

令和 2 年 6 月 9 日現在

機関番号：32660

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K00058

研究課題名(和文) 欠測値をもつ多次元データに対する統計的手法の開発とその応用

研究課題名(英文) Development of statistical procedure of multi-dimensional data with missing value and its application

研究代表者

瀬尾 隆 (Seo, Takashi)

東京理科大学・理学部第一部応用数学科・教授

研究者番号：00266909

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：多次元データの成分の一部がランダムに欠測を持つ場合の統計的手法(統計的仮説検定)について研究成果を得た。特に平均ベクトルの検定における検定統計量の分布の漸近展開および検定統計量の改良、そして、平均ベクトルばかりでなく分散共分散行列も同時考える同時検定、部分平均ベクトルの検定、プロフィール分析(2つの群の平均ベクトルが平行であるかどうかなどの検定問題)について、尤度比検定統計量の分布の導出などの研究成果を得た。データについては、主に単調型欠測データを中心に、一般欠測データの場合についても成果を得ることができた。また成長曲線モデルの下でのパラメータの推定やAIC型情報量規準を与えた。

研究成果の学術的意義や社会的意義

統計的仮説検定問題においては、検定統計量の帰無仮説の下での分布、つまり、上側パーセント点を与える必要がある、しかしながら、多次元データを取り扱う多くの問題では、完全データの下でさえ、その分布の導出は容易ではなく、漸近展開などを用いた近似上側パーセント点を与える研究が多い。またその結果を用いて検定統計量を改良することによって、取り扱いやすいカイ二乗分布への近似がよい変換検定統計量を与える研究がある。そのような背景の下、本研究では、多次元欠測データの場合に同様の理論的結果を与えたことは学術的に大きな意義がある。さらに実データにも適用できるものであり、社会的意義もあると思われる。

研究成果の概要(英文)：We obtained some results on the statistical procedures (the statistical hypothesis testing problems) in the case where multi-dimensional data have some missing observations at random. In particular, we derived asymptotic expansions of the distribution of the test statistics for the test of mean vector and the improved test statistic, and we also obtained the distributions of likelihood ratio test statistics of the simultaneous test for the mean vector and covariance matrix, the test of the sub-mean vector, and the profile analysis (the test problem of whether the two mean vectors are parallel). With regard to the data, we were able to obtain results mainly for monotonic missing data and also for general missing data. Moreover, parameter estimation and AIC type information criterion under the growth curve model are given under the monotone missing data.

研究分野：数理統計学

キーワード：欠測データ 多変量解析 尤度比検定 漸近展開

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

近年、あらゆるデータが電子化され膨大なデータを容易に取り扱うことが可能となってきた。一方、膨大であるがゆえにデータが何らかの理由で欠測(欠損)することが多くあり、欠測が生じた場合の統計的推測理論や統計的仮説検定法の開発が重要な課題のひとつとなっている。この分野における研究は以前より数多く行われているが不十分な点や未解決な問題も多く、課題として残されている。

### 2. 研究の目的

上記の背景のもと、平均ベクトルや分散共分散行列に関する検定問題を軸に多変量解析における多変量正規性の検定法、プロフィール分析、成長曲線モデルの下での検定問題などに焦点を絞り、新たな統計的手法の開発、また、それらの手法の理論的な性質や特性を調べることを目的とする。

(1) 単調型欠測データの下での平均ベクトルの検定問題について  $T^2$  型検定統計量や尤度比検定統計量の帰無分布を導出する。特に、一般単調欠測データの場合や多標本問題への拡張を取り扱う。

(2) プロフィール分析や部分平均ベクトルの検定問題について、単調型欠測が生じた場合の尤度比検定および、その修正尤度比検定を導出する。

(3) 多変量正規性の検定問題、成長曲線(GMANOVA)モデルの下での推定・検定問題について、データが単調欠測している場合の理論を構築する。

### 3. 研究の方法

(1) 学術論文や専門書を通して先行研究の結果の理解ばかりでなくその導出法なども理解・整理し、また、研究代表者のこれまでの研究成果も再確認して、現在考えている問題に適用する。

(2) 手計算および、Mathematica などを用いて統計量の計算や計算チェックを行う。

(3) 導出した研究結果の近似精度の数値的評価を行うためにあらゆるパラメータに対して大規模なモンテカルロ・シミュレーションを行う。

### 4. 研究成果

#### (1) 平成 29 年度

成分の一部がランダムに欠測しているデータベクトルによって構成されるデータ行列に対して、特に単調欠測データ(データベクトルの成分の値が欠測するとそれ以降の成分もすべて欠測するデータのこ)の仮定の下で平成 29 年度は大きく分けて主に以下の 4 つの内容について研究成果を得た。

(a) 平均ベクトルの検定問題で、1 標本問題及び 2 標本問題を含む多標本問題(MANOVA)について、以前から尤度比検定統計量の近似上側パーセント点の導出と修正尤度比検定統計量や変換統計量などを議論し投稿中であったが、修正や改良を行い、海外の査読付き学術雑誌に論文として掲載された。また、欠測値を含まないが高次元データの下での関連する議論を行い、いくつかの成果を得て、今後欠測値を含む場合につながる成果を得た。

(b) 多標本問題に対する部分平均ベクトルの検定問題に関して、完全データの場合の議論がこれまでになかったことから、単調欠測データの場合を議論する前に完全データの場合の尤度比検定統計量の導出とその分布について、近似上側パーセント点を与え、モンテカルロシミュレーションにより数値的評価を与えた。さらに、単調欠測データの場合については 2 ステップ単調欠測データの下で議論を行い、いくつかの成果を得た。

(c) プロフィール分析に関するもので、一様共分散構造をもつ平行性仮説検定問題について、尤度比検定統計量とその帰無分布の導出に成功した。

(d) 平均ベクトルにある種の構造がある場合の問題であり、具体的には経時データなどの場合に考える成長曲線モデルの下でのパラメータの推定や検定に関する問題について、2 ステップ単調欠測データについて議論を行い、平均パラメータや分散共分散行列の最尤推定量の導出を試みいくつかの成果を得た。

#### (2) 平成 30 年度

平成 29 年度に引き続き、多次元データにおいてデータベクトルの成分が一部分欠測している下で、主に以下の内容について研究成果を得た。

(a) 平均ベクトルの検定については、平成 29 年度は主に尤度比検定統計量について議論したが、平成 30 年度は、2 標本問題を中心に  $T^2$  型検定統計量について、検定統計量の分解を利用した分布関数とその上側パーセント点の近似を導出することに成功した。この結果は漸近展開法によるもので、非常に良い近似となっている。関連して、一般欠測データの下での 2 標本問題における多変量正規母集団の同等性検定や 2 ステップ単調欠測データの下での部分平均ベクトルの検定問題、さらに欠測値を含まない高次元データの下での一般化分散の同等性検定などについても研究し、いくつかの成果を得た。

(b) プロフィール分析については、平成 29 年度、平行性仮説検定問題に注目し、単調欠測データの下で、一様共分散構造をもつ多変量正規分布に従う場合の尤度比検定統計量とその正確な帰無分布を導出することに成功したが、それを受けて、平成 30 年度は水準差の信頼区間について議論し、精度の良い近似信頼区間の導出に成功した。この水準差の推定は、平行性仮説が採択された場合、すなわち、プロフィール間で平行性が認められた場合に問題となるもので、本研究では、単調欠測データの下で 2 群の場合に対する近似信頼区間を与えた。

(c) 単調欠測データの下での成長曲線モデルに関する研究については、平成 29 年度に引き続いて研究を進め、成長曲線モデルに関するパラメータの推定やその性質及び AIC 型選択規準について考え、いくつかの理論的結果を得た。

### (3) 平成 31 年度

最終年度はこれまでの研究成果を引き続き行い、まとめるとともに、主に以下の成果を得た。

(a) 平均ベクトルと分散共分散行列の同時検定について議論し、こちらは単調でない一般欠測データの下で数値的反復法による最尤推定量の導出により、尤度比検定などを与えることに成功した。

(b) 単調欠測データの下で 1 標本問題であるが従来の検定統計量を分解し独立な統計量の和になるように修正した新たな検定統計量を提案し、その帰無分布の漸近展開をもとにカイ二乗近似の良い変換統計量を与えた。

(c) 単調欠測データの下での成長曲線モデルに対する平均パラメータと分散共分散行列の最尤推定量については、これまでの結果を論文としてまとめ学術雑誌に投稿した。また、データが単調型欠測構造をもつ場合の AIC 型モデル選択基準については、平成 30 年度に引き続いて研究を進め 1 群の場合の成長曲線モデルに対してモデルの良さを評価する基準である AIC 型リスクを考え、分散共分散行列を既知とした場合の正確な補正項が 2 倍の独立パラメータ数となることを示すことができた。

## 5. 主な発表論文等

〔雑誌論文〕 計11件（うち査読付論文 11件／うち国際共著 5件／うちオープンアクセス 4件）

1. 著者名 Tamae Kawasaki, Nobumichi Shutoh, Takashi Seo	4. 巻 12
2. 論文標題 On the asymptotic distribution of $T^2$ -type statistic with two-step monotone missing data	5. 発行年 2018年
3. 雑誌名 Journal of Statistical Theory and Practice	6. 最初と最後の頁 657-668
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/15598608.2018.1450795	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 Ayaka Yagi, Takashi Seo, Muni Srivastava	4. 巻 47
2. 論文標題 Testing equality of mean vectors in a one-way MANOVA with monotone missing data	5. 発行年 2018年
3. 雑誌名 Communications in Statistics - Theory and Methods	6. 最初と最後の頁 5534-5546
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/03610926.2017.1397168	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 Zofia Hanusz, Rie Enomoto, Takashi Seo, Kazuyuki Koizumi	4. 巻 47
2. 論文標題 A Monte Carlo comparison of Jarque-Bera type tests and Henze-Zirkler test of multivariate normality	5. 発行年 2018年
3. 雑誌名 Communications in Statistics - Simulation and Computation	6. 最初と最後の頁 1439-1452
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/03610918.2017.1315771	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 Masashi Hyodo, Hiroki Watanabe, Takashi Seo	4. 巻 168
2. 論文標題 On simultaneous confidence interval estimation for the difference of paired mean vectors in high-dimensional settings	5. 発行年 2018年
3. 雑誌名 Journal of Multivariate Analysis	6. 最初と最後の頁 160-173
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jmva.2018.07.008	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 Yagi Ayaka, Seo Takashi, Srivastava Muni S.	4. 巻 37
2. 論文標題 A modified likelihood ratio test for a mean vector with monotone missing data	5. 発行年 2017年
3. 雑誌名 Discussiones Mathematicae Probability and Statistics	6. 最初と最後の頁 39 ~ 55
掲載論文のDOI (デジタルオブジェクト識別子) 10.7151/dmps.1194	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Rie Enomoto, Zofia Hanusz, Ayako Hara, Takashi Seo	4. 巻 49
2. 論文標題 Multivariate normality test using normalizing transformation for Mardia's multivariate kurtosis	5. 発行年 2020年
3. 雑誌名 Communications in Statistics - Simulation and Computation	6. 最初と最後の頁 684-698
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/03610918.2019.1661476	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 Tamae Kawasaki, Toshiki Naito, Takashi Seo	4. 巻 9
2. 論文標題 $T^2$ type test statistic and simultaneous confidence intervals for two sub-mean vectors	5. 発行年 2020年
3. 雑誌名 International Journal of Statistics and Probability	6. 最初と最後の頁 1-8
掲載論文のDOI (デジタルオブジェクト識別子) 10.5539/ijsp.v9n1p1	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 Takatoshi Sugiyama, Masashi Hyodo, Hiroki Watanabe, Shin-ichi Tsukada, Takashi Seo	4. 巻 55
2. 論文標題 Test for equality of generalized variance in high-dimensional and large sample settings	5. 発行年 2019年
3. 雑誌名 SUT Journal of Mathematics	6. 最初と最後の頁 139-154
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hiroki Watanabe, Masashi Hyodo, Yuki Yamada, Takashi Seo	4. 巻 49
2. 論文標題 Estimation of misclassification probability for a distance-based classifier in high-dimensional data	5. 発行年 2019年
3. 雑誌名 Hiroshima Mathematical Journal	6. 最初と最後の頁 175-193
掲載論文のDOI (デジタルオブジェクト識別子) 10.32917/hmj/1564106544	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hiroki Watanabe, Takashi Seo, Masashi Hyodo	4. 巻 55
2. 論文標題 An estimator of misclassification probability for multi-class Euclidean distance classifier in high-dimensional data	5. 発行年 2019年
3. 雑誌名 SUT Journal of Mathematics	6. 最初と最後の頁 11-23
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Ayaka Yagi, Takashi Seo, Zofia Hanusz	4. 巻 48
2. 論文標題 Improved simplified $T^2$ test statistics for a mean vector with monotone missing data	5. 発行年 2019年
3. 雑誌名 Communications in Statistics - Simulation and Computation	6. 最初と最後の頁 1606-1622
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/03610918.2017.1419261	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

〔学会発表〕 計24件 (うち招待講演 0件 / うち国際学会 7件)

1. 発表者名 Hiroki Watanabe, Masashi Hyodo, Takashi Seo
2. 発表標題 Inference on High-dimensional Mean Vectors Under Alternative Hypothesis
3. 学会等名 DSSV (Data Science, Statistics & Visualization) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Tamae Kawasaki, Takashi Seo
2. 発表標題 Tests for mean vector using approximate degrees of freedom with two-step monotone missing data
3. 学会等名 ISI2019-62nd World Statistics Congress (国際学会)
4. 発表年 2019年

1. 発表者名 杉山 高聖, 兵頭 昌, 渡邊 弘己, 塚田 真一, 瀬尾 隆
2. 発表標題 高次元大標本枠組みにおける一般化分散の同等性検定
3. 学会等名 統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 八木 文香, 瀬尾 隆, 藤越 康祝
2. 発表標題 3-step単調欠測データをもつ成長曲線モデルに関するAIC型選択規準
3. 学会等名 日本数学会
4. 発表年 2019年

1. 発表者名 野村 玲実, 八木 文香, 瀬尾 隆
2. 発表標題 一般欠測データの下での多変量正規母集団の同等性検定
3. 学会等名 日本計算機統計学会第33回シンポジウム
4. 発表年 2019年

1. 発表者名 八木 文香, 小野沢 瑞季, 瀬尾 隆
2. 発表標題 単調欠測データにおける平均ベクトルに対する新たな検定統計量
3. 学会等名 日本数学会
4. 発表年 2019年

1. 発表者名 川崎 玉恵, 瀬尾 隆
2. 発表標題 2-step単調欠測データのもとでの部分平均ベクトルの検定に対する修正尤度比検定統計量
3. 学会等名 科研費シンポジウム「融合する統計科学」
4. 発表年 2018年

1. 発表者名 佐伯 悠一郎, 八木 文香, 瀬尾 隆, 百武 弘登
2. 発表標題 単調欠測データが一様共分散構造を持つ場合の平行性仮説検定と水準差の信頼区間
3. 学会等名 科研費シンポジウム「融合する統計科学」
4. 発表年 2018年

1. 発表者名 野村 玲実, 八木 文香, 瀬尾 隆
2. 発表標題 一般欠測データの下での2標本問題における多変量正規母集団の同等性検定
3. 学会等名 科研費シンポジウム「融合する統計科学」
4. 発表年 2018年



1. 発表者名 八木 文香, 瀬尾 隆, 藤越 康祝
2. 発表標題 単調型欠測をもつ成長曲線モデルにおける推定
3. 学会等名 日本数学会
4. 発表年 2018年

1. 発表者名 八木 文香, 瀬尾 隆, 藤越 康祝
2. 発表標題 単調欠測データをもつ成長曲線モデルに関するAIC型選択規準
3. 学会等名 統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 佐伯 悠一郎, 八木 文香, 瀬尾 隆, 百武 弘登
2. 発表標題 一様共分散構造をもつ単調欠測データの下での平行性仮説検定と水準差について
3. 学会等名 統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 林 大将, 兵頭 昌, 瀬尾 隆
2. 発表標題 高次元データのための平均ベクトルの差の信頼区間について
3. 学会等名 日本計算機統計学会
4. 発表年 2018年

1. 発表者名 八木 文香, 瀬尾 隆
2. 発表標題 単調欠測データにおける平均ベクトル間の検定と同時信頼区間
3. 学会等名 日本計算機統計学会
4. 発表年 2018年

1. 発表者名 Tamae Kawasaki, Takashi Seo
2. 発表標題 Likelihood ratio tests for sub-mean vectors with two-step monotone missing data in two-sample problem
3. 学会等名 Joint International Society for Clinical Biostatistics and Australian Statistical Conference (国際学会)
4. 発表年 2018年

1. 発表者名 Rie Enomoto, Zofia Hanusz, Ayako Hara, Takashi Seo
2. 発表標題 Multivariate normality test using normalizing transformation of Mardia's multivariate kurtosis
3. 学会等名 International Conference on Trends and Perspectives in Linear Statistical Inference (国際学会)
4. 発表年 2018年

1. 発表者名 内藤 寿輝, 川崎 玉恵, 瀬尾 隆
2. 発表標題 部分平均ベクトルにおける $T^2$ 型検定統計量とその同時信頼区間の構成について
3. 学会等名 日本計算機統計学会
4. 発表年 2017年

1. 発表者名 佐伯 悠一郎, 八木 文香, 瀬尾 隆, 百武 弘登
2. 発表標題 多標本問題における一様共分散構造をもつ単調欠測データの下での平行性仮説検定
3. 学会等名 日本計算機統計学会
4. 発表年 2017年

1. 発表者名 八木 文香, 瀬尾 隆
2. 発表標題 Transformed $T^2$ -type Statistics for Testing Equality of Two Mean Vectors with Monotone Samples
3. 学会等名 日本数学会
4. 発表年 2017年

1. 発表者名 八木 文香, 瀬尾 隆, 藤越 康祝
2. 発表標題 2-step単調型欠測をもつ成長曲線モデルにおける推定
3. 学会等名 統計関連学会連合大会
4. 発表年 2017年

1. 発表者名 Tamae Kawasaki, Takashi Seo
2. 発表標題 An F approximation to the distribution of T-squared type test statistic with two-step monotone missing data
3. 学会等名 The CEN- ISBS Vienna 2017 Joint Conference on Biometrics & Biopharmaceutical Statistics (国際学会)
4. 発表年 2017年

1. 発表者名 Tamae Kawasaki, Takashi Seo
2. 発表標題 Testing equality of two mean vectors on subvectors with two-step monotone missing data
3. 学会等名 61st World Statistics Congress (国際学会)
4. 発表年 2017年

1. 発表者名 Ayaka Yagi, Takashi Seo
2. 発表標題 Simultaneous Confidence Intervals for Pairwise Comparisons Among Mean Vectors with Monotone Missing Data
3. 学会等名 The Xth International Conference on Multiple Comparison Procedures (国際学会)
4. 発表年 2017年

1. 発表者名 八木 文香, 瀬尾 隆
2. 発表標題 Transformations of Simplified $T^2$ Statistic with Monotone Missing Data
3. 学会等名 日本計算機統計学会
4. 発表年 2017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
---------	---------------------------	-----------------------	----