

令和 5 年 6 月 15 日現在

機関番号：15201

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K03419

研究課題名（和文）大標本かつ高次元なデータに対する統計的推測法の理論研究と応用

研究課題名（英文）Theoretical research and application of statistical inference methods for large-sample, high-dimensional data

研究代表者

山田 隆行（Yamada, Takayuki）

島根大学・学術研究院理工学系・准教授

研究者番号：60510956

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：高次元データに対する多変量一般化線形モデルにおける平均パラメータの両側線形仮説検定問題の研究成果についてまず説明する。母集団に正規分布を含む一般分布を仮定したもとで、平均のL2ノルムに基づく検定統計量の確率分布について、次元数と標本サイズを共に大きくする高次元漸近枠組みにおける極限分布を導出した。また、関連研究として行った、正規母集団の完全独立性の研究成果を説明する。相関係数のL2ノルムに基づく検定統計量の帰無仮説のもとでの確率分布について、上記の高次元漸近枠組みにおいて漸近展開を導出し、その展開式を用いて検定規準の修正を与えた。

研究成果の学術的意義や社会的意義

多変量一般化線形モデルにおける平均パラメータの両側線形仮説検定の問題は平均の同質性の仮説検定問題などを特別な場合を含むなど、より一般的な仮説についての検定問題である。これについて検定規準を与えることができたことが学術的に意義がある。また完全独立性の検定の帰無仮説のもとでの漸近展開は、今まで次元数が標本サイズを超える場合に対しての結果がなかったため、そういった場合にも分布の漸近展開を与えることができると、それをういた仮説検定の補正を与えたことは学術的に意義があると考えられる。

研究成果の概要（英文）：We first describe our research results on the two-tailed linear hypothesis testing problem for the mean parameter in multivariate generalized linear models for high-dimensional data. Under the assumption of a general distribution including a normal distribution for the population, we derived the limit distribution for the probability distribution of the test statistic based on the L2 norm of the mean in a high-dimensional asymptotic framework where both the dimension and sample size goes toward infinity together. Next, we describe the results of a related study on the complete independence of normal populations. For the probability distribution under the null hypothesis of the test statistic based on the L2 norm of the correlation coefficient, we derived the asymptotic expansion in the above higher dimensional asymptotic framework and gave modifications to the test criterion using the expansion formula.

研究分野：統計的推論

キーワード：多変量解析 統計的推論 漸近論 高次元データ

## 1. 研究開始当初の背景

高次元データについての多変量統計解析の推測法(高次元推測法と呼称する)の研究は主に母集団分布に正規分布を仮定した下で始まった。これらの研究では主に、検定統計量や推定量について高次元漸近枠組み： $n$  (標本サイズ) ,  $p$  (次元数) ,  $p/n \rightarrow c \in (0, \infty)$  における漸近近似分布を導出することで行われている。例として、Bai and Saranadasa (1996, *Statistica Sinica*)では平均ベクトルについての検定を、Ledoit and Wolf (2002, *Ann. Statist.*)は共分散行列に関するいくつかの検定を扱っている。

近年、母集団分布に正規分布を含むような分布(一般化した分布と呼称する)を仮定した下での研究が行われている。その多くは、Chen and Qin (2010, *Ann. Statist.*)で扱われている母集団分布の仮定の下での成果である。しかし、その仮定は多変量  $t$  分布のような楕円分布族の仮定を含んでいない。

ここで強調したいこととして、高次元推測法の理論を体系化・拡張できないかということである。

例えば、2群の判別分析に対する推測法の理論研究は線形判別法(W判別法)について主に研究されている(Hyodo and Kubokawa [2014, *J. Multivariate Anal.*], Cai and Liu [2011, *J. American Statist. Assoc.*], Fujikoshi [2000, *J. Multivariate Anal.*]など)。一方で、最大尤度判別法(Z判別法)に関する推測法の研究成果は知られていない。この2つの判別関数を包括するものに対して推測法の理論研究を行うことで、線形判別法で得られた研究成果を体系化できる。

## 2. 研究の目的

本研究の目的は、高次元データに対する推測法の理論について体系化・拡張を行うことである。そのために、課題(A)と(B)に取り組む。

- (A) 多変量解析の推測法について、既存の方法を包括する手法の構築。
- (B) 多変量解析において、一般化された帰無仮説に対する検定法の構築。

課題(A)、(B)では、既存の問題より抽象化されたものを扱うことで、推測法の特徴を明らかにしようとしている。母集団分布に正規分布を仮定して研究を行うが、そこで得られた成果を一般化分布の下での研究に拡張する。その際に、多変量  $t$  分布のような楕円分布族も含むような一般化分布において研究を行う。

## 3. 研究の方法

以下の課題を中心に、まず母集団分布に正規分布を仮定したもとで研究を行う。その後、母集団分布を一般化した分布した下での研究に拡張する。それぞれの課題に共通していることは、

- (1) 統計量を見出す。
  - (2) 見出した統計量について高次元漸近枠組みの下での漸近近似分布を導出する。
- である。

## 4. 研究成果

### (1) 2群の判別分析に対する推測法の理論研究

2群の判別問題について、次元数が標本サイズの合計を超えない下で、線形判別法(W判別法)と最大尤度判別法(Z判別法)の誤判別確率の漸近近似を与えた研究成果を紹介する。なお本研究は母集団に正規分布を仮定した下で行った成果である。まず2つの判別関数を包括する統計量を考えた。これにより、係数の取り方で線形判別関数と最大尤度判別関数それぞれの確率分布を統一的に扱うことを可能にした。包括型の統計量について、標本サイズの合計と次元数を共に大きくする漸近枠組みの下で、統計量の分布関数の漸近展開を与えた。この漸近展開より、線形判別関数と最大尤度判別関数それぞれを用いた場合の誤判別確率の近似を与えることができる。この近似に対する推定を与えることで、誤判別確率の推定量を提案した。数値実験より、近似精度に優れることを確認した。

### (2) 高次元 GMANOVA model における一般化多変量線型仮説の検定問題

観測値ベクトルに GMANOVA モデルを仮定した場合の平均に関する一般化多変量線型仮説に対する検定問題を扱った。この問題は、パラメータや係数行列の取り方によって、平均ベクトルの検定や一元配置多変量分散分析などを扱うことができるため、いくつかの問題を総合的に扱うことができる問題である。帰無仮説からの乖離具合を平均行列に関するフロベニウスノルムで測った場合の不偏推定量を提案し、それに基づく検定統計量の極限帰無分布を、誤差ベクトルに一般化分布を仮定した下で導出した。また、対立仮説の下での漸近極限分布を導出し、提案検定統計量に基づく検定法の検出力が1に収束するような平均はどのような場合であるかを調べた。数値実験より、検定法の第1種過誤の確率の近似精度の検証および実際の検出力がどのようになるのかを検証した。

### (3) 多変量正規分布の完全独立性についての検定問題

正規母集団に対する高次元データの完全独立性の検定に対する研究について報告する。標本相関係数の2乗に基づく検定統計量について、帰無仮説が真であるという仮定の下で次元数と標本サイズが共に大きくなる漸近枠組みにおいて確率分布の漸近展開を1次の項まで導出し、その展開式を用いて検定規準の改良を行った。数値実験を通して、改良された検定規準の第1種過誤の確率の近似精度が、従来の極限分布に基づく検定規準の近似精度より良くなる事を検証した。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 Yamada, T., Himeno, T., Tillander, A. and Pavlenko, T.	4. 巻 -
2. 論文標題 Test for mean matrix in GMANOVA model under heteroscedasticity and non-normality for high-dimensional data	5. 発行年 2023年
3. 雑誌名 Theory of Probability and Mathematical Statistics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 該当する

1. 著者名 Yamada Takayuki	4. 巻 -
2. 論文標題 High-dimensional asymptotic expansion of the null distribution for Schott's test statistic for complete independence of normal random variables	5. 発行年 2022年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 1~17
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/03610926.2022.2094414	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 Yamada Takayuki, Sakurai Tetsuro, Fujikoshi Yasunori	4. 巻 51
2. 論文標題 High-dimensional asymptotic results for EPMCs of W- and Z- rules	5. 発行年 2020年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 2385 ~ 2413
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/03610926.2020.1774060	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 -

〔学会発表〕 計7件（うち招待講演 1件/うち国際学会 2件）

1. 発表者名 玉谷充、山田隆行
2. 発表標題 正規母集団から得られた高次元データの完全独立性に対するMaoの検定の一般化
3. 学会等名 日本計算機統計学会 第35回大会
4. 発表年 2021年

1. 発表者名 玉谷充、山田隆行
2. 発表標題 高次元小標本における特徴選択手法について
3. 学会等名 日本分類学会 2021年度シンポジウム
4. 発表年 2021年

1. 発表者名 山田隆行、姫野哲人
2. 発表標題 高次元非正規母集団の下での平均に対する検定統計量の Edgeworth 展開
3. 学会等名 統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 Takayuki Tamada, Tetsuto Himeno
2. 発表標題 Testing identity and sphericity for covariance matrix for high-dimensional data under general distribution
3. 学会等名 simstat2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Takayuki Yamada, Tetsuro Sakurai, Yasunori Fujikoshi
2. 発表標題 Asymptotic robustness for error rate of 2 group discriminant analysis for large dimensional case
3. 学会等名 CMStatistics2019 (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 山田隆行、姫野哲人
2. 発表標題 一般化された母集団分布からの大標本高次元データに基づく共分散行列の構造についての検定
3. 学会等名 2018年度 統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 山田隆行、櫻井哲朗、藤越康祝
2. 発表標題 2群の線形判別法に関する誤判別確率の高次元漸近ロバストネスについて
3. 学会等名 2018年度 統計関連学会連合大会
4. 発表年 2018年

〔図書〕 計0件

〔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