

平成 30 年 5 月 24 日現在

機関番号：32670

研究種目：基盤研究(C) (一般)

研究期間：2013～2017

課題番号：25330043

研究課題名(和文) 行列値統計量に関わる統計的決定理論の深化と統合

研究課題名(英文) Development on statistical inference for matrix-valued statistics

研究代表者

今野 良彦 (Yoshihiko, Konno)

日本女子大学・理学部・教授

研究者番号：00205577

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)：(1) 両側切断データは左側切断と右側切断の間においてのみデータが観測されるサンプリングである。この両側切断データのもとで、分布関数の Nonparametric maximum 推定量を self-consistency algorithm によって求める方法を提案し、漸近的な挙動を調べた。(2) サンプリングメカニズムを正しく設定しないと母集団の分布の母数の推測にたいしては偏りが生じる。この論文では、生存データ解析で重要な両側切断データを扱い、特殊指数分布族における未知母数の最尤推定量の漸近的な性質について評価を行った。

研究成果の概要(英文)：(1) Doubly truncated data consist of samples whose observed values fall between the right- and left- truncation limits. With such samples, the distribution function of interest is estimated using the nonparametric maximum likelihood estimator (NPML) that is obtained through a self-consistency algorithm.

(2) Biased sampling affects the inference for population parameters of interest if the sampling mechanism is not appropriately handled. This paper considers doubly-truncated data arising in lifetime data analysis in which samples are subject to both left- and right-truncations. To correct for the sampling bias with doubly-truncated data, maximum likelihood estimator (MLE) has been proposed under a parametric family called the special exponential family (Efron and Petrosian, in J Am Stat Assoc, 1999).

研究分野：数理統計学、多変量解析

キーワード：ウィシャート分布 行列値統計量 統計的決定理論 最尤推定量 両側切断データ

1. 研究開始当初の背景

観測変数間に相関がある多次元のデータの統計的解析の手法と理論の体系である多変量統計的解析は、1959年に刊行された T.W. Anderson による成書(最新版は3版)を基礎として、その手法の体系と理論は精緻化や一般化が行われてきた。一方、計算機の計算能力やデータ収集・蓄積能力の飛躍的な向上により、標本数よりも観測変量の次元が大きなデータ(いわゆる高次元データ)や T.W. Anderson の著書にある古典的なアプローチ(古典的なアプローチという)を超えた複雑な構造を持つデータに対する統計手法が様々なアプローチから提案されている。これらのデータに対する統計手法の導出とその最適理論を考える上で、古典的なアプローチでは未解決な本質的な困難が存在する。データ間の相関構造を推測するための統計量である経験分散共分散行列が特異(その行列式がゼロになり、逆行列を持たない)となることである。さらに、スパースな大規模母数行列をいかに adaptive に推測するかという点である。これらの困難を解決するための統計手法の構築と手法の精度評価を目的とした研究は、現代統計科学のひとつの大きな流れとなっている。さらに、Compressed sensing や画像認識といった観点からも注目されている。本研究では、近年の計算機環境の劇的な変化を見据えつつ、分野横断的な数理理論を援用し、行列値統計量に関わる統計的決定理論の深化を目指す。

2. 研究の目的

研究目的(概要)に上げた(1)から(5)までの各項目をどこまであきらかにしようとするかは以下のとおりである。

(1) 等質錐上のウィシャート分布の推測離論: 多変量解析の理論において中心的役割を果たすウィシャート分布は正値対称行列の空間に値を取る分布である。この分布は、Wishart(1928, Biometrika)による確率密度関数の一般形の導出に始まり、ランダム行列理論の重要なモデルでもある。等質錐(推移的な作用群をもつ凸錐体)に値を取る分布としてウィシャート分布族を Graczyk and Ishi は非常に見通しよく拡張されている。彼らの理論の本質は、Euclid 空間から等質錐がなす空間への2次写像による多変量正規分布の像としてウィシャート分布族を定義し、等質錐の双対錐によりウィシャート分布族を母数付けしたことにある。結果として、誘導された標本空間(統計量が値を取る空間)と母数空間(直観的には、多変量正規分布の分散共分散行列の逆行列の成す空間)との間の双対構造から自然に導入される群がそれぞれの空間に推移的に作用する。これは、甘利により提唱された情報幾何で論じられる構造と同じものであり、自然指数分布族の観点とも同じである。彼らのウィシャート分布モ

デルは、ガウジアングラフィカルモデルやモノトーンパターン欠損値データに対する多変量正規モデルなど見かけ上は異なるモデルを統一的に扱う枠組みを提示する。そこで、古典的なアプローチでは、ウィシャート分布の母数に関わり個別に扱われてきた統計的推測理論を統一的な観点から研究を進めることを目指す。すなわち、尤度比検定統計量の漸近分布論、不変推定法、ベイズ推定法などを統一的に論ずるための一般化された多変量解析推測理論の新たな枠組み構築することを目指す。

(2) 等質錐上の新たな統計モデル: Graczyk and Ishi の理論は多変量正規分布の測度に基づくものである。統計手法の頑健性を考察する上で、正規分布の混合分布の確率測度の2次写像の像として得られる統計モデルは重要である。このような観点から等質錐上の多変量統計モデルの構築を目指す。

(3) 多様体上の統計モデル: 複素平面上のメビウス変換により、実数上に分布する t 分布から円周上の分布族が定義できる。これは実数上で定義された確率測度を変換で誘導するという考え方に基づいて行われている。この観点からみると課題(1)と(2)の研究の着想が、コンパクトな多様体(円周は端的な例)上の確率分布族の導出に重要な役割を果たすことが期待される。円周上の確率分布族の多変量バージョンへの拡張を調和解析学の観点から目指した。さらに、拡張されたウィシャート分布族の理論に基づき、多様体上の統計モデルの統一的な研究を目指す。

(4) 低ランク行列母数の推定問題: 多変量正規分布の未知母数(平均ベクトルや分散共分散行列など)の推定問題は基本的で重要である。観測データに基づいて未知母数を推定する基本的な方法是最尤推定法である。一般に、最尤法は漸近的有効性を持つことが知られている。しかし、標本数を有限としたときに、平均自乗誤差などの自然な危険関数のもとで推定問題を考えると最尤推定量は非許容的であり、最尤推定量は最適でないことが知られている。高次元データの解析に有用な多変量回帰モデルにおいて、回帰係数行列がスパース(低ランク)であるという仮定のもとで回帰係数行列の推定理論を扱うことは重要な問題である。すなわち、ノイズにより観測される行列値統計量はフルランクであるが、低ランクの回帰係数行列をより高い精度でどのように推定するかという問題を統計的決定問題の枠組みで考えることである。Stein の不偏リスク推定(SURE 法)を一般化することにより、スパース(低ランク)な回帰係数行列の adaptive な推定手法を構築し、

その理論的な性質を検討する。さらに、考案した手法の精度の数値的な評価も行う。

- (5) 経験分散共分散行列が特異な場合の多変量正規分布の平均ベクトルの縮小推定法：特異なウィシャート行列に関する部分積分の公式を Konno は導出した。これにより、経験分散共分散行列の一般化逆行列の期待値の評価が可能になった。これは、実多変量正規モデルだけでなく複素多変量正規分布モデルも含む結果である。この結果を用いて、経験分散共分散行列が特異な場合の実および複素多変量正規分布の平均ベクトルの縮小推定量を導出することを目指す。さらに、高次元データに基づく判別分析や共通平均問題等の関連した問題に対しても縮小推定量の導出を試みる。

3. 研究の方法

独創性の高い共著研究論文の評価の確立した国際的な学術雑誌に発表するために、専門性の高い研究者(連携研究者と研究協力者)の知見の交換と研究討論を手段として、それぞれのアイデアを本質的に融合させて研究を進めていく。さらに、本研究に関連する最新の研究動向を的確に把握するための研究情報収集をしっかりと行うことが肝要である。本研究期間の前半は理論的な研究を中心に行なっていく。理論的な研究成果の進捗に合わせて順次数値的な検証も進めていく。国内と海外の専門家によって構成されている研究体制は、本研究目的を達成するために、必要かつ妥当なものである。本研究と関連する同研究プロジェクトおよび共著論文発表を含めた研究交流を密に行っている研究者による体制である。

4. 研究成果

(1) 両側切断データとは、右側切断と左側切断の間でデータが観測されるサンプリングである。このサンプリングメカニズムのもとで、分布関数のノンパラメトリック最尤推定量を self-consistency アルゴリズムによって求め、その漸近的な挙動について評価を行った。求めたノンパラメトリック最尤推定量は表現形式が複雑であるので、実際の計算を敢行することは難しい。その困難を解決するために、ブートストラップ法の適応が考えられる。本研究では、ノンパラメトリック推定量の漸近分散共分散行列を陽な形で求めた。さらに、区間推定法や適合度検定の手続きを提案した。

(2) 両側切断データのもとで特殊指数分布族を考え、未知母数の最尤推定量の漸近的な挙動について評価をおこなった。本研究では、古典的な極限定理を不同一独立分布の仮定のもとでの極限定理に置き換えることで、漸近分布について成果を得た。さらに、標準誤差の漸近表現を用いて、区間推定法の提案を行った。

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 2 件)

1. Emura, Takeshi, Yu-Ya-Husan., Konno Yoshihiko, Konno, Asymptotic inference for maximum likelihood estimators under the special exponential family with double-truncation. *Statistical Papers*, vol 58 (2017) No.3, 877-909..

2. Emura, Takeshi, Konno Yoshihiko, Mchima, Hirofumi, Statistical Inference based on the nonparametric maximum likelihood estimator under double-truncation, *Lifetime Data Analysis* 21(2015), No. 3, 397-418.

〔学会発表〕(計 0 件)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ等
<http://mcm-www.jwu.ac.jp/~konno/research.html>

6. 研究組織

(1) 研究代表者

今野 良彦 (KONNO, Yoshihiko)

日本女子大学・理学部・教授

研究者番号：00205577

(2) 研究分担者

()

研究者番号：

(3)連携研究者
()

研究者番号：

(4)研究協力者
江村 剛志 (EMURA, Takeshi)