論文審査の結果の要旨

論文題目：Variable selection problem in mixed effects models with application to small area estimation
題目和訳：　混合効果モデルにおける変数選択問題と小地域推定への応用
氏名：　川久保　友超

本論文は，線形混合モデルに代表される混合効果モデルにおける変数選択に関して次のような問題に取り組んでいる．（A）線形混合モデルにおいて条件付赤池情報量規準(cAIC)が，真のモデルを含んでいるという仮定の下で導出されてきたのに対して，真のモデルを含まない場合への拡張，（B）予測モデルの共変量の値が観測モデルの共変量の値と異なる場合を共変量シフトと呼んでいるが，（A）で導出した規準の共変量シフトの場合への拡張，（C）地域別死亡率などの離散データ解析に用いられるモデルとして指数型分布族の混合効果モデルが考えられるが，このモデルにおける条件付 AIC の導出，（D）線形回帰モデル及び線形混合モデルにおける新たなベイズ的情報量規準の導出．こうした問題に関して，新たな情報量規準を理論的に導くとともにシミュレーション実験及び実データの解析を通して提案手法の良さを数値的に調べたものである．具体的には以下のような章立てで構成されている．

第 1 章　序
第 2 章　混合効果モデルと条件付き AIC
第 3 章　線形混合モデルにおける修正条件付き AIC
第 4 章　共変量シフトのもとでの条件付き AIC と小地域推定への応用
第 5 章　自然指数型分布族の混合効果モデルにおける条件付き AIC
第 6 章　ベイズ周辺尤度を用いた AIC に類似の規準
第 7 章　線形混合モデルにおける条件付き AIC に類似の規準

線形混合モデルは，共変量に基づいた固定効果の項と確率変動する変量効果及び観測誤差の項から構成されるモデルで，Henderson により 1950 年代に提唱され家畜育種の分野で利用されていた．近年，官庁統計・政府統計において，特に小地域の統計分析を行う際，地域の差異（地域効果）を変量とするモデルを利用すると推定精度を高めることができるため，線形混合モデルの考え方が広く使われるようになった．線形混合モデルにおける共変量である説明変数の選択問題については，様々なアプローチが提案されてきたが，Vaida and Blanchard (2005) は条件付赤池情報量（cAI）という概念を導入した．これは変量効果を所与とした条件付尤度にもとづいた期待カルバック・ライブラー・ダイバージェンスに関連した，モデルのリスクとも言うべきものである．cAI とそこから導かれる条件付赤池情報量規準（cAIC）は，クラスターごとの推測に関心がある場合は，変量効果を積分消去した周辺尤度にもとづいた従来の赤池情報量規準(AIC)よりも適切であると考えられ，小地域推定のような変量効果の予測を含んだ問題において有用であることが示唆されている．第 1 章では，変量効果モデルの変数選択に関する先行研究を cAIC を中心にレビューし，第 2 章では，変量効果モデルの一般論とその最も基本的なモデルのクラスとして線形混合モデルを説明し，変数選択の問題設定と，Vaida and Blanchard (2005) の提案した c AIC の説明を与えている．

第 3 章から続く 5 つの章で，オリジナルな研究成果がまとめられている．第 3 章では，線形混合モデルにおける Vaida and Blanchard (2005) の cAIC の問題点を指摘した上で，cAIC を修正した規準として修正条件付き赤池情報量規準（McAIC）を提案している．Vaida and Blanchard (2005) の問題点は，候補モデルが真のモデルを含んでいるという仮

定のもとで，cAI の不偏推定量を求めている点である．その結果，真のモデルを含んでいないモデルにおいて, cAIC は cAI の推定量として大きなバイアスを生じてしまう. そこで, 真のモデルを含んでいる場合も含んでいない場合も含めて cAI の漸近不偏推定量を求めることによって修正条件付 AIC(McAIC)という新たな情報量規準を導出している．また，赤池型の情報量規準にもとづいたモデル平均化推定量に McAIC を用いると，cAIC にもとづいたモデル平均化推定量よりも予測精度が良くなることを数値実験で示している．

第 4 章では，推定に用いるモデルの共変量の値と予測に用いるモデルの共変量の値が異なる状況を共変量シフトと呼んでいるが，こうした状況における線形混合モデルの変数選択問題を考えている．共変量シフトの下での cAI を定義した上で，候補モデルが真のモデルを含んでいるという仮定のもとで cAI の不偏推定量として共変量がシフトした場合の cAIC (CScAIC) を導出している．ここで真のモデルを含んでいる場合のみを扱うこうした規準については， CScAIC の尤度部分が cAI のあまり良い推定量とは言えず, 真のモデルを含んでいるという仮定が崩れると，非常に大きなバイアスを生んでしまう欠点があり，共変量シフトの状況下ではいっそう問題となることを指摘している．そこで，真のモデルを含んでいる場合も含んでいない場合も含め cAI の漸近不偏推定量となる新たな変数選択規準を導出している．さらに，共変量シフトの状況は小地域推定における有限母集団の平均の推定問題で重要となることに言及し，数値実験で提案手法の有用性を示している．

第 5 章では，自然指数型分布族にもとづいた混合効果モデルにおける cAIC を導出している．自然指数型分布族にもとづいた混合効果モデルのクラスは，死亡数などの計数データや二値データのモデリングに有用なポアソン・ガンマモデル，二項ベータモデルといった非線形混合モデルを含んでいる. これらのモデルは解析的に周辺尤度を導出できるため, 変数選択規準として周辺尤度にもとづいた従来の周辺分布に基づいた mAIC が当然利用可能である．しかしながら，クラスターごとの推測に関心がある場合は，mAIC は適切でない．そこで，非線形混合モデルにおける cAI を定義し，その漸近不偏推定量として cAIC を導出している．漸近不偏推定量の構成法として，3 通りの方法を提案している．1 つは，完全に解析的な方法で，これは超母数が Godambe and Thompson (1989) から示唆される推定方程式で推定されたもとで推定量の確率展開を行うことでなされる．さらにこの手法は非線形なリンク関数を多項式近似するため，クラスター内のサンプルサイズがクラスター数 m に対して一定のスピードで発散するという仮定（仮定 A）も必要となる．仮定はやや強いが，計算負荷のあまりかからない手法といえる．2 つ目の手法は，数値微分，数値積分を用いた手法であり, 仮定 A は必要でない. これは超母数の推定量の確率展開を行い，漸近バイアス，漸近分散を解析的に求めるが，バイアス補正にモンテカルロ近似にもとづいた数値積分および数値微分を用いる手法である．3 つ目の手法は，パラメトリック・ブートストラップを用いる手法である．仮定 A も超母数の推定法の制約もないが，ブートストラップ法でモンテカルロ積分やバイアス補正を行い，ブートストラップ繰り返しのそれぞれのステップで超母数の推定が必要なため，計算負荷が他の 2 つの手法に比べ大きい．

第 6 章では，ベイズ周辺尤度にもとづいた予測密度のリスクを，頻度論の立場から測った情報量規準を提案している．具体的には，線形回帰モデルの変数選択規準を，回帰係数に事前分布を入れることで導出している．提案手法には 3 つの利点がある．1 つは，この手法は頻度論とベイジアンの折衷であり，ベイズモデルのリスクを頻度論の立場から測っているため，事前分布の特定化の誤りの影響を受けにくい．2 つ目は，規準の構築に無情報事前分布を用いることができる点である．回帰係数に一様事前分布を仮定すると，得られる規準は Shi and Tsai (2002) の残差情報量規準(RIC)と一致するものが得られる．3 つ目は, 提案手法が変数選択の一致性を持つ点である. 誤差項ベクトルの共分散の構造を様々なケースで数値実験を行い，提案手法の有用性を示している．

第 7 章では，線形混合モデルの変数選択問題において，変量効果を所与とした条件付尤度の期待カルバック・ライブラー・ダイバージェンスにもとづいた予測密度のリスクを，様々な予測密度で比較し，導出される情報量規準の性質を議論している．予測密度としてプラグイン予測密度を考えると，そこから得られる情報量規準は Vaida and Blanchard (2005) の cAIC に一致する．しかし，予測密度をプラグイン予測密度に制約する必然性はなく，ある観点からプラグイン予測密度より優れた予測密度は存在する．そこで，以下の 2 つの予測密度を考えている．1 つ目は，ベイズ予測密度関数であり，これは期待カルバック・ライブラー・ダイバージェンスを最小にするという意味で，最も優れた予測密度であることが知られている．ここから導かれる規準は，Akaike (1980) の予測尤度や Kitagawa (1997) の予測情報量規準(PIC)と呼ばれるものである．ここでは，線形混合モデルにおける PIC として，回帰係数を未知パラメータと考えたもの PIC1 と，回帰係数に事前分布を仮定したもの PIC2 の 2 通りを考え，比較している．2 つ目は，ベイズ周辺尤度にもとづいた予測密度であり，これは第 6 章で導出した規準の，変量効果の予測を目的とした場合の変形と見ることができる．数値実験を通してそれぞれの性質を調べている．

論文の評価

線形回帰モデルの変数選択は古くから研究されてきた問題で，自由度調整済み決定係数，マローズの Cp 規準，赤池情報量規準(AIC)が代表的であり，近年では様々なベイズ的変数選択法も提案されてきた．一方．変量効果を組み込んだ線形混合モデルの有用性が官庁統計の小地域推定などで認識されてくると，線形混合モデルに共変量の選択問題が議論されようになる．変量効果を積分消去すると，周辺分布は共分散行列に分散成分を組み込んだ線形回帰モデルに帰着されるので，そうした周辺尤度から AIC を導出して変数選択を行うことができる．しかし，小地域推定では変量効果の予測量を用いて推定精度の改善を図っているため，その点を考慮した変数選択規準が望まれる．そこで導入されたものが Vaida and Blanchard(2005)の条件付 AIC(cAIC)という規準である．その後 cAIC の様々な拡張などがいくつかの論文の中でなされてきたが，一つの大きな弱点があった．それは，いずれの論文も真のモデルを含んでいるという大前提のもとで議論されてきたことにある．これに対して，本論文では，真のモデルを含まない場合も含めて条件付 AIC を提案している．これを修正 cAIC(McAIC)と呼んでいるが，条件付情報量(cAI)の高次漸近不偏推定量を，真のモデルを含むか含まないかに関係なく導出している．このことにより，真のモデルを含まない場合，従来の cAIC が大きなバイアスを生じていたのに対して提案した McAIC は漸近的にバイアスのない規準になっている．そして，この性質はモデル平均化推定量を構成する際に重要であることを指摘して，McAIC に基づいたモデル平均化推定量の予測誤差が小さくなることを数値的に示した．これは cAIC に対する重要な貢献を示すものである．

官庁統計で現れる小地域推定は有限母集団の枠組みで議論されることがある．この場合，観測モデルの共変量の値と予測分布の共変量の値との間にずれが生じてしまう．これを共変量シフトと呼び，本論文では，共変量シフトが存在するときの cAIC の導出をおこなっている．しかも真のモデルを含む場合だけでなく真のモデルを含まない場合も含めて，cAI の高次漸近不偏推定量を求めることによって cAIC の導出を行っている．共変量シフトを考慮した提案手法が優れていることを数値的に示している．線形混合モデルの枠組みで共変量シフトを導入した議論は独創的であり，その導出を解析的に行っている点や，真のモデルを含まない場合への対応もなされている点，そして小地域推定で扱う有限母集団モデルに応用例が存在している点が評価される．

本論文では，連続分布だけでなく離散分布の cAIC の導出も扱っている．特に 2 項・ベータ分布やポアソン・ガンマ分布を含む指数型分布族に基づいた変量効果モデルを扱い，共変量の変数選択規準として cAIC の導出を行っている．こうしたモデルは地域別死亡率

の推定など計数データに基づいた小地域推定に利用可能である．しかし，このモデルでcAIC を解析的に求めることは極めて困難な労作業を伴うが，本論文では辛抱強く漸近展開を行って cAIC を求めている．また数値微分やブートストラップに基づいた簡便な方法での条件付 AIC の導出も行っている．こうした内容は極めて独創的であり，離散データの地域統計の解析に有用である．

情報量規準は一般に予測リスク関数の一部に対する不偏推定量もしくは高次漸近不偏推定量を与えていると解釈することができる．AIC は当然こうした考え方から導かれる規準であるが，本論文ではベイズ情報量規準(BIC)もまたこの考え方から導出できることを指摘することにより，この方法が一つの統一的な枠組みを与えていることを明確にしている．そして，回帰係数ベクトルに一様分布を仮定すると Shi and Tsai (2002)の残差情報量規準(RIC)が得られ，未知の超母数をもつ正規分布を仮定すると Bozdogan (1987)の規準を含んだもので，BIC と AIC とを組み合わせた規準が得られることを示している．このような規準の一致性を証明するとともに数値的にも一致性を確認している．以前に提案されてきた情報量規準を統一的な枠組みで捉え直した点が興味深いとともに，新たな経験ベイズ的情報量規準を導出してその一致性を証明し，数値的な比較を通して良さを明らかにしている点が評価できる．さらに，最後に，同様な枠組みを線形混合モデルに当てはめ，ベイズ的予測情報量規準(PIC)やベイズ的周辺尤度に基づいた情報量規準を導出し提案している．この枠組みから当然 cAIC も導かれるので，統一的に議論できる枠組みを提供している点や，そこから新たなベイズ的条件付情報量規準を導出している点が評価できる．その一致性については示されていないが，そのパフォーマンスが優れていることが数値的に確かめられており，これまで提案されてきた様々な変数選択規準に比べ説得力のあるものになっている．

以上，説明してきたように，本論文は，線形混合モデルに代表される混合効果モデルにおける変数選択のための情報量規準に関して，川久保友超氏自身の独創的で優れた研究成果をまとめたものである．その中で理論的に導出された変数選択規準は小地域推定などの応用分野において利用価値の高いものであり，この分野における貴重な貢献であると高く評価できる．従って，論文審査委員会は全員一致で，川久保友超氏が博士（経済学）の学位を授与されるにふさわしいという結論に達した。

審査委員
久保川達也
國友直人
矢島美寛
大森裕浩
下津克己
加藤賢悟