倫治 稲津 佑
Tomoharu Sato, Yu Inatsu
Department of Mathematics,
Graduate school of Hiroshima university

§1. はじめに

実データ解析において, 医学や経済学, その他の分野で相関のあるデータを解析する方法が議論されてきた. その中でも経時データとは, 同一個体からの時間経過に伴う繰り返し測定データのことで, 一般的には同一個体からのデータには相関があり, 異個体間のデータは独立であるという特徴を持つ. 相関のある経時データを解析する手法の一つとして, Liang and Zeger (1986) で提案された一般化推定方程式 (GEE) というものが知られている. これは, 一般化線形モデル (Nelder and Wedderburn, 1972) におけるスコア方程式を多次元に拡張したもので, 相関構造を考慮するため, 作業用相関行列を用いて解析する手法である. この作業用相関行列が GEE を特徴づけるものとなっており, 作業用相関は解析者が自由に決めてもよいものであるが, それが必ずしも真の相関構造を選ばなくてもよく, 間違った相関を選んでも良い性質を持つ推定量を得ることができるということが知られている. さらに, GEE では同時分布を完全に特定する必要もない. これらが理由で, GEE は広く解析に用いることが可能であると言える.

一方で, いくつかのモデルの中から最適なモデルを選ぶ「モデル選択」も非常に重要な問題であり, GEE を用いた解析においてもモデル選択は重要であると言える. 通常モデル選択では「リスク関数」によってモデルの当てはまりの良さを測り, リスク関数を最小にするモデルを選択する. このときリスク関数の漸近不偏な推定量を用いることでモデル選択規準を考えることが多い. 例えば, 最も有名な赤池情報量規準 (AIC)(Akaike, 1973, 1974) は期待カルバック-ライブラー情報量 (Kullback and Leibler, 1951) に基づ

くリスクの漸近不偏な推定量として提案されている. AIC は尤度関数により, シンプルに AIC $= -2 \times$ (最大対数尤度) $+ 2 \times$ (独立パラメータ数) として得られる. さらに, Nishii (1984), Rao (1988) で提案された AIC の一般的な拡張である GIC もあり, モデル選択において広く適応でき, 多くの論文で様々な性質が議論されてきた.

しかしながら, GEE を用いた解析では, 同時分布を完全に特定しないために尤度に基づくモデル選択規準を考えることができないため, AIC や GIC と同様には考えることができない. しかし GEE を用いた解析でも AIC や GIC のようなモデル選択規準はすでに研究されてきていた. たとえば, Pan (2001) では疑似尤度に基づく規準として QIC を提案している. 疑似尤度は Wedderburn (1974) によるものである. しかし QIC は独立な疑似尤度に基づくリスクを考えていたため, 相関構造を無視した状況でモデル選択を行っていた. さらに, Cantoni *et al.* (2005) で提案された GC_p は Mallow's C_p (Mallows, 1973) を一般的に拡張したものである. Hin and Wang (2009) や Gosho *et al.* (2011) で提案された CIC は相関構造を選択していた. しかしながらこれらのモデル選択規準はどれも, 相関構造を考慮に入れずに導出されたものであり, 経時データの特徴である相関を反映できていないと言える.

そこで, Inatsu and Imori (2013) では相関構造を考慮に入れたモデル選択規準 PMSEG を提案した. リスクを共分散行列で基準化した予測平均二乗誤差で定義し, 相関構造を決めるパラメータ $\boldsymbol{\alpha}$ と尺度母数 ϕ を既知としたもとで漸近バイアスを計算しバイアス補正を行った. 相関パラメータ $\boldsymbol{\alpha}$ や尺度母数 ϕ は実解析の面では未知である場合が多いため, Inatsu (2014) では, 相関パラメータ $\boldsymbol{\alpha}$ を未知とし尺度母数 ϕ は既知のままで推定を行い, バイアス評価をし, モデル選択規準を導出した.

本研究では, 変数の最適な部分集合を決定すること, すなわち「変数選択」に焦点を当てており, 相関構造を考慮に入れた変数選択規準を提案することを目標とし, 先行研究である Inatsu and Imori (2013) と Inatsu (2014) をさらに拡張した変数選択規準の提案を行う. 相関パラメータ $\boldsymbol{\alpha}$ に加え尺度母数 ϕ も未知として推定を行い, バイアス評価を行う. このとき, 興味のあることは未知パラメータの推定がリスクの推定量の漸近バイアスに影響があるのかないのかということと, バイアス計算に必要な十分条件の違いがないかの確認であった. リスク関数として共分散行列で基準化された予測平均二乗誤差 (the prediction mean squared error, PMSE) を用い, 提案するモデル選択規準を PMSEG (the prediction mean squared error in the GEE) と呼ぶ.

§2. GEE 推定量の確率展開

y_{ij} を i 番目の観測者の j 時点での応答変数, $\boldsymbol{x}_{f,ij}$ を l 次元の説明変数ベクトルとする. ここで, $i=1,\ldots,n$; $j=1,\ldots,m$ とする. 同一個体からの応答変数は相関をもち, 異個体間の応答変数は独立であるとする. 各 $i=1,\ldots,n$ に対して, i 番目の観測者の応答変数ベクトルを $\boldsymbol{y}_i=(y_{i1},\ldots,y_{im})'$, 説明変数行列を $\boldsymbol{X}_{f,i}=(\boldsymbol{x}_{f,i1},\ldots,\boldsymbol{x}_{f,im})'$ とし, $\boldsymbol{X}_i=(\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{im})'$ を $\boldsymbol{X}_{f,i}$ の $m\times p$ 部分行列とする. Liang and Zeger (1986) では, y_{ij} の周辺密度に以下のような一般化線形モデル (GLM) を用いた.

$$f(y_{ij};\boldsymbol{x}_{ij},\boldsymbol{\beta},\phi)=\exp\left[\{y_{ij}\theta_{ij}-a(\theta_{ij})\}/\phi+b(y_{ij},\phi)\right]. \tag{2.1}$$

ここで, $a(\cdot)$, $b(\cdot)$ は既知の関数, θ_{ij} は未知の位置母数, ϕ は未知または既知の尺度母数である. GLM の枠組みでは, 位置母数 $\theta_{ij}=u(\eta_{ij})=\theta_{ij}(\boldsymbol{\beta})$ で $u(\cdot)$ は既知の関数, $\eta_{ij}=\boldsymbol{x}_{ij}'\boldsymbol{\beta}$ で $\boldsymbol{\beta}$ は未知のパラメータである. 本論文では尺度母数 ϕ を未知として議論を行う. またパラメータ空間 Θ は式 (2.1) で与えられる指数型分布族に属する分布の自然パラメータ空間 (Xie and Yang, 2003) とする. さらに, Θ の内部を Θ° と表すことにする. このとき, Θ は凸で, 関数 $a(\cdot)$ は Θ° 上で無限階微分可能で, y_{ij} のすべてのモーメントが存在する. この設定の下では, 平均と分散は以下のようになる.

$$\mu_{ij}(\boldsymbol{\beta})=\mathrm{E}[y_{ij}]=\dot{a}(\theta_{ij}),\sigma_{ij}^2(\boldsymbol{\beta})=\mathrm{Cov}[y_{ij}]=\ddot{a}(\theta_{ij})\phi\equiv\nu(\mu_{ij}(\boldsymbol{\beta})).$$

この状況下で, y_{ij} の期待値はリンク関数によってモデル化される. すなわち, リンク関数 $g(t)=(\dot{a}\circ u)^{-1}(t)$ と線形予測子 η_{ij} によって $g(\mu_{ij}(\boldsymbol{\beta}))=\eta_{ij}=\boldsymbol{x}_{ij}'\boldsymbol{\beta}$ とモデル化される. 関数 u を恒等関数 $u(s)=s$ とする. このとき, $g(t)=\dot{a}^{-1}(t)$ を自然リンク関数と呼ぶ. $\boldsymbol{x}_{f,ij}$ を用いたモデルをフルモデル, $\boldsymbol{x}_{ij}$ を用いたモデルを候補モデルとする. y_{ij} の従う真の分布の密度関数は式 (2.1) でかけるとする. すなわち, 真のモデルは候補モデルのひとつであるとする.

以下のように記号を定義する. $\boldsymbol{\mu}_i(\boldsymbol{\beta})=(\mu_{i1}(\boldsymbol{\beta}),\ldots,\mu_{im}(\boldsymbol{\beta}))'$, $\boldsymbol{A}_i(\boldsymbol{\beta})=\mathrm{diag}(\sigma_{i1}^2(\boldsymbol{\beta}),\ldots,\sigma_{im}^2(\boldsymbol{\beta}))$, $\boldsymbol{\Delta}_i(\boldsymbol{\beta})=\mathrm{diag}(\partial\theta_{i1}/\partial\eta_{i1},\ldots,\partial\theta_{im}/\partial\eta_{im})$ とおく. ここで, $\mathrm{diag}(a_1,\ldots,a_m)$ は $m\times m$ の対角行列で第 (i,i) 成分が a_i である行列を表す. さらに, $\boldsymbol{D}_i(\boldsymbol{\beta})=\partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}=\boldsymbol{A}_i(\boldsymbol{\beta})\boldsymbol{\Delta}_i(\boldsymbol{\beta})\boldsymbol{X}_i$, $\boldsymbol{\Sigma}_i(\boldsymbol{\beta})=\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}_0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\phi$ とし, $\boldsymbol{R}_0$ は真の相関行列とする. ここで, すべての i に対して, $\boldsymbol{y}_i$ の相関行列は等しく $\boldsymbol{R}_0$ であるとする. $\boldsymbol{R}(\boldsymbol{\alpha})$ を作業用相関行列とする. 作業用相関行列は自由に決めることができるが, す

べての候補モデルで共通と仮定する. $\boldsymbol{R}(\boldsymbol{\alpha})$ は局外パラメータ $\boldsymbol{\alpha}$ を含む. 局外パラメータ空間 $\mathcal{A}$ を以下で定義する.

$$\mathcal{A} = \{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)' \in \mathbb{R}^s | \boldsymbol{R}(\boldsymbol{\alpha}) \text{ は正定値行列 } \}$$

状況に応じて, 作業用相関行列は自由に決めることができるが, その例をいくつか提示する.

[1]independence: $(\boldsymbol{R})_{jk} = 0, (j \neq k)$.

[2]exchangeable: $(\boldsymbol{R})_{jk} = \alpha, (j \neq k)$.

[3]autoregressive: $(\boldsymbol{R})_{jk} = (\boldsymbol{R})_{kj} = \alpha^{j-k}, (j > k)$.

[4]1-dependence: $(\boldsymbol{R})_{jk} = (\boldsymbol{R})_{kj} = \alpha, (j = k+1)$.

[5]unstructured: $(\boldsymbol{R})_{jk} = (\boldsymbol{R})_{kj} = \alpha_{jk}, (j > k)$.

また, $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\phi(\boldsymbol{\beta})$ とおく. もし $\boldsymbol{R}(\boldsymbol{\alpha}) = \boldsymbol{R}_0$ なら, $\boldsymbol{V}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}) = \boldsymbol{\Sigma}_i(\boldsymbol{\beta}_0) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}_0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\phi_0 = \mathrm{Cov}[\boldsymbol{y}_i]$ となる. ここで $\boldsymbol{\beta}_0$ は $\boldsymbol{\beta}$ の真値である. $\boldsymbol{\alpha}$ の次元は選択した作業用相関行列によって変わる.

多くの場合 $\boldsymbol{\alpha}$ は未知のパラメータである. また $\boldsymbol{\alpha}$ は局外パラメータであるが, $\boldsymbol{\beta}$ の推定のために推定しなければならない. その際, 通常データから推定する. $\boldsymbol{\alpha}$ の推定量ベクトルを $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}) = (\hat{\alpha}_1(\boldsymbol{\beta}, \hat{\phi}), \ldots, \hat{\alpha}_s(\boldsymbol{\beta}, \hat{\phi}))'$ とし, $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \phi_0) \xrightarrow{a.s.} \boldsymbol{\alpha}_0 \in \mathcal{A}^\circ$ と仮定する. ここで $\mathcal{A}^\circ$ は $\mathcal{A}$ の内部である. さらに, 本論文では ϕ の推定量として以下を用いる.

$$\hat{\phi} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\ddot{a}(\hat{\theta}_{ij})}$$

さらに, $\hat{\phi} \xrightarrow{p} \phi_0$ と仮定する. このとき, Liang and Zeger (1986) で提案された GEE は以下で与えられる.

$$\boldsymbol{q}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i'(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = \boldsymbol{0}_p. \tag{2.2}$$

ここで, $\boldsymbol{0}_p$ は p 次元の零ベクトルであり, $\boldsymbol{V}_i(\boldsymbol{\beta}) = \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)$ である. $\boldsymbol{\beta}_0$ の推定量は式 (2.2) の解として定義する. さらに $\boldsymbol{\alpha}_0$ と ϕ_0 は未知の場合, GEE は以下で与えられる.

$$s_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i'(\boldsymbol{\beta})\boldsymbol{\Gamma}_i^{-1}(\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = \boldsymbol{0}_p. \tag{2.3}$$

ここで, $\boldsymbol{\Gamma}(\boldsymbol{\beta}) = \boldsymbol{V}_i(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}))$ である. そして式 (2.3) の解 $\hat{\boldsymbol{\beta}}$ を $\boldsymbol{\beta}_0$ の推定量とし, GEE 推定量と呼ぶ.

$\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi$ は未知なので以下のような反復計算によって推定を行う.

Step1. $\boldsymbol{\alpha}$ の初期値 $\hat{\boldsymbol{\alpha}}^{<0>}$ を決める.

Step2. $\hat{\boldsymbol{\beta}}^{<k>} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}^{<k>})$ とする. ここで, $\hat{\boldsymbol{\beta}}^{<k>}$ は $\hat{\boldsymbol{\alpha}}^{<k>}$ を式 (2.3) に代入して $\boldsymbol{\beta}$ について解いた解とする.

Step3. $\hat{\phi}^{<k+1>}$を $\boldsymbol{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}^{<k>})$ により推定する.

Step4. $\hat{\boldsymbol{\alpha}}^{<k+1>} = \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}^{<k>}, \hat{\phi}^{<k+1>})$ を計算する.

Step5. Step2 から Step4 を値が収束するまで反復する.

$\boldsymbol{\alpha}$ の推定量としてモーメント推定量を用いることで, C9〜C13 の条件を満たすことが Inatsu (2013) により示されている.

また, 手順 4 において $\boldsymbol{\alpha}$ の推定量は作業用相関行列によって異なる. 以下で $\boldsymbol{\alpha}$ の推定量の例を挙げる.

$$\text{Exchangeable} : \hat{\alpha} = \frac{1}{nm(m-1)} \sum_{i=1}^{n} \sum_{j>k} \hat{r}_{ij}\hat{r}_{ik}/\hat{\phi}.$$

$$\text{Autoregressive} : \hat{\alpha} = \frac{1}{n(m-1)} \sum_{i=1}^{n} \sum_{j=1}^{m-1} \hat{r}_{ij}\hat{r}_{i,j+1}/\hat{\phi}.$$

$$1-\text{dependence} : \hat{\alpha} = \frac{1}{n-1} \sum_{i=1}^{n-1} \hat{\alpha}_i, \hat{\alpha}_i = \frac{1}{n} \sum_{i=1}^{n} \hat{r}_{ij}\hat{r}_{i,j+1}/\hat{\phi}.$$

$$\text{Unstructured} : \hat{\alpha}_{jk} = \frac{1}{n} \sum_{i=1}^{n} \hat{r}_{ij}\hat{r}_{ik}/\hat{\phi}.$$

第 3 節で新しい変数選択規準の提案をするため, ここでは必要な $\hat{\boldsymbol{\beta}}$ の展開を行う. 本節では, $\hat{\boldsymbol{\beta}}$ を n^{-1} のオーダーまで展開する. 表記の簡単のため, 関数の表記の際に $(\boldsymbol{\beta})$ を省略し, $\mu_{ij}(\boldsymbol{\beta})$ を μ_{ij} のように表記する. また, $\boldsymbol{\beta}$ の関数において, $\boldsymbol{\beta}$ の真値 $\boldsymbol{\beta}_0$ を代入したものは添え字に 0 を追加し, $\mu_{ij}(\boldsymbol{\beta}_0)$ を $\mu_{ij,0}$ のように表記し, $\boldsymbol{\beta}$ の推定量 $\hat{\boldsymbol{\beta}}$ を代入したものはハット記号を付して, $\mu_{ij}(\hat{\boldsymbol{\beta}})$ を $\hat{\mu}_{ij}$ のように表記する. さらに, GEE 推定量の漸近性質を保証するために, 以下のような正則条件を仮定する (Xie and Yang, 2003).

C1. すべての列 $\{\boldsymbol{x}_{ij}\}$ はあるコンパクト集合 $\mathcal{X}$ に含まれる.

C2. $\boldsymbol{\beta}_0$ は許容集合 $\mathcal{B}$ の内点, すなわち, $\boldsymbol{\beta}_0 \in \mathcal{B}^{\circ}$.
ここで $\mathcal{B}$ は次のように定義される. $\mathcal{B} = \{\boldsymbol{\beta} | u(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) \in \Theta, \boldsymbol{x}_{ij} \in \mathcal{X}\}$.

C3. 任意の $\boldsymbol{\beta} \in \mathcal{B}$ に対して, $\boldsymbol{x}_{ij}'\boldsymbol{\beta} \in g(\mathcal{M})$ であり, $\mathcal{M}$ は Θ° の $\dot{a}$ による像である.

C4. $u(\eta_{ij})$ は 4 階連続微分可能で, $g(\mathcal{M}^{\circ})$ 上で $\dot{u}(\eta_{ij}) > 0$ である.

C5. $\boldsymbol{H}_n, \boldsymbol{M}_n$ を以下のように定義する.

$$\boldsymbol{H}_n = \sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{D}_i, \boldsymbol{M}_n = \sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{\Sigma}_i \boldsymbol{V}_i^{-1} \boldsymbol{D}_i.$$

このとき, $n \to \infty$ で, $\boldsymbol{H}_{n,0}, \boldsymbol{M}_{n,0}$ はともに正定値行列である.

C6. $\liminf_{n\to\infty} \lambda_{\min}(\boldsymbol{H}_{n,0}/n) > 0$. ここで, $\lambda_{\min}(\boldsymbol{A})$ は対称行列 $\boldsymbol{A}$ の最小固有値である.

C7. $\boldsymbol{\beta}_0$ の近傍 N_0 において, ある定数 $c_0 > 0$ と n_0 が存在して, 任意の大きさ 1 の p 次元ベクトル $\boldsymbol{\lambda}$ に対して, $n \geq n_0$ のとき以下が成り立つ.

$$P\left(-\boldsymbol{\lambda}' \frac{\partial \boldsymbol{s}_n}{\partial \boldsymbol{\beta}'} \boldsymbol{\lambda} \geq nc_0\right) = P\left(-\boldsymbol{\lambda}' \boldsymbol{\Upsilon}_n \boldsymbol{\lambda} \geq nc_0\right) = 1, (\boldsymbol{\beta} \in N_0).$$

C8. GEE は $n \to \infty$ のとき唯一解を持つ.

C1, C2, C3 は GLM の枠組みにおいて必要な条件であり, C4, C5 はリスクの推定量の漸近バイアスの計算に必要な条件である. さらに, C1, C6〜C8 の条件は, 強一致性や漸近正規性, GEE 推定量の唯一性のための条件であり, Xie and Yang (2003) の条件を修正したものである. さらに, 以下の条件を加える.

C9. $\boldsymbol{\alpha}_0$ のコンパクトな近傍 $U_{\boldsymbol{\alpha}_0}$ が存在し, $\mathrm{vec}\{\boldsymbol{R}^{-1}(\boldsymbol{\alpha})\}$ は $U_{\boldsymbol{\alpha}_0}$ 上で 3 階連続微分可能である.

C10. $\boldsymbol{\beta}_0$ のコンパクトな近傍 $U_{\boldsymbol{\beta}_0}$ が存在し, $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ は $U_{\boldsymbol{\beta}_0}$ 上で 3 階連続微分可能である.

C11. 任意の $\boldsymbol{\beta} \in U_{\boldsymbol{\beta}_0}$ に対して, $\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}), \hat{\boldsymbol{\alpha}}^{(2)}(\boldsymbol{\beta}), \hat{\boldsymbol{\mu}}^{(3)}(\boldsymbol{\beta}) = O_p(1)$ である. ただし,

$$\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}) = \frac{\partial \hat{\boldsymbol{\alpha}}}{\partial \boldsymbol{\beta}'}, \hat{\boldsymbol{\alpha}}^{(2)}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}), \hat{\boldsymbol{\alpha}}^{(3)}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \hat{\boldsymbol{\alpha}}^{(2)}(\boldsymbol{\beta}),$$

である.

C12. $\sqrt{n}(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0) = O_p(1)$ である. また, 有界な $s \times p$ 次元非確率行列 $\boldsymbol{\mathcal{H}}$ が存在して, $(\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{H}}) = O_p(n^{-1/2})$ である.

C13.

$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{h}_{i,0}\right]=O(n^{-1}),$$
$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{j}_{i,0}\right]=O(n^{-1}),$$
$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{h}_{i,0}\right]=O(n^{-1}),$$
$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{D}_{i,0}\boldsymbol{h}_{i,0}\right]=O(n^{-1}),$$
$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{j}_{i,0}\right]=O(n^{-1}),$$
$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{D}_{i,0}\boldsymbol{j}_{i,0}\right]=O(n^{-1}).$$

$\boldsymbol{h}_{1,0},\boldsymbol{j}_{1,0},\boldsymbol{A}_{f,i,0}^{*},\boldsymbol{b}_{f,0}$ はバイアス計算に出てくる項である.

また, 行列 $W=(\omega_{ij})$ のベクトル $\boldsymbol{\beta}$ での微分を以下で定義する.

$$\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{W}=\left(\frac{\partial\boldsymbol{W}}{\partial\beta_1},\cdots,\frac{\partial\boldsymbol{W}}{\partial\beta_p}\right),\frac{\partial\boldsymbol{W}}{\partial\beta_k}=\left(\frac{\partial\omega_{ij}}{\partial\beta_k}\right).$$

条件 C9, C10, C11, C12, C13 は局外パラメータである $\boldsymbol{\alpha}$ の推定の影響を無視するために必要な条件である. さらに C5 の条件により, $\boldsymbol{H}_{n,0}=O(n)$ である.

上記の条件の下で, $\hat{\boldsymbol{\beta}}$ の確率展開を行う. $\hat{\boldsymbol{s}}_n=\boldsymbol{0}_p$ なので, この方程式を $\boldsymbol{\beta}=\boldsymbol{\beta}_0$ の周りでテイラー展開すると, GEE は以下のように展開できる.

$$\begin{aligned}\boldsymbol{0}_p=&\boldsymbol{s}_{n,0}+\frac{\partial\boldsymbol{s}_n}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\otimes\boldsymbol{I}_p\}\left(\frac{\partial}{\partial\boldsymbol{\beta}}\otimes\frac{\partial\boldsymbol{s}_n}{\partial\boldsymbol{\beta}'}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\=&\boldsymbol{s}_{n,0}-\boldsymbol{\mathcal{D}}_{n,0}(\boldsymbol{I}_p+\boldsymbol{\mathcal{D}}_{1,0}+\boldsymbol{\mathcal{D}}_{2,0})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\otimes\boldsymbol{I}_p\}\boldsymbol{L}_1(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0).\end{aligned}$$

$\boldsymbol{\beta}^*$ は $\boldsymbol{\beta}_0$ と $\hat{\boldsymbol{\beta}}$ の間にあるベクトルであり, $\boldsymbol{I}_p$ は p 次元の単位行列である. さらに, $\boldsymbol{L}_1(\boldsymbol{\beta}^*)$, $\boldsymbol{\mathcal{D}}_{n,0}$, $\boldsymbol{\mathcal{D}}_{1,0}$, $\boldsymbol{\mathcal{D}}_{2,0}$ は以下で与えられる.

$$\boldsymbol{L}_1(\boldsymbol{\beta}^*) = \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_n}{\partial \boldsymbol{\beta}'}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}, \boldsymbol{\mathcal{D}}_{n,0} = \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \boldsymbol{\Gamma}_{i,0}^{-1} \boldsymbol{D}_{i,0},$$

$$\boldsymbol{\mathcal{D}}_{1,0} = -\mathcal{D}_{n,0}^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{\Gamma}_i^{-1}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right) \{\boldsymbol{I}_p \otimes (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\},$$

$$\boldsymbol{\mathcal{D}}_{2,0} = -\mathcal{D}_{n,0}^{-1} \sum_{i=1}^{n} \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{D}_i^{-1}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right) [\boldsymbol{I}_p \otimes \{\boldsymbol{\Gamma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}.$$

ここで, $\boldsymbol{L}_1(\boldsymbol{\beta}^*) = O_p(n)$, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \boldsymbol{\mathcal{D}}_{1,0}, \boldsymbol{\mathcal{D}}_{2,0} = O_p(n^{-1/2})$ である. さらに, $\boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}}_0)$ の展開は以下で与えられる.

$$\boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}}_0) = \boldsymbol{R}^{-1}(\boldsymbol{\alpha}_0) + \boldsymbol{R}^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}(\boldsymbol{\alpha}_0) - \boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}^{-1}(\boldsymbol{\alpha}_0) + O_p(n^{-1}).$$

テイラーの定理より,

$$|\boldsymbol{R}(\boldsymbol{\alpha}_0) - \boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0)| \le \left|\frac{\partial}{\partial \boldsymbol{\alpha}} \otimes \boldsymbol{R}(\boldsymbol{\alpha})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}\right| |\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0| = O_p(n^{-1/2}),$$

であるので, $\boldsymbol{R}(\boldsymbol{\alpha}_0) - \boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0) = O_p(n^{-1/2})$ である. さらに

$$\begin{aligned}
\boldsymbol{\mathcal{D}}_{n,0} &= \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \boldsymbol{\Gamma}_{i,0}^{-1} \boldsymbol{D}_{i,0} \\
&= \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}}_0) \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0) \boldsymbol{D}_{i,0} \\
&= \boldsymbol{H}_{n,0} + O_p(n^{1/2}),
\end{aligned}$$

$\boldsymbol{s}_{n,0} = \boldsymbol{q}_{n,0} + O_p(1)$ を用いると $\hat{\boldsymbol{\beta}}$ は以下のように展開できる.

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{H}_{n,0}^{-1} \boldsymbol{q}_{n,0} + O_p(n^{-1}) = \boldsymbol{b}_{1,0} + O_p(n^{-1}).$$

また,

$$\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right) - \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] = O_p(n^{-1/2}),$$

である.

以上の結果を用いて, GEE も以下のように展開できる.

$$\begin{aligned}\boldsymbol{s}_{n,0} =&\boldsymbol{H}_{n,0}(\boldsymbol{I}_p + \boldsymbol{G}_{1,0} + \boldsymbol{G}_{2,0} + \boldsymbol{G}_{3,0} + \boldsymbol{h}_{1,0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &- \frac{1}{2}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\{\boldsymbol{\mathcal{S}}_{1,0} + (\boldsymbol{L}_1(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{S}}_{1,0})\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &- \frac{1}{6}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\left\{\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_n}{\partial \boldsymbol{\beta}'}\right)\right\}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}.\end{aligned} \tag{2.4}$$

ここで, $\boldsymbol{\beta}^{**}$ は $\boldsymbol{\beta}_0$ と $\hat{\boldsymbol{\beta}}$ の間にあるベクトル, $\boldsymbol{\mathcal{S}}_{1,0} = \mathrm{E}[\boldsymbol{L}_1(\boldsymbol{\beta}_0)]$ であり, $\boldsymbol{\mathcal{S}}_{1,0} = O_p(n)$, $\boldsymbol{L}_1(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{S}}_{1,0} = O_p(n^{1/2})$ である. また, 式 (2.4) の最終項は $O_p(n^{-1/2})$ である.

さらに $\boldsymbol{\mathcal{C}}_{1i}$, $\boldsymbol{\mathcal{C}}_{2i}$, $\boldsymbol{\mathcal{C}}_{3i}$, $\boldsymbol{G}_{1,0}$, $\boldsymbol{G}_{2,0}$, $\boldsymbol{G}_{3,0}$, $\boldsymbol{h}_{1,0}$, $\boldsymbol{j}_{1,0}$ を以下で定義する.

$$\begin{aligned}&\boldsymbol{\mathcal{C}}_{1i} = \boldsymbol{D}_i'\boldsymbol{A}_i^{-1/2}\boldsymbol{R}^{-1}(\boldsymbol{\alpha}_0), \boldsymbol{\mathcal{C}}_{2i} = \boldsymbol{D}_i'\boldsymbol{A}_i^{-1/2}, \boldsymbol{\mathcal{C}}_{3i} = \boldsymbol{R}^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_i^{-1/2}, \\ &\boldsymbol{G}_{1,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\mathcal{C}_{1i,0}\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{A}_i^{-1/2}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\{\boldsymbol{I}_p \otimes (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}, \\ &\boldsymbol{G}_{2,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{\mathcal{C}}_{2i}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)[\boldsymbol{I}_p \otimes \{\boldsymbol{\mathcal{C}}_{3i,0}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}], \\ &\boldsymbol{G}_{3,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{2i,0}\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right][\boldsymbol{I}_p \otimes \{\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}], \\ &\boldsymbol{h}_{1,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{1i,0}\{\boldsymbol{R}(\boldsymbol{\alpha}_0) - \boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{\mathcal{C}}_{1i,0}'\boldsymbol{b}_{1,0}, \\ &\boldsymbol{j}_{1,0} = \boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{1i.,0}\{\boldsymbol{R}(\boldsymbol{\alpha}_0) - \boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{\mathcal{C}}_{3i,0}(\boldsymbol{y}_{i,0} - \boldsymbol{\mu}_{i,0}).\end{aligned} \tag{2.5}$$

ここで, $\boldsymbol{G}_{1,0}$, $\boldsymbol{G}_{2,0}$, $\boldsymbol{G}_{3,0} = O_p(n^{-1/2})$, $\boldsymbol{h}_{1,0}, \boldsymbol{j}_{1,0} = O_p(n^{-1})$ である.

式 (2.5) を用いると, $\hat{\boldsymbol{\beta}}$ は以下のように展開できる.

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{b}_{1,0} + \boldsymbol{b}_{2,0} + O_p(n^{-3/2}).$$

ここで, $\boldsymbol{b}_{2,0} = \boldsymbol{H}_{n,0}^{-1}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_p)\boldsymbol{\mathcal{S}}_{1,0}\boldsymbol{b}_{1,0}/2 - \boldsymbol{G}_{1,0}\boldsymbol{b}_{1,0} - \boldsymbol{G}_{2,0}\boldsymbol{b}_{1,0} - \boldsymbol{G}_{3,0}\boldsymbol{b}_{1,0} + \boldsymbol{h}_{1,0} + \boldsymbol{j}_{1,0}$, $\boldsymbol{b}_{1,0} = O_p(n^{-1/2}), \boldsymbol{b}_{2,0} = O_p(n^{-1})$ である.

§3. 主結果

本節では, 新しい変数選択規準を提案する. モデルの当てはまりの良さ Risk を共分散行列で基準化された PMSE に基づいたリスク関数によって測り, 以下のように与える.

$$\mathrm{Risk} = \mathrm{PMSE} - mn = \mathrm{E}_y\left[\mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)\right]\right] - mn.$$

ここで, $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})'$ は m 次元の確率変数ベクトルで, $\boldsymbol{y}_i$ とは独立に同一の分布に従うものとする. $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$ のとき, Risk は最小値 0 をとる. すなわち, PMSE は最小値 mn をとる. PMSE を最小にするモデルを良いモデルと考えるが, PMSE は未知であるため推定しなければならない.

$\boldsymbol{R}_0$ と $\boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, $\boldsymbol{\mathcal{L}}^*(\boldsymbol{\beta})$ を以下で定義する.

$$\boldsymbol{R}_0(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'\boldsymbol{A}_i^{-1/2}/\hat{\phi},$$

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_1))'\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_2)\boldsymbol{R}_0^{-1}(\boldsymbol{\beta}_2)\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_2)(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_1))\hat{\phi}^{-1}(\boldsymbol{\beta}_2),$$

$$\boldsymbol{\mathcal{L}}^*(\boldsymbol{\beta}) = \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i).$$

このとき, PMSE の推定量として $\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ が考えられる. ここで $\hat{\boldsymbol{\beta}}_f$ はフルモデルにおける GEE 推定量, すなわち, $\hat{\boldsymbol{\beta}}_f$ は以下の方程式の解で与えられる.

$$\boldsymbol{s}_{f,n}(\boldsymbol{\beta}_f) = \sum_{i=1}^{n}\boldsymbol{D}_i'(\boldsymbol{\beta}_f)\boldsymbol{V}_i^{-1}(\boldsymbol{\beta}_f, \boldsymbol{\alpha}_f)(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_f)) = \boldsymbol{0}_l,$$

ここで, $\boldsymbol{D}_i(\boldsymbol{\beta}_f) = \boldsymbol{A}_i(\boldsymbol{\beta}_f)\boldsymbol{\Delta}(\boldsymbol{\beta}_f)\boldsymbol{X}_{f,i}, \boldsymbol{V}_i(\boldsymbol{\beta}_f, \boldsymbol{\alpha}_f) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_f)\boldsymbol{R}_i(\boldsymbol{\alpha}_f)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_f)$ であり, $\boldsymbol{R}_i(\boldsymbol{\alpha}_f)$ は正定値の作業用相関行列. さらに, $\boldsymbol{\beta}_f$ はフルモデルにおける l 次元の未知のパラメータベクトルである. また $\boldsymbol{R}(\boldsymbol{\alpha}_f)$ はすべての候補モデルにおいて等しい. 以下簡単のために $\boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_2) = \boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_2)$, $\boldsymbol{\mathcal{L}}^*(\boldsymbol{\beta}_0) = \boldsymbol{\mathcal{L}}^*$ と表すことにする.

$\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ は PMSE の漸近不偏な推定量ではないため, 漸近バイアスを評価した上で新しい変数選択規準を提案する. $\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ で PMSE を推定したときのバイアスは以下で与えられる.

$$
\begin{aligned}
\mathrm{Bias} =& \mathrm{PMSE} - \mathrm{E}_y[\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)] \\
=& \{\mathrm{Risk} - \mathrm{E}_y[\boldsymbol{\mathcal{L}}^*(\hat{\boldsymbol{\beta}})]\} + \{\mathrm{E}_y[\boldsymbol{\mathcal{L}}^*(\hat{\boldsymbol{\beta}})] - \mathrm{E}_y[\boldsymbol{\mathcal{L}}^*]\} \\
& + \{\mathrm{E}_y[\boldsymbol{\mathcal{L}}^*] - \mathrm{E}_y[\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}_f)]\} + \{\mathrm{E}_y[\boldsymbol{\mathcal{L}}^*(\hat{\boldsymbol{\beta}}_f)] - \mathrm{E}_y[\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)]\} \\
=& \mathrm{Bias1} + \mathrm{Bias2} + \mathrm{Bias3} + \mathrm{Bias4}.
\end{aligned}
$$

以下で Bias1, Bias2, Bias3, Bias4 それぞれを評価する.

まず Bias3 は以下のようになる.

$$
\begin{aligned}
\mathrm{Bias3} =& \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f)\}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] \\
=& mn - \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f)(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right].
\end{aligned}
$$

これは, 候補モデルに依らない値なので, Bias3 の計算は変数選択において無視してもよい.

同様に, Bias1 は以下のように展開できる.

$$
\begin{aligned}
\mathrm{Bias1} =& \mathrm{E}_y\left[\mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)\right] - \sum_{i=0}^{n}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)\right] \\
=& \mathrm{E}_y\Bigg[\mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
& - \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{mu}}_i)\Bigg] \\
=& \mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0})\right] + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
& - \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] - 2\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{mu}}_i)\right] \\
& - \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
=& 2\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0})\right]. \qquad (3.1)
\end{aligned}
$$

Bias1 の展開には $\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0}$ の展開が必要なので, 以下のように展開する.

$$\begin{aligned}\hat{\boldsymbol{\mu}}_i =&\boldsymbol{\mu}_{i,0} + \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{1}{2}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_m\}\left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\\ &+ \frac{1}{6}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_m\}\left\{\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}\right)\right\}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{***}}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}\\ =&\boldsymbol{\mu}_{i,0} + \boldsymbol{D}_{i,0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{1}{2}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_m\}\boldsymbol{D}_{i,0}^{(1)}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O_p(n^{-3/2}),\end{aligned}$$

$$\boldsymbol{D}_{i,0}^{(1)} = \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \boldsymbol{D}_i\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

ここで, $\boldsymbol{\beta}^{***}$ は $\boldsymbol{\beta}_0$ と $\hat{\boldsymbol{\beta}}$ の間にあるベクトルである. $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ の確率展開を代入することで $\hat{\boldsymbol{\mu}}_i$ も以下のように展開できる.

$$\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0} + \left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\} + O_p(n^{-3/2}). \tag{3.2}$$

式 (3.1) と式 (3.2) により以下の結果を得る.

$$\begin{aligned}\frac{1}{2}\mathrm{Bias1} =&\mathrm{E}_y\left[\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0})\right]\\ =&\mathrm{E}_y\left[\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{i,0}\right]\\ &+ \mathrm{E}_y\left[\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right]\\ &+ \mathrm{E}_y[O_p(n^{-1/2})].\end{aligned}$$

異個体間のデータは独立なため, $\mathrm{E}[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})]=0,(i\neq j)$ であるので,

$$
\begin{aligned}
&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{i,0}\right] &(3.3)\\
=&\mathrm{E}_y\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{j,0}'\boldsymbol{V}_{j,0}^{-1}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})\right]\\
=&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}'\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\right]\\
=&\mathrm{E}_y\left[\mathrm{tr}\left\{\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}'\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\right\}\right]\\
=&\mathrm{E}_y\left[\mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\}\right]\\
=&\mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\boldsymbol{V}_{i,0}^{-1}\mathrm{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\right]\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\}\\
=&\mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\boldsymbol{V}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\}\\
=&\mathrm{tr}\left\{\boldsymbol{I}_p\right\}\\
=&p, &(3.4)
\end{aligned}
$$

となる.

また, 各 $i,j,k(\text{not } i=j=k)$ に対して, $\mathrm{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\otimes(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'(\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0})\right]=\boldsymbol{0}_m$ であるので,

$$
\begin{aligned}
&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0}+\frac{1}{2}(\boldsymbol{b}_{1,0}'\otimes\boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right]\\
=&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2i,0}+\frac{1}{2}(\boldsymbol{b}_{1i,0}'\otimes\boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1i,0}\right\}\right]\\
=&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}(\boldsymbol{b}_{2i,0}-\boldsymbol{h}_{1,0}-\boldsymbol{j}_{1,0})+\frac{1}{2}(\boldsymbol{b}_{1i,0}'\otimes\boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1i,0}\right\}\right]\\
&+\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\{\boldsymbol{D}_{i,0}(\boldsymbol{h}_{1,0}+\boldsymbol{j}_{1,0})\}\right].
\end{aligned}
$$

ここで, $\boldsymbol{b}_{1i,0} = \boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}'\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})$, $\boldsymbol{b}_{2i,0} = \boldsymbol{H}_{n,0}^{-1}(\boldsymbol{b}_{1i,0}' \otimes \boldsymbol{I}_p)\boldsymbol{\mathcal{S}}_{1,0}\boldsymbol{b}_{1i,0}/2 - \boldsymbol{G}_{1i,0}\boldsymbol{b}_{1i,0} - \boldsymbol{G}_{2i,0}\boldsymbol{b}_{1i,0} - \boldsymbol{G}_{3i,0}\boldsymbol{b}_{1i,0} + \boldsymbol{h}_{1,0} + \boldsymbol{j}_{1,0}$ であり,

$$
\begin{aligned}
\boldsymbol{G}_{1i,0} &= -\boldsymbol{H}_{n,0}^{-1}\mathcal{C}_{1i,0}\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{A}_i^{-1/2}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\{\boldsymbol{I}_p \otimes (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\},\\
\boldsymbol{G}_{2,0} &= -\boldsymbol{H}_{n,0}^{-1}\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{\mathcal{C}}_{2i}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)[\boldsymbol{I}_p \otimes \{\boldsymbol{\mathcal{C}}_{3i,0}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}],\\
\boldsymbol{G}_{3,0} &= -\boldsymbol{H}_{n,0}^{-1}\boldsymbol{\mathcal{C}}_{2i,0}\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right][\boldsymbol{I}_p \otimes \{\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}],
\end{aligned}
$$

である. 仮定 C13 より

$$
\begin{aligned}
&\boldsymbol{D}_{i,0}(\boldsymbol{b}_{2i,0} - \boldsymbol{h}_{1,0} - \boldsymbol{j}_{1,0}) + (\boldsymbol{b}_{1i,0}' \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1i,0}/2 = O_p(n^{-2}),\\
&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\{\boldsymbol{D}_{i,0}(\boldsymbol{h}_{1,0} + \boldsymbol{j}_{1,0})\}\right] = O(n^{-1}),
\end{aligned}
$$

であるので,

$$
\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right] = O(n^{-1}),
$$

となる.

正則条件のもとで, 期待値の極限と極限の期待値は等しい. さらに, 統計量のモーメントは n^{-1} の整級数で展開できることが知られている (Hall, 1992). ゆえに, 以下のように Bias1 の漸近展開が得られる.

$$
\mathrm{Bias1} = 2p + O(n^{-1}).
$$

さらに, 同様の漸近展開により

$$
\mathrm{Bias2} + \mathrm{Bias4} = O(n^{-1}),
$$

が得られる. Inatsu and Sato (2017)

以上の結果より漸近バイアスは以下のように展開できる.

$$
\mathrm{Bias} = 2p + \mathrm{Bias3} + O(n^{-1}).
$$

Bias3 はフルモデルにおける推定量により決まる値で候補モデルには依存しないものなので, 変数選択規準を Bias − Bias3 の推定量として, 以下のように定義する.

$$\mathrm{PMSEG} = \mathcal{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f) + 2p.$$

PMSEG は"the prediction mean squared error in the GEE"の意味である. AIC と同様に, 説明変数の個数に対応した罰則項が加わった形になり, 自然な変数選択規準であると言える. 今回, 時点数 m は個体すべて共通で固定していたが, 個体数 n に加え, 時点数 m が大きい場合の変数選択規準の導出や, リンク関数, 作業用相関行列の選択などが今後の課題である.

参考文献

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (eds. B. N. Petrov & F. Csáki), 267-281, Akadémiai Kiadó, Budapest.

[2] Akaike, H. (1974). A new look at the statistical model identification. IEEE *Trans. Automatic Control*, **AC-19**, 716-723.

[3] Inatsu, Y. (2014). Model selection criterion based on the prediction mean squared error in generalized estimating equations. *Master's Thesis, Department of Mathematics, Graduate School of Science, Hiroshima University, Hiroshima.*

[4] Inatsu, Y. & Imori, S. (2013). Model selection criterion based on the prediction mean squared error in generalized estimating equations. *TR13-10, Statistical Research Group, Hiroshima University, Hiroshima.*

[5] Gosho, M., Hamada, C. & Yoshimura, I. (2011). Modifications of QIC and CIC for selecting a Working Correlation Structure in the Generalized Estimating Equation Method. *Japanese Journal of Biometrics*, **32**, 1-12.

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

[7] Hin, L. Y. & Wang, Y. G. (2009). Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine*, **28**, 642-658.

[8] Kullback, S. & Libler, R. (1951). On information and sufficiency. *Ann. Math. Statist.*, **22**, 79-86.

[9] Liang, K. Y. & Zegerm S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

[10] Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, **15**, 661-675.

[11] Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc. ser. A.* **135**, 370-384.

[12] Nishii, R. (1984). Asymptotic Properties of Criteria for Selecting of Variables in Multiple Regression. *Ann. Statist.*, **12**, 758-765.

[13] Pan. W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, **57**, 120-125.

[14] Rao, C. R. & Wu, Y. (1989). A strongly Consistent Procedure for Model Selection in a Regression Problem. *Biometrika*, **76**, 369-374.

[15] Xie, M. & Yang, Y. (2003). Asymptotics for generalized estimating equations with large cluster sizes. *Ann. Statist.*, **31**, 310-347.