

令和 4 年 8 月 29 日現在

機関番号：34419

研究種目：若手研究

研究期間：2019～2021

課題番号：19K14595

研究課題名(和文)非正規性をもつ単調欠測データに対する汎用的かつ高精度な統計解析法の理論研究と応用

研究課題名(英文)Studies on general and nearly exact statistical methods for monotone missing data under nonnormality and their applications

研究代表者

首藤 信通 (Shutoh, Nobumichi)

近畿大学・理工学部・准教授

研究者番号：00634099

交付決定額(研究期間全体)：(直接経費) 2,500,000円

研究成果の概要(和文)：本研究課題では、母集団分布として対称な非正規分布を含む分布族(楕円分布族)を仮定し、得られたデータが欠測している状況下で適用可能な統計解析法を考える。この設定の下で、多くのデータに適用可能かつ標本サイズが十分に得られていない状況でも高精度に機能する手法の理論的導出と応用を与えるものである。特に、観測対象が徐々に脱落するような状況であられる単調欠測データに着目し、(1) 平均ベクトルに関する仮説検定とBartlett修正、(2) (1)で得られた仮説検定の近似検出力の提案、(3) 判別分析における変量の冗長性に関する尤度比検定とその数値的考察に関する研究成果を与えた。

研究成果の学術的意義や社会的意義

本研究は、データの多様化に対応すべく、多変量正規分布および対称な非正規分布を含む分布族(楕円分布族)から得られた欠測データに対して適用可能な統計解析法を求めようとするものである。特に、平均ベクトルに対する仮説検定や、どの変量がデータ解析に必要なかを解析する冗長性検定は関心がもたれることが多いため、これらの検定問題に対する検定統計量の導出、理論的修正、および数値的考察を与えた。

研究成果の概要(英文)：In this research, assuming that missing datasets were drawn from the elliptically contoured distributions including some non-normal distributions, we considered the statistical analysis which could be applied to the datasets. In these settings, our goal was to obtain the statistical analysis which performed well even if the sample size was not so large and to consider the applications to the real datasets. In particular, we focused on missing data whose missing patterns were monotone type owing to dropouts and obtained the following results: (i) the likelihood ratio based test for a mean vector and its Bartlett-type correction, (ii) the approximated power of the test stated in (i), and (iii) the likelihood ratio test for the redundancy of the variables in linear discriminant analysis and its simulation studies.

研究分野：数理統計学

キーワード：欠測データ 楕円分布族 仮説検定 尤度比検定 判別分析

## 様式 C - 19、F - 19 - 1、Z - 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・産業分野で必要となるデータ解析である。特に医療分野において患者に対する研究を行うときは、医師の判断によって患者の検査が中止となる場合、患者が通院せず検査を行えない場合など、さまざまな原因によってデータセットの一部において欠測が生じる。また、患者への侵襲に対する倫理的配慮から再度の測定は行われない場合が多い。以上のように、データセットに欠測をもちながらデータ解析を行わざるを得ない場合があり、欠測データ解析に関する理論研究が半世紀以上にわたって行われている状況にある。

当該研究の難しさのひとつは、たとえ多変量正規母集団を仮定したとしても、欠測データに基づいて得られる最尤推定量は陽に書くことが困難である点にある。しかし、その特別な場合、すなわち、欠測データメカニズムが MCAR、かつ、標本が測定時点を経るごとに測定から脱落することによってあらわれる欠測パターンとして知られる単調欠測データの場合においては、最尤推定量を陽に書くことができる。したがって、多変量正規性をもち、欠測データメカニズムが MCAR の単調欠測データを仮定した下で研究成果が多く与えられている。なお、一般の仮定の下で得られている研究成果もあるが、計算機を援用した極限分布に基づく方法が大多数であり、正確な確率分布に基づく研究成果や、分布の修正を施すような研究成果は稀少である。

実用上の問題としては、欠測データメカニズムが MCAR でない場合や、母集団分布が対称性をもつ確率分布であったとしても多変量正規分布でない場合が多く考えられる。また、標本サイズが十分に確保されない状況も考えられるため、欠測データ解析の理論研究に関して改良の余地が多く残されていると考えられる。

### 2. 研究の目的

上記の研究開始当初の背景から、本研究では母集団分布に多変量正規分布と対称な非正規分布を含む楕円分布族を仮定し、その母集団から単調欠測データが得られている下で、特に重要と考えられる仮説検定問題を扱うこととする。また、提案する仮説検定法について標本サイズが極めて大きい状況でなくても比較的機能するよう、検定統計量に対して理論的な修正を施すことにより、少ない標本サイズでも精度を高く保つ仮説検定の導出を行うことを研究の目的とした。本研究の取り組みは、ある確率分布族に対して統一的に適用可能な研究成果の導出を目指しているものであり、欠測データ解析における理論研究に対する貢献といえる。

### 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) 専門書、過去に公表された学術論文、学会参加による研究動向調査を行う。
- (2) (1)の結果に基づき、改良を行うべき仮説検定問題を選択し、検定統計量の導出方針を確認する。
- (3) (2)において必要となるベクトルや行列に関する微分公式の導出を行う。
- (4) (3)を利用して検定統計量の導出を行う。
- (5) 行列代数に関する公式等を利用・導出し、(4)で得られた検定統計量を可能な限り単純な形で書き換える。
- (6) 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。

(7) (6)で得られた導出方針に従って,(5)で得られた検定統計量の漸近分布を求める.必要となる期待値に関する補題を示す.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.

(8) (7)で漸近帰無分布が得られている場合は,極限分布への収束が改良できるような検定統計量に対する単調変換(Bartlett 型修正)を与える.(7)で漸近検出力が得られている場合は,分布のパラメータ等によって漸近検出力がどのように変化するかについて考察を与える.

(9) (8)で得られた結果について,大規模なモンテカルロ・シミュレーションによって数値的評価を与える.

(10)(1)~(9)の結果を学術論文にまとめ,研究成果について他の研究分野の研究者等から応用上の観点によるレビューを受ける.

#### 4. 研究成果

本研究課題では(1)平均ベクトルに関する仮説検定問題,(2)判別分析における一部の変量の冗長性検定問題に焦点を当て,以下の研究成果を得ている.

##### (1)平均ベクトルに関する仮説検定問題(2019~2020 年度)

楕円分布族から 2-step 単調欠測データ(欠測パターン数が 2 の単調欠測データ)が得られているものとし,各欠測パターンにおいて標本ベクトルの観測されている成分の周辺分布の平均ベクトルや尺度パラメータは共通,尖度パラメータは異なるとする.したがって,この設定は欠測データメカニズムが MCAR でない状況となる.

この設定の下で,平均ベクトルに関する 2 種の検定統計量に対し,漸近帰無分布を求め,Bartlett 型修正を与えた.数値実験の結果からも標本サイズがそれほど大きくない状況でも,第 1 種の過誤を名目上の有意水準とほぼ等しく制御可能であることを確認している.

さらに,上記の検定統計量の 1 つに対して,漸近検出力関数の導出に関する研究を遂行した.局所対立仮説の下で検定統計量の漸近分布の導出は困難を極めたため,検定統計量の期待値と極限分布の期待値を等値して得られる分布の近似による方法を得ている.数値実験の結果からは,標本サイズが大きい状況では分布の近似が比較的機能し,検出力の近似として用いることができるものの,標本サイズが少ない状況では精度が保てず,実用上の課題が残っている.

##### (2)判別分析における一部の変量の冗長性検定問題(2021 年度)

2 つの多変量正規母集団からそれぞれ 2-step 単調欠測データが得られているものとし,各欠測パターンの標本ベクトルにおいて観測されている成分について周辺分布の平均ベクトルや尺度パラメータは共通であるとする.したがって,この設定は欠測データメカニズムが MCAR である状況となる.

この設定の下で,(i)一部の標本ベクトルで欠測している変量が冗長であるという帰無仮説,(ii)すべての標本ベクトルで観測されている変量の一部が冗長であるという 2 つの帰無仮説について,尤度比検定統計量の構成に取り組んだ.(i)については検定統計量に対し,完全データ部分の標本サイズ,一部の変量が欠測している部分の標本サイズと,極限分布を帰無分布の近似として用いた場合の第 1 種の過誤の近似精度の関係を数値的に考察した.その結果,完全データ部分の標本サイズが大きくなるにつれて近似精度が向上することが確認されている.(ii)については得られた検定統計量の導出過程について再度確認を行っている最中である.

なお,(1)で得られている帰無仮説の下での Bartlett 型修正に関する研究成果については,

統計科学の理論研究について扱う海外専門誌である Communications in Statistics - Theory and Methods に掲載されている。(2)の議論については欠測データ解析の分野における変数選択に関わる重要な問題であり、当初計画していた研究課題に先んじて取り組むこととした。欠測データ解析における変数選択問題は研究成果の導出に関して理論的障壁があるが、(2)で得られた結果はその糸口となる重要な結果であると考えている。

5. 主な発表論文等

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

1. 著者名 Shutoh, N.	4. 巻 50
2. 論文標題 Effect of nonnormality on tests for a mean vector with missing data under an elliptically contoured pattern-mixture model	5. 発行年 2021年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 4448-4469
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/03610926.2020.1716254	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4件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 岡田 悠希, 首藤 信通
2. 発表標題 2-step単調欠測データに基づく判別関数の係数に対する仮説検定
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 米口 貴誠, 首藤 信通, 兵頭 昌
2. 発表標題 楢円母集団から得られた2-step単調欠測データに基づく平均ベクトルの尤度比検定の検出力について
3. 学会等名 日本計算機統計学会 第34回シンポジウム
4. 発表年 2020年

1. 発表者名 Shutoh, N.
2. 発表標題 Effect of nonnormality on the distribution of the statistics for testing a mean vector with two-step monotone missing data
3. 学会等名 10th International Workshop on Simulation and Statistics (SimStat2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Shutoh, N.
2. 発表標題 Asymptotic distribution of the linear discriminant function with two-step monotone sample
3. 学会等名 12th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics2019) (国際学会)
4. 発表年 2019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
研究協力者	兵頭 昌 (Hyodo Masashi)	神奈川大学・経済学部・准教授  (32702)	
研究協力者	米口 貴誠 (Yoneguchi Takaaki)	神戸大学・大学院海事科学研究科・大学院生  (14501)	
研究協力者	岡田 悠希 (Okada Yuki)	神戸大学・大学院海事科学研究科・大学院生  (14501)	

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
---------	---------