

科学研究費助成事業 研究成果報告書

令和元年6月13日現在

機関番号：14501

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K17642

研究課題名(和文)単調欠測データに基づく検定統計量の高次元大標本漸近理論とその応用

研究課題名(英文)High-Dimensional and Large-Sample Asymptotic Theory for the Test Statistics with Monotone Missing Data and Its Application

研究代表者

首藤 信通 (Shutoh, Nobumichi)

神戸大学・海事科学研究科・講師

研究者番号：00634099

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：観測対象の脱落などによって観測される単調欠測データに焦点を当て、主に欠測データメカニズム(標本ベクトルと欠測確率の同時分布の構造)に関する仮説検定、平均ベクトルに関する仮説検定に対する検定統計量に改良を与え、標本サイズがそれほど大きくない状況下でもほぼ正確に機能する仮説検定を提案した。また、平均ベクトルに関する仮説検定問題については、漸近検出力に関する考察を与えた。いずれの研究成果についても、数値実験による先行研究との比較を行い、提案手法の有用性を確認した。

研究成果の学術的意義や社会的意義

欠測データ(データセットの一部に欠損があるデータ)はさまざまな研究分野や産業分野においてしばしば観測される。特に、再測定が難しい場面で必要とされる欠測データ解析法において、欠測データメカニズム(標本ベクトルと欠測確率の同時分布の構造)や、母集団分布のパラメータに対する統計的推測は非常に重要である。本研究課題では上記の問題に対する仮説検定を改良し、数値実験により提案手法は標本サイズがそれほど大きくない状況下でも機能する仮説検定であることを確認した。また、仮説検定の性能を表す検出力についても考察を行った。これは応用例が多くみられる欠測データ解析法の改良と性能評価に寄与する結果であると考えられる。

研究成果の概要(英文)：For the statistical inference based on monotone missing data particularly observed by dropout of samples, the test statistics for the missing data mechanism and those for a mean vector were modified via Bartlett-type correction. The asymptotic power functions of the test statistics for a mean vector were also derived. Finally, the simulation results implied that the proposed methods outperformed the existing procedures in aspects of controlling type I error more exactly.

研究分野：数理統計学

キーワード：欠測データ 仮説検定 統計的漸近理論 検出力関数

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

欠測データ解析はさまざまな研究分野，産業分野において，特に再測定を行うことが容易でない状況下での応用例が多く，その理論的進展は極めて重要である．数理統計学の研究分野においては，特に観測対象の脱落によって生じる単調欠測データに対し，多変量正規母集団，大標本漸近枠組の下で研究成果が多く報告されている．

ただし，単調欠測データの場合，欠測パターン数が多くなるにつれて平均ベクトルや共分散行列の最尤推定量が多くの分割をもつようになることから，欠測データメカニズムに関する仮説検定問題などにみられるような非常に重要な仮説検定問題に対しても，理論的改良が十分になされていない状況にある．また，既に研究成果として公表されている結果の多くは次元数固定の大標本漸近枠組の下で得られたものである．

2. 研究の目的

研究開始当初の背景から，本研究課題では多変量正規母集団から得られた単調欠測データを仮定し，特に重要であると考えられる仮説検定問題を扱う．これらの仮説検定問題に対し，さまざまな漸近枠組（高次元，大標本）の下で高い精度を保つ仮説検定法の提案を行うことを本研究課題の目的とする．

本研究課題で扱う問題は，(a) 欠測データメカニズムに関する仮説検定問題，(b) 母集団分布のパラメータ（平均ベクトル，分散共分散行列）に関する仮説検定問題である．(a)，(b) に対し，検定統計量を定義し，仮定した漸近枠組の下で検定統計量の漸近帰無分布（帰無分布に対する漸近展開）を導出する．導出した漸近帰無分布の結果を利用し，Bartlett 型修正検定統計量を求めることにより，帰無仮説の下で検定統計量の極限分布への収束の高速化を図る．ただし，Bartlett 型修正検定統計量を求めることができないような場合には，Cornish-Fisher 展開による棄却限界値の改良を行う．また，局所対立仮説の下でも漸近分布の導出が可能な場合は，漸近検出力に対する考察を与える．

上記の結果により，比較的小標本の状況下でも有意水準をほぼ正確に制御可能な仮説検定方式を提案し，先行研究との比較を多くの設定の下で大規模なモンテカルロ・シミュレーションによって行い，提案手法の有用性について数値的に検証する．また，データ例への応用を目指す．

3. 研究の方法

以下の手順にしたがい，研究計画を遂行した．

- (1) 先行研究で提案された仮説検定法の問題点を確認するため，専門書，過去に公表された学術論文，学会発表等における研究動向の調査を行う．
- (2) (1)の結果を踏まえて，改良を行うべき仮説検定法に照準を定め，検定統計量の導出過程を確認する．
- (3) (2)で得られた単調欠測データに基づく検定統計量は多くの分割ベクトル，分割行列を含み，複雑な表現となるため，行列代数に関する補題を用いることにより，可能な限り検定統計量の表現を単純化する．新たに補題が必要となる場合はその証明を与える．

- (4) 過去に得られている検定統計量の漸近分布の導出方針を比較・検討し，本研究課題で扱う問題における具体的な漸近分布の導出方針を定める．
- (5) (4)で得られている漸近分布の導出方針にしたがい，(3)で表現が単純化された検定統計量に対する摂動展開を与える．漸近分布の導出に必要な確率変数ベクトル，確率変数行列の期待値に関する補題を証明する．
- (6) (5)で得られた摂動展開と期待値に関する補題を用いて，(3)で得られた検定統計量の漸近特性関数を得る．また，反転公式を利用することにより，検定統計量の漸近帰無分布または漸近検出力を求める．
- (7) (6)で得られた漸近帰無分布を用いて，検定統計量の近似上側パーセント点を与える．Bartlett 型修正検定統計量を提案する場合は，漸近分布関数の形に応じて検定統計量の極限分布への収束を高速化するような単調変換を求め，変換を施した検定統計量に基づく近似検定方式を提案する．(6)で漸近検出力を得た場合は，検定統計量の検出力に関する考察を与える．
- (8) (7)で得られている近似検定方式に対し，多くのパラメータの設定の下で大規模なモンテカルロ・シミュレーションを実行することにより，その性能について先行研究との比較を行いながら数値的評価を与える．また，数値実験の結果から得られたさらなる課題や，実際の欠測データへの応用について検討を行う．

4．研究成果

本研究課題では(a)欠測データメカニズムに関する仮説検定問題，(b)母集団分布のパラメータに関する仮説検定問題に焦点を当て，以下の研究成果を得ている．

(a) 欠測データメカニズムに関する仮説検定問題 (2016 年度)

多変量正規母集団から得られた欠測パターン数 2 の単調欠測データを仮定し，先行研究で与えられている MCAR (完全にランダムな欠測データメカニズム) の成立に関する尤度比検定統計量の漸近帰無分布の導出に成功し，Bartlett 型修正検定統計量を提案した．数値実験の結果から，標本サイズが十分に大きくない状況下でも本研究課題で提案する検定統計量の有意水準は名目上の有意水準とほぼ等しく，この点において先行研究で得られている仮説検定方式より有用であることを確認している．

(b) 母集団分布のパラメータに関する仮説検定問題 (2016 ~ 2018 年度)

平均ベクトルに関する仮説検定問題について，多変量正規母集団から得られた欠測パターン数 2 の単調欠測データ (欠測データメカニズムは MCAR を仮定) に基づいて構成される Hotelling の T^2 型検定統計量および尤度比検定統計量の漸近帰無分布の導出に成功し，Bartlett 型修正検定統計量を提案した．数値実験の結果からも(a)と同様の有用性を確認した．また，局所対立仮説の下で，Hotelling の T^2 型検定統計量，尤度比検定統計量の漸近検出力関数の導出に成功した．数値実験の結果から，多くのケースで Hotelling の T^2 型検定統計量の検出力よりも尤度比検定統計量の検出力が高いこと，その結果は概ね漸近検出力関数の大小との整合性があることがそれぞれ確認された．

当初，次元数大の漸近枠組を仮定した研究成果の導出を予定しており，その下での漸近分布の導出方針に関する基礎研究やそれらを応用した結果の導出も行っていたが，研究調査の結果，次元数大の状況下では多変量正規性の仮定は自然でないために，高次元の完全データ（欠測がないデータ）に対する近年の研究成果は非正規性の仮定の下で得られていることが判明した．そのため，高次元における議論に先んじて，非正規母集団を仮定した研究成果の導出を行う研究計画に変更した．

具体的には，楕円母集団から得られた欠測パターン数 2 の単調欠測データ（欠測データメカニズムは MCAR を仮定）の下で，多変量正規性を仮定して得られている Hotelling の T^2 型検定統計量の漸近帰無分布の導出を行った．得られた漸近帰無分布の結果と数値実験の結果から帰無分布への非正規性の影響を考察することができ，両面において当該検定統計量の有意水準が名目上の有意水準よりも過大となるか過小となるかについての傾向は，母集団分布の尖度パラメータの正負によって変化することが確認された．

なお，(a)で得られている研究成果は数理統計学および統計科学分野の海外専門誌である *Statistica Neerlandica* に出版されており，当該雑誌における 2017 年～2018 年のダウンロード数が Top 20 の学術論文となっている．また，(b)に挙げた単調欠測データの下での検定統計量の検出力に関する研究成果は先行研究において十分な議論がなされていないため，当該研究課題における研究成果として重要な結果であると考えられる．

5．主な発表論文等

〔雑誌論文〕(計 4 件)

Hyodo, M., Shutoh, N. (to appear). Asymptotic power comparison of T^2 -type test and likelihood ratio test for a mean vector based on two-step monotone missing data, to appear in *Communications in Statistics - Theory and Methods*, 査読有.

Kawasaki, T., Shutoh, N., Seo, T. (2018). On the asymptotic distribution of T^2 -type statistic with two-step monotone missing data, *Journal of Statistical Theory and Practice*, 12, 657-668, 査読有. DOI: 10.1080/15598608.2018.1450795

Shutoh, N., Nishiyama, T., Hyodo, M. (2017). Bartlett correction to the likelihood ratio test for MCAR with two-step monotone sample, *Statistica Neerlandica*, 71, 184-199, 査読有. DOI: 10.1111/stan.12106

〔学会発表〕(計 5 件)

首藤 信通, Effect of nonnormality for the null distribution of T^2 -type test statistic based on two-step monotone sample under elliptical distributions, 2018 年度統計関連学会連合大会, 2018 年 9 月 11 日, 中央大学 後樂園キャンパス.

首藤 信通, 西山 貴弘, 兵頭 昌, Bartlett correction to the likelihood ratio test for MCAR with two-step monotone missing data, 科研費シンポジウム「統計科学における数理と理論の最先端について」, 2016 年 10 月 29 日, 北海道大学大学院経済学研究科.

首藤 信通, 単調欠測データに基づく線形判別分析と仮説検定法（特別講演）, 2016 年日

本数学会秋季総合分科会，2016年9月17日，関西大学．

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

なし

6．研究組織

(1)研究分担者

なし

(2)研究協力者

なし

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、研究者個人に帰属されます。