

令和 6 年 5 月 7 日現在

機関番号：22701

研究種目：若手研究

研究期間：2020～2023

課題番号：20K19759

研究課題名(和文) 変数効果分布の誤特定に対して頑健な一般化準線形混合効果モデルの開発

研究課題名(英文) Development of a generalized quasi-linear mixed-effects model robust to misspecification of the random effects distribution

研究代表者

三枝 祐輔 (Saigusa, Yusuke)

横浜市立大学・附属病院・講師

研究者番号：30806469

交付決定額(研究期間全体)：(直接経費) 1,500,000円

研究成果の概要(和文)：本研究の成果として、準線形モデリングを用いて一般化線形混合効果モデルを拡張したモデルを提案した。さらに、提案モデルの推定アルゴリズムと提案モデルに適用可能な情報量規準を与えた。また、変数効果分布の誤特定下において、提案したモデルが予測を改善することを数値シミュレーションおよび実データ解析によって明らかにした。

これらの研究成果について2020年度統計関連学会連合大会、応用統計学会2021年度年会で口頭発表を行った。さらに、研究成果をまとめた論文が査読付き国際学術雑誌のStatistical Methods in Medical Researchに掲載された。

研究成果の学術的意義や社会的意義

学術的意義について、本研究の主目的である準線形モデリングを用いた一般化線形混合効果モデル(GLMM)の拡張は、非線形モデルの新たなクラスの提案と位置づけられる。これまでに準線形モデリングがGLMMに適用された報告はないため、新規の方法論に関する検討である。

社会的意義について、本研究の成果は従来のクラスターデータ解析や時系列データ解析における選択肢を広げ、新たな科学的発見に寄与することが期待される。したがって本研究の成果は統計学のみならず他の科学領域における研究にも資すると期待される。特に医学データ解析を通じて、提案手法が治療効果の予測を改善することを示した。

研究成果の概要(英文)：As a result of this study, a model was proposed that extends the generalized linear mixed-effects model using quasi-linear modeling. Additionally, an estimation algorithm for the proposed model and an applicable information criterion were provided. It was also demonstrated, through numerical simulations and real data analysis, that the proposed model improves predictions under the misspecification of the random effects distribution. These research findings were presented orally at the Japanese Joint Statistical Meeting in 2020 and the Japanese Society of Applied Statistics Annual Meeting in 2021. Furthermore, a paper summarizing the research results was published in the peer-reviewed international academic journal Statistical Methods in Medical Research.

研究分野：統計学

キーワード：一般化線形混合効果モデル 準線形モデリング 変数効果分布の誤特定

## 1. 研究開始当初の背景

一般化線形混合効果モデル (GLMM) は、医学、疫学、生態学などの領域におけるクラスターデータ解析で最も広く適用されている解析手法の一つである。GLMM は、関心のあるデータにおける応答が連続データの場合だけでなく、2 値やカウントなどの離散データの場合にも適用することができる。また、変量効果と呼ばれる確率変数パラメータをもつことによって、クラスター間の異質性を表現できる点が特徴である。

1990 年代以降、GLMM の適用に関する問題が指摘され、いまなお数多くの論文で議論されている。GLMM を用いた解析のほとんどは、各クラスターに特異的な変量効果が正規分布に従うという仮定の下で行われている。その理由は複数あり、推定のための周辺尤度関数が比較的容易に近似できること、複雑なプログラミングを必要としない統計解析ソフトにおいて実装、普及されていること、慣習的に使われてきた背景があることなどである。しかしながら、変量効果の確率分布は潜在的で観測することができないため、変量効果が正規分布に従うという仮定が崩れているかどうか、すなわちモデルの誤特定が起きているかどうかを疑うことは自然な問題である。先行研究より、変量効果分布の誤特定が起きている状況においては、GLMM の回帰係数の最尤推定量は漸近的に無視できないバイアスをもつことが知られている (Heagerty and Kurland, 2001; Agresti et al., 2004; Litière et al., 2008)。さらに、GLMM の平均構造 (あるいは固定効果) に関する統計学的検定の第 1 種の過誤と第 2 種の過誤が影響を受けることや、回帰パラメータの信頼区間の被覆確率が低下することも報告されている (Litière et al., 2007; Marquart and Haynes, 2019)。この問題に関連し、Haung (2009) や Drikvandi et al. (2017) は変量効果分布の誤特定の有無を調べるための診断法を提案した。Garcia and Ma (2016) や Wei et al. (2019) では誤特定に対して頑健な推定量が提案された。特に、平均構造の線形性を保持して変量効果に関する制約を弱めたときのモデルを用いた解析に関する研究として、Zhang et al. (2008) や Fabio et al. (2012) は変量効果が歪んだ分布に従うと仮定したモデルを提案した。Follman and Lambert (1989) や Molenberghs and Verweke (2005) は変量効果が混合正規分布に従うと仮定したモデルを与えた。Zhang and Davidian (2001) や Chen et al. (2002) は、変量効果が Gallant and Nychka (1987) によって与えられた平滑密度をもつという仮定の下でセミパラメトリックモデルによる推定法を提案した。

## 2. 研究の目的

本研究の主たる目的は、準線形モデリングを用いて GLMM を一般化し、変量効果の誤特定に対して頑健なモデルを開発することである。具体的な目標を以下に示す。

**解釈可能性** 新たな統計モデルを用いた解析手法が各科学領域の研究者に広く受け入れられるためには、解析結果の解釈が容易に得られることが肝要である。そのため、新たなモデルを用いた解析結果がどのように解釈可能かを検討し、各領域における実際のクラスターデータ解析にどのように貢献するかを示す。

**変量効果分布の誤特定に対する頑健性** GLMM には、変量効果分布の誤特定下の解析結果を信頼できないという欠点が存在する。この問題に対して、新たなモデルを用いた解析手法が変量効果分布の誤特定に対して頑健性を有するかを検討する。

**推定アルゴリズム** 本研究で検討する新たなモデルは非線形混合効果モデル

(NLMM) に分類される。そのため、NLMM推定のために用意されている既存の計算プログラムが使用可能か検討する。もし使用可能でない場合は、新たに推定アルゴリズムを構築する。

**予測性能の評価指標** 新たなモデルを用いた解析手法の予測性能やデータに対する当てはまりの良さを評価する目的で、GLMMに対する評価指標が適用可能か検討する。もし、既存の指標が適用できない場合は、新たに評価指標を検討する。

### 3. 研究の方法

**解釈可能性** 準線形モデリングの先行研究を参考に新たなモデルから容易な解釈が得られるか検討し、実データ解析を通じて解析結果からどのような知見が得られるか調べる。

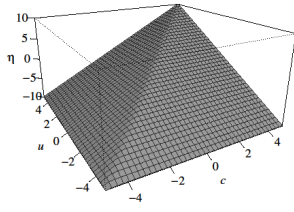
**変量効果分布の誤特定に対する頑健性** Agresti et al. (2004)などを参考に、変量効果が正規分布に比べて歪みや広い裾をもつ分布（指数分布、二項分布、混合分布など）に従うときのシミュレーションを行う。モデルの当てはまりの良さや変量効果が与えられた下での応答変数の条件付き期待値とその推定値の差などを用いて、変量効果分布の誤特定に対する頑健性の強さを評価する。

**推定アルゴリズム** NLMM解析における計算プログラムは、統計ソフトRのnlmeパッケージ (Lindstorm and Bates, 1990)、マルコフ連鎖モンテカルロ法を用いるWinBUGS (Lunn et al., 2000)、自動微分を用いるAD Model Builder (Fournier et al., 2012) など用意されているので、それらが使用可能か調べる。もし使用が難しい場合は、GLMMに対する推定アルゴリズムを与えたBreslow and Clayton (1993) やPinheiro and Bates (1995) を参考に、尤度関数や準尤度関数の近似から新たに推定アルゴリズムを構築する。

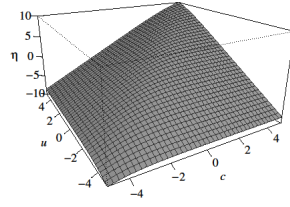
**予測性能の評価指標** 混合効果モデルの予測性能評価に関して、Vaida and Blanchard (2005) は2つの立場を示した：1つ目は母集団における推論に関心があり、クラスター単位の効果には関心がないという立場、2つ目は各クラスターにおける推論に関心があるという立場である。前者には周辺赤池情報量規準 (mAIC)、後者には条件付き赤池情報量規準 (cAIC) を用いることが推奨されている。本研究では、mAICとcAICが新たなモデルの評価に適用可能か調べる。特にcAICの適用においては、Lu et al. (2007) によって与えられたGLMMの複雑さを測る指標が利用可能か検討する。もし適用可能でない場合は、先行研究を参考に、新たに予測性能の評価指標を提案する。

### 4. 研究成果

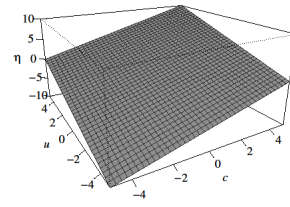
本研究においては、GLMMにおける線形予測子に代わって固定効果項と変量効果項のKolmogorov-Nagumo平均を用いて（準線形予測子として）、準線形混合効果モデル(GQLMM)を提案した。準線形予測子はチューニングパラメータ $\tau$ をもつ。特に $\tau$ が0に近づくときの極限において準線形予測子は線形予測子に一致するため、GQLMMはGLMMを含む。GQLMMがもつ回帰曲面（準線形予測子）は図1のようにチューニングパラメータ $\tau$ の値に応じて変化した。GLMMにおいては固定効果と変量効果はどの標本に対しても等しく作用する構造を仮定する。一方でGQLMMにおいては固定効果と変量効果の影響の度合いが標本ごとに異なる構造を仮定することが図1から確認できた。



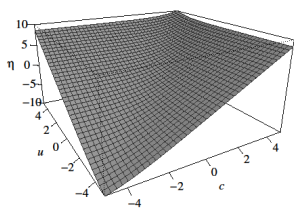
(a)  $\tau = 10$



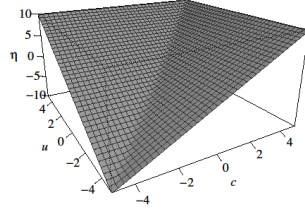
(b)  $\tau = 1$



(c)  $\tau = 0$



(d)  $\tau = -1$



(e)  $\tau = -10$

図 1. いくつかのチューニングパラメータ $\tau$ の値に対する固定効果項と変量効果項 ( $u, c$ ) と準線形予測子の曲面 ( $\eta$ ) の関係

GQLMM の推定アルゴリズムについては、準尤度関数 (Breslow and Clayton, 1993) と制限付き尤度関数 (Harville, 1977) に基づくパラメータ推定法を与えた。さらに、モデルの当てはまりの良さを評価する目的で、GQLMM に適用可能な条件付き赤池情報量規準を与えた。

提案手法の性能評価について、変量効果が誤特定された場合のシミュレーション実験において既存の GLMM よりも頑健にパラメータ推定を行えることを示した。また、Koch et al. (1989) の呼吸器疾患データの解析を行い、GLMM の予測を改善し、既存手法とは異なる解釈が得られることを示した。

詳細は Saigusa et al. (*Stat Meth Med Res* 31(7), 1280-1291, 2022) を参照されたい。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Yusuke Saigusa, Shinto Eguchi, Osamu Komori	4. 巻 31
2. 論文標題 Generalized quasi-linear mixed-effects model	5. 発行年 2022年
3. 雑誌名 Statistical Methods in Medical Research	6. 最初と最後の頁 1280-1291
掲載論文のDOI（デジタルオブジェクト識別子） 10.1177/09622802221085864	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 三枝祐輔, 小森理, 江口真透
2. 発表標題 準線形予測関数をもつ一般化混合効果モデルについて
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 三枝祐輔, 小森理, 江口真透
2. 発表標題 Generalized mixed model with quasi-linear predictor
3. 学会等名 2021年度応用統計学会年会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------