

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：32661

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500361

研究課題名(和文) 行列型パラメータをもつ統計的モデルに対するベイズ推定理論と応用についての研究

研究課題名(英文) A study on theory and applications of Bayesian estimation for parameter matrices in multivariate statistical models

研究代表者

津熊 久幸 (TSUKUMA, Hisayuki)

東邦大学・医学部・講師

研究者番号：50424685

交付決定額(研究期間全体)：(直接経費) 2,000,000円、(間接経費) 600,000円

研究成果の概要(和文)：この研究では、行列型配列データを解析するときに必要な行列型パラメータをもつ統計的モデルのベイズ推定問題を扱った。統計的決定理論の観点、及び実データへの応用の観点から最適性を有するベイズ推定手法や縮小推定手法の開発を主な目標とした。共分散行列のミニマクス推定のための最も不利な事前分布や、母数制約のある精度行列の一般化ベイズ推定量の性質、平均行列の縮小推定量の非許容性などについての成果が得られた。

研究成果の概要(英文)：This study addresses the Bayesian estimation problems of parameter matrices in multivariate statistical models which are used for analyses of two dimensional array data. The principal aim of the study is to develop optimal Bayes estimators and shrinkage-type estimators not only from the point of view of statistical decision theory, but also from that of practical application. Some results are obtained for a least favorable sequence of prior distributions in covariance estimation, dominance properties of a generalized Bayes estimator of a bounded precision matrix, inadmissibility of shrinkage estimators for a mean matrix, and others.

研究分野：数理統計学

科研費の分科・細目：情報学・統計科学

キーワード：統計的推測 ベイズ推定 統計的決定理論 縮小推定 スタイン現象 共分散行列

1. 研究開始当初の背景

科学・情報技術の発展とともに、地球観測データやゲノム関連データ、画像データのような大規模データを解析する必要性が高まり、複雑なデータ構造を柔軟に表現するための統計的モデルと、より体系的で統一的な多変量推測統計学の理論構築が喫緊の課題となっている。

一般に、大規模データの構造を記述する統計的モデルの解析は、困難をともなう場合が多い。例えば、パラメータ数が標本数より大きいときには、古典的な多変量推測手法を適用できない場合もある。また、データ数が増大することにより、統計量の算出における計算量の幾何級数的な増大といった問題などもある。

大規模データの解析において、ベイズ統計学は体系的な理論を構築するための方法を提供すると思われる。ベイズ流の統計理論は、近年、世界的に実用化が進んでおり、日本においても基礎及び応用研究が活発に進められている。ベイズ統計学の理論的な研究は様々な学術分野と関連があり、実際に、数理統計学、多変量解析、多変数積分論、群論、調和解析、情報通信理論、ゲーム理論などとの関係がある。

ベイズ推定は、複雑なモデルを仮定したとしても、マルコフ連鎖モンテカルロ法などから計算機技術を駆使し比較的容易におこなえるようになってきた。しかし、行列型パラメータの推定問題における統計的決定理論の観点からの議論は少なく、平均行列の推定問題以外のベイズ理論的な結果はあまり得られていない。

行列型パラメータの事後平均のようなベイズ推定量は、行列変数の関数の多重積分を用いて表現される。一方、統計的決定理論では推定量のリスク関数の評価が必要になる。したがって、事後平均のようなベイズ推定量とそのリスクに関しては数学的な取り扱いの難しさがあり、活発に研究されてこなかったのだと思われる。

以上のような背景から、大規模データを記述する統計モデル、より限定的には行列型のパラメータで特徴づけられる統計モデルの推測問題に関する決定理論的な研究は、未開拓の研究分野の一つとなっていた。

2. 研究の目的

本研究では、大規模データ、特に行列型配列データを解析する上で必要となる行列型パラメータをもつ統計的モデルに対するベイズ推定手法や関連する問題について、次のような目的で研究を進めた。

- (1) さまざまな推定問題における、統計的決定理論の観点から最適性を有するベイズ推定方式(一般化ベイズ法、経験ベイズ法、階層的ベイズ法など)の開発
- (2) 古典的な多変量推測統計理論における

未解決問題の再考

(3) 実データへの応用や推測アルゴリズムの効率など、関連する統計学的・数学的な諸問題についての研究

(4) 一般化ベイズ推定量や経験ベイズ推定量の、大標本における漸近的な性質と小標本における安定性の追求

(5) 新たなベイズ推定法と、最尤推定法や縮小型推定法との数値的比較検討と実データへの応用、及び推定手法(事前分布)のさらなる改良

3. 研究の方法

本研究の目的を達成するために、次のような(1)、(2)のアプローチをとった。

(1) 問題点の整理や準備など

地球観測データやゲノム関連データなどに対して、行列型パラメータをもつ統計モデルを用いたデータ解析の実例はいくつか存在する。したがって、それらの中から標準的な実データをいくつか選択して、従来の様々な推定手法の問題点を抽出した。またシミュレーションによる検証もおこない、推定量の誤った使用による危険性の評価も検討した。これらの結果から、事前分布の設定などベイズ推定への適用方法について考察した。

統計的モデルを特徴付ける行列型パラメータには、平均行列(回帰係数行列)、共分散行列、相関行列、非心度行列など様々なものがあり、これらが推定の対象となる。このようなパラメータの多変量推測統計について、古典的な教科書や最新の論文などの文献から統計的決定理論に関連する問題点の整理や、従来の推定手法の改良すべき点などについて検討をおこなった。

2 乗損失関数に関する行列型パラメータのベイズ推定量は、行列変数関数の多重積分で表されることになる。一般にベイズ推定量は最尤推定量と同様に、比較的ゆるい正則条件のもとで、一致性をもつことが知られている。また推定量の性質を論じる上で、代数的なアプローチである不変性を考慮することも重要である。ベクトル型パラメータの推定問題では不変性に注目することによって、決定理論の観点からよい縮小推定量やベイズ推定量が発見された。同じように行列型パラメータの推定問題でも、推定量や確率分布、事前分布に関する不変性から優位なベイズ推定量の導出が期待できる。

このようなことから、よりよい推定量を導出するために、ベイズ推定量(一般化ベイズ推定量、経験ベイズ推定量、階層的ベイズ推定量など)の数学的な特徴や正則条件などについて調査をおこなった。

近年、標本数がパラメータ数より少ない場合の推定論や仮説検定論、モデル選択、変数選択のような多変量解析の研究が盛んに行われている。これらの研究の基礎となっているものは、古典的多変量推測統計の技法で

あり、標本数がパラメータ数より多い場合の研究成果の拡張によるところが大きい。したがって、古典的な多変量推測統計における決定理論の未解決問題が解決すれば、対応する標本数がパラメータ数より少ない場合の問題も解決の糸口が見えてくると思われる。

そこで古典的な多変量推測統計分野における未解決問題を再考し、大規模データや行列型配列データへの応用や数値的検証などについて考察した。

(2) 決定理論的な最適性をもつ推定量の導出を目指し、以下の問題に挑戦した。

多変量正規分布モデルの推定問題における、統計的決定理論の観点から最適性をもつベイズ推定手法の開発

成長曲線モデルや変量効果モデル、楕円型分布族モデルなど、より複雑なモデルに対するベイズ推定手法の考案

縮小型推定など、大規模データや行列型配列データに対するベイズ推定法から派生する推測問題の考察

ベイズ推定手法の数値実験による比較検討、実データへの適用と検証、及びベイズ推定手法(事前分布)のさらなる改良

4. 研究成果

(1) 変量効果モデルにおける予測問題では、予測量の族を制限することによって、母数制約のある精度行列(共分散行列の逆行列)を2乗損失関数の下での推定問題に帰着できる場合がある。そこで、母数制約のある精度行列の推定問題に対して、ベイズ推定法を適用することを考えた。

母数制約のある精度行列のベイズ推定のために、制約された母数空間上に一様に分布する事前分布(無情報事前分布)を考え、その事前分布に関する一般化ベイズ推定量(事後平均)が2乗損失関数の下で不偏推定量を改良することが証明できた。また、この一般化ベイズ推定量の改良を考え、階層的ベイズ推定量などの改良型推定量を導出することができた。

上記の結果を利用して、母数制約のある精度行列の推定問題に関連する変量効果モデルの予測問題において、最良線形不偏予測量を改良する一般化ベイズ予測量を構成することに成功した。この結果は、ある特定の量効果モデルのみの結果であるが、今後、より複雑な変量効果モデルのベイズ予測についての研究につながることを期待される。

(2) 多変量正規分布モデルの共分散行列をエントロピー損失関数の下で推定する問題において、正の対角成分をもつ下三角行列全体がなす変換群についての最良な共変推定量はミニマクス性を持つことが知られている。これは共変推定量とミニマクス性との関連性についての結果から示されていたが、推定量のミニマクス性を証明するためには最

も不利な事前分布の列を利用する方法もある。

共分散行列の推定問題において、最も不利な事前分布の列の構成は長い間未解決の問題となっていたが、この未解決問題を解決することに成功した。この事前分布の列は、楕円型分布モデルの尺度行列の最良な不変推定量のミニマクス性の証明にも応用できることを示した。

また、共分散行列に母数制約があるときの推定問題において、最良共変推定量がミニマクス性を保存することを、最も不利な事前分布の列を利用して証明した。

(3) 共分散行列のミニマクス推定量の構成は、正の対角成分をもつ下三角行列全体がなす変換群について不変な推定量を考えることによって達成される。このミニマクス推定量の改良は直交行列全体がなす変換群について不変な推定量を考えることが一般的である。これに対して、正の対角成分をもつ下三角行列の部分群に関する不変な推定量に着目し、この不変推定量の中からミニマクス推定量を見つけることを考えた。

ここで考えた下三角行列の部分群は交換子部分群であり、対角成分が1である下三角行列の全体からなる群である(この交換子部分群を(I)とする)。部分群(I)による尺度変換に関して不変な推定量の族を考え、最も不利な事前分布の列を用いてエントロピー損失関数の下でミニマクス性を有する推定量を発見することができた。また、交換子部分群(I)に関する不変推定量の中から、そのミニマクス推定量を改良するものが存在することを示すことにも成功した。

下三角行列全体がなす群は可解群であるため、交換子部分群の列が存在する。したがって、交換子部分群(I)から生成される交換子部分群が存在するが、この部分群は対角成分が1で対角成分の1つ下の成分が0である下三角行列全体からなる群である(この交換子部分群を(II)とする)。このとき、交換子部分群(II)に関する不変な推定量を考えることができるが、(II)に関する不変推定量の中で(I)に関するミニマクスな不変推定量を改良するものがあるかどうかは分かっておらず、今後の課題として残された。

(4) 共分散行列が未知の行列型正規分布モデルにおける平均行列の推定問題を不変な2乗損失関数の下で扱った。経験ベイズ法より導出される行列型縮小推定量を元にして、より広い推定量の族を考えた。この推定量の族は、平均行列のサイズや、次元の大きさ(共分散行列のサイズ)のすべての大小関係を含む形で表現され、これまでに発見されている様々な縮小推定量を含むという特徴がある。また、その推定量の族のリスク関数が、平均行列のサイズや次元の大きさに依存しない統一的方法によって評価できるところも

利点となっている。

推定量の族のリスク関数について部分積分法を用いて統一的に評価し、その評価式から不偏推定量を改良する推定量を導く方法をいくつか紹介した。

縮小推定量には、過度な縮小によって推定量の符号が変わってしまうという問題がある。この問題を克服する方法の一つである打ち切り手法について考察し、打ち切り型推定量が元の縮小推定量のリスクを減少させることを解析的に示すことに成功した。また、数値実験を通じて、打ち切り型推定量が元の縮小推定量のリスクを大きく減少させることを実証した。

今後の課題として、不偏推定量を改良する一般化ベイズ推定量や階層的ベイズ推定量などが平均行列のサイズや次元の大きさの様々な大小関係について統一的に構成できるかどうかがある。また、統計モデルの次元が大きくなるにつれ、推定量の性能が低下するため、高次元の下でもある程度性能が保証される推定手法の開発が望まれる。

(5) 実データに対する古典的な多変量解析には、データから算出される共分散行列の逆行列を利用するものもある。しかし、データから算出される共分散行列の固有値がゼロに近い値をとるようなときは、その解析結果の信頼性は保証できないであろう。

そこで、未知の共分散行列が特異である場合の多変量正規分布モデルに関する推測、具体的には平均ベクトルの推定問題を統計的決定理論の観点から議論した。ここで、特異な共分散行列をもつ多変量正規分布からの標本より算出される標本共分散行列もまた特異であることに注意する。

平均ベクトルの推定問題は特異な共分散行列を重みとする2乗損失関数の下で考察した。まず、ある事前分布の列を考えることにより、最尤推定量のミニマクス性を示した。また、縮小型推定量の族に関するリスクの不偏推定量を導出し、その族がミニマクスになるための条件を求めた。平均ベクトルの次元や標本数、共分散行列の階数の間のすべての大小関係を含む形で、リスクの不偏推定量が統一的に記述できることも示した。

ミニマクス性のための条件を利用して、ミニマクスなベイズ推定量を導出した。また、打ち切り型推定量に関する解析的な結果を与えることにも成功した。

5. 主な発表論文等

〔雑誌論文〕(計 7件)

Tsukuma, H. (2014). Bayesian estimation of a bounded precision matrix, *Journal of Multivariate Analysis*, 査読有, Vol.127, 160-172. DOI:10.1016/j.jmva.2014.02.0160047-259X

Tsukuma, H. (2014). Minimax covariance

estimation using commutator subgroup of lower triangular matrices, *Journal of Multivariate Analysis*, 査読有, Vol.124, 333-344. DOI: 10.1016/j.jmva.2013.11.007

Tsukuma, H. and Kubokawa, T. (2014). Minimality in estimation of restricted and non-restricted scale parameter matrices, accepted for publication in *Annals of the Institute of Statistical Mathematics*. 査読有

Tsukuma, H. and Kubokawa, T. (2014). Estimation of the mean vector in a singular multivariate normal distribution, 査読無, Discussion Paper Series, CIRJE-F-930, Faculty of Economics, The University of Tokyo. <http://www.cirje.e.u-tokyo.ac.jp/research/dp/2014/2014cf930.pdf>

Tsukuma, H. and Kubokawa, T. (2014). A unified approach to estimating a normal mean matrix in high and low dimensions, 査読無, Discussion Paper Series, CIRJE-F-926, Faculty of Economics, The University of Tokyo. <http://www.cirje.e.u-tokyo.ac.jp/research/dp/2014/2014cf926.pdf>

Tsukuma, H. (2012). Simultaneous estimation of restricted location parameters based on permutation and sign-change, *Statistical Papers*, 査読有, Vol.53, 915-934. DOI: 10.1007/s00362-011-0396-z

Tsukuma, H. (2011). Simultaneous estimation of restricted means via the Gauss divergence theorem, *Sankhya Series A*, 査読有, Vol.73, 110-124. http://sankhya.isical.ac.in/search/73a1/A08051_f.pdf

〔学会発表〕(計 1件)

津熊久幸 (2014)、行列型パラメータの推定問題における統計的決定理論、日本数学会 2014 年度年会、会場：学習院大学 (東京都豊島区)、2014 年 3 月 17 日発表

〔その他〕

ホームページ：

<http://www.lab2.toho-u.ac.jp/med/minfo/tsukuma/>

6. 研究組織

(1) 研究代表者

津熊 久幸 (TSUKUMA, Hisayuki)

東邦大学・医学部・講師

研究者番号：50424685

(2) 研究分担者

該当なし