様 式 Ｃ－１９、Ｆ－１９－１、Ｚ－１９ （共通）

科学研究費助成事業　　研究成果報告書

平成 ３０ 年　 ６ 月 １２ 日現在

機関番号：３２６６１
研究種目：基盤研究(C)（一般）
研究期間：2015～2017
課題番号：１５Ｋ００００５５

研究課題名（和文）非正則ウィシャート行列を用いた推測問題の理論と応用に関する研究

研究課題名（英文）A study on statistical inference based on singular Wishart matrices and its applications

研究代表者
津熊　久幸（TSUKUMA, Hisayuki）

東邦大学・医学部・准教授

研究者番号：５０４２４６８５

交付決定額（研究期間全体）：（直接経費）　　1,900,000円

研究成果の概要（和文）：高次元・小標本の多変量正規分布モデルにおいて，非正則ウィシャート行列に基づく推定法について統計的決定理論の観点から研究を行った。共分散行列をスタイン型損失関数の下で推定する問題に関して，いくつかの新たな推定手法を提案した。また，行列型正規分布の平均行列の推定に関連した予測確率密度関数のベイズ推定問題を扱い，カルバック・ライブラー情報量に関して許容的かつミニマックスな予測密度の導出に成功した。

研究成果の概要（英文）：This study addressed a high-dimensional normal distribution model with small sample size and considered shrinkage estimation based on singular Wishart matrices from a decision-theoretic point of view. Some new estimators of covariance matrix were proposed under a Stein-type loss. Also, this study considered the problem of estimating predictive densities related to estimation of a normal mean matrix, and succeeded in deriving proper Bayes minimax predictive densities relative to the Kullback-Leibler loss.

研究分野：数理統計学

キーワード：多変量推測統計　統計的決定理論　スタイン現象　縮小型推定　ベイズ推定　高次元モデル

１．研究開始当初の背景

(1) 近年の科学技術や情報技術における急速な進歩にともない，地球環境データや生体データ，画像データなどといった大規模なデータの測定が容易となり，その有効活用が模索されている。統計学的な手法としては，これまでに，非正則な統計モデル，スパースモデリング，機械学習，空間データ解析など様々なアプローチが提案されてきた。また，古典的な多変量解析手法（小次元大標本のモデルにおける手法）の延長での議論もなされており，国内外の多くの研究者らによって，高次元小標本モデル（未知パラメータの次元が標本サイズより大きいモデル）を仮定した統計的仮説検定のような理論的方法論をはじめ，モデル選択法，2 標本問題，判別分析，主成分分析などについて，いくつかの成果が得られていた。

(2) 統計的決定理論は 1950 年前後から研究が積み重ねられてきた分野であるが，その最大の成果は縮小推定量の発見であろう。多変量正規分布モデルにおいて，平均ベクトルの最尤推定量は最小分散不偏推定量であり，一致性やミニマクス性などの優れた性質を併せ持っている。しかし，平均ベクトルが 3 次元以上のときに，2 乗損失関数のもとで最尤推定量が非許容的になるという意外な事実が発見され，縮小推定量と呼ばれる最尤推定量よりよいミニマックス推定量が具体的に与えられた。

(3) 縮小推定量の発見以降，縮小推定やベイズ推定などの決定理論の研究は理論的な側面で大きく発展しており，共分散行列が未知の正規分布モデルへの拡張や平均行列の推定問題，尺度パラメータの同時推定問題など，数多くの成果が得られている。また，James-Stein 縮小推定量は階層ベイズ推定量や経験ベイズ推定量として特徴づけられるが，ベイズ推定法は一般に小標本のときに安定した推定量を提供するため，多岐にわたるデータの解析において有効と考えられている。実際，野球選手の打撃成績の予測研究，胃がんの都市別死亡率の推定への応用などの先行研究があり，実用に十分耐えうることが示されている。

(4) 高次元かつ小標本を仮定した多変量正規分布モデルに関連する分布理論は 1990 年代から徐々に研究されるようになった。また，高次元・小標本を前提にした多変量正規分布モデルの縮小型推定法の研究に関しては，2010 年前後から，共分散行列（精度行列）の推定問題における縮小型推定量の導出と応用が国内外の研究者によって議論され始め，平均ベクトルや平均行列の縮小推定における成果も得られ始めていた。

(5) 高次元正規分布モデルの縮小型推定法の先行研究で基礎となっている技法は，従来からある部分積分法を利用してリスク関数を評価する方法であり，非正則ウィシャート行列を用いた推測手法となっていた。この手法は上記の先行研究以外の高次元モデルにおける様々な推定問題にも応用可能と予想されるが，先行研究以外の縮小推定理論に関する結果は知られていなかった。また，高次元小標本モデルのベイズ推定に関する統計的決定理論の観点からの議論はほとんど存在せず，高次元小標本モデルにおけるベイズ推定理論，特に統計的決定理論の観点からのベイズ推定は，さらなる努力と成果の積み重ねが待たれるところとなっていた。

２．研究の目的

以上の学術的背景及び本研究代表者の研究成果を踏まえ，非正則ウィシャート行列を利用する統計的高次元モデルのベイズ推定法や縮小推定法に関連して，以下の問題の解決・発展を目的とした。

(1) 高次元・小標本の多変量（行列型）正規分布モデルを特徴づけるパラメータである平均ベクトルや平均行列，共分散行列に関する縮小型推定法やベイズ推定法の開発

(2) 予測確率密度関数のベイズ推定量の統計的決定理論に関連する性質の解明。統計的決定論の観点から良い性質をもつベイズ推定量を得るための，事前分布の特徴づけ

(3) モンテカルロ・シミュレーションや実データへの応用による，縮小型推定法やベイズ推定法と，標準的な推定法である最尤推定法との数値的比較検討。また，データが欠損している場合や外れ値を含む場合，測定誤差を含む場合などの高次元小標本問題に関連する様々な統計モデルへの理論と応用の展開

３．研究の方法

本研究の目的を達成するために，以下のようなアプローチをとった。

(1) 先行研究の調査，問題点の整理，準備

① 株価データや地球観測データ，ゲノム関連データなどについて，統計的高次元モデルを用いたデータ解析の実例があるので，それらの中から標準的な実データをいくつか選択して，高次元小標本データの特徴抽出を行った。また，モンテカルロ・シミュレーションによる検証もおこない，先行研究で提案されている推定手法の誤った使用による危険性の評価なども検討した。

② 先行研究における理論的な成果について調査・整理し，その成果に至るまでにどのような数学的・統計学的な理論の展開がなされているか調べた。

③ 事後平均のようなベイズ推定量は尤度関

数と事前分布の密度関数の多重積分で表現されることになり，高次元になればなるほど，その理論的な取扱いは困難なものとなる。そこで，その困難さを回避し，最適性をもつ推定量の導出についての方向性を考えるために，ベイズ推定量及び縮小推定量の数学的な構造を考察した。

④ 高次元モデルの下で統計的決定理論の観点からの最適性を有する推定量を構築する上での予想される問題点の整理を行った。また，データ解析へ応用する際に予想される問題点を検討した。

(2) 新たな推定手法の開発と数値的検証

① 数学・理論統計学の知識を援用しながら，高次元モデルのパラメータに関する，ミニマックス性や許容性のような最適性をもつ新たなベイズ推定手法や縮小推定手法の開発及び改良に取り掛かった。

② 本研究で確立された新たな縮小型推定法やベイズ推定手法について，モンテカルロ・シミュレーションを用いた数値的な比較検討を行った。また，ベイズ推定値を算出する際に問題となる，アルゴリズムの収束可能性や効率性，解の安定性などの観点からの考察も行った。

(3) 実データへの応用と手法の改良

① 新たな推定手法の性能評価と，先行研究における成果との比較検証のために，先行研究を参考にして，本研究で得られた推定手法の実データへの応用を試みた。

② モンテカルロ・シミュレーションや実データへの応用した結果から，本研究で得られた新しい推定手法のさらなる改良方法を探った。また，それらの結果から，統計的モデルの仮定や事前分布の設定，損失関数の選択なども含めて推定法の構築と運用方法について再考察した。

４．研究成果

(1) 従来の低次元モデルにおける共分散行列の推測問題では，共分散行列について正則性を仮定して理論や応用についての研究がなされてきた。正則な共分散行列の推定問題に対して統計的決定理論の観点からのアプローチをとる場合，推定量の性能評価のためにスタイン型のエントロピー損失関数がしばしば利用されてきたが，このスタイン型損失関数は，共分散行列とその推定量に対して正則性を前提としている。高次元モデルや，共分散行列の正則性を仮定しないモデルの下ではウィシャート行列がランク落ちする場合があるため，不偏推定量のような通常利用される推定量も同様にランク落ちし，推定量の性能評価のためにスタイン型損失関数を利用することができなかった。

そこで，高次元モデルや非正則な共分散行列をともなうモデルの下で非正則なウィシャート行列を推定問題に利用するために，スタイン型の損失関数を高次元モデルの下でも利用できるように一般化した。さらに，リスク関数の評価方法も工夫し，低次元モデルの下で得られている推測手法を高次元モデルの解析に応用するための方法を考案した。この方法より，様々な推定手法のリスク関数の比較ができるようになり，推定精度の高い手法の確立に成功した。

また，株価データを用いた最適なポートフォリオ・ウェイトの推定への応用を行い，共分散行列の改良型推定量を流用して精度の高いウェイト推定量が得られることが確認された。

(2) 多変量正規分布モデルの共分散行列の推測問題における代表的な推定量の改良方法の一つとして，補助統計量に含まれる共分散行列の情報を引き出して，推定量を改良するという技法が提案されている。この方法によって，低次元正規分布モデルの下では，不偏推定量や三角不変なミニマックス推定量を改良する優れた性能をもつ推定量がいくつか提案されてきた。しかし，高次元正規分布モデルにおける共分散行列の推定問題では補助統計量に含まれる情報を活用する方法は考案されていなかった。

補助統計量として，位置母数に関連する統計量が利用できる場合の高次元正規分布モデルを本研究の研究対象とした。推定量の性能評価のための損失関数として，研究成果(1)で考案したスタイン型損失関数を利用し，共分散行列の推定量の改良を考えた。非正則ウィシャート分布に関連する部分積分公式などを利用して，パラメータの次元や標本サイズの様々な大小関係の下で統一的に扱えるか試してみた。その結果，高次元モデルの下でも，補助統計量を利用して推定精度を高めることができて，しかもある程度までは統一的に扱えることがわかった。また，数値実験を通じて，性能の優れた推定量が構築できることを確認した。

また，2 乗損失関数においても，スタイン型損失関数の下での結果と同様な改良が可能であることが確認された。本研究では，次元数や標本サイズに関する完全に統一的な取り扱いは出来なかったため，この点については未解決のまま残されることとなった。

(3) 多変量解析の分野では，これまでに，行列型正規分布に従う確率行列に対する QR 分解（岩澤分解）や特異値分解，ウィシャート行列に対するバートレット分解（コレスキー分解）やスペクトル分解など様々な行列の分解法が利用され，多変量確率分布に基づく分布理論や統計モデル理論，統計的推測理論，確率演算アルゴリズムなどの発展において非常に重要な役割を演じてきた。本研究では，共分散行列の推定問題に対して岩澤分解・バートレット分解の応用を試みた。

共分散推定の推定問題をスタイン型損失

関数の下で議論するとき，共分散行列やその推定量（十分統計量）に対して岩澤分解（バートレット分解）を繰り返し施す。こうすると，元の共分散行列の推定問題は，複数個の平均ベクトルと分散の同時推定問題に分解できることがわかった。この結果から，平均ベクトルと分散に関する既存の改良型推定量を使って，新しい共分散行列の推定量を構成することに成功した。

共分散行列を分解することによって得られる複数個の平均ベクトルと分散は，共分散行列のパラメータ変換の結果と捉えることができる。このパラメータ変換と同様な変換をウィシャート行列に適用したとき，その成分の同時モデルは一般化された自己回帰モデルや移動平均モデルと解釈できることも分かった。また，データが欠損している場合の共分散行列の推定問題において，情報を回復し推定量の精度を高めることも可能であるなど，様々な方面への応用が可能なのではないかと思われた。

(4) 予測密度関数のベイズ推定問題は，近年，統計的決定理論の分野で注目されており，日本人による貢献の大きい分野の一つと言える。先行研究では許容性やミニマックス性といった性質をもつベイズ予測が活発に議論されており，モデルの拡張など研究の広がりを見せている。

先行研究での成果の多くは，小次元かつ大標本を前提とした，ベクトル型正規分布に関連したものである。本研究では，行列型配列を成す大規模データへの活用を考慮し，その基礎となるであろう行列型正規分布モデルに対するベイズ予測を議論した。モデルにおける分散構造は既知と仮定し，予測密度の精度を評価するためにカルバック・ライブラー情報量を損失関数として議論を進めた。また，階層的事前分布は，あるハイパーパラメータで特徴づけられた行列型正規分布を第１段階事前分布とし，第２段階事前分布，すなわちハイパーパラメータの事前分布として多変量ベータ分布を用いた。

最大の成果は，上記の階層モデルから，許容的かつミニマックスな予測密度関数の導出に成功したことである。この結果は行列型平均の推定問題にも関連しており，そこで議論した方法論はより複雑かつ現代的な統計的モデルへの拡張の可能性を示唆するものと考えている。

５．主な発表論文等

〔雑誌論文〕（計 7 件）

① Tsukuma, H. and Kubokawa, T. (2017). Proper Bayes and minimax predictive densities related to estimation of a normal mean matrix, Journal of Multivariate Analysis, 159, 138-150. 査読有.
DOI: 10.1016/j.jmva.2017.05.004

② Tsukuma, H. (2016). Estimation of a high-dimensional covariance matrix with the Stein loss, Journal of Multivariate Analysis, 148, 1-17. 査読有.
DOI: 10.1016/j.jmva.2016.02.012

③ Tsukuma, H. (2016). Minimax estimation of a normal covariance matrix with the partial Iwasawa decomposition, Journal of Multivariate Analysis, 145, 190-207. 査読有.
DOI: 10.1016/j.jmva.2015.12.013

④ Tsukuma, H. and Kubokawa, T. (2016). Unified improvements in estimation of a normal covariance matrix in high and low dimensions, Journal of Multivariate Analysis, 143, 233-248. 査読有.
DOI: 10.1016/j.jmva.2015.09.016

⑤ Tsukuma, H. and Kubokawa, T. (2015). Minimaxity in estimation of restricted and non-restricted scale parameter matrices, Annals of the Institute of Statistical Mathematics, 67, 261-285. 査読有.
DOI: 10.1007/s10463-014-0449-x

⑥ Tsukuma, H. and Kubokawa, T. (2015). A unified approach to estimating a normal mean matrix in high and low dimensions, Journal of Multivariate Analysis, 139, 312-328. 査読有.
DOI: 10.1016/j.jmva.2015.04.003

⑦ Tsukuma, H. and Kubokawa, T. (2015). Estimation of the mean vector in a singular multivariate normal distribution, Journal of Multivariate Analysis, 140, 245-258. 査読有.
DOI: 10.1016/j.jmva.2015.05.016

〔その他〕

ホームページ等

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・准教授
研究者番号：50424685

(2)研究分担者

該当なし