

令和 5 年 6 月 5 日現在

機関番号：32661

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K11201

研究課題名（和文）ベイズ法を利用した多変量推測統計の新たな理論展開と応用に関する研究

研究課題名（英文）Study on modern Bayesian methods and applications in multivariate statistical analysis

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） 2,200,000円

研究成果の概要（和文）：独立変数に測定誤差を含む線形モデルの回帰係数の推定問題において、許容的推定量を導出し、最小二乗推定量の改良方法をいくつか提案した。多変量歪正規分布モデルにおける共分散行列の推定問題において、最良一般尺度不変推定量を改良する推定量を発見した。行列型正規分布モデルの平均行列や共分散行列に関する小標本の下での推定問題を再考し、楕円型分布モデルや成長曲線モデルの推定問題への応用について研究した。

研究成果の学術的意義や社会的意義

本研究の成果は統計的決定理論の発展やベイズ統計学の理論的な広がりにも寄与し、現代的な推測問題における統計学的手法の実用化を支える基礎理論の補完および改良にも役立つと想像される。また、本研究は回帰分析や分散分析などの古典的な多変量データ解析に応用され得るだけでなく、高次元小標本データに代表される現代的な多変量データの解析への応用の可能性も秘めている。

研究成果の概要（英文）：In estimation of slope parameter in linear model with measurement error, an admissible estimator is derived and some methods are proposed for improvement on least squares estimator. In estimation of covariance matrix in multivariate skew-normal distribution model, some estimators are given for improving the best scale-invariant estimator. We revisit the problem of estimating mean matrix and covariance matrix in matrix-variate normal distribution model with small sample setting, and their prior works are applied to estimation in elliptically contoured distribution model and growth curve model.

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 ベイズ推論 高次元モデル

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

(1) 統計学が解決すべき現代的な推測問題の代表例として、極端に多くの属性をもつデータで、かつデータの個数が非常に少ないときのデータ解析がある。このようなデータは高次元小標本データ、または単に高次元データと呼ばれている。高次元データに対して統計モデルを考えると、属性の個数を p 、データの個数を n とすると、 $n < p$ という大小関係が成り立つ。このことから高次元データによる推測問題は「 $n < p$ 問題」と呼ばれることもある。 $n < p$ 問題は、近年、活発に研究されており、高次元データを効率よく分析するための統計学的手法が数多く開発されるようになった。開発された統計学的手法の理論的な正当化については、不偏性や一致性などの基準に拠ることが多い。しかし、不偏性や一致性は推定量の精度や収束速度の観点から必ずしも好ましいと言えない場合もある。このため、不偏性や一致性以外の基準による新しい統計学的手法の開発が待たれていた。

(2) 現代的な推測問題における統計学的手法の理論的正当化のために、統計的決定理論に関連する基準を利用するのの一つの方法と考えられる。統計的決定理論は 1950 年前後から研究されてきた分野であり、その興味深い成果として正規分布モデルの推測問題に関するものが挙げられる。分散が 1 で、未知の平均をもつ正規分布が複数あるとき、その複数の平均を同時に推定する問題を、平均 2 乗誤差の和による損失関数の下で考える。この推定問題は、共分散が単位行列で未知の平均ベクトルをもつ多変量正規分布モデルにおける平均ベクトルの推定問題を、2 乗損失関数の下で考えることと同値である。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、直交変換に関して共変推定量であるなど、推定量としての優れた性質を併せもつ。この平均ベクトルの最尤推定量について、統計的決定理論の観点からは、平均ベクトルがどのような次元であってもミニマックス性をもち、2 次元以下であれば許容的、3 次元以上のときは非許容的になるという意外な事実が発見され、1960 年頃に最尤推定量を優越する縮小型推定量が具体的に与えられた。その後、部分積分法による推定量のリスク関数の簡明な評価方法が考案され、より優れた推定量の導出や他の推定問題への拡張などが行われている。

位置母数や尺度母数に関連する $n < p$ 問題は最近になって議論されるようになった。 $n < p$ という状況の下でのベクトル型多変量正規分布の平均ベクトルの縮小型推定手法が開発され、その行列型多変量正規分布への拡張などの成果も得られている。尺度母数の $n < p$ 問題として、多変量正規分布モデルの共分散行列に関する結果もいくつか得られているが、位置母数や尺度母数以外の母数の推測問題への広がりはあまりなかった。

統計的決定理論では損失関数に依存するリスク関数から手法の優劣を評価するため、大標本低次元モデルにおける推測問題のような古典的な問題においては様々な損失関数を基に多様な推測手法が生み出されてきた。このことから現代的な問題においても、決定理論的なアプローチによって不偏性や一致性などの観点からは想像もできない推測手法の構築が期待された。また、現代的な推測問題における各手法について、多変量推測理論や統計的決定理論の観点からの詳細な考察も十分行われていると言えないのが実状であった。

(3) 推測統計において、重要な役割を担う数学的な道具としてベイズ理論がある。近年のベイズ理論の研究は非常に学際的となっており、機械学習や人工知能、金融工学、ゲーム理論など様々な学術分野への広がりを見せている。実用化も急速に進んでおり、インターネット関連技術への活用もみられる。このようなことから、ベイズ理論は技術革新の強力な道具であるといっても過言ではない。

ベイズ理論に基づく統計学的手法では母数に関する事前情報を積極的に活用するものであり、これまでに理論と応用の両面で活発に研究されてきた。統計的決定理論の分野では許容性やミニマックス性を有する推定量の構成が目標の一つとなるが、1970 年頃に多変量正規分布モデルの平均ベクトルに対して許容的ミニマックスな縮小型ベイズ推定量が発見されている。21 世紀に入ってから大きな進展がみられた統計的決定理論に関連する代表的な推測問題として、確率密度関数の予測問題(予測密度の推定問題)がある。確率密度関数の予測問題はベイズ理論と密接に関係しており、ベクトル型多変量正規分布の確率密度関数の予測問題では、推定量のミニマックス性や許容性についての議論、経験ベイズ推定量の考案、ベクトル型多変量正規分布についての成果を行列型多変量正規分布の密度推定へ拡張することなどの成果が得られていた。

ベイズ理論は多変量推測統計でも重要な役割を担っているが、現代的なベイズ決定理論の応用研究は少なく、様々な推測問題への理論展開の余地も多分に残されていた。現代的な問題と古典的な問題との関連性を考察することは未解決問題を検討する上で有益と思われるが、その関連性についてベイズ理論を利用した統計的決定理論の観点からの議論も行われていない。また、ベイズ理論に基づく推定量の数学的な構造や統計学的な性質についても詳細には明らかにされていない。本研究によってベイズ法や縮小推定法に関連した新たな知見が得られれば、統計的決定理論の発展はもちろん、ベイズ法の理論的な広がりや急速な実用化を支える基礎理論の補完および改良にも役立つと想像された。

2. 研究の目的

多変量推測統計の現代的問題において、統計的決定理論の観点からの最適性をもつ推定法の開発と実データへの応用、および未解決問題の検討について、事前情報を利用するベイズ流の推測手法を利用した解決を本研究の目的とした。具体的な研究課題として、次のものを計画した。

(1) 高次元正規分布モデルに関する新たな推定手法の開発

正規性をもつモデルでのベクトル型平均や行列型平均、共分散行列、精度行列(共分散行列の逆行列)の推定問題に対する新たなベイズ推定法や縮小推定法の開発や、先行研究で得られている成果の拡張を行う。

(2) 非正規モデルや、複雑な統計モデルに関する推測手法の開発

楕円型分布族や指数型分布族のような非正規モデルでの推測問題や、複雑な構造をともなう母数をもつ統計モデルについての推測問題を広く扱い、新たな推測手法を開発する。

(3) 予測確率密度関数の推定問題についての研究

楕円型分布族のような非正規分布モデルにおいて、高次元小標本データによる予測確率密度関数の推定手法を探索する。

(4) 多変量推測統計に関する未解決問題の再考や発掘

多変量推測統計や統計的決定理論において、古くから知られている未解決問題を再考する。古典的な未解決問題と、現代的な推測問題との数学的な構造の関連性を明らかにする。

(5) 高次元モデルに関連する諸問題の解決

母数空間が制約された場合の推定問題や、階層的な構造を持つ統計モデルのような高次元モデルに関連する話題についても研究対象とした。

3. 研究の方法

本研究では理論と応用の両面から推測問題を扱うが、いずれの場合も基本的または個別的な問題から研究を始め、可能であれば手法の統一化や理論の体系化といった一般的な成果につなげていくことも意図している。理論研究に対しては、確率論や代数学など数学の諸分野の結果を援用して研究を進めた。また、応用研究では、文献にある標準的なデータから考察を始め、最新の官庁統計や金融データなどインターネット上のオープンデータを用いて研究を行った。具体的な研究の方法は以下の通りである。

(1) 先行研究の調査、問題点の整理、準備

先行研究には高次元モデルを用いたデータ解析の実例が数多く存在する。それらの中から標準的な実データを選び、高次元小標本モデルに関する問題点を抽出した。また、先行研究で提案されている推測手法について、モンテカルロ・シミュレーションによる検証もおこない、推測手法の特徴も調査した。

先行研究における理論的な成果について、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。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

数学・理論統計学の知識を援用しながら、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。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。

(3) 実データへの応用と手法の改良

本研究で新たに得られた推定法の性能評価、および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、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。モンテカルロ・シミュレーションや実データへ適用した検証結果から、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。

4. 研究成果

(1) 多変量回帰モデルにおいて統計量の縮約を行うと、回帰係数行列に関する十分統計量と共分散行列に関する十分統計量に分けられる。この2つの十分統計量は独立であり、回帰係数行列に関する十分統計量は回帰係数行列を平均とする行列型正規分布、共分散行列に関する十分統計量はウィシャート分布にそれぞれ従う。多変量回帰モデルは単純な平均・分散モデルを含み、多変量分散分析モデルや成長曲線モデルなど、実用の上で重要なモデルも含んでいる。このような背景から、行列型正規分布モデルにおける推測問題は古くから研究の対象とされており、文献においては縮小型推定による研究成果も数多く見受けられる。この行列型正規分布モデルについて、特に平均行列の推定問題に関する再考を行った。

行列型正規分布モデルの平均行列の推定問題は、これまで平均行列のサイズやサンプルサイズの大小関係によって個別に扱われてきたが、これを統一的に扱う技法が最近発見された。この統一的技法を利用して、行列値をとる2乗損失関数に関する縮小推定法を議論し、最尤推定量を改良するための条件を具体的に与えた。また、楕円型分布モデルへの拡張を考え、行列型正規分布モデルの下での統一的推定手法が楕円型分布モデルでもそのまま利用できることを確認した。さらに、成長曲線モデルにおける回帰係数行列に関して、実数値をとる2乗損失関数の下での推定問題を扱い、モデルのバイリニアな構造を利用して、行列型正規分布モデルの下での統一的推定手法が適用可能であることが確かめられた。

(2) 独立変数に測定誤差を含む線形回帰モデルにおいて回帰係数や切片パラメータの推定問題を中心に扱い、いくつかの成果が得られた。具体的には、まず、平均2乗誤差をリスク関数としたときの許容的推定量の導出に成功したことが挙げられる。この導出には、ある事前分布に関する事後平均について、リスクの有限性を示すことから接近した。研究成果の2つ目は、ある推定量の族に関する平均2乗誤差の改良である。最小2乗推定量を含む広い推定量の族を考え、この推定量の族について平均2乗誤差の基準の下で改良する方法を統一的に確立することができた。

上記以外の成果としては、測定誤差モデルにおける最小2乗推定量のバイアスを補正する方法を考案したことが挙げられる。測定誤差を伴う線形回帰モデルでの推定問題は最尤推定法に関連した議論が多いが、最尤推定量は有限な積率を持たず、バイアスや平均2乗誤差についての議論ができない。そこで、積率が存在する最小2乗推定量に着目し、その構造に明らかにして、マクローリン展開を利用したバイアスの改良を考えた。高次のバイアス補正についても成功しており、最尤法のような大標本を前提にした手法では難しい、小標本の下での推定精度の改良法について提案することができた。

(3) 近年注目されている多変量歪正規分布モデルにおける共分散行列の推定問題を考察した。歪正規分布は非対称正規分布などとも呼ばれ、正規分布に歪みを表す母数を付加した確率分布として特徴づけられる。多変量歪正規分布モデルでの共分散行列の推定は局外母数の存在によって不偏推定量や定数リスクをもつ不変推定量の構成が難しく、これまでに十分な研究がなされていなかった。この推定問題について、本研究ではエントロピー型損失関数の下で議論し、いくつかの成果を得ることができた。

対称な多変量正規分布モデルの下では、一般の尺度変換に関する最良不変推定量は三角行列に関する尺度変換に関する最良不変推定量によって一様に改良されることが知られている。これに対して、多変量歪正規分布モデルでは必ずしも最良三角不変推定量が最良一般尺度不変推定量を改良するとは限らないことを示した。具体的には、多変量歪正規分布が大きく歪んでいるときに最良三角不変推定量が最良一般尺度不変推定量より良くないことが理論的に確認できた。また、共分散行列の最良一般尺度不変推定量を一様に改良する方法についても議論した。最良一般尺度不変推定量は上方バイアスをもち、このことは縮小型推定量で改良されることを示唆している。実際、経験ベイズ法による縮小型推定量や、標本平均に含まれる情報を利用した縮小型推定によって改良されることを証明した。

多変量歪正規分布モデルにおける共分散行列の逆行列、すなわち精度行列の推定問題も考察し、精度行列に関する最良一般尺度不変推定量が拡大型推定量によって改良されることをエントロピー型損失関数の下で示した。多変量歪正規分布モデルの共分散行列や精度行列の推定問題は非心ウィシャート分布の尺度行列の推定問題に置き換えられることも示しており、他の推定問題への拡張が可能であることも確認できた。

(4) 共分散が既知の多変量正規分布モデルの平均ベクトルの推定問題において、ミニマックス性を有するジェームズ・スタイン推定量（以下、JS推定量）のリスクの改良は2乗損失関数のもとで議論することが多い。具体的な改良型推定量としては過度な縮小を避けるために考えられた打ち切り型推定量や、階層型ベイズ推定量などがこれまでに提案されている。

JS推定量を改良するために、ここでは直交変換について不変な推定量の族に着目することにした。この推定量の族の中で、JS推定量のバイアスを補正することを意図し、JS推定量を拡大するような推定量に注目した。2乗損失関数に関するリスクをポアソン変量の期待値で表して行う評価法によって、JS推定量のリスクを改良する簡便な推定量を発見することができた。その改良型推定量のリスクに関する数値実験を行い、JS推定量や打ち切り推定量などのリスクを減じることが示された。

拡大型推定量についてはJS推定量より分散が大きくなることや、非許容的であることなどの欠点があるものの、平均ベクトルの次元が大きいときに有効な改良方法と思われた。また、拡大型推定量について現代的な高次元小標本の多変量正規分布モデルへの適用は今後の課題として残されたが、共分散が未知の場合における古典的な小次元大標本の多変量正規分布モデルでは拡大型推定量でJS推定量の改良が可能であることが確認できた。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Tsukuma Hisayuki, Kubokawa Tatsuya	4. 巻 49
2. 論文標題 Estimation of a covariance matrix in multivariate skew-normal distribution	5. 発行年 2020年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 1174 ~ 1200
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/03610926.2018.1554137	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tsukuma Hisayuki	4. 巻 2
2. 論文標題 Exact finite-sample bias and MSE reduction in a simple linear regression model with measurement error	5. 発行年 2019年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 1-29
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s42081-018-0025-3	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計0件

〔図書〕 計1件

1. 著者名 Tsukuma, Hisayuki, Kubokawa, Tatsuya	4. 発行年 2020年
2. 出版社 Springer	5. 総ページ数 112
3. 書名 Shrinkage Estimation for Mean and Covariance Matrices	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------