

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 5月11日現在

機関番号：14401

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500275

研究課題名（和文）

シミュレーションに基づく最適な混合効果モデルの選定：理論と応用

研究課題名（英文）

Selecting an optimal mixed effect model based on simulation: theory and applications

研究代表者

坂本 亘（SAKAMOTO WATARU）

大阪大学・大学院基礎工学研究科・准教授

研究者番号：70304029

研究成果の概要（和文）：

混合効果モデルは、個体ごとに繰り返し測定されたデータの解析に有用であり、医学・生物学など様々な分野に応用されている。しかしながら、最適なモデルを選択する方法の研究はこれまで十分には行われていなかった。本研究では、理論上適切でかつ実用に即したモデル選定の規準を提案した。この規準は乱数を用いて近似的に計算することができる。提案した規準は、儉約な（パラメータの少ない）モデルを誤って選択する危険を減らすことを確認した。

研究成果の概要（英文）：

Mixed effect models are useful for analysis of data which are measured repeatedly for each individual, and have been applied to various fields such as medical and biological sciences. However, how to select an optimal mixed effect model has not been fully examined. We proposed a criterion to select such an optimal model, which would be appropriate in theory and practically more useful. The criterion can be computed approximately by using random numbers. We confirmed that the proposed criterion should lessen risk of choosing a parsimonious model wrongly.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	900,000	270,000	1,170,000
2010年度	900,000	270,000	1,170,000
2011年度	700,000	210,000	910,000
年度			
年度			
総計	2,500,000	750,000	3,250,000

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：

モデル選択, 情報量規準, 制限付き最尤推定, モンテカルロ法, パラメータ制約モデル

## 1. 研究開始当初の背景

混合効果モデルは、経時対応データ、メタ・アナリシス（多水準モデル）など、医学・生物学をはじめとする様々な分野で適用され、その方法論が盛んに研究されている。例えば、線形混合効果モデル

$$y = x^T \beta + z^T b + \varepsilon$$

( $y$ : 応答変数;  $x, z$ : 説明変数 (ベクトル);  $\varepsilon$ : 誤差項) では、応答が固定効果 ( $x^T \beta$ ) とランダム効果 ( $z^T b$ ) の両方で説明され、パラメータ  $\beta$  と確率変数  $b$  についての推測を行う。固定効果に関する推測はパラメータ  $\beta$  の推定に帰着され、通例は最尤推定法が用いられる。これに対して、ランダム効果に関する推測は確率変数  $b$  の分散に含まれるパラメータの推定に帰着され、主に制限付き最尤推定法や Bayes 流接近法が用いられる。

複数の固定効果やランダム効果が複雑に入り混じった混合効果モデルでは、それらの多数の候補となる組み合わせの中から最適なモデルを選定することが、データ解析の過程の中で自然に要請される。しかしながら、既存の尤度比検定や情報量規準などによるモデル選定の方法論は、混合効果モデルにそのまま適用することができない。その理由として、ランダム効果の分散パラメータの推定問題では、モデルのパラメータの真値がパラメータ空間の内点にあるという正則条件の一つを満たさないために、漸近正規性などの大標本特性が保証されない、という点が挙げられる。

例えば、上記の線形混合効果モデルで、 $b$  は平均 0、分散行列  $\sigma_b^2 \Sigma$  ( $\Sigma$  は既知で正定値の相関行列) の正規分布に従うとすると、ランダム効果が存在するか否かについての以下の仮説検定を考える：

(帰無仮説)

$$H_0: \sigma_b^2 = 0 \quad (\text{すなわち } b \equiv 0)$$

(対立仮説)

$$H_A: \sigma_b^2 > 0 \quad (\text{すなわち } b \text{ は確率変数})$$

この状況で、対数尤度比統計量の帰無仮説下の漸近分布が理論的に得られるのは実験研究での特定のデザインのみに限られており、一般には単一のカイ二乗分布による漸近近似が成立しない。そこで、帰無分布をシミュレーションに基づいて数値的に生成するのが一般的である。先行研究では、ランダム

効果の分散パラメータが単一の場合に限り、帰無分布の効率的な生成方法が開発されていた。

混合効果モデルの選定に関する研究は、上で述べたような検定に基づく方法か、あるいは既存の情報量規準をそのまま類推により適用しているものが殆どであったが、その妥当性については十分に研究されていないという状況であった。

## 2. 研究の目的

混合効果モデルの選定問題では、固定効果、ランダム効果の様々な候補の組み合わせの中から最適なものを選択するための包括的な基準ないし方法論を整備しその評価を行うことが重要である。そのような方法は大きく以下の三つに分けることができる。

- 検定に基づく方法 (従来の方法)
- 情報量規準を用いる方法
- Bayes 流接近法

本研究では、検定に基づく方法の問題点を探るとともに、情報量規準を用いる方法の検討に重点をおいた。

情報量規準は、Kullback-Leibler 情報量の最小化 (平均対数尤度の最大化) を狙い、平均対数尤度を最大対数尤度によって推定するときに生じる偏りを考慮したものである。AIC (赤池情報量規準) をはじめとする従来の情報量規準では、その偏りの漸近的評価を行うことにより、補正項が導出されてきた。しかしながら、混合効果モデルにおけるランダム効果項の選択問題では、1. で述べたのと同様に、情報量規準の漸近的性質の論拠となっている、正則条件の一つが満たされない場合がある。そこで、シミュレーションに基づいて偏りを評価する方法を新たに構築することを目指した。

### 3. 研究の方法

#### (1) 尤度比検定に基づく方法の検討

尤度比検定統計量の帰無分布をシミュレーションに基づいて数値的に生成することにより、 $p$  値などの裾確率を求めることができる。そこで、経時対応データの事例に対して、ベキ変換スプライン混合効果モデルを提案し、シミュレーションに基づく尤度比検定を適用することを検討した。

#### (2) 平均対数尤度を最大対数尤度によって推定するとき生じる偏りの評価

文献事例のデータに線形混合効果モデルをあてはめ、推定されたモデルを用いて、その偏りをシミュレーションに基づいて評価した。

#### (3) パラメータ制約付きモデルに対する偏り補正項の導出

モデルのパラメータの真値がパラメータ制約付き空間の端にある場合に、偏り補正項の近似式を導出することを検討した。さらに、実際のデータ解析で、シミュレーションに基づいて補正項を効率的に計算するための方法を検討した。

### 4. 研究成果

#### (1) 尤度比検定に基づく方法の検討

ベキ変換スプライン混合効果モデルに対して、階層的に仮説を立て、尤度比検定によりモデルを選択する方法を例示した。しかしながら、尤度比検定では、階層構造の段階ごとに帰無分布をシミュレートすることが必要であるなどの問題があることが分かった。

#### (2) 平均対数尤度を最大対数尤度によって推定するとき生じる偏りの評価

真のモデルよりも過剰なパラメータをもつモデルをあてはめようとした場合に、その偏りは既存の情報量規準の補正項よりも小さくなることを、シミュレーションで確認した。これは前述のとおり、混合効果モデル正則条件が成り立たないことに起因すると考えられる。

#### (3) パラメータ制約付きモデルに対する偏り補正項の導出

平均対数尤度および最大対数尤度の2次近似に基づいて、偏り補正項の近似式を得た。パラメータの推定量がパラメータ空間の内

部にある確率が漸近的に1に収束するならば、この偏り補正項は既存の情報量規準に対する補正項に一致する。

次に、この補正項をモンテカルロ近似により得ることを提案した。提案法は、偏りのブートストラップ推定の着想に基づいているが、実際にブートストラップを実施する必要はなく、打ち切りのある多変量正規分布に従う乱数を生成すればよい。したがって、既存のブートストラップ情報量規準よりもはるかに計算時間が短縮される。

さらに、提案した情報量規準の性能を、シミュレーション研究により、既存の情報量規準と比較した。提案した規準は、既存のものよりも偏りが小さくなり、また誤って儉約なモデルを選択する危険がより小さくなることを確認した。

なお、これまでの研究では統計解析環境 R を用いてきたが、さらなる研究のためには計算速度の飛躍的向上が不可欠であると考えており、現在、計算プログラムの Fortran への実装を試みているところである。反復回数を増やしてシミュレーションの精度を上げるとともに、前提条件の組み合わせを増やして包括的なシミュレーションを行うことを計画している。

また、スプライン回帰モデルは一種の線形混合効果モデルによって表現されることから、スプライン回帰モデルに対するモデル選択の方法についても検討したい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 10 件)

- (1) Sakamoto, W.: Selecting variance structure in mixed effect models by information criteria based on Monte Carlo approximations.  
Joint Meeting of the 2011 Taipei International Statistical Symposium and 7th Conference of the Asian Regional Section of the IASC, Academia Sinica, Taipei, Taiwan, 2011 年 12 月 18 日
- (2) 高瀬貴夫・坂本亘・後藤昌司: 経時対応 2 値データ解析と妥当性診断.  
日本行動計量学会第 39 回大会, 岡山理科大学, 2011 年 9 月 13 日
- (3) 尾崎寿昭・藤澤正樹・坂本亘・後藤昌司: 形状不変モデルに基づく曲線分解.  
日本行動計量学会第 39 回大会, 岡山理科大学, 2011 年 9 月 14 日
- (4) 山口祐介・坂本亘・白旗慎吾・後藤昌司: メタ・アナリシスにおける交互作用効果の評価: シミュレーションに基づく方法.  
日本計算機統計学会第 25 回大会, 函館市亀田福祉センター, 2011 年 5 月 7 日
- (5) 坂本亘: シミュレーションの前と後: 統計的観点.  
医学統計研究会・特定主題シンポジウム 2010 「臨床評価過程におけるシミュレーションとその実際」, ファイザー株式会社 (東京都), 2010 年 10 月 30 日
- (6) Sakamoto, W.: Selecting an optimal mixed effect model based on information criteria.  
COMPSTAT 2010, Conservatoire National des Arts et Métiers, Paris, France, 2010 年 8 月 24 日
- (7) 山口祐介・坂本亘・白旗慎吾・後藤昌司: 要約統計量からのデータの復元とその評価.  
医学統計学および疫学の方法と実際に関するシンポジウム, 山梨大学, 2010 年 7 月 2 日

- (8) 坂本亘: 混合効果のあてはめに適切な応答変換の選択.  
医学統計研究会秋季セミナーかごしま 2009, 宝山ホール (鹿児島県), 2009 年 9 月 12 日
- (9) Sakamoto, W.: Selecting an appropriate transformation of responses for fitting a linear or additive mixed model.  
The 57th Session of the International Statistical Institute (ISI 2009), International Convention Centre, Durban, South Africa, 2009 年 8 月 20 日
- (10) 坂本亘: 混合効果モデルのあてはめに適切な応答変換の選択.  
医学統計研究会夏季セミナー 2009, 大分大学, 2009 年 8 月 7 日

6. 研究組織

(1) 研究代表者

坂本 亘 (SAKAMOTO WATARU)  
大阪大学・大学院基礎工学研究科・准教授  
研究者番号: 70304029

(2) 研究分担者

( )

研究者番号:

(3) 連携研究者

( )

研究者番号: