数理解析研究所講究録
第1910巻 2014年 43-57

経験ベイズモデルにおける条件付赤池情報量規準

東京大学・大学院経済学研究科* 川久保 友超
Yuki Kawakubo
Graduate School of Economics
University of Tokyo

概要

線形混合モデルおよび一般化線形混合モデルにおける変数選択規準として，条件付赤池情報量規準（conditional AIC, cAIC）はその重要性が認められてきた．cAICは条件付尤度にもとづいた期待Kullback-Leibler情報量の推定に関係している．同様の観点はPoisson-Gammaモデルや Binomial-Betaモデルを含む自然指数型分布族にもとづいた経験ベイズモデルにおいても重要であり，これらのモデルの変数選択規準としてcAICを導出する．

キーワード：Akaike information criterion, conditional AIC, empirical Bayes, variable selection.

1 はじめに

母数効果と変量効果を混合させた線形混合モデルと，そこから導かれる経験最良線形不偏予測量（empirical best linear unbiased predictor, EBLUP）の性質は，理論と応用の双方から長年にわたって研究され続けてきた．応用分野としては計量経済学におけるパネルデータ分析，生物統計や医療統計における経時データ分析，官庁統計における小地域推定などが主に知られている．線形混合モデルにおいて，良い予測を与えるための説明変数の選択は，重要な問題である．従来からの手法の1つは，Akaike (1973, 1974) によって提案された赤池情報量規準（Akaike information criterion, AIC）である．AICを線形混合モデルに用いる際は，変量効果について積分消去した周辺尤度にもとづいた情報量を考

〒113-0033 東京都文京区本郷 7-3-1, Email: y.k.5.58.2010@gmail.com

えていることになる．しかし，特定の変量効果（すなわちそれが意味するクラスターや小地域）における予測の観点からは，AIC が望ましいとは言えない．

そこで Vaida and Blanchard (2005) は特定の変量効果における予測の観点から，条件付赤池情報量規準（conditional AIC, cAIC）を提案した．情報量規準 cAIC は，変量効果を所与とした条件付密度にもとづいた期待 Kullback-Leibler 情報量の推定に関係している．その後 cAIC は，線形混合モデルにおける変数選択規準として，Liang et al. (2008), Greven and Kneib (2010), Srivastava and Kubokawa (2010), Kubokawa (2011), Kawakubo and Kubokawa (2014) 等で理論的な研究が進んだ．詳しくは Müller et al. (2013) が良いサーベイ論文となっている．さらに cAIC は，応答変数の分布を正規分布から指数型分布族に拡張した線形混合モデル，すなわち一般化線形混合モデルにおける変数選択規準としても，Donohue et al. (2011), Yu and Yau (2012), Yu, Zhang and Yau (2013) 等によって提案されている．

上記のように線形混合モデル，および一般化線形混合モデルに対する変数選択規準として cAIC はその重要性が認められてきた．それは特定の変量効果における予測の観点から，周辺尤度にもとづいた従来の AIC よりも良いとされるからである．これは，変量効果を所与とした上で現在の観測ベクトルと独立同一分布に従う将来の値の予測を行うという，ベイズ予測を考えていることになる．同様に，Poisson-Gamma モデルや Binomia-Beta モデルのような，平均にランダムネスを持たせた階層型のモデルにおいても，ランダムネスを止めたうえでの将来の値の予測，すなわちベイズ予測を考えることが望ましい．

そこで本研究では，自然指数型分布族（natural exponential family, NEF）と，その平均パラメータに共役事前分布を入れた経験ベイズモデルに対し，cAIC で変数選択を行うことを提案する．本論文の残りの構成は以下の通りである．まず第 2 節で，本研究で取り扱うモデルと cAIC について説明を行う．次に第 3 節では，cAIC におけるペナルティ項の評価を行い，第 4 節で数値実験の結果を報告する．なお本研究は今後の研究につながる予報（annaouncement）としての位置づけであることを記しておく．

2 モデルと条件付赤池情報量規準

2.1 モデル

$y_1,\ldots,y_k$ は互いに独立な確率変数とし，θ_i を所与とした y_i の条件付分布が以下のような自然指数型分布族（natural exponential family, NEF）に入るとする．

$$y_i|\theta_i \sim f(y_i|\theta_i,n_i)=\exp\left[n_i(\theta_i y_i-\psi(\theta_i))+c(y_i,n_i)\right]. \tag{2.1}$$

ここで θ_i が自然母数，n_i は尺度母数で既知とする．(2.1) にしたがう確率変数 y_i の例として，以下のような状況を考える．

クラスター内で独立同一分布にしたがう確率変数の標本平均が観測されている状況でのモデリングを考える．クラスター $i=1,\ldots,k$ に対して，$Z_{i1},\ldots,Z_{i,n_i}$ が θ_i を所与として独立同一分布にしたがい，以下のような 1 パラメータの自然指数型分布族に入るとする．

$$P(Z_{ij}\in A)=\int_A \exp\{\theta_i z-\psi(\theta_i)\}\,\mathrm{d}F(z). \tag{2.2}$$

ただし θ_i は自然母数，F は実数上の Stieltjes 測度である．ここで $y_i=(Z_{i1}+\cdots+Z_{i,n_i})/n_i$ と定義すると，θ_i を所与とした y_i の条件付き分布は (2.1) にしたがう．このような問題設定は，NEF を用いたモデリングで小地域の平均の信頼区間を提案した Ghosh and Maiti (2008) で用いられており，$k\to\infty, n_i\to\infty$ の漸近理論のもとで議論されている．本研究においても $k\to\infty, n_i\to\infty$ の条件下で変数選択規準を提案する．

指数型分布族の性質より，θ_i を所与としたときの y_i の平均を μ_i とすると，

$$\mu_i=E(y_i|\theta_i)=\psi'(\theta_i).$$

さらに $\psi''(\theta_i)=Q(\mu_i)$ と定義すると，θ_i を所与としたときの y_i の分散は，

$$Var(y_i|\theta_i)=\frac{\psi''(\theta_i)}{n_i}=\frac{Q(\mu_i)}{n_i},$$

で与えられる．$Q(\mu_i)$ は variance function と呼び，以後本論文では $Q(\mu_i)$ として高々二次関数であるような分布のクラスに制限して議論を進めることとする．すなわち同時に 0 とならない既知の v_0,v_1,v_2 に対して $Q(x)=v_0+v_1x+v_2x^2$ と定める．このような分布のクラスは，natural exponential family with quadratic variance function (NEF-QVF) と呼ばれ，Morris (1982, 1983) により導入され，その性質が調べられた．

次に自然母数 θ_i に，以下のような（2.1）に対する共役事前分布を入れる．

$$\theta_i|\lambda, m_i \sim \pi(\theta_i|\lambda, m_i) = \exp\left[\lambda(m_i\theta_i - \psi(\theta_i))\right] C(\lambda, m_i). \tag{2.3}$$

このとき，μ_i の事前分布における平均と分散は，

$$E(\mu_i|\lambda, m_i) = m_i, \quad Var(\mu_i|\lambda, m_i) = \frac{Q(m_i)}{\lambda - v_2},$$

で与えられる．ここで p 次元の補助変数 $\boldsymbol{x}_i$ が利用可能なとき，平均 m_i の予測量として $\boldsymbol{x}_i^t\boldsymbol{\beta}$ を利用したい．ただし $\boldsymbol{\beta}$ は未知パラメータで p 次元ベクトルとする．しかし平均 m_i のとりうる値には制約がある場合が多く，$\boldsymbol{x}_i^t\boldsymbol{\beta}$ との間は適切なリンク関数 $h(\cdot)$ を用いて $m_i = h^{-1}(\boldsymbol{x}_i^t\boldsymbol{\beta})$ という構造を入れる．以後簡単化のために，リンク関数は以下のような正準リンク関数を考える．

$$m_i = h^{-1}(\boldsymbol{x}_i^t\boldsymbol{\beta}) = \psi'(\boldsymbol{x}_i^t\boldsymbol{\beta}).$$

$\psi'(\cdot)$ は NEF において自然母数 θ_i と平均母数 μ_i を結ぶ関数であり，その意味で自然である．またこのモデルにおける未知パラメータを $\boldsymbol{\eta} = (\boldsymbol{\beta}^t, \lambda)^t$ とおく．

y_i を所与とした θ_i の事後分布は，

$$\pi(\theta_i|y_i, \lambda, m_i) = \exp\left[(n_i + \lambda)(\widehat{\mu}_i^B\theta_i - \psi(\theta_i))\right] C(n_i + \lambda, \widehat{\mu}_i^B) \tag{2.4}$$

となる．ただし $\widehat{\mu}_i^B$ は二乗損失のもとでの μ_i のベイズ推定量 $E(\mu_i|y_i)$ であり，

$$\widehat{\mu}_i^B = \widehat{\mu}_i(y_i, \boldsymbol{\eta}) = \frac{n_i y_i + \lambda m_i}{n_i + \lambda}, \quad m_i = \psi'(\boldsymbol{x}_i^t\boldsymbol{\beta}), \tag{2.5}$$

で与えられる．$\widehat{\mu}_i(y_i, \boldsymbol{\eta})$ における $\boldsymbol{\eta}$ をその推定量 $\widehat{\boldsymbol{\eta}}$ でおきかえると，μ_i の経験ベイズ（empirical Bayes, EB）推定量

$$\widehat{\mu}_i^{EB} = \widehat{\mu}_i(y_i, \widehat{\boldsymbol{\eta}}) = \frac{n_i y_i + \hat{\lambda}\hat{m}_i}{n_i + \hat{\lambda}}, \quad \hat{m}_i = \psi'(\boldsymbol{x}_i^t\widehat{\boldsymbol{\beta}}), \tag{2.6}$$

が得られる．（2.5）式で与えられるベイズ推定量 $\widehat{\mu}_i^B$ は，データ y_i と事前分布の平均 m_i の加重平均となっており，m_i はクラスター間で共通のパラメータ $\boldsymbol{\beta}$ の入った平均構造 $\boldsymbol{x}_i^t\boldsymbol{\beta}$ と結ばれている．$\boldsymbol{\beta}$ は全データを用いた $\boldsymbol{y}$ の周辺分布から推定するため，クラスター数 k が大きいときには安定した推定が期待される．よって（2.6）式の経験ベイズ推定量 $\widehat{\mu}_i^{EB}$ は，μ_i の予測量として望ましい精度をもつことが期待される．

以下 NEF-QVF とその共役事前分布で構成されるモデルの例を 3 つ挙げる．

[1. Fay-Herriot model] 分散既知の正規分布モデルである．$y_i \sim \mathcal{N}(\mu_i, n_i^{-1})$ すなわち y_i が平均 μ_i，分散 n_i^{-1} の正規分布にしたがっているとき，y_i の pdf は，(2.1) において $\theta_i = \mu_i, \psi(\theta_i) = \theta_i^2/2$ としたものになり，variance function は $v_0 = 1, v_1 = v_2 = 0$ である．また μ_i の共役事前分布は，平均 m_i，分散 λ^{-1} の正規分布であり，$\mu_i \sim \mathcal{N}(m_i, \lambda^{-1})$ と表記する．m_i と予測量 $\boldsymbol{x}_i^t\boldsymbol{\beta}$ とのリンクは恒等関数で，$m_i = \boldsymbol{x}_i^t\boldsymbol{\beta}$ である．このモデルは簡便的ではあるものの，Fay and Herriot (1979) で提案されて以降，小地域推定の地域レベルモデルとして最もよく利用されている．

[2. Poisson-Gamma model] $Z_{i1}, \ldots, Z_{i,n_i}$ が独立に平均 μ_i の Poisson 分布にしたがっているとき，$y_i = (Z_{i1} + \cdots + Z_{i,n_i})/n_i$ のしたがう分布の pmf は，(2.1) において $\theta_i = \log \mu_i = h(\mu_i), \psi(\theta_i) = e^{\theta_i}$ としたものになり，variance function は $v_1 = 1, v_0 = v_2 = 0$ である．また μ_i の共役事前分布は，形状母数 λm_i，尺度母数 $1/\lambda$ のガンマ分布であり，$\mu_i \sim Ga(\lambda m_i, 1/\lambda)$ と表記する．m_i と予測量 $\boldsymbol{x}_i^t\boldsymbol{\beta}$ とのリンクは，$m_i = \psi'(\boldsymbol{x}_i^t\boldsymbol{\beta}) = \exp(\boldsymbol{x}_i^t\boldsymbol{\beta})$ である．

[3. Binomial-Beta model] $Z_{i1}, \ldots, Z_{i,n_i}$ が独立に平均 μ_i のベルヌーイ分布にしたがっているとき，$Z_{i1} + \cdots + Z_{i,n_i}$ は試行数 n_i，平均 μ_i の二項分布にしたがう．このとき $y_i = (Z_{i1} + \cdots + Z_{i,n_i})/n_i$ のしたがう分布の pmf は，(2.1) において $\theta_i = \log\{\mu_i/(1-\mu_i)\} = h(\mu_i), \psi(\theta_i) = \log(1+e^{\theta_i})$ としたものになり，variance function は $v_0 = 0, v_1 = 1, v_2 = -1$ である．また μ_i の共役事前分布は，母数 $\lambda m_i, \lambda(1-m_i)$ のベータ分布であり，$\mu_i \sim Be(\lambda m_i, \lambda(1-m_i))$ と表記する．m_i と予測量 $\boldsymbol{x}_i^t\boldsymbol{\beta}$ とのリンクは，$m_i = \psi'(\boldsymbol{x}_i^t\boldsymbol{\beta}) = \exp(\boldsymbol{x}_i^t\boldsymbol{\beta})/\{1 + \exp(\boldsymbol{x}_i^t\boldsymbol{\beta})\}$ である．

2.2 条件付赤池情報量規準

Vaida and Blanchard (2005) は Kullback-Leibler divergence を損失関数とした plug-in 予測密度のリスクの一部として conditional Akaike Information (cAI) を定義し，その（漸近）不偏推定量として条件付赤池情報量規準（conditional Akaike Information Criterion, cAIC）を提案した．そこで 2.1 節で説明したモデルにおいて，補助変数 $\boldsymbol{x}_i$ の選択を行うための規準として，cAIC を提案したい．

$\tilde{y}_i$ を θ_i を所与として y_i と独立同一分布にしたがう将来の確率変数とする．このとき plug-in 予測密度の対数は，

$$\log f(\tilde{y}_i|\hat{\theta}_i^{EB}) = n_i(\hat{\theta}_i^{EB}\tilde{y}_i - \psi(\hat{\theta}_i^{EB})) + c(\tilde{y}_i, n_i), \tag{2.7}$$

で与えられる．ただし $\hat{\theta}_i^{EB} = \psi'^{-1}(\hat{\mu}_i^{EB}) = h(\hat{\mu}_i^{EB})$ とする．(2.7) 式の第 2 項 $c(\tilde{y}_i, n_i)$

はモデルに依存しないので，2.1 節のモデルにおける cAI を以下のように定義する．

$$\begin{aligned} cAI &= -2E^{\tilde{\boldsymbol{y}},\boldsymbol{y},\boldsymbol{\theta}}\left[\sum_{i=1}^{k} n_i(\hat{\theta}_i^{EB}\tilde{y}_i - \psi(\hat{\theta}_i^{EB}))\right] \\ &= -2E^{\boldsymbol{y},\boldsymbol{\theta}}\left[\sum_{i=1}^{k}\int n_i(\hat{\theta}_i^{EB}\tilde{y}_i - \psi(\hat{\theta}_i^{EB}))\mathrm{d}F(\tilde{y}_i|\theta_i,\boldsymbol{\eta})\right] \\ &= -2E^{\boldsymbol{y},\boldsymbol{\theta}}\left[\sum_{i=1}^{k} n_i(\hat{\theta}_i^{EB}\mu_i - \psi(\hat{\theta}_i^{EB}))\right]. \end{aligned} \tag{2.8}$$

これを $-2\sum_{i=1}^{k} n_i(\hat{\theta}_i^{EB} - \psi(\hat{\theta}_i^{EB}))$ で推定したときのバイアスは，以下のようになる．

$$\begin{aligned} & E\left[-2\sum_{i=1}^{k} n_i(\hat{\theta}_i^{EB} - \psi(\hat{\theta}_i^{EB}))\right] - cAI \\ &= -2E\left[\sum_{i=1}^{k} n_i\hat{\theta}_i^{EB}(y_i - \mu_i)\right] (\equiv -2B). \end{aligned} \tag{2.9}$$

そこで 2.1 節のモデルにおける条件付赤池情報量規準（cAIC）を，cAI のバイアス補正した推定量として以下のように提案する．

$$cAIC = -2\sum_{i=1}^{k} n_i(\hat{\theta}_i^{EB}y_i - \psi(\hat{\theta}_i^{EB})) + 2\widehat{B}, \tag{2.10}$$

ただし $\widehat{B}$ は B の漸近不偏推定量であり，バイアス補正項もしくはペナルティ項と呼ぶ．次節でペナルティ項 B の漸近近似および漸近不偏推定量を与える．

3 ペナルティ項の近似と推定

ペナルティ項 B を正確に評価することは難しいため，大きい k に対して B の 2 次漸近近似を与えることとする．$\hat{\theta}_i^{EB} = h\{\widehat{\mu}_i(y_i,\widehat{\boldsymbol{\eta}})\}$ を $\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta}$ まわりで以下のようにテイラー展開する．

$$h(\widehat{\mu}_i^{EB}) = h(\widehat{\mu}_i^{B}) + \frac{\partial h(\widehat{\mu}_i^{B})}{\partial\boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + \frac{1}{2}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t\frac{\partial^2 h(\widehat{\mu}_i^{B})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + o_p(k^{-1}),$$

これを用いて（2.9）式で与えられるペナルティ項 B は以下のように展開される．

$$\begin{aligned}B=&\sum_{i=1}^{k}n_iE\left[h(\widehat{\mu}_i^B)(y_i-\mu_i)\right]+\sum_{i=1}^{k}n_iE\left[(y_i-\mu_i)\frac{\partial h(\widehat{\mu}_i^B)}{\partial\boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\right],\\&+\frac{1}{2}\sum_{i=1}^{k}n_iE\left[(y_i-\mu_i)(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})^t\frac{\partial^2h(\widehat{\mu}_i^B)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\right]+o(1)\\\equiv&B_1+B_2+2^{-1}B_3+o(1),\end{aligned}\tag{3.1}$$

ただし $B_1=O(k),B_2=O(1),B_3=O(1)$ である．（3.1）のそれぞれの項は，以下に続く小節で評価していく．

3.1 B_1 の評価

本小節では以下で与えられる B_1 を評価する．

$$B_1=E\left[\sum_{i=1}^{k}n_ih(\widehat{\mu}_i^B)(y_i-\mu_i)\right].\tag{3.2}$$

ベイズ推定量 $\widehat{\mu}_i^B=(n_iy_i+\lambda m_i)/(n_i+\lambda)$ は y_i の線形関数で書かれているにも関わらず，平均母数と自然母数を結ぶ関数 $h(\cdot)$ は，正規分布を除いてほとんどの自然指数型分布族に入る分布では非線形関数である．そのため B_1 を正確に評価することは難しい．

そこで大きい n_i のもとでは $\widehat{\mu}_i^B$ と μ_i が近いことを利用し，（3.2）式における $h(\widehat{\mu}_i^B)=h(\mu_i+\widehat{\mu}_i^B-\mu_i)$ を展開する．その結果，$h(\widehat{\mu}_i^B)$ は確率変数 y_i と μ_i における線形関数の多項式で近似され，B_1 の漸近的な近似が得られる．

ここでは B_1 の評価の例として Poisson-Gamma モデルをとりあげる．同モデルにおいては関数 $h(\cdot)$ は対数関数であるため，$h(\widehat{\mu}_i^B)=\log(\widehat{\mu}_i^B)$ は以下のように展開される．

$$\begin{aligned}h(\widehat{\mu}_i^B)=&\log\left\{\mu_i\left(1+\frac{\widehat{\mu}_i^B-\mu_i}{\mu_i}\right)\right\}\\=&\theta_i+\frac{\widehat{\mu}_i^B-\mu_i}{\mu_i}-\frac{1}{2}\left(\frac{\widehat{\mu}_i^B-\mu_i}{\mu_i}\right)^2+\frac{1}{3}\left(\frac{\widehat{\mu}_i^B-\mu_i}{\mu_i}\right)^3+\cdots.\end{aligned}$$

ここで，$r\geq4$ に対して $E\left[\{(\widehat{\mu}_i^B-\mu_i)/\mu_i\}^r(y_i-\mu_i)\right]=O(n_i^{-3})$ であるため，ある正数

$\delta_i > 0$ について $n_i = O(k^{1/2+\delta_i})$ であるとき，以下が成り立つ．

$$B_1 = \sum_{i=1}^{k} n_i E[\theta_i(y_i - \mu_i)] + \sum_{i=1}^{k} n_i E\left[\frac{\widehat{\mu}_i^B - \mu_i}{\mu_i}(y_i - \mu_i)\right] - \frac{1}{2}E\left[\left(\frac{\widehat{\mu}_i^B - \mu_i}{\mu_i}\right)^2 (y_i - \mu_i)\right]$$
$$+ \frac{1}{3}\sum_{i=1}^{k} n_i E\left[\left(\frac{\widehat{\mu}_i^B - \mu_i}{\mu_i}\right)^3 (y_i - \mu_i)\right] + o(1), \tag{3.3}$$

以下の式が成り立つことに気を付けて，(3.3) 式の各々の項は正確に評価できる．

$$\widehat{\mu}_i^B - \mu_i = \frac{n_i}{n_i + \lambda}(y_i - \mu_i) - \frac{\lambda}{n_i + \lambda}(\mu_i - m_i).$$

いくらかのモーメント計算を経て，以下の補題を得ることができる．

補題 3.1 各々の $i = 1, \ldots, k$ について，ある正数 $\delta_i > 0$ に対して $n_i = O(k^{1/2+\delta_i})$ であるとする．このとき，Poisson-Gamma モデルにおいて，ペナルティ項の一部である (3.2) 式の B_1 は，以下のように 2 次のオーダーまで近似される．

$$B_1 = B_{11}(\boldsymbol{\eta}) + o(1),$$

ただし，

$$B_{11}(\boldsymbol{\eta}) = k - \sum_{i=1}^{k} \frac{2\lambda^2 m_i - \lambda}{2n_i(\lambda m_i - 1)}. \tag{3.4}$$

である．

(3.4) 式で与えられる $B_{11}(\boldsymbol{\eta})$ について，$B_{11}(\widehat{\boldsymbol{\eta}})$ は以下のように展開される．

$$B_{11}(\widehat{\boldsymbol{\eta}}) = B_{11}(\boldsymbol{\eta}) + \frac{\partial B_{11}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + \frac{1}{2}\mathrm{tr}\left[\frac{\partial^2 B_{11}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}^t}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t\right] + o_p(1),$$

よって B_1 のバイアス補正された推定量として，

$$\widehat{B}_1 = B_{11}(\widehat{\boldsymbol{\eta}}) - B_{12}(\widehat{\boldsymbol{\eta}}) - B_{13}(\widehat{\boldsymbol{\eta}}), \tag{3.5}$$

を提案する．ただし，

$$B_{12}(\boldsymbol{\eta}) = \frac{\partial B_{11}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^t} E(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}),$$
$$B_{13}(\boldsymbol{\eta}) = \frac{1}{2}\mathrm{tr}\left[\frac{\partial^2 B_{11}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}^t} E[(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t]\right].$$

である．ここで $\boldsymbol{\eta}$ が次小節で説明する推定方程式で推定される場合，$E[\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}]=\boldsymbol{U}^{-1}(\boldsymbol{a}+2^{-1}\boldsymbol{b})+o(k^{-1})$ および $E[(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})^t]=\boldsymbol{U}^{-1}+o(k^{-1})$ が成り立つ．ただし $\boldsymbol{U}$ および $\boldsymbol{a},\boldsymbol{b}$ はそれぞれ (3.6), (3.7) で与えられる．Poisson-Gamma モデルにおいては，$\partial B_{11}(\boldsymbol{\eta})/\partial\boldsymbol{\eta}$，$(\partial^2 B_{11}(\boldsymbol{\eta}))/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^t)$ は，

$$\frac{\partial B_{11}(\boldsymbol{\eta})}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{k}\frac{\lambda^2 m_i}{2n_i(\lambda m_i-1)^2}\boldsymbol{x}_i,\quad \frac{\partial B_{11}(\boldsymbol{\eta})}{\partial\lambda}=\sum_{i=1}^{k}\frac{1}{n_i}\left\{\frac{1}{2(\lambda m_i-1)^2}-1\right\},$$

および

$$\frac{\partial^2 B_{11}(\boldsymbol{\eta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^t}=-\sum_{i=1}^{k}\frac{\lambda^2 m_i(\lambda m_i+1)}{2n_i(\lambda m_i-1)^3}\boldsymbol{x}_i\boldsymbol{x}_i^t,\quad \frac{\partial^2 B_{11}(\boldsymbol{\eta})}{\partial\boldsymbol{\beta}\partial\lambda}=-\sum_{i=1}^{k}\frac{\lambda m_i}{n_i(\lambda m_i-1)^3}\boldsymbol{x}_i,$$

$$\frac{\partial^2 B_{11}(\boldsymbol{\eta})}{\partial\lambda^2}=-\sum_{i=1}^{k}\frac{m_i}{n_i(\lambda m_i-1)^3}.$$

で与えられる．

3.2 B_2 および B_3 の評価

本小節においては，超パラメータ $\boldsymbol{\eta}$ が Godambe and Thompson (1989) で示唆される推定方程式で推定されるとき，B_2 および B_3 の解析的な漸近近似を与える．$\boldsymbol{g}_i=(g_{1i},g_{2i})^t, g_{1i}=y_i-m_i, g_{2i}=(y_i-m_i)^2-\phi_i Q(m_i)$ とし，また

$$\begin{aligned}\boldsymbol{D}_i^t=&E\left(-\frac{\partial\boldsymbol{g}_i^t}{\partial\boldsymbol{\eta}}\right)\\ =&Q(m_i)\begin{bmatrix}\boldsymbol{x}_i & Q'(m_i)\phi_i\boldsymbol{x}_i\\ 0 & -(1+v_2/n_i)(\lambda-v_2)^{-2}\end{bmatrix},\\ \boldsymbol{\Sigma}_i=&\mathbf{Cov}\,(\boldsymbol{g}_i)=\begin{bmatrix}\mu_{2i} & \mu_{3i}\\ \mu_{3i} & \mu_{4i}-\mu_{2i}^2\end{bmatrix},\end{aligned}$$

とする．ただし $\mu_{ri}=E[(y_i-m_i)^r]$ である．このとき，Ghosh and Maiti (2004) は $\boldsymbol{s}(\boldsymbol{\eta})\equiv\sum_{i=1}^{k}\boldsymbol{D}_i^t\boldsymbol{\Sigma}_i^{-1}\boldsymbol{g}_i$ として $\boldsymbol{s}(\boldsymbol{\eta})=\boldsymbol{0}$ なる推定方程式を導出した．この手法は Ghosh and Maiti (2008), Kubokawa *et al.* (2014) などでも用いられている．Godambe and Thompson (1989) の推定方程式は Wedderburn (1974) により提案された疑似尤度法の拡張であり，$\boldsymbol{s}(\boldsymbol{\eta})$ は拡張された “疑似スコア関数” といえる．この文脈で，“疑似フィッ

シャー情報量" は,

$$E(\boldsymbol{s}\boldsymbol{s}^t) = \sum_{i=1}^{k} \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} E(\boldsymbol{g}_i \boldsymbol{g}_i^t) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{D}_i$$
$$= \sum_{i=1}^{k} \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{D}_i (\equiv \boldsymbol{U}), \tag{3.6}$$

で与えられ, $\hat{\boldsymbol{\eta}}$ の漸近分散は $E(\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}}^t) = \boldsymbol{U}^{-1} + o(k^{-1})$ である.

Ghosh and Maiti (2004) にならい,

$$\boldsymbol{J}_r = \mathbf{Cov}\,(\boldsymbol{s}, \frac{\partial s_r}{\partial \boldsymbol{\eta}}), \quad \boldsymbol{K}_r = E\left(\frac{\partial^2 s_r}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^t}\right), \quad r = 1, \ldots, p+1 \tag{3.7}$$
$$\boldsymbol{a}^t = [\mathrm{tr}\,(\boldsymbol{U}^{-1}\boldsymbol{J}_1), \ldots, \mathrm{tr}\,(\boldsymbol{U}^{-1}\boldsymbol{J}_{p+1})], \quad \boldsymbol{b}^t = [\mathrm{tr}\,(\boldsymbol{U}^{-1}\boldsymbol{K}_1), \ldots, \mathrm{tr}\,(\boldsymbol{U}^{-1}\boldsymbol{K}_{p+1})],$$

と定義する. ただし $\boldsymbol{s} = (s_1, \ldots, s_{p+1})^t$ とし, また $\boldsymbol{J}_r$ および $\boldsymbol{K}_r$ の具体的な表現は本論文では割愛する. さらに, 行列 $\boldsymbol{U}^{-1}$ を以下のように分割する.

$$\boldsymbol{U}^{-1} = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{bmatrix}, \quad \boldsymbol{U}^{-1} = \begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{U}_{12}^t & U_{22} \end{bmatrix},$$

ただし $\boldsymbol{U}_1$ および $\boldsymbol{U}_2$ は $(p, p+1)$ および $(1, p+1)$ 行列であり, $\boldsymbol{U}_{11}, \boldsymbol{U}_{12}$ は $(p,p), (p,1)$ 行列, U_{22} はスカラーである. このとき B_2 の評価は以下の補題で得られる.

補題 3.2 ペナルティ項の一部である (3.1) 式中の B_2 は, 以下のように近似される.

$$B_2 = B_{21}(\boldsymbol{\eta}) + B_{22}(\boldsymbol{\eta}) + o(1).$$

ただし, $B_{21}(\boldsymbol{\eta})$ および $B_{22}(\boldsymbol{\eta})$ は以下で与えられる.

$$B_{21}(\boldsymbol{\eta}) = \sum_{i=1}^{k} \frac{n_i \lambda^2}{(n_i+\lambda)^2} Q(m_i) \boldsymbol{x}_i^t \boldsymbol{U}_1 \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} \begin{bmatrix} \xi_{2i} \\ \xi_{3i} - \phi_i Q(m_i) \xi_{1i} \end{bmatrix}$$
$$- \sum_{i=1}^{k} \frac{n_i^2 \lambda}{(n_i+\lambda)^3} \boldsymbol{U}_2 \boldsymbol{D}_i^t \boldsymbol{\Sigma}_i^{-1} \begin{bmatrix} \xi_{3i} \\ \xi_{4i} - \phi_i Q(m_i) \xi_{2i} \end{bmatrix},$$
$$B_{22}(\boldsymbol{\eta}) = \sum_{i=1}^{k} \frac{n_i \lambda^2}{(n_i+\lambda)^2} Q(m_i) \xi_{1i} \boldsymbol{x}_i^t \boldsymbol{U}_1 (\boldsymbol{a} + 2^{-1}\boldsymbol{b}) - \sum_{i=1}^{k} \frac{n_i^2 \lambda}{(n_i+\lambda)^3} \xi_{2i} \boldsymbol{U}_2 (\boldsymbol{a} + 2^{-1}\boldsymbol{b}),$$

ここで $r = 1, \ldots, 4$ に対して $\xi_{ri} = E[h'(\hat{\mu}_i^B)(y_i - m_i)^r]$ であり, $n_i = O(k^{1/2+\delta}), \delta > 0$ の場合においては剰余項が $O(n_i^{-1})$ のオーダーとなるような近似がなされれば十分で,

Poisson-Gamma モデルでは，

$$\xi_{1i} = -\frac{1}{\lambda m_i - 1} + O(n_i^{-1}), \quad \xi_{2i} = \frac{m_i}{\lambda m_i - 1} + O(n_i^{-1}),$$
$$\xi_{3i} = -\frac{m_i}{\lambda(\lambda m_i - 1)} + O(n_i^{-1}), \quad \xi_{4i} = \frac{3\lambda m_i^2 - 2m_i}{\lambda^2(\lambda m_i - 1)} + O(n_i^{-1}), \tag{3.8}$$

である．

次に B_3 の評価を以下の補題で与える．

補題 3.3 ペナルティ項の一部である（3.1）式中の B_3 は，以下のように近似される．

$$B_3 = B_{31}(\boldsymbol{\eta}) + 2B_{32}(\boldsymbol{\eta}) + o(1).$$

ただし，$B_{31}(\boldsymbol{\eta})$ および $B_{32}(\boldsymbol{\eta})$ は以下で与えられる．

$$B_{31}(\boldsymbol{\eta}) = \sum_{i=1}^{k} \frac{n_i\lambda^2}{(n_i+\lambda)^2} Q(m_i)Q'(m_i)\xi_{1i}\boldsymbol{x}_i^t\boldsymbol{U}_{11}\boldsymbol{x}_i,$$
$$B_{32}(\boldsymbol{\eta}) = \sum_{i=1}^{k} \frac{n_i^2\lambda}{(n_i+\lambda)^3} Q(m_i)\xi_{1i}\boldsymbol{x}_i^t\boldsymbol{U}_{12},$$

ここで $r = 1, \ldots, 4$ に対して $\xi_{ri} = E[h'(\widehat{\mu}_i^B)(y_i - m_i)^r]$ であり，$n_i = O(k^{1/2+\delta}), \delta > 0$ の場合においては剰余項が $O(n_i^{-1})$ のオーダーとなるような近似がなされれば十分で，Poisson-Gamma モデルでは（3.8）式で与えられる．

以上の補題 3.2 および 3.3 より，B_2 および B_3 の推定量として，

$$\widehat{B}_2 = B_{21}(\widehat{\boldsymbol{\eta}}) + B_{22}(\widehat{\boldsymbol{\eta}}), \tag{3.9}$$
$$\widehat{B}_3 = B_{31}(\widehat{\boldsymbol{\eta}}) + 2B_{32}(\widehat{\boldsymbol{\eta}}), \tag{3.10}$$

を提案する．

定理 3.1 ペナルティ項 B の推定量として，(3.5), (3.9) および (3.10) を用いて

$$\widehat{B} = \widehat{B}_1 + \widehat{B}_2 + 2^{-1}\widehat{B}_3,$$

を考える．このとき，$\widehat{B}$ は B の 2 次漸近不偏推定量である．すなわち

$$E(\widehat{B}) = B + o(1),$$

が成り立つ．また，

$$cAIC = -2\sum_{i=1}^{k} n_i(\hat{\theta}_i^{EB} y_i - \psi(\hat{\theta}_i^{EB})) + 2\widehat{B},$$

について，

$$E(cAIC) = cAI + o(1),$$

が成り立つ．

4 数値実験

本節では，提案したcAICと従来のAIC（marginal AIC, mAICと呼ぶ）のパフォーマンスを，正しいモデルを選ぶ割合と，それぞれの規準で選んだ最良なモデルにおける予測誤差を測ることで比較する．2節で説明したPoisson-Gammaモデルにおいて，$k = 150, \lambda = 10, n_1 = \cdots = n_k = 75$ と設定し，超パラメータの推定は推定方程式を用いる．またmAICは-2倍の周辺尤度に，2倍の未知パラメータ数をペナルティとした値とする．説明変数 $\boldsymbol{x}_i$ の選択問題を考えるが，モデル j は $\omega = \{1, \ldots, p_\omega\}$ の部分集合で，そのモデルの含む $\boldsymbol{x}_i$ の非ゼロ要素番号を示す．ここでは $p_\omega = 5, j_\alpha = \{1, \ldots, \alpha\}, \alpha = 1, \ldots, 5$ という状況，すなわちフルモデルが5次元で，入れ子型の候補モデル族を考える．真のモデルは $j_* = \{1, 2, 3\}, \boldsymbol{\beta} = (1, 1, 1, 0, 0)^t$ とする．実験の繰り返しは100回行い，2つの規準がそれぞれのモデルを選んだ回数を数える．また予測誤差は，それぞれの規準で選ばれた最良なモデルにおける経験ベイズ推定量で μ_i を予測したときの2乗誤差，すなわち1回の実験ごとに計算される $\sum_{i=1}^{k}(\widehat{\mu}_i^{EB} - \mu_i)^2$ の値の平均を測る．これらの結果をまとめたものが表1である．

表1に示されているように，真のモデル $\{1, 2, 3\}$ を選んだ割合はcAICが80 %，mAICは77 %と大差はないもののcAICの方がやや良いパフォーマンスを示している．また予測誤差もcAICの方が小さい．cAICは予測誤差の小さいモデルを選ぶための規準であり，従来のAIC (mAIC) よりも小さい予測誤差を達成するモデルが選ばれるのは，理論に整合的である．しかし，cAICとmAICに大きなパフォーマンスの差がないのも事実である．考えられる1つの要因としては，mAICはペナルティ項がパラメータの値に依存していないのに対し，cAICはペナルティ項がパラメータの推定値に依存しており，推定の不安定さが情報量規準の精度にも影響を与える点である．言い換えるとそれはmAICの利点であるとも言える．モデルの設定に合わせて変数選択規準をつくると，複雑な規準と

model	cAIC	mAIC
$\alpha = 1$	0	0
$\alpha = 2$	0	0
$\alpha = 3$	0.8	0.77
$\alpha = 4$	0.1	0.12
$\alpha = 5$	0.1	0.11
予測誤差	1.7471	1.7473

表 1　cAIC と mAIC のおいてそれぞれのモデルを選んだ割合と，最良なモデルにおける予測誤差．真のモデルは $\alpha = 3$.

なってしまう例は多くあるが，本研究で提案した cAIC もその一例といえ，今後の課題である．

謝辞

本研究を進めるにあたって，久保川達也教授（東京大学）には有益な助言をいただき，また日頃の研究に対する叱咤激励を賜った．RIMS 共同研究における研究集会 “Asymptotic Statistics and Its Related Topics” においては，赤平昌文名誉教授（筑波大学），小池健一准教授（同）に大変お世話になった．ここに記して感謝したい．

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*, (B.N. Petrov and Csaki, F, eds.), 267-281, Akademia Kiado, Budapest.

[2] Akaike, H. (1974). A new look at the statistical model identification. System identification and time-series analysis. *IEEE Trans. Autom. Contr.*, **AC-19**, 716-723.

[3] Donohue, M.C., Overholser, R., Xu, R., and Vaida, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika*, **98**, 685-700.

[4] Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*,

74, 269-277.

[5] Ghosh, M. and Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, **91**, 95-112.

[6] Ghosh, M. and Maiti, T. (2008). Empirical Bayes confidence intervals for means of natural exponential family-quadratic variance function distributions with application to small area estimation. *Scandinavian J. Statist.*, **35**, 484-495.

[7] Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with Discussion). *J. Statist. Plan. Infer.*, **22**, 137-152.

[8] Greven, S. and Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, **97**, 773-789.

[9] Kawakubo, Y. and Kubokawa, T. (2014). Modified conditional AIC in linear mixed models. *J. Multivariate Anal.*, to appear.

[10] Kubokawa, T. (2011). Conditional and unconditional methods for selecting variables in linear mixed model. *J. Multivariate Anal.* **102**, 641-660.

[11] Kubokawa, T., Hasukawa, M. and Takahashi, K. (2014). On measuring uncertainty of benchmarked predictors with application to disease risk estimate. *Scandinavian J. Statist.*, to appear.

[12] Liang, H., Wu, H., and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika*, **95**, 773-778.

[13] Morris, C. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.*, **10**, 65-80.

[14] Morris, C. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.*, **11**, 515-529.

[15] Müller, S., Scealy, J.L. and Welsh, A.H. (2013). Model selection in linear mixed models. *Statist. Science*, **28**, 135-167.

[16] Srivastava, M.S. and Kubokawa, T. (2010). Conditional information criteria for selecting variables in linear mixed models. *J. Multivariate Anal.*, **101**, 1970-1980.

[17] Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92**, 351-370.

[18] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

[19] Yu, D. and Yau, K.K.W. (2012). Conditional Akaike information criterion for generalized linear mixed models. *Comput. Statist. Data Anal.*, **56**, 629-644.

[20] Yu, D., Zhang, X. and Yau, K.K.W. (2013). Information based model selection criteria for generalized linear mixed models with unknown variance component parameters. *J. Multivariate Anal.*, **116**, 245-262.