

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 4 日現在

機関番号：12601

研究種目：基盤研究(C)

研究期間：2009～2013

課題番号：21540114

研究課題名(和文) 多変量推測理論の新たな展開とその応用に関する研究

研究課題名(英文) New developments of theories in multivariate statistical inference and their applications

研究代表者

久保川 達也 (Kubokawa, Tatsuya)

東京大学・経済学研究科(研究院)・教授

研究者番号：20195499

交付決定額(研究期間全体)：(直接経費) 3,500,000円、(間接経費) 1,050,000円

研究成果の概要(和文)：本研究課題では、多変量解析モデルの推定・検定・予測・変数選択など新たな推測手法の開発とそれに伴う理論展開を行った。特に、従来の手法に欠点があったり利用可能でない場合においてそれを解決する手法の開発を目指した。中でも、(1) 小地域推定の平均2乗誤差及び信頼区間の高次漸近補正、(2) 連続及び離散混合モデルにおける小地域推定のベンチマーク問題の理論展開とその応用、(3) 線形混合モデルの検定についてのパートレット補正と変数選択規準の開発、(4) 線形判別分析に関する高次元手法の理論展開、(5) 多次元母数の推定における最適性理論の新たな展開に取り組み、有用で有効な推測手法の開発を行った。

研究成果の概要(英文)：In this research project, new procedures were derived in estimation, hypothesis testing, prediction and variable selection in multivariate statistical models as well as theories for justifying the new procedures were developed. In particular, we treated problems, models or situations where conventional methods had drawbacks for use or were not available, and then we tried to derive and suggest procedures which could resolve the problems. Of these, the following multivariate statistical issues were addressed and new theories were developed with applications: (1) higher order asymptotic corrections in mean squared errors and confidence intervals in small area estimation, (2) benchmark problems in small area estimation under continuous and discrete mixed models, (3) Bartlett corrections and methods for variable selection in linear mixed models, (4) high dimensional procedures in linear discrimination and (5) optimality in estimation of multi-dimensional parameters.

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：多変量解析 線形混合モデル 小地域推定 統計的決定論 パートレット補正 高次元解析 漸近不偏性 ベイズ推測

1. 研究開始当初の背景

多変量解析手法の有用性は広く認識され、現実のデータ解析の場面で役立ってきた。しかし、ゲノムデータやファイナンスデータなど近年注目されている高次元データは、次元がデータ数よりはるかに大きいので、従来の推測手法では解析することができない。こうした状況から、近年では高次元の多変量解析の理論研究が世界的に注目され、この未知の分野への研究が挑戦的に行われている。

多変量推測理論の中で注目されてきた理論研究にスタイン問題がある。これは、正規分布の平均ベクトルの推定において、1次元及び2次元のときには通常の標本平均が許容的になるのに対して、3次元以上のときには非許容的となり、縮小推定量によって改良されるという驚くべき現象が成り立つことをいう。この興味深い現象とそれに関連する問題の理論研究が活発に研究されてきた。縮小推定の有用性は理論研究に留まらず、応用分野においても経験ベイズ推定として役立っている。官庁統計で注目される小地域推定において利用される経験最良不偏予測量(EBLUP)も縮小推定の変形である。また地域別死亡率や疾病地図の作成に利用される経験ベイズ推定も縮小推定の形をしている。このように、縮小推定に関する研究も、理論的な側面や応用的側面から新たな展開がなされている。特に官庁統計などの分野では、新たな解析の必要性が新たな手法の開発とその理論展開が要請される状況を生み出している。例えば、ベンチマーク問題は小地域推定の新たな研究課題を提供している。

本研究課題では、多変量解析の新たな理論展開と、従来の推測手法を改良する新たな解析方法の研究開発及び現実のデータ解析への適用を目標に取り組んだ。

2. 研究の目的

本研究課題では、多変量モデルにおいて従来の推測手法を改良する新たな統計手法を開発し、その新たな手法が従来の手法に比べて理論的に優越していることや有効性・最適性に関する理論の展開を行い、シミュレーション実験による数値的な比較及び現実のデータ解析での有用性を示すことを目的とした。特に、従来の統計手法をそのまま用いたのでは問題があるような統計モデルにおいて、従来の手法の欠陥を解決していかも従来の手法を理論上優越する新たな手法の開発やそのような導出の理論構築を行った。

(1) 線形混合モデルを利用した小地域推定について

線形混合モデルの特徴は、観測値を共変量を用いて回帰するときに個体や地域の違いを変量として組み入れ、それらの背後に共通な

確率分布を想定して個体や地域の差異を推定している点である。全体の特性値だけでなく個体や地域ごとの特性値への関心が高まるにつれ、個々の差異を変量として捉えた線形混合モデルについての研究が盛んになり、このモデルの研究が始まった家畜育種学の分野はもとより医学・生物学分野から経済・教育など社会科学の分野、特に官庁統計分野での小地域推定において利用されている。小地域推定の問題に注目すると、問題点は、人口が疎な地域の平均を推定する際にはデータの不足から推定誤差が大きくなってしまふ点にあり、周辺地域のデータをプールすることによって推定精度の改善が図られる。そのために利用されるのが線形混合モデルであり、そこから導かれる経験最良線形不偏予測量(EBLUP)が推定精度を改善した推定法として利用される。推定精度を改善するために用いられた EBLUP が実際のどの程度の推定誤差があるのかを見積もることは現実のデータを応用する場面で重要である。そのためには平均2乗誤差(MSE)の推定を与えることと信頼区間を構成することが考えられる。

従来の研究では誤差分散が既知の簡単なモデルや不偏推定量に基づいた簡単な母数推定の方法が扱われてきたが、実際に応用するには誤差分散が未知のモデルを扱う必要があり、また最近是最尤推定、制限最尤推定、ベイズ推定などより一般的な推定手法が用いられる。変量効果が複数組み入れられたモデルなど、より一般的な線形混合モデルの枠組みで EBLUP の MSE の2次漸近不偏推定量を導出し、また信頼区間の構成を行う。

次に、線形混合モデルにおいて共変量を用いて回帰する項に対して検定及び変数選択の問題を考える。回帰係数の仮説検定に関しては、通常のF検定が有意水準をみたまないという欠点があることが指摘されている。この問題に対して、ここでは次の2つのアプローチでその解決をはかる。1つはバートレット型補正を行うことによって2次漸近的に有意水準をみたま検定手法の導出を行うことであり、もう1つは赤池情報量規準(AIC)や条件付きAICを導出することである。

さらに、近年では、ベンチマーク問題が話題になってきた。これは、予算を各市町村に配分する際に、各地域の EBLUP に基づいて配分してしまうと、その合計が標本平均と一致しないという問題で、EBLUP に基づいて各地域への配分額をどのように決定するのか、そのための予測法の開発を行う。

ベンチマーク問題は、正規線形混合モデルに限らず、離散分布に拡張した一般化線形混合モデルにおいても問題になる。このモデルは特定疾患による死亡数などの離散データを解析するのに広く利用されているが、EBLUP

もしくは経験ベイズ推定値が縮小し過ぎてしまう傾向性があることが指摘されてきた。この性質を改善するために、平均だけでなく分散についてもベンチマーク制約をいれた推定法が望まれる。またそのときの推定誤差の評価を行う。

(2) 高次元多変量解析の展開について

多変量解析の現代的流れの一つは、ゲノムデータの解析に代表されるような高次元データを解析するための新たな理論作りである。データの次元が観測数を大きく超えるため古典的な多変量解析を利用することができない。ここでは、高次元線形回帰モデルにおいて変数選択のための情報量規準の導出と高次元での線形判別分析法の開発とその性質の理論研究を行う。

(3) 多次元母数の推定における最適性理論の新たな展開について

統計的決定論におけるミニマックス性や許容性に関するいくつかの問題に取り組む。この分野で有名な問題が平均ベクトルのスタイン問題で、これは、平均ベクトルの標準的な推定量である標本平均がミニマックスであるものの3次元以上では非許容的になってしまうという現象をいう。実際、縮小型推定量を用いて改良することができる。このように通常の推定量が非許容的になってしまうことは、母数空間が制約されている場合にも起こりうる。通常の推定量がミニマックスになるか否かを含め、こうした問題の理論研究を行う。

3. 研究の方法

(1) 線形混合モデルを利用した小地域推定について

一般の正規線形混合モデルを扱い、分散成分などの母数推定に関しては最尤推定量、制限最尤推定量、モーメント推定量などを含んだ一般的な推定量を扱い、その確率展開を仮定した上で、EBLUPのMSEの2次及び3次の漸近展開とその不偏推定量を導出する。同じ設定のもとで、EBLUPに基づいた信頼区間について、2次及び3次の信頼性をもった信頼区間の構成を行う。また、より具体的なモデルとして、枝分かれ誤差分散回帰モデルを有限母集団の枠組みで考察し、信頼区間の漸近補正を導出する。

線形混合モデルの回帰係数の検定問題については、通常のF検定が有意水準をみたさないという欠点があることが指摘されている。ここでは、棄却域の閾値を調整することによって2次漸近近似の意味で有意水準をみたす検定統計量を導出することを試みる。線形混合モデルの検定問題は、共分散構造が埋め込

まれた一般的な共分散行列をもつ多変量線形モデルの検定問題として捉えることができるので、そうした一般的な問題設定での尤度比検定のバートレット補正の方法を参考にして研究を進める。

線形混合モデルの説明変数の選択問題については、赤池情報量規準(AIC)よりも条件付きAICの方が変量効果や地域効果の予測を考慮に入れた方法として望ましいことが指摘されている。しかし、従来は共分散行列にかなり特殊な仮定を入れた上で、条件付きAICを求めているため、分散成分の選択などを含めることができない。そこで、かなり一般的な線形混合モデルにおいて、分散成分の推定量に確率展開を仮定して、バイアス項の2次漸近不偏推定量の導出を行い、一般的な条件付きAICを与える。

ベンチマーク問題についても、一般的な正規線形混合モデルにおいて、平均と分散の両方について一般的なベンチマーク制約を入れて、制約付き経験ベイズ推定を求めることによって、ベンチマーク推定量の導出を行う。また、その推定量の誤差評価を行う。

離散分布のベンチマーク問題については、2次分散関数をもつ自然指数型分布族を仮定して、ベンチマーク制約のもとでの制約付き経験ベイズ推定量の導出と誤差評価を行う。この分布族は、これまでの正規分布族や、ポアソン-ガンマ分布、2項-ベータ分布を含んでおり、死亡率や疾病地区の作成に用いられる分布族を含んでいる。

(2) 高次元多変量解析の展開について

高次元の多変量解析の問題として、2群の線形判別関数の問題を扱う。この問題では、共分散行列の逆行列を推定する必要があるが、高次元の場合、標本分散共分散行列の逆行列は不安定になってしまう。そこで、リッジ型の推定量で置き換えることによって安定した推定値が得られ、その結果、安定した判別分析が可能になると期待できる。その点を理論的に示すには、誤判別確率の漸近展開を行う必要がある。この漸近展開を用いて、誤判別確率の2次漸近不偏推定量の導出も可能になる。

(3) 多次元母数の推定における最適性理論の新たな展開について

分散が未知の正規分布における平均ベクトルの推定に関するスタイン問題については、一般化ベイズ推定量の広いクラスが提案され、そのミニマックス性が示されてきた。しかし従来の方法は微分不等式に基づいた条件を用いるためミニマックス性の条件が狭いものになってしまう。そこで、より一般的な積分不等式に基づいた条件を導出し、より広いミニ

マックス性の条件を与える。

非ミニマックス性の代表的な例として、いくつかの分布の平均が正に制約されているときに、その和を推定する問題がある。このとき、通常の標本平均もしくはその打ち切り推定量の和で推定したのではミニマックスにならないことが知られている。この問題について、許容的でミニマックスな一般化ベイズ推定量の導出を試みる。

4. 研究成果

(1) 小地域推定の MSE 及び信頼区間の高次漸近補正

まず、簡単な Fay-Herriot モデルと枝分かれ誤差分散回帰モデルについて、有限母集団モデルの枠組みで、EBLUP に基づいた信頼区間の信頼係数が 2 次漸近的に 95% になるように、補正項を解析的に求め、京急線沿いの地価公示価格データへ応用した。

次に、一般的な正規線形混合モデルにおいて、共分散行列に埋め込まれた未知母数の推定量に一般的な確率展開を仮定し、EBLUP の MSE の 2 次、3 次の漸近近似を与え、それに基づいて MSE の 2 次、3 次の漸近不偏推定量を解析的に求めた。これに関連して高次の信頼性をもつ信頼区間の構成を行った。従来は 2 次までの漸近近似が提供されてきたが、ここでは 3 次の漸近近似とそのための条件を明確にした。また最尤推定量と制限付き最尤推定量の確率展開を与えた。

以上の結果はテラー展開による解析的な方法に基づいており、具体的なモデルと特定の母数推定の方法に応じて偏微分やモーメントを個々に計算する必要があり、2 次近似を適用する上で不便である。そこで、パラメトリック・ブートストラップ法を提案し、MSE の 2 次漸近不偏推定量と信頼区間の 2 次漸近補正を提案した。

EBLUP の MSE の推定に関して、小地域のデータを与えたときの条件付き MSE の推定についても、この分野で著名な研究者である G. Datta, J. N. K. Rao らとともに、2 次漸近不偏推定量を求めた。

(2) 線形混合モデルの検定についてのバートレット補正

線形混合モデルの回帰係数の検定問題については、通常の F 検定が有意水準をみたさないという欠点があることが指摘されている。そこでバートレット補正を行うのが常套手段であるが、実は Rothenberg (1984) によりかなり一般的な枠組みでワルド検定、スコア検定、尤度比検定のバートレット補正が求められて

いる。しかし彼の結果は共分散行列の母数推定に最尤推定を用いている点とバートレット補正を具体的に計算するのが困難であることが問題点としてあげられる。線形混合モデルもこの一般的な枠組みに入るが、最尤推定以外にも制限付き最尤推定やモーメント推定量など様々な推定量が用いられており、これらについては Rothenberg の手法を用いることができない。そこで、一般的な推定量の確率展開を仮定し、ワルド検定、スコア検定、尤度比検定のバートレット補正を解析的に導出した。またバートレット補正をパラメトリック・ブートストラップ法により求める方法についても提案した。

(3) 線形混合モデルにおける変数選択規準

線形混合モデルの説明変数の選択問題については、赤池情報量規準(AIC)よりも条件付き AIC の方が変数効果や地域効果の予測を考慮に入れた方法として望ましいことが指摘されている。しかし、共分散行列にかなり特殊な仮定を入れた上で条件付き AIC を求めているため、分散成分の選択などを含めることができない。そこで、かなり一般的な線形混合モデルにおいて、分散成分の推定量に確率展開を仮定して、バイアス項の 2 次漸近不偏推定量の導出を行い、一般的な条件付き AIC を導出した。この情報量規準は、テラー展開を用いて導出されるので、分散成分に関する偏微分等を含むことになり、複雑なモデルに対しては計算が困難になってしまう。そこでテラー展開を用いる代わりにパラメトリック・ブートストラップ法を利用した簡便な方法によりペナルティー項を導出し、得られた情報量規準の 2 次漸近不偏性を証明した。またその有用性を数値実験により示した。また経験ベイズの考え方を導入した変数選択基準を導出し一貫性を証明した。

(4) 線形混合モデルにおけるベンチマーク問題

一般的な正規線形混合モデルにおけるベンチマーク問題について、平均と分散の両方に一般的なベンチマーク制約を入れた制約付き経験ベイズ推定量を求め、その MSE の漸近近似を求めて性質を論じた。特に平均のベンチマーク問題については、制約付き経験ベイズ推定量と制約無しの経験ベイズ推定量の違いが MSE の 2 次の項に現れる一方、分散のベンチマーク問題については、それらの違いが 1 次の項に現れてくることを指摘した。また、制約付き経験ベイズ推定量の MSE の 2 次漸近不偏推定量をパラメトリック・ブートストラップ法を用いて与えた。またその結果を家計

調査データの解析へ応用した。

(5) 離散混合分布族におけるベンチマーク問題

(4)の正規線形混合モデルにおいて地域レベルの集約データを解析するためのFay-Herriotモデルを、離散分布を含んだ混合モデルに拡張し、その一般的なモデルのもとで、平均と分散のベンチマーク制約に対する制約付き経験ベイズ推定量を求め、その平均2乗誤差の2次漸近評価と2次不偏推定の導出を行った。この一般論の応用として、ポアソン-ガンマ分布、2項-ベータ分布を含むような、2次分散関数をもつ自然指数型分布族を取り上げ、制約付き経験ベイズ推定量の漸近不偏推定方法を提案した。また埼玉県各市町村別胃ガン死亡率データに適用して提案した方法の有用性を示した。

(6) 高次元多変量解析の展開

高次元の多変量解析の中心的な話題である2群の線形判別分析に関して、共分散行列のリッジ推定量を用いた線形判別分析の誤判別確率に対する漸近展開を導出し、リッジ推定を用いたことによる有効性を示した。また誤判別率の2次漸近不偏推定量を導出した。大標本での誤判別確率の漸近展開は非常に古くから研究されてきたが、次元が高くなると近似が悪くなることが知られてきた。また次元が高くなると共分散行列の逆行列の推定が不安定になるという問題点も知られてきた。ここで導出された手法は、これらの問題を同時に解決するものであり、数値的にも従来のものよりかなり改善されていることが示された。

さらに、誤判別確率の2次漸近不偏推定量が変数選択規準としても利用可能なことを示し、変数選択規準としての最適性も証明した。

(7) 多次元母数の推定における最適性理論の新たな展開

分散が未知の正規分布における平均ベクトルの推定に関するスタイン問題について、縮小推定量がミニマックスになるための積分不等式を導出し、一般化ベイズ推定量のミニマックス性について従来条件が緩和できることを示した。

平均が正に制約されているK個の正規分布の平均の和の推定問題を考えると、Kが大ききときには一般化Bayes推定量や最尤推定量がミニマックスにならない。それぞれの推定量についてミニマックス性が成り立つための必要十分条件を導出した。また許容的でミニマックスな推定量を導出するとともにいくつかの方向への拡張を行った。

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)

[雑誌論文] (計 21 件)
以下の論文はすべて査読有り。

(1) M. Hyodo and T. Kubokawa (2014). A variable selection criterion for linear discriminant rule and its optimality in high dimensional and large sample setting. *Journal of Multivariate Analysis*, 123, 364-379.

(2) T. Kubokawa and W. E. Strawderman (2013). Dominance properties of constrained Bayes and empirical Bayes estimators. *Bernoulli*, 19, Issue 5B, 2200-2221.

(3) T. Kubokawa (2013). Constrained empirical Bayes estimator and its uncertainty in normal linear mixed models. *Journal of Multivariate Analysis*, 122, 377-392.

(4) M. Kojima and T. Kubokawa (2013). Bartlett adjustments for hypotheses testing in linear models with general error covariance matrices. *Journal of Multivariate Analysis*, 122, 162-174.

(5) T. Kubokawa, E. Marchand, W. E. Strawderman and J.-P. Turcottex (2013). Minimality in predictive density estimation with parametric constraints. *Journal of Multivariate Analysis*, 116, 382-397.

(6) T. Kubokawa, M. Hyodo and M. S. Srivastava (2013). Asymptotic expansion and estimation of EPMC for linear classification rules in high dimension. *Journal of Multivariate Analysis*, 115, 496-515.

(7) M. S. Srivastava and T. Kubokawa (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, 115, 204-216.

(8) T. Kubokawa and B. Nagashima (2012). Parametric bootstrap methods for bias correction in linear mixed models. *Journal of Multivariate Analysis*, 106, 1-16.

(9) T. Kubokawa (2012). Minimax estimation of linear combinations of

restricted location parameters.
Contemporary Developments in Bayesian
Analysis and Statistical Decision Theory:
A Festschrift for William E. Strawderman",
IMS Collections, Vol. 8, 24-41, Institute
of Mathematical Statistics.

(10) T. Kubokawa (2011). On measuring
uncertainty of small area estimators with
higher order accuracy. Journal of the
Japan Statistical Society, 41, No.2,
93-119.

(11) T. Kubokawa and W.E. Strawderman
(2011). A unified approach to
non-minimality of sets of linear
combinations of restricted location
estimators. Journal of Multivariate
Analysis, 102, Issue 10, 1429-1444.

(12) T. Kubokawa (2011). Conditional and
unconditional methods for selecting
variables in linear mixed models. Journal
of Multivariate Analysis, 102, Issue 3,
641-660.

(13) H. Tsukuma and T. Kubokawa (2011).
Modifying estimators of ordered positive
parameters under the Stein loss. Journal of
Multivariate Analysis, 102, Issue 1,
164-181.

(14) T. Kubokawa and W.E. Strawderman
(2011). Non-minimality of linear
combinations of restricted location
estimators and the related problems.
Journal of Statistical Planning and
Inference, 141, Issue 6, 2141-2155.

(15) G.S. Datta, T. Kubokawa, J.N.K. Rao
and I. Molina (2011). Estimation of mean
squared error of model-based small area
estimators. Test, an Official Journal of
the Spanish Society and Operations
Research, 20, No.2, 367-388.

(16) T. Kubokawa and M.S. Srivastava (2010).
An empirical Bayes information criterion
for selecting variables in linear mixed
models. Journal of the Japan Statistical
Society, 40, No.1, 111-130.

(17) T. Kubokawa (2010). A review of
linear mixed models and small area
estimation. Journal of the Statistical
Research, 44, No.1, 31-55.

(18) T. Kubokawa and N. Erdembat (2010).
On testing linear hypothesis in a nested
error regression model. Communication in

Statistics - Theory and Methods, 39,
1552-1562.

(19) T. Kubokawa (2010). Corrected
empirical Bayes confidence intervals in
nested error regression models. Journal
of the Korean Statistical Society, 39, No. 2,
221-236.

(20) M.S. Srivastava and T. Kubokawa (2010).
Conditional information criteria for
selecting variables in linear mixed models.
Journal of Multivariate Analysis, 101,
1970-1980.

(21) T. Kubokawa (2009). Integral
inequality for minimality in the Stein
problem. Journal of the Japan Statistical
Society, 39, No.2, 155-175.

[学会発表] (計 4 件)

- (1) 川久保友超, 久保川達也, M. Ghosh.
(2013年9月10日). 小地域推定における
ベンチマーク問題と乗法モデルにおける
推定誤差評価. 2013年度統計関連学会連
合大会. 大阪大学
- (2) 川久保友超, 久保川達也. (2013年9月10
日). 線形混合モデルに対する変数選択規
準 cAIC の修正. 2013年度統計関連学会
連合大会. 大阪大学
- (3) 小島将裕, 久保川達也. (2013年9月9日).
一般線形回帰モデルの回帰係数に関する
検定統計量のバートレット補正. 2013年
度統計関連学会連合大会. 大阪大学
- (4) 久保川達也. (2011年9月6日). 縮小推定
とその周辺. 2011年度統計関連学会連合
大会. 九州大学

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ホームページ等

<http://www.tatsuya.e.u-tokyo.ac.jp/>

6. 研究組織

(1) 研究代表者

久保川達也 (KUBOKAWA, Tatsuya)
東京大学・大学院経済学研究科・教授
研究者番号: 20195499