Liang and Zeger (*Biometrika* 1986) の Theorem 2 の証明について

Kengo Nagashima

平成 28 年 7 月 4 日

1 はじめに

本稿では Liang and Zeger (1984) の記法をほぼそのまま採用した ($\mathbf{G}_i = \mathrm{diag}\{a''(\theta_{it})\}$ だけ記号がかぶっているので変更した)．また，論文中のタイポや記述が曖昧な部分は独自に補った．証明の補足部分については，原典を引用しているパラグラフには (原典)，独自に補足したパラグラフには (補足) と記載した．

2 Generalized Estimating Equations (GEE)

$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{it}, \dots, Y_{in_i})^{\mathrm{T}}$ は $n_i \times 1$ 次元の結果変数ベクトル，$\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \dots, \mathbf{Y}_K^{\mathrm{T}})^{\mathrm{T}}$，$\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})^{\mathrm{T}}$ は $p \times 1$ 次元のデザインベクトル，$\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})^{\mathrm{T}}$ は $n_i \times p$ 次元のデザイン行列，$\boldsymbol{\beta}$ は $p \times 1$ 次元のモデルパラメータベクトルとし，$i = 1, \dots, K$ は個体を表わす添え字，$t = 1, \dots, n_i$ は各個体の時点を表わす添え字とする．Y_{it} の周辺密度には以下を仮定する;

$$f(Y_{it} = y_{it}) = \exp[\{y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})\}\phi],$$

ただし，$\theta_{it} = h(\eta_{it})$，$\eta_{it} = \mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\beta}$ とする．このとき，

$$\mathrm{E}(Y_{it}) = a'(\theta_{it}), \mathrm{Var}(Y_{it}) = a''(\theta_{it})/\phi,$$

である．以下では，一般性を失わず $n_i = n$ として話を進める．

上述の仮定の下で，$\mathbf{R}(\boldsymbol{\alpha})$ を $n \times n$ 次元の作業相関行列とし，

$$\mathbf{V}_i = \mathbf{G}_i^{1/2}\mathbf{R}(\boldsymbol{\alpha})\mathbf{G}_i^{1/2}/\phi,$$

とする，ただし，$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_l, \dots, \alpha_s)^{\mathrm{T}}$ は $s \times 1$ 次元のパラメータベクトルである．もし $\mathbf{R}(\boldsymbol{\alpha})$ が実際の $\mathbf{Y}_i$ の相関行列に等しいとき，$\mathbf{V}_i$ は $\mathrm{Cov}(\mathbf{Y}_i)$ と等しくなる．

これを用い，p 組の推定方程式を以下で定義する;

$$\sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{S}_i = \mathbf{0}, \tag{2.1}$$

ただし，

$$\mathbf{D}_i = \frac{\partial a'(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^{\mathrm{T}}} = \frac{\partial a'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}^{\mathrm{T}}}\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} = \mathbf{G}_i\boldsymbol{\Delta}_i\mathbf{X}_i,$$

は $n\times p$ 次元行列，$\mathbf{G}_i=\mathrm{diag}\{a''(\theta_{it})\}$ は $n\times n$ 次元行列，$\boldsymbol{\Delta}_i=\mathrm{diag}(\delta_{it})=\mathrm{diag}(\partial\theta_{it}/\partial\eta_{it})$ は $n\times n$ 次元行列，$\mathbf{S}_i=(S_{i1},\ldots,S_{in})^{\mathrm{T}}=\mathbf{Y}_i-a'(\boldsymbol{\theta}_i)$ は $n\times 1$ 次元ベクトル，$a'(\boldsymbol{\theta}_i)=(a'(\theta_{i1}),\ldots,a'(\theta_{in}))^{\mathrm{T}}$，$\boldsymbol{\theta}=(\theta_{i1},\ldots,\theta_{in})^{\mathrm{T}}$，および $\boldsymbol{\eta}=(\eta_{i1},\ldots,\eta_{in})^{\mathrm{T}}$ である．

(2.1) 式は $\boldsymbol{\alpha}$ と ϕ に依存する．そのため，$\boldsymbol{\alpha}$ を $\hat{\boldsymbol{\alpha}}(\mathbf{Y},\boldsymbol{\beta},\phi)$ に，ϕ を $\hat{\phi}(\mathbf{Y},\boldsymbol{\beta})$ にプラグインした推定方程式,

$$\sum_{i=1}^{K}\mathbf{U}_i[\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta},\hat{\phi}(\boldsymbol{\beta})\}]=\mathbf{0}, \tag{2.2}$$

に基づく推測を行う方法を GEE (Generalized Estimating Equations) と呼ぶ．通常，$\hat{\boldsymbol{\alpha}}$ については $(\boldsymbol{\beta},\phi)$ が既知のもとで $\sqrt{K}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})=O_p(1)$ が成り立つ推定量が用いられる．$\hat{\phi}$ についても同様に $\boldsymbol{\beta}$ が既知のもとで $\sqrt{K}(\hat{\phi}-\phi)=O_p(1)$ が成り立つ推定量が用いられる．(2.2) 式を $\boldsymbol{\beta}$ について解いた解を $\hat{\boldsymbol{\beta}}_G$ と定義する．$\hat{\boldsymbol{\beta}}_G$ は Theorem 2 に述べた性質を満たす．

3 Theorem 2

Theorem 2. *Under mild regularity conditions and given that:*

1. $\hat{\boldsymbol{\alpha}}$ *is* $\sqrt{K}$*-consistent given* $\boldsymbol{\beta}$ *and* ϕ*;*
2. $\hat{\phi}$ *is* $\sqrt{K}$*-consistent given* $\boldsymbol{\beta}$*; and*
3. $\frac{\partial\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)}{\partial\phi}\le H(\mathbf{Y},\boldsymbol{\beta})$ *which is* $O_p(1)$*,*

then $\sqrt{K}(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})$ *is asymptotically multivariate Gaussian with zero mean and covariance matrix* $\mathbf{V}_G$ *given by*

$$\begin{aligned}\mathbf{V}_G&=\lim_{K\to\infty}K\left(\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1}\left\{\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{Y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i\right\}\left(\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1}\\&=\lim_{K\to\infty}\left(\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1}\left\{\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{Y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i\right\}\left(\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1}.\end{aligned}$$

4 Theorem 2 の証明＋補足

Proof. (原典) まず，$\boldsymbol{\alpha}^*(\boldsymbol{\beta})=\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta},\hat{\phi}(\boldsymbol{\beta})\}$ とおき，適当な正則条件の下で $\sqrt{K}(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})$ は,

$$-\left[\sum_{i=1}^{K}\frac{\delta}{\delta\boldsymbol{\beta}^{\mathrm{T}}}\mathrm{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}/K\right]^{-1}\left[\sum_{i=1}^{K}\mathrm{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}/\sqrt{K}\right], \tag{4.1}$$

で近似できる，ただし,

$$\begin{aligned}\frac{\delta\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\delta\boldsymbol{\beta}^{\mathrm{T}}}&=\frac{\partial\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}+\frac{\partial\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\partial\boldsymbol{\alpha}^{*\mathrm{T}}}\frac{\partial\boldsymbol{\alpha}^*(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\\&=\mathbf{A}_i+\mathbf{B}_i\mathbf{C},\end{aligned} \tag{4.2}$$

である．

(補足 1) $\sum_{i=1}^{K}\mathbf{U}_i\{\hat{\boldsymbol{\beta}}_G,\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}=\mathbf{0}$ を $\hat{\boldsymbol{\beta}}_G=\boldsymbol{\beta}$ の周りで Taylor 展開すると，

$$
\begin{aligned}
\sum_{i=1}^{K}\mathbf{U}_i\{\hat{\boldsymbol{\beta}}_G,\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}&=\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}+\sum_{i=1}^{K}\frac{\delta}{\delta\boldsymbol{\beta}^{\mathrm{T}}}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})+o_p(1)\\
\Leftrightarrow \mathbf{0}&=\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}+\sum_{i=1}^{K}\frac{\delta}{\delta\boldsymbol{\beta}^{\mathrm{T}}}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})+o_p(1)\\
\Leftrightarrow (\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})&=-\left[\sum_{i=1}^{K}\frac{\delta}{\delta\boldsymbol{\beta}^{\mathrm{T}}}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}\right]^{-1}\left[\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}\right]+o_p(1)\\
\Leftrightarrow \sqrt{K}(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})&=-\left[\sum_{i=1}^{K}\frac{\delta}{\delta\boldsymbol{\beta}^{\mathrm{T}}}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}/K\right]^{-1}\left[\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}/\sqrt{K}\right]+o_p(1),
\end{aligned}
$$

である．原典では $\delta/\delta\boldsymbol{\beta}^{\mathrm{T}}$ と記号を分けているが，本質的にはベクトル関数を Taylor 展開しているだけのはずなので，(4.1) 式の第一項を偏微分した結果である (4.2) 式に直接置き換えて書いても問題ないと思われる．

(原典) $\sum_{i=1}^{K}\mathbf{B}_i=o_p(K)$，$\mathbf{C}=O_p(1)$ は簡単に分かる．また $\sum_{i=1}^{K}\mathbf{A}_i/K$ は $K\to\infty$ において $\lim_{K\to\infty}-\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i/K$ に確率収束する．

(補足 2) 定義より，

$$
\frac{\partial\mathbf{S}_i}{\partial\boldsymbol{\beta}^{\mathrm{T}}}=\frac{\partial\mathbf{Y}_i-a'(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}=\mathbf{0}-\mathbf{D}_i,
$$

を用いれば，

$$
\begin{aligned}
\frac{\partial\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}&=\frac{\partial\{\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\}(\boldsymbol{\beta})\mathbf{S}_i}{\partial\boldsymbol{\beta}^{\mathrm{T}}}+\frac{\partial\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\mathbf{S}_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\\
&=\frac{\partial\{\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\}(\boldsymbol{\beta})\mathbf{S}_i}{\partial\boldsymbol{\beta}^{\mathrm{T}}}-\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\mathbf{D}_i,
\end{aligned}
\tag{4.3}
$$

である．ただし，$\partial\{\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\}(\boldsymbol{\beta})\mathbf{S}_i/\partial\boldsymbol{\beta}^{\mathrm{T}}$ は $\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}$ だけを $\boldsymbol{\beta}$ の関数と見なした偏微分であり，

$$
\begin{pmatrix}\frac{\partial\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}}{\partial\beta_1}\mathbf{S}_i & \frac{\partial\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}}{\partial\beta_2}\mathbf{S}_i & \cdots & \frac{\partial\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}}{\partial\beta_p}\mathbf{S}_i\end{pmatrix},
$$

を意味する．ここで，

$$
\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\mathbf{S}_i=\hat{\phi}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{G}_i\mathbf{G}_i^{-1/2}\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}\mathbf{G}_i^{-1/2}\mathbf{S}_i,
$$

の右辺第一項の一部を展開すると，

$$
\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{G}_i\mathbf{G}_i^{-1/2}=\{\delta_{it}\sqrt{a''(\theta_{it})}x_{ijt}\}_{1\le j\le p,1\le t\le n},
$$

となる．ここで，$p\times n$ 次元の行列

$$
\mathbf{J}=\begin{pmatrix}1&1&\cdots&1\\1&1&\cdots&1\\\vdots&\vdots&\ddots&\vdots\\1&1&\cdots&1\end{pmatrix},
$$

および

$$a_{\sup}=\sup_{i,j,t}\{\delta_{it}\sqrt{a''(\theta_{it})}x_{ijt}\},\ b_{\sup}=\sup_{i,j,t}\left\{\frac{\partial\delta_{it}\sqrt{a''(\theta_{it})}x_{ijt}}{\partial\beta_j}\right\},$$
$$a_{\inf}=\inf_{i,j,t}\{\delta_{it}\sqrt{a''(\theta_{it})}x_{ijt}\},\ b_{\inf}=\inf_{i,j,t}\left\{\frac{\partial\delta_{it}\sqrt{a''(\theta_{it})}x_{ijt}}{\partial\beta_j}\right\},$$

とおいておく．(4.3) 式の二行目第一項と第二項の各要素は明らかに $\mathbf{S}_i$ の線型関数であり，$\mathrm{E}(\mathbf{S}_i)=\mathbf{0}$ であり，$\hat{\boldsymbol{\alpha}}$ の $\sqrt{K}$-consistency (条件 1) より $\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}\xrightarrow{P}\mathbf{R}(\boldsymbol{\alpha})^{-1}$，$\hat{\phi}$ の $\sqrt{K}$-consistency (条件 2) より $\hat{\phi}\xrightarrow{P}\phi$, 後述するが $\partial\hat{\phi}/\partial\beta_j=O_p(1)$, 連続写像定理および大数の弱法則より，

$$\left(\frac{\partial\hat{\phi}}{\partial\beta_j}a_{\sup}+\hat{\phi}b_{\sup}\right)\mathbf{JR}(\boldsymbol{\alpha}^*)^{-1}\frac{1}{K}\sum_{i=1}^K\mathbf{G}_i^{-1/2}\mathbf{S}_i\xrightarrow{P}\mathbf{0},$$

$$\hat{\phi}a_{\sup}\mathbf{J}\frac{\partial\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}}{\partial\beta_j}\frac{1}{K}\sum_{i=1}^K\mathbf{G}_i^{-1/2}\mathbf{S}_i\xrightarrow{P}\mathbf{0},$$

$$\hat{\phi}a_{\sup}\mathbf{JR}(\boldsymbol{\alpha}^*)^{-1}\frac{1}{K}\sum_{i=1}^K\frac{\partial\mathbf{G}_i^{-1/2}}{\partial\beta_j}\mathbf{S}_i\xrightarrow{P}\mathbf{0},$$

であり，$a_{\inf},b_{\inf}$ の組み合わせについても同様に $\mathbf{0}$ に収束する．すると，

$$\begin{aligned}\left(\frac{\partial\hat{\phi}}{\partial\beta_j}a_{\inf}+\hat{\phi}b_{\inf}\right)\mathbf{JR}(\boldsymbol{\alpha}^*)^{-1}\frac{1}{K}\sum_{i=1}^K\mathbf{G}_i^{-1/2}\mathbf{S}_i&\le\frac{1}{K}\sum_{i=1}^K\frac{\partial\hat{\phi}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{G}_i^{1/2}}{\partial\beta_j}\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}\mathbf{G}_i^{-1/2}\mathbf{S}_i\\&\le\left(\frac{\partial\hat{\phi}}{\partial\beta_j}a_{\sup}+\hat{\phi}b_{\sup}\right)\mathbf{JR}(\boldsymbol{\alpha}^*)^{-1}\frac{1}{K}\sum_{i=1}^K\mathbf{G}_i^{-1/2}\mathbf{S}_i\end{aligned}$$

と挟み撃ちの原理より

$$\frac{1}{K}\sum_{i=1}^K\frac{\partial\hat{\phi}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{G}_i^{1/2}}{\partial\beta_j}\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}\mathbf{G}_i^{-1/2}\mathbf{S}_i\xrightarrow{P}\mathbf{0},$$

である．他の場合も同様であり

$$\frac{1}{K}\sum_{i=1}^K\frac{\partial\{\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\}(\boldsymbol{\beta})\mathbf{S}_i}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\xrightarrow{P}\mathbf{0},$$

が成り立つ．また，$\hat{\boldsymbol{\alpha}}$ と $\hat{\phi}$ の $\sqrt{K}$-consistentcy (条件 1–2) より，$K\to\infty$ において，

$$\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\mathbf{D}_i\xrightarrow{P}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i,$$

である．よって，

$$\frac{1}{K}\sum_{i=1}^K\mathbf{A}_i\xrightarrow{P}\lim_{K\to\infty}-\frac{1}{K}\sum_{i=1}^K\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i,$$

が成り立つ．

$\partial\hat{\phi}/\partial\beta_j = O_p(1)$ の証明については，(補足 4) の結果がほぼそのまま適用できる．ここで，

$$r_{it} = \frac{Y_{it} - a'(\theta_{it})}{\sqrt{a''(\theta_{it})}},$$

とおく．ϕ の推定量は

$$\hat{\phi}^{-1} = \frac{1}{nK-p}\sum_{i=1}^{K}\sum_{t=1}^{n} r_{it}^2,$$

を用いて得られる．したがって，$\partial r_{it}^2/\partial\beta_j = \partial r_{it}r_{it}/\partial\beta_j$ と見れば，(補足 4) と同様の議論より，$\partial r_{it}^2/\partial\beta_j = O_p(1)$ と言える．

(補足 3) 定義より，

$$\frac{\partial\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\partial\boldsymbol{\alpha}^{*\mathrm{T}}} = \frac{\partial\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}(\boldsymbol{\alpha}^*)\mathbf{S}_i}{\partial\boldsymbol{\alpha}^{*\mathrm{T}}},$$

である．前補足とほぼ同様であり，各要素

$$\mathbf{D}_i^{\mathrm{T}}\frac{\partial\hat{\mathbf{V}}_i^{-1}(\boldsymbol{\alpha}^*)}{\partial\alpha_l^*}\mathbf{S}_i,$$

で考えれば良い．ここで，相関行列の非対角要素が高々α_l^c (c は適当な正整数) であることと，逆行列の微分の公式から $\mathbf{R}^{-1}/\partial\alpha_l^* = -\mathbf{R}^{-1}(\mathbf{R}/\partial\alpha_l^*)\mathbf{R}^{-1}$ であることより，$\partial\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}/\partial\alpha_l^* = O_p(1)$ である．$\hat{\boldsymbol{\alpha}}$ の $\sqrt{K}$-consistentcy (条件 1)，大数の弱法則および連続写像定理より，

$$\hat{\phi}a_{\mathrm{inf}}\frac{\partial\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}}{\partial\alpha_l^*}\frac{1}{K}\sum_{i=1}^{K}\mathbf{G}_i^{-1/2}\mathbf{S}_i \xrightarrow{P} \mathbf{0},$$

$$\hat{\phi}a_{\mathrm{sup}}\frac{\partial\mathbf{R}(\boldsymbol{\alpha}^*)^{-1}}{\partial\alpha_l^*}\frac{1}{K}\sum_{i=1}^{K}\mathbf{G}_i^{-1/2}\mathbf{S}_i \xrightarrow{P} \mathbf{0},$$

である．挟み撃ちの原理を用いれば，

$$\frac{1}{K}\sum_{i=1}^{K}\frac{\partial\mathbf{D}_i^{\mathrm{T}}\hat{\mathbf{V}}_i^{-1}\mathbf{S}_i}{\partial\alpha_l^*} \xrightarrow{P} \mathbf{0},$$

となる．したがって，

$$\frac{1}{K}\sum_{i=1}^{K}\mathbf{B}_i = o_p(1),$$

である．

(補足 4) 一般の $\mathbf{R}(\boldsymbol{\alpha})$ の (u,v) 要素の推定量は

$$R_{uv}^* = \frac{1}{K-p}\sum_{i=1}^{K}r_{iu}r_{iv} = \frac{1}{K-p}\sum_{i=1}^{K}\frac{Y_{iu}Y_{iv} - a'(\theta_{iu})Y_{iv} - a'(\theta_{iv})Y_{iu} + a'(\theta_{iu})a'(\theta_{iv})}{\sqrt{a''(\theta_{iu})a''(\theta_{iv})}},$$

の単純な関数により構成されている．よって，

$$\begin{aligned}\frac{\partial r_{iu}r_{iv}}{\partial\beta_j} =& \frac{-a''(\theta_{iu})Y_{iv} - a''(\theta_{iv})Y_{iu} + a''(\theta_{iu})a'(\theta_{iv}) + a'(\theta_{iu})a''(\theta_{iv})}{\sqrt{a''(\theta_{iu})a''(\theta_{iv})}} \\ &- \frac{Y_{iu}Y_{iv} - a'(\theta_{iu})Y_{iu} - a'(\theta_{iv})Y_{iv} + a'(\theta_{iu})a'(\theta_{iv})}{2\{a''(\theta_{iu})a''(\theta_{iv})\}^{3/2}\{a'''(\theta_{iu})a''(\theta_{iv}) + a''(\theta_{iu})a'''(\theta_{iv})\}},\end{aligned}$$

である．ここで，大数の弱法則から，

$$\frac{1}{K-p}\sum_{i=1}^{K}\frac{a''(\theta_{iu})\{a'(\theta_{iv})-Y_{iv}\}}{\sqrt{a''(\theta_{iu})a''(\theta_{iv})}}\xrightarrow{P}0,$$

$$\frac{1}{K-p}\sum_{i=1}^{K}\frac{a''(\theta_{iv})\{a'(\theta_{iu})-Y_{iu}\}}{\sqrt{a''(\theta_{iu})a''(\theta_{iv})}}\xrightarrow{P}0,$$

である．また，

$$\frac{1}{K-p}\sum_{i=1}^{K}\frac{Y_{iu}Y_{iv}-a'(\theta_{iu})Y_{iu}-a'(\theta_{iv})Y_{iv}+a'(\theta_{iu})a'(\theta_{iv})}{2\{a''(\theta_{iu})a''(\theta_{iv})\}^{3/2}\{a'''(\theta_{iu})a''(\theta_{iv})+a''(\theta_{iu})a'''(\theta_{iv})\}}=\frac{1}{K-p}\sum_{i=1}^{K}c_k r_{iu}r_{iv},$$

と書け (c_k は実定数)，

$$\frac{1}{K-p}\left|\sum_{i=1}^{K}c_k r_{iu}r_{iv}\right|\leq\frac{d}{K-p}\left|\sum_{i=1}^{K}r_{iu}r_{iv}\right|,$$

を満たす様な実数 $d<\infty$ が取れるため，$\hat{\boldsymbol{\alpha}}$ の $\sqrt{K}$-consistency より，$\mathbf{C}$ についても $O_p(1)$ となることが分かる．また，他の場合も r_{it} の積によって定義された似たような推定量が用いられるため，同様の議論が成り立つ．原典では $\sum_{i=1}^{K}\mathbf{B}_i$ と $\mathbf{C}$ の漸近的な結果は簡単に分かると書いてあるが，あまり簡単ではないように思える (ただし，私が勉強不足なだけで，もっとエレガントに証明できるかもしれない)．

(原典) (4.1) 式の第二項について，$\boldsymbol{\alpha}^*(\boldsymbol{\beta})=\boldsymbol{\alpha}$ の周りで Taylor 展開すると，

$$\begin{aligned}\sum_{i=1}^{K}\frac{\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}}{\sqrt{K}}&=\sum_{i=1}^{K}\frac{\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})}{\sqrt{K}}+\left\{\frac{1}{K}\frac{\partial\sum_{i=1}^{K}\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}\right\}\sqrt{K}(\boldsymbol{\alpha}^*(\boldsymbol{\beta})-\boldsymbol{\alpha})+o_p(1)\\&=\mathbf{A}^*+\mathbf{B}^*\mathbf{C}^*+o_p(1),\end{aligned}\tag{4.4}$$

である．$\partial\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^{\mathrm{T}}$ は期待値 $\mathbf{0}$ の $\mathbf{S}_i$ の線型関数であるため，$\mathbf{B}^*=o_p(1)$ が成り立つ．また，条件 1–3 のもとで，

$$\begin{aligned}\mathbf{C}^*&=\sqrt{K}\left[\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta},\hat{\phi}(\boldsymbol{\beta})\}-\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)+\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)-\boldsymbol{\alpha}\right]\\&=\sqrt{K}\left\{\frac{\partial\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi^*)}{\partial\phi}(\hat{\phi}-\phi)+\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)-\boldsymbol{\alpha}\right\}=O_p(1),\end{aligned}\tag{4.5}$$

が成り立つ．結果的に，$\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})\}/\sqrt{K}$ は漸近的に $\mathbf{A}^*$ と同等であることが示された．

(補足 5) 定義より，

$$\frac{1}{K}\frac{\partial\sum_{i=1}^{K}\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}=\frac{1}{K}\sum_{i=1}^{K}\frac{\partial\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}(\boldsymbol{\alpha})\mathbf{D}_i^{\mathrm{T}}}{\partial\boldsymbol{\alpha}^{\mathrm{T}}},$$

である．これは期待値 $\mathbf{0}$ の確率変数 $\mathbf{S}_i$ の線型関数になっているので，

$$\mathrm{E}\left[\mathbf{D}_i^{\mathrm{T}}\frac{\partial\mathbf{V}_i^{-1}}{\partial\alpha_l}\mathbf{S}_i\right]=\mathrm{E}(\mathbf{T}_{il})=\mathbf{0},$$

がそれぞれの (i,l) について成り立つ．ここで $\mathbf{T}_i = (\mathbf{T}_{i1},\dots,\mathbf{T}_{is})$ とおけば，あとは大数の弱法則より，

$$\mathbf{B}^* = \frac{1}{K}\sum_{i=1}^{K}\mathbf{T}_i = o_p(1),$$

が示される．

(補足 6) (4.5) 式の一行目から二行目については，平均値の定理を用いれば，

$$\frac{\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta},\hat{\phi}(\boldsymbol{\beta})\} - \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)}{\hat{\phi}-\phi} = \frac{\partial\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi^*)}{\partial\phi},$$

ただし，ϕ^* は $\hat{\phi}$ と ϕ の中点である．これを変形して，

$$\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta},\hat{\phi}(\boldsymbol{\beta})\} - \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi) = \frac{\partial\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi^*)}{\partial\phi}(\hat{\phi}-\phi),$$

を得る．ここで，

$$\mathbf{C}^* = \frac{\partial\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi^*)}{\partial\phi}\sqrt{K}(\hat{\phi}-\phi) + \sqrt{K}\{\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)-\boldsymbol{\alpha}\},$$

である．条件 3 では $\partial\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\phi)/\partial\phi \le H(\mathbf{Y},\boldsymbol{\beta})$，すなわち ϕ とは無関係な関数で抑えられることを仮定していることに注意すれば (実際，$\hat{\alpha}$ は $\phi\times(R^*_{uv}$ の関数) または (R^*_{uv} の関数) のいずれかの場合で $\sqrt{K}$-consistent な推定量が構成できる)，連続写像定理と条件 1–3 を用いて，

$$\mathbf{C}^* \le H(\mathbf{Y},\boldsymbol{\beta})O_p(1) + O_p(1) = O_p(1)O_p(1) + O_p(1) = O_p(1),$$

となる．

(原典) $\mathbf{A}^*$ は平均ベクトル $\mathbf{0}$，分散共分散行列が，

$$\lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{Y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i,$$

の正規分布に従う．

(補足 7) 確率変数ベクトル $\mathbf{Y}_i$ について，期待値は $\mathrm{E}(Y_{it}) = a'(\theta_{it})$，分散は存在することを仮定している．また，

$$\sum_{i=1}^{K}\frac{\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})}{\sqrt{K}} = \sum_{i=1}^{K}\frac{\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{S}_i}{\sqrt{K}}$$

である．よって，

$$\mathrm{E}\left\{\sum_{i=1}^{K}\frac{\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{S}_i}{\sqrt{K}}\right\} = \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{E}(\mathbf{S}_i) = \frac{1}{\sqrt{K}}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{E}\{\mathbf{Y}_i-\mathrm{E}(\mathbf{Y}_i)\} = \mathbf{0},$$

および，

$$\mathrm{Cov}\left\{\sum_{i=1}^{K}\frac{\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{S}_i}{\sqrt{K}}\right\} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{S}_i)\mathbf{V}_i^{-1}\mathbf{D}_i = \frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{Y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i,$$

である．したがって，多変量中心極限定理より，$\mathbf{A}^*$ は平均ベクトルが $\mathbf{0}$，分散共分散行列が

$$\lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{Y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i,$$

の多変量正規分布に分布収束する．以上より Theorem 2 は示された．

(補足 8) (4.1) 式第一項は $-[K^{-1}\sum_{i=1}^{K}\mathbf{A}_i+\mathbf{B}_i\mathbf{C}]^{-1}$ であり，$K\to\infty$ において，$K^{-1}\sum_{i=1}^{K}\mathbf{A}_i$ は $\lim_{K\to\infty}-K^{-1}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i$ に確率収束し，また，$K^{-1}(\sum_{i=1}^{K}\mathbf{B}_i)\mathbf{C}=o_p(1)O_p(1)=o_p(1)$ である．したがって，(4.1) 式第一項は $K\to\infty$ において，

$$\mathbf{M}_0^{-1}=\left(\lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1},$$

に確率収束する．また，(4.1) 式第二項は平均ベクトルが $\mathbf{0}$，分散共分散行列が

$$\mathbf{M}_1=\lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{Y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i,$$

の多変量正規分布に分布収束する．(4.1) 式第二項の収束先を $\mathbf{Z}\sim N(\mathbf{0},\mathbf{M}_1)$ と書けば，連続写像定理より，

$$\sqrt{K}(\hat{\boldsymbol{\beta}}_G-\boldsymbol{\beta})\approx (4.1)\,\text{式}\xrightarrow{d}\mathbf{M}_0^{-1}\mathbf{Z}\sim N(\mathbf{0},\mathbf{M}_0^{-1}\mathbf{M}_1\mathbf{M}_0^{-1}),$$

である．積の極限と極限の積は交換可能であるため，

$$\mathbf{M}_0^{-1}\mathbf{M}_1\mathbf{M}_0^{-1}=\\ \lim_{K\to\infty}\left(\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1}\left\{\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathrm{Cov}(\mathbf{Y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i\right\}\left(\frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1},$$

が成り立つ． □

5 考察

相関構造の誤特定によらず一致性がある，という主張については，(4.1) 式第二項が漸近的には $\mathbf{A}^*$ だけ評価すればよいという部分に依る．上述の通り，$\mathrm{E}(\mathbf{A}^*)=\mathbf{0}$ であり，これは $\mathbf{V}_i^{-1}$ の構造には関係なく成り立つためである．

また，本文中にも言及があるように，$\hat{\boldsymbol{\beta}}_G$ の推定においては $\hat{\boldsymbol{\alpha}}$ や $\hat{\phi}$ の推定精度が問題となる．実際には K が十分大きくなく，n が大きく，無構造の作業相関行列を仮定した場合などは，正規近似がうまくいかずに信頼区間の被覆確率が名義よりもかなり小さくなってしまうことは容易に想像できる．また，$\hat{\phi}$ の $\sqrt{K}$-consistency が成り立つためには Y_{it} の四次モーメントが有限であることを仮定する必要がある．

原典では正則条件についてはほぼ触れられていない．$\mathbf{M}_0$ や $\mathbf{M}_1$ が極限を含んだ形で示されている点はかなり気がかりである．私は未読だが，GEE の漸近的な結果や正則条件を議論している文献をいくつか挙げておく (Crowder, 1986; Li, 1997; Xie and Yang, 2003; Balan and Schiopu-Kratina, 2005, 時間を見つけて読んでいきたい)．補足 1 の剰余項が

$o_p(1)$ となるためには，$\mathbf{U}_i$ の二階偏微分が存在しかつパラメータベクトルの真値の近傍で確率有界でなければならない．

また，Pepe and Anderson (1994) や Pan et al. (2000) にあるように，周辺分布の期待値にある仮定が満たされない場合，対角以外の作業相関行列を用いると推定方程式が不偏でなくなる場合があることが知られているため，注意が必要だろう．

6 推定アルゴリズムの概要

1. 独立の場合などから，初期値 $\hat{\boldsymbol{\beta}}_r$ $(r=1)$ を決める．
2. スケールパラメータがある場合はその推定量 $\hat{\phi}_r$ を，現在の $\hat{\boldsymbol{\beta}}_r$ の値により求める．
3. 作業相関行列の推定量 $\hat{\boldsymbol{\alpha}}_r$ を，現在の $\hat{\boldsymbol{\beta}}_r$ と $\hat{\phi}_r$ の値により求める．
4. $\mathbf{D}_i$, $\mathbf{V}_i$, $\mathbf{S}_i$ を，現在の $\hat{\boldsymbol{\beta}}_r$, $\hat{\phi}_r$, $\hat{\boldsymbol{\alpha}}_r$ の値により求める．
5. 以下により，$\hat{\boldsymbol{\beta}}_{r+1}$ に更新する

$$\hat{\boldsymbol{\beta}}_{r+1} = \hat{\boldsymbol{\beta}}_r + \left\{ \sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i \right\}^{-1} \left\{ \sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{S}_i \right\}.$$

6. 収束するまで 2–5 を繰り返す．

参考文献

Balan RM, Schiopu-Kratina I. Asymptotic results with generalized estimating equations for longitudinal data. *The Annals of Statistics* 2005; **33**(2): 522–541.

Crowder M. On consistency and inconsistency of estimating equations. *Econometric Theory* 1986; **2**(3): 305–330.

Li B. On consistency of generalized estimating equations. *Institute of Mathematical Statistics Lecture Notes - Monograph Series* 1997, 115–136.

Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**(1): 13–22.

Pan W, Louis TA, Connett JE. A note on marginal linear regression with correlated response data. *The American Statistician* 2000; **54**(3): 191–195.

Pepe MS, Anderson GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics – Simulation and Computation* 1994; **23**(4): 939–951.

Xie M, Yang Y. Asymptotics for generalized estimating equations with large cluster sizes. *The Annals of Statistics* 2003; **31**(1): 310–347.