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　２０　日現在

機関番号：　１７７０１
研究種目：　若手研究(B)
研究期間：　2014～2017
課題番号：　２６８０００８８

研究課題名（和文）高次元データの多変量解析についての統計的推測に関する研究とその応用

研究課題名（英文）Multivariate statistical inference for high-dimensional data and its application

研究代表者
山田　隆行（YAMADA, Takayuki）

鹿児島大学・共通教育センター・講師

研究者番号：６０５１０９５６

交付決定額（研究期間全体）：（直接経費）　　2,800,000円

研究成果の概要（和文）：まず、母集団分布に正規分布を仮定してよいかを調べる問題について、高次元データに適用できる正規性の検定を多変量3次モーメントの推定量に基づき提案した。
２つめの結果は、伝統的な2群の線形判別法について次元数が標本サイズより小さいが、共に大きい場合の推測論に基づくものである。2群の判別には2種類の誤判別が存在するがその確率は判別境界の位置に依存する。母集団に正規分布を仮定した場合に，片方の誤判別の確率を事前に設定した水準となるようにするための判別境界点を与えた。

研究成果の概要（英文）：Firstly, we propose an estimate of multivariate 3rd moment which is well defined for the case that the dimensionality of the observation vector is larger than the sample size. As an application, we apply to testing the multivariate normality.

Secondary, we propose a cut-off point for the classical linear discriminant rule in 2 groups which one of two types of expected probability of misclassification takes pre-setting level. It is derived by the asymptotic distribution for the studentized linear discriminant function under the assumption that the population has multivariate normal distribution. The asymptotic distribution which we dealt is under the high-dimensional asymptotic framework that the dimension and the sample size go to infinity together while the ratio of the dimension to the sample size converges to a constant in [0,1).

研究分野：　数理統計学

キーワード：　多変量解析　高次元データ　漸近理論

１．研究開始当初の背景

多変量解析における統計的推測は，伝統的には母集団分布に多変量正規分布を仮定した下で発展・体系化してきた．この多くは観測項目の数(次元数)が 10 程度で標本サイズをそれ以上確保できる場合を前提とした話である．しかし，近年の科学技術の急速な発展から，多くの観測項目からなる多次元データ(高次元データとよぶこととする)を採取・蓄積が可能となり，特に観測項目の数が標本サイズより大きい高次元データ(例えば，遺伝子研究におけるマイクロアレイデータや量的形質座位の解析，画像データの解析)の統計解析の必要性が高まっている．しかし高次元データに対しては伝統的な多変量解析はほとんど使うことができない．

このような状況の中，高次元データについての多変量解析の統計的推測(高次元推測とよぶこととする)の研究は行われてきた．母集団分布に多変量正規分布を仮定したもとで行われてきた．実際に解析を行う場合，データが正規分布に従うように事前処理を行う．しかし，中には正規性を許容できないものがある．実際には正規分布に従っていないデータに正規性の仮定のもとで提案された手法を適用した場合，推定量が外れ値の影響を受けたり，検定のサイズが非常に大きくなったりと深刻な問題を引き起こす．

これを解決するために検定法の補正が行われてきた．母集団分布に正規分布を仮定したもとで有効とされる検定が正規分布を含むような分布族に対しても有効で在ることを示す分布に対する頑健性の研究も進展している．

　このような背景を踏まえて，より広い範囲で使える解析手法の開発，およびその推測理論の研究は重要な課題である．

２．研究の目的

　研究の背景を踏まえ，本研究で扱う高次元データについての多変量解析の統計的推測の問題としては，

(1) 分布に対して頑健な推定・検定，

(2) 母集団分布の正規性の検定，

(3) 比較する母集団分布が異なるという仮定の下での統計推測，

に取り組む．

３．研究の方法

　本研究で扱う高次元データについての多変量解析の統計的推測の問題としては，

(1) 分布に対して頑健な推定・検定，

(2) 母集団分布の正規性の検定，

(3) 比較する母集団分布が異なるという仮定の下での統計推測，

に取り組む．

(1)に関しては，多変量正規分布を含むような広い分布族の仮定の下での検定統計量の高次元漸近分布の導出を行う．

(2)に関しては，正規分布からの違いを示すのに十分な特徴量を与え，その推定量を基に検定統計量を与える．

(3)に関しては，平均ベクトルの検定や判別問題を扱う．

以上の研究を次元数と標本サイズを共に大きくする漸近枠組み(高次元漸近枠組み)の下で漸近分布を導出することで行う．

遺伝学におけるマイクロアレイデータと量的形質座位の解析等にも応用する．

４．研究成果

・研究目的(1)の成果について

第１に多変量成長曲線モデルの下での一般化仮説検定に対する成果を報告する．ある個体の変化を多数の計測点において調べるが、調査対象の個体数が計測点の数より少ない状況を考えている．その計測値たちが計測点によって変化しないなど，計測値ベクトルについて線形で表現できる形の仮説に

対しての検定問題を扱っている。伝統的には個体数が計測点の数より大きい場合に考えられてきたが，そこでの統計量は今回の状況では定義できないものとなっている。新たな検定統計量として、Yamada and Himeno (2015, JMVA)の結果を基に検定統計量を導きし，漸近正規性を示すことで検定規準を提案している．数値実験より検定のサイズと経験検出力についての精度を調べ，実用に足るものであることを確認した．日本計算機統計学会第 31 回シンポジウムにおいて成果発表を行い，現在論文の形に纏めている最中である．

第 2 に判別分析の誤判別確率の推定問題に関係する研究成果について報告する．当初考えていた問題として，次元数は標本サイズより小さいが共に大きい場合に，2 群の線形判別分析の誤判別確率に関する推定問題とそれに関する応用問題を研究する予定であった．2 次判別関数について誤判別確率の漸近近似を導出したが，数値実験を通して検証した近似精度が悪く実用に足るものではなかったために，現在保留中である．研究期間終了間際に母集団平均ベクトルの差に 0 となるものが少ない(non-sparse)という限定的な条件下で線形判別法の誤判別確率が漸近的に母集団に正規分布を仮定したものと同じくなる（漸近的頑健性）ことを示し，数値実験を通して近似精度が悪くないことがわかった．この成果を報告することが今後の展望である．関係する周辺の研究として，母集団に正規分布を仮定した場合の 2 群の線形判別法について，2 種類の誤判別のうちの片方を事前に設定した水準となるようにするための判別境界点を導出した．数値実験の結果とともに論文としてまとめたものが学術誌に掲載されている．

・研究目的(2)の成果について

古典的な多変量の歪度の定義は、1 次元の場合の歪度 2 乗をとったものであったことに違和感をもった．そのため，本研究ではベクトルのアダマール積を使って 1 次元の歪度の自然な拡張となるような 3 次モーメントを定義した．この推定量を導出し，3 次モーメントが 0 であるかどうかの診断規準を提案した．本研究成果は実際に正規分布には従っていないものを正規分布に従っていると誤診断する危険性を含んでいるため，今後改良していく必要がある．本研究成果については論文としてまとめて投稿中である．

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] Yamada, T. Interval estimation in two-group discriminant analysis under heteroscedasticity for large dimension, Accepted in Communication in Statistics-Theory and Methods.査読有.
DOI:doi.org/10.1080/03610926.2017.1400060

[2] Yamada, T., Himeno, T. and Sakurai, T. Asymptotic cut-off point in linear discriminant rule to adjust the misclassification probability for large dimensions, Hiroshima Mathematical Journal, 査読有, 4 巻, 2017 年, 319-348, https://projecteuclid.org/euclid.hmj/1509674450

〔学会発表〕（計 3 件）

[1] 姫野哲人，山田隆行．一般化した分布の仮定の下での高次元 MANOVA 問題．統計関連学会連合大会 (2017).

[2] 山田隆行，姫野哲人．不等分散を仮定した高次元成長曲線モデルにおける一般化線形仮説の検定について．日本計算機統計学会第 31 回シンポジウム (2017).

[3] Yamada, T. and Himeno, T. Estimation of multivariate 3rd moment for

high-dimensional data and its application for testing multivariate normality. ISI2017 61st World Statistics Congress (2017).

〔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（計 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〕
特にありません

６．研究組織
(1)研究代表者
山田　隆行(YAMADA, Takayuki)
鹿児島大学・共通教育センター・講師
研究者番号：60510956