

平成 30 年 5 月 15 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26330036

研究課題名(和文) 多変量解析の現代的諸問題に対する推測理論の新たな展開に関する研究

研究課題名(英文) Research on new developments of theory of statistical inference in several modern problems in multivariate analysis

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(経済学部)・教授

研究者番号：20195499

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究課題では、多変量統計推測手法が直面する諸問題に対して数理統計の立場から新たな解決策を導出し、その有効性・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て実施した。特に、(A)線形混合モデル・一般線形混合モデルを利用した小地域推定の新展開、(B)小地域推定におけるベンチマーク問題、(C)線形混合モデルにおける変数選択問題、(D)高次元多変量推測における縮小推定量法の有用性、(E)多次元母数の推測に関する最適性理論の展開、の5つのテーマについて研究成果を得た。

研究成果の概要(英文)：In this research project, I addressed several problems in multivariate data analysis, derived new statistical methods to solve the problems and developed theory on their usefulness and optimality. The performances of the suggested methods were investigated through simulation experiments and data analysis. In particular, I obtained research results for the following five topics: (A) New developments in small area estimation based on linear mixed models and generalized linear mixed models, (B) Benchmark problems in small area estimation, (C) Variable selection in linear mixed models, (D) Use of shrinkage estimation methods in high-dimensional multivariate analysis, and (E) Developments of optimality theory in multivariate statistical inference.

研究分野：数理統計学

キーワード：線形混合モデル 小地域推定 高次元解析 統計的決定理論 共分散行列 縮小推定 経験ベイズ 階層ベイズ

1. 研究開始当初の背景

多次元データの解析は応用上極めて有用で、そのための統計的推測手法の理論研究は活発に議論されている。特に興味あるトピックは小地域推定と高次元解析である。

小地域推定に利用されるモデルが線形混合モデルで、従来の線形回帰モデルに確率変動する項(変量効果)を加えたものである。例えば、個体や地域に関する観測値に共変量が存在する場合、その共変量で回帰するモデルが線形回帰モデルとなるが、それに個体や地域などの差異を変量として組み込み、それらの背後に共通な確率分布を想定するモデルが考えられる。このような線形混合モデルを用いると変量効果を予測してやることで個体や地域の差異を推定することができる。全体の平均的な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、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着目すると、問題点は、人口が疎な地域の平均をその地域の標本平均で推定すると、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。この問題に線形混合モデルを用いると、線形回帰の項(固定効果)はデータをプールして安定した推定値を与える機能がある一方、地域毎の変量効果は地域毎の標本平均を安定した推定値の方向へ近づける(縮小する)作用を生み出す。こうして導かれる個々の地域の推定方法は経験最良線形不偏予測量(EBLUP)と呼ばれ、個々の地域の標本平均をプールされ安定した推定値の方向へ縮小することにより、地域の差異を反映し、しかもより精度の高い推定値を与えることが可能となる。ここで、経験最良線形不偏予測量によって実際どの程度の推定誤差の改善が図られるのかを見積もることは現実のデータ解析の場面で重要である。そのためには平均2乗誤差に対する推定値を与えること、もしくは信頼区間を構成することが考えられる。

現実のデータを解析する際、従来の方法では不十分な様々問題に直面する。その一つが、正のデータを正規分布を用いて解析することの妥当性である。また、すべての地域で同じ誤差分散を用いていることの妥当性も問題である。さらに、小地域の経験最良線形不偏予測量の値を合算して全体平均を計算してみると全体の標本平均の値と一致しないという問題点もある。これはベンチマーク問題と呼ばれ、例えば全体の予算を各地域の値に応じて配分することを考えると、経験最良線形不偏予測量の値に基づいて配分することはできないことになる。この問題に対して理論的側面から解決策を図る必要がある。

多変量解析の別のトピックとして高次元

解析がある。これは多変量解析の現代的流れの一つであり、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が活発に議論されている。判別分析などの多変量解析の手法には共分散行列の逆行列の推定が組み込まれているが、共分散行列の次元がデータ数より大きいときには標本共分散行列は逆行列を持たない。こうした問題をどのように解決するかが問題となる。

2. 研究の目的

本研究課題では、多変量統計推測手法が直面する諸問題に対して数理統計の立場から新たな解決策を導出し、その有効性・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て実施した。また理論的及び応用的側面から従来の手法との比較検討を行った。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統一理論の構築を行った。

具体的には次の5つのトピックを扱い新たな統計手法の開発とそれに伴う理論展開を行った。

- (A) 線形混合モデル・一般化線形混合モデルを利用した小地域推定理論の新展開
- (B) 小地域推定におけるベンチマーク問題
- (C) 混合モデルにおける変数選択の理論展開
- (D) 高次元多変量推測手法における新たな理論展開
- (E) 多次元母数の推測に関する最適性理論の展開

3. 研究の方法

(A) については、正のデータを実数直線上へ変換する関数を考え、変換母数の一致推定量の導出を行った。また地域毎に誤差分散が異なる問題については、線形混合モデルに分散の変量効果を組み込んだモデルを考え、経験ベイズ及び階層ベイズの枠組みで推定手法の導出に取り組んだ。(B) ベンチマーク問題については、全体の平均や分散が予め決められた値になるという制約条件の下でベイズ解を求めることから始めた。この解から制約付き経験ベイズ推定量が導出されるので、その推定誤差の漸近的評価を与えた。

(C) については、混合モデルにおける変数選択の方法として最近提案されたものに条件付き赤池情報量規準があり、真のモデルを含んでいないときにはバイアスを生ずることが知られている。Fujikoshi-Satoh の考え方を適用してバイアス補正を行うことを考えた。(D) については、共分散行列の次元がデータ数より大きいときには標本共分散行列は逆行列を持たないため、リッジ型の推

定量もしくは線形縮小推定量の逆行列が用いられる。そこで正規性を仮定した上で導出されるベイズ推定量から自然な線形縮小推定量の形を与え、ノンパラメトリックな枠組みで線形縮小推定量の中に登場する重み関数を推定する方法を用いた。(E)については、多変量正規分布の平均ベクトルの Stein 推定に関連して、この推定量のミニマックス性が高次元の場合にも拡張できることを示した。また共分散行列の James-Stein 推定量のミニマックス性についても、Kiefer の定理のような高度な一般論を使うことなく、基本的な方法で容易に示すことができることを示した。この方法を用いることにより、母数が制約されている場合への拡張も可能になる。

4. 研究成果

個々のトピックについて研究成果を説明する。

(A) 線形混合モデル・一般化線形混合モデルを利用した小地域推定理論の新展開

まず、収入や支出など正の値をとるデータを Fay-Herriot モデルという正規線形混合モデルに当てはめる問題を考察した。通常は対数変換を行うが必ずしも相応しいとは限らない。そこで Box-Cox 変換を用いることが考えられるが変換母数の最尤推定量が一致性をもたないことが知られている。これに対して双対変換を導入した。双対変換については変換母数の最尤推定量が一致性をもつことを示し、この一致推定量に基づいた経験最良線形不偏予測量とその平均 2 乗誤差の漸近 2 次不偏推定量の導出を行い、数値実験を通して有用性を示した。

次に、小地域毎に誤差分散が異なっているときの推測問題を扱った。分散の異質性を表現するために分散の逆数にガンマ分布を仮定する変量分散混合モデルを提案した。このモデルを用いると、小地域平均のベイズ推定量の中に平均の縮小推定と分散の縮小推定の両方が現れることを指摘し、最尤法を用いて母数推定を行い、それらの漸近共分散行列を求めて経験ベイズ推定量の平均 2 乗誤差の漸近 2 次近似を与えた。またブートストラップ法に基づいて、その平均 2 乗誤差の漸近 2 次不偏推定量を導出し、シミュレーション実験と京浜急行線沿いの宅地公示価格データの解析を通して提案手法の良さを示した。さらに、分散の異質性を組み込むためのガンマ分布について、その超母数に事前分布を想定する階層的なフルベイズモデルを提案し、マルコフ連鎖モンテカルロ法のアルゴリズムを導出するとともに事後分布の積分可能性を証明した。

地域毎に分散が異なる問題を扱う別の方法として、各地域の分散が地域に依存した共変量の関数として表されるモデルを考えることができる。こうした分散関数を用いる方

法の良さは分散が共通の関数とパラメータで表されている点にあり、上述のような変量分散モデルのような階層的な事前分布を導入する必要がない点にある。この枠組みにおいて経験最良線形不偏予測量を導出し、その推定誤差の漸近 2 次近似とその 2 次不偏推定量の解析的な導出を行った。京浜急行線沿いの宅地公示価格データに当てはめてみると、最寄り駅に近いほど価格の分散が大きくなっており、分散関数が駅からの距離の関数で表現できることがわかり、提案手法の有用性を示すことができた。

一般化線形混合モデルにおける経験ベイズ推定量について、通常平均 2 乗誤差と条件付き平均 2 乗誤差に関する研究を行い、2 次分散関数をもつ指数型分布族において両者の違いを明確にした。その結果、ベータ分布やポアソン分布については 1 次の項に違いが生じ、正規分布のときには 2 次の項に違いが生ずることを示した。

その他、新たな展開として変量効果の不確実性を考慮したモデルとして spike-slab 事前分布を用いたベイズ的モデルを考え、経験ベイズ法やフルベイズ法に基づいた推定手法の導出と数値的検討を行った。

(B) 小地域推定におけるベンチマーク問題について

ベンチマーク問題の先行研究は主に平均に関するもので平均と分散に関するベンチマーク問題についての研究はあまりなされていない。そこで一般的な混合モデルの設定で、平均と分散に制約を入れた制約付きベイズ推定量を求めることによって、平均と分散の両方についてのベンチマーク問題の一般的な解を与えた。その制約付き経験ベイズ推定量の平均 2 乗誤差の評価とその推定方法を求め、家計調査の地域別ベータや埼玉県の間ガン死亡率データ、戦前の乳児死亡率に関する石川県の市町村別データなどに当てはめ、提案する手法の妥当性を調べた。

次に、家計調査の支出など正の値を取るデータについて平均のベンチマーク問題を取り上げた。この問題は、現実的に起こりうる設定であるが、平均 2 乗誤差のもとでは制約付きベイズ推定量の解を陽に求めることができず、その解の存在性すら懐疑的である。そこで、カルバック・ライブラー情報量に基づいた損失関数を導入し、乗法モデルのもとで極めて自然な形のベンチマーク推定量を求めることができることを示し、推定誤差の漸近的な評価を行った。

(C) 混合モデルにおける変数選択の理論展開

混合モデルの変数選択法として条件付き赤池情報量規準が提案されているが、未解決の問題として、候補のモデルが真のモデルを含んでいないときに条件付き赤池情報量規準がバイアスをもってしまうという欠点が

あることが知られている。この点を修正するために条件付き赤池情報量規準のバイアスの補正項を導出し、修正条件付き赤池情報量規準なる手法を提案し、その良さを数値的に確認した。また、共変量が観測モデルと予測モデルで異なる場合を扱い、こうした状況に適応した条件付き情報量規準の導出を行った。

(D) 高次元多変量推測手法における新たな理論展開

まず、共分散行列の推定に関して、リッジ係数の最適推定を与える統一的な方法として線形縮小推定法を与えた。これは、正規分布に基づいたベイズ推定量から線形縮小推定量の形を導き、リッジ係数もしくは重み付き関数を高次元の枠組みでノンパラメトリックに推定する方法である。この方法で求めた推定量が先行研究で提案されたものよりも数値的に優れていることを示した。次に、ファクターモデルが予想される状況においてファクターモデルを想定した推定量の方向へ縮小する共分散行列の推定量を提案し、高次元での収束性を調べるとともに数値的な性質を調べた。

(E) 多次元母数の推測に関する最適性理論の展開

この課題では、統計的決定論の視点に立って、これまで解決されてこなかった推測問題の最適性について研究した。まず、分散比の改良型推定量の導出に関する統一理論を構築し母数制約問題との関係を明らかにした。次に、共分散行列の推定においてミニマックス推定を与える事前分布の列を構成した。これは、古くから知られる Kiefer の定理に対して、共分散行列の推定の枠組みでより簡便な証明方法を提供できたことを意味する。さらに、平均ベクトルの推定問題における Stein 型縮小推定量の改良結果を、高次元と低次元を統一するかなり一般的な枠組みにまで拡張できることを示した。

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 27 件)

① S. Sugasawa and T. Kubokawa (2017). Bayesian estimators in uncertain nested error regression models. *Journal of Multivariate Analysis*, 153, 52-63.
DOI: 10.1016/j.jmva.2016.09.011 査読有

② S. Sugasawa, H. Tamae and T. Kubokawa (2017). Bayesian estimators for small area models shrinking both means and variances. *Scandinavian Journal of Statistics*, 44, 150-167.

DOI: 10.1111/sjos.12246 査読有

③ T. Kubokawa, Eric Marchand and W.E. Strawderman (2017). On predictive density estimation for location families under integrated absolute error loss. *Bernoulli*, 23, 3197-3212.

DOI:10.3150/16-BEJ842 査読有

④ H. Tsukuma and T. Kubokawa (2017). Proper Bayes and minimax predictive densities related to estimation of a normal mean matrix. *Journal of Multivariate Analysis*, 159, 138-150.

DOI: 10.1016/j.jmva.2017.05.004 査読有

⑤ S. Sugasawa and T. Kubokawa (2017). Transforming response values in small area prediction. *Computational Statistics and Data Analysis*, 114, 47-60.

DOI: 10.1016/j.csda.2017.03.017 査読有

⑥ S. Sugasawa and T. Kubokawa (2017). Heteroscedastic nested error regression models with variance functions. *Statistica Sinica*, 27, 1101-1123.

DOI: 10.5705/ss.202015.0318 査読有

⑦ T. Kubokawa, E. Marchand and W.S. Strawderman (2017). A unified approach to estimation of noncentrality parameters, the multiple correlation coefficient, and mixture models. *Mathematical Methods of Statistics*, 26, 134-148.

DOI: 10.3103/S106653071702003X 査読有

⑧ S. Sugasawa, T. Kubokawa and K. Ogasawara (2017). Empirical uncertain Bayes methods in area-level models. *Scandinavian Journal of Statistics*, 44, 684-706.

DOI: 10.1111/sjos.12271 査読有

⑨ T. Kubokawa, S. Sugasawa, M. Ghosh and S. Chaudhuri (2016). Prediction in heteroscedastic nested error regression models with random dispersions. *Statistica Sinica*, 26, 465-492.

DOI:10.5705/ss.202014.0070 査読有

⑩ Y. Ikeda, T. Kubokawa and M.S. Srivastava (2016). Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions. *Computational Statistics and Data Analysis*, 95, 95-108.

DOI: 10.1016/j.csda.2015.09.011 査読有

⑪ H. Tsukuma and T. Kubokawa (2016). Unified improvements in estimation of a

normal covariance matrix in high and low dimensions. *Journal of Multivariate Analysis*, 143, 233-248.
DOI: 10.1016/j.jmva.2015.09.016 査読有

⑫ S. Sugasawa and T. Kubokawa (2016). On conditional prediction errors in mixed models with application to small area estimation. *Journal of Multivariate Analysis*, 148, 18-33.
DOI: 10.1016/j.jmva.2016.02.009 査読有

⑬ Y. Ikeda and T. Kubokawa (2016). Linear shrinkage estimation of large covariance matrices using factor models. *Journal of Multivariate Analysis*, 152, 61-81.
DOI: 10.1016/j.jmva.2016.08.001 査読有

⑭ T. Kubokawa (2016). Development of shrinkage methods in estimation - high dimensional analysis and small area estimation -. *Journal of the Japan Statistical Society, J-Series*, 46, 43-67.
DOI: 10.11329/jjssj.46.43 査読有

⑮ H. Tsukuma and T. Kubokawa (2015). Minimality in estimation of restricted and non-restricted scale parameter matrices. *Annals of the Institute of Statistical Mathematics*, 67, 261-285.
DOI: 10.1007/s10463-014-0449-x 査読有

⑯ S. Sugasawa and T. Kubokawa (2015). Parametric transformed Fay-Herriot model for small area estimation. *Journal of Multivariate Analysis*, 139, 295-311.
DOI: 10.1016/j.jmva.2015.04.001 査読有

⑰ H. Tsukuma and T. Kubokawa (2015). A unified approach to estimating a normal mean matrix in high and low dimensions. *Journal of Multivariate Analysis*, 139, 312-328.
DOI: 10.1016/j.jmva.2015.04.003 査読有

⑱ H. Tsukuma and T. Kubokawa (2015). Estimation of the mean vector in a singular multivariate normal distribution. *Journal of Multivariate Analysis*, 140, 245-258.
DOI: 10.1016/j.jmva.2015.05.016 査読有

⑲ T. Kubokawa, Eric Marchand and W.E. Strawderman (2015). On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, 142, 57-74.
DOI: 10.1016/j.jmva.2015.07.013 査読有

⑳ M. Ghosh, T. Kubokawa and Y. Kawakubo (2015). Benchmarked empirical

Bayes methods in multiplicative area-level models with risk evaluation. *Biometrika*, 102, 647-659.
DOI: 10.1093/biomet/asv010 査読有

㉑ T. Kubokawa, M. Hasukawa and K. Takahashi (2014). On measuring uncertainty of benchmarked predictors with application to disease risk estimate. *Scandinavian Journal of Statistics*, 41, 394-413.
DOI: 10.1111/sjos.12040 査読有

㉒ T. Kubokawa (2014). General dominance properties of double shrinkage estimators for ratio of positive parameters. *Journal of Statistical Planning and Inference*, 147, 224-234.
DOI: 10.1016/j.spi.2013.11.009 査読有

㉓ T. Kubokawa and A. Inoue (2014). Estimation of covariance and precision matrices in high dimension. *Electronic Journal of Statistics*, 8, 130-158.
DOI: 10.1214/14-EJS878 査読有

㉔ 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.
DOI: 10.1016/j.jmva.2013.10.005 査読有

㉕ Y. Kawakubo and T. Kubokawa (2014). Modified conditional AIC in linear mixed models. *Journal of Multivariate Analysis*, 129, 44-56.
DOI: 10.1016/j.jmva.2014.03.017 査読有

㉖ M.S. Srivastava, H. Yanagihara and T. Kubokawa (2014). Tests for covariance matrices in high dimension with less sample size. *Journal of Multivariate Analysis*, 130, 289-309.
DOI: 10.1016/j.jmva.2014.06.003 査読有

㉗ T. Kubokawa, Eric Marchand and W.E. Strawderman (2014). On improved shrinkage estimators for concave loss. *Statistics and Probability Letters*, 96, 241-246.
DOI: 10.1016/j.spl.2014.09.024 査読有

[学会発表] (計 12 件)

① 久保川達也「混合効果モデルを利用した小地域推定」日本統計学会春季大会, 2016

② 菅澤翔之助, 久保川達也「分散関数を用い

た Heteroscedastic Nested Error Regression モデル」統計関連学会連合大会，2015

③玉江大将，久保川達也「小地域推定における平均と分散の同時縮小推定法」統計関連学会連合大会，2015

④玉江大将，久保川達也「平均と分散の同時縮小推定における変量効果のロバストな予測について」統計関連学会連合大会，2015

⑤伊藤翼，久保川達也「高次元共分散行列の逆行列の推定問題とその応用」統計関連学会連合大会，2015

⑥久保川達也「縮小推定の展開—小地域推定と高次元問題—」統計関連学会連合大会，2015

⑦久保川達也「線形混合モデルと EBLUP を利用した小地域推定の新展開」統計関連学会連合大会，2014

⑧菅澤翔之助，久保川達也「経験ベイズ推定量の条件付き MSE とその小地域推定への応用」統計関連学会連合大会，2014

⑨川久保友超，久保川達也「線形混合モデルにおける予測情報量規準」統計関連学会連合大会，2014

⑩菅澤翔之助，久保川達也「変換線形混合モデルを用いた小地域推定」統計関連学会連合大会，2014

⑪川久保友超，菅澤翔之助，久保川達也「共変量の値が変化する状況下での線形混合モデルの変数選択とその小地域推定への応用」統計関連学会連合大会，2014

⑫池田祐樹，久保川達也「追加的な共変量をもちいた高次元共分散行列の推定」統計関連学会連合大会，2014

〔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 なし

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

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

なし