

平成 29 年 5 月 31 日現在

機関番号：32660

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330050

研究課題名(和文) 高次元および欠損データにおける多変量統計手法の開発とその応用

研究課題名(英文) Development of Multivariate Statistical Procedure for High Dimensional and Missing Data and Its Application

研究代表者

瀬尾 隆 (SEO, Takashi)

東京理科大学・理学部第一部数理情報科学科・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)：欠損データに対する統計的推測理論と検定法の開発について、特に、平均ベクトルの検定問題と平均ベクトルと分散共分散行列の同時検定問題を取り扱い、多母集団かつ一般の単調欠測データの下で、漸近展開論を用いて新たな修正尤度比検定統計量などを導出した。また、高次元データの下での判別分析についてもいくつかの新たな研究成果を得た。多変量正規性検定については、欠損データではない完全データの下で、モンテカルロ・シミュレーションによる従来の検定法を比較した研究成果を与えた。

研究成果の概要(英文)：On the development of statistical inference theory and test procedure for missing data, in particular, test for mean vectors and simultaneous test for mean vectors and covariance matrices were treated. New modified likelihood ratio test statistics under the general step monotone missing data and multi-sample problem were derived using asymptotic expansion theory. The discriminant analysis under the high dimensional data were considered and obtained some new results. For the multivariate normality test, the results by comparing with previous test procedures under the complete data were given by Monte Carlo simulation.

研究分野：数理統計学

キーワード：統計科学 統計数学 統計理論 多変量解析法

1. 研究開始当初の背景

近年、大規模データを取り扱う統計的推測理論や統計的検定論の開発が統計科学の分野において重要な問題のひとつとなっている。そのような背景のもと、データ数(標本の大きさ)ばかりでなく、変数の数(次元数)も大きくなる高次元データの場合の統計解析の理論的開発が重要である。また、データ数が増えてくるとデータが何らかの理由で欠損することが多く、このようなデータが欠損した場合の統計解析の理論的開発も重要な問題の一つである。このような問題は以前より研究されては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、欠測データに対する、平均ベクトルの検定、分散共分散行列の検定、そして平均ベクトルと分散共分散行列の同時検定に焦点を絞る。さらに、判別分析・多変量正規性検定についても取り扱う。具体的には以下の内容を研究目的とする。

- (1) 研究代表者自身の先行研究の結果を受けて、平均ベクトルの検定について、欠測の構造は単調欠測データとし、母集団数を1母集団から多母集団まで拡張し、検定統計量の導出とその帰無分布について議論し、具体的な検定手法を与える。
- (2) 分散共分散行列の検定や平均ベクトルと分散共分散行列の同時検定についても上記と同様の議論を行う。
- (3) 判別分析については、高次元データの下で議論し、また、多変量正規性検定に関する問題も取り扱う。

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ばかりでなく導出法などを整理し、また、研究代表者の過去の研究成果なども再確認して、現在考えている問題に適用する。
- (2) 手計算および、数式処理ソフトを用いて統計量の計算や計算チェックを行う。
- (3) 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い、提案する近似値の近似精度の数値的評価を行う。

4. 研究成果

- (1) 平成 26 年度
- (a) 2 標本問題における平均ベクトルの同等待性検定について、3 ステップ単調欠測データの下での T^2 型検定統計量の導出と帰無分布に対する線形補間による近似上側パーセント点を与えた。結果はサンプル

数が少ない場合にも良い近似となるものであり、モンテカルロ・シミュレーションにより、数値的評価を与えている。さらに、欠測の型が異なる 2 ステップ単調欠測データに適用することに成功している。また 1 母集団の場合であるが、一般の単調欠測データの場合に拡張することに成功している。

- (b) 平均ベクトルと分散共分散行列の同時検定について、2 ステップ単調欠測データの下で、尤度比検定統計量の帰無分布についていくつの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、その精度について検証している。さらには、上記の問題は母集団が 1 つの場合であったが、母集団が 2 つの場に拡張することにも成功している。

- (c) 判別分析については、高次元データの下での 2 群判別について新たな誤判別確率の漸近近似を導出した。

(2) 平成 27 年度

- (a) 2 ステップ単調欠測データの下での T^2 型検定統計量について、摂動法を用いて、1 次モーメント(期待値)と 2 次モーメントに対する漸近展開を導出することに成功し、バイアス修正を与えることに成功した。

- (b) 前年度に引き継ぐ形で平均ベクトルと分散共分散行列の同時検定について、2 ステップ単調欠測データの下で議論し、母集団の数が 2 個以上の一般の場合の結果を与えることに成功した。

- (c) 前年度に引き続き、高次元データの下でのユークリッド距離による判別関数を用いて、条件付き誤判別確率の分布の漸近正規性を与え、期待誤判別確率に対する近似同時区間を与えた。

(3) 平成 28 年度

最終年度はこれまでの研究成果をまとめるとともに、以下の成果を得た。

- (a) 単調欠測データの下での平均ベクトルの検定について、母集団数が複数個の場合である多標本問題へ拡張することに成功し、成果として、尤度比検定統計量の帰無分布についてサンプル数が増えるにつれてより速くカイ二乗分布に近づく修正尤度比検定統計量の導出に成功した。また、1 標本の場合ではあるが、ホテル型 T^2 統計量にも同様の議論を行うことを試み、完全ではないが同様な変換統計量を与えることに成功した。

- (b) 平均ベクトルと分散共分散行列の同時検定に関連して、分散共分散行列の構造についての検定問題について、1 標本問題かつ単調欠測データの下で、尤度比検定統計量とその近似帰無分布について結果を得た。
- (c) 多変量正規性検定問題については、欠測データの下での議論をする予定であったが、その前段階として完全データの下で提案されているいくつかの検定統計量について性質や検出力などを調べる必要があり、その比較などを論文としてまとめ、査読付き論文として学術雑誌に掲載された。

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 12 件)

Zofia Hanusz, Rie Enomoto, Takashi Seo and Kazuyuki Koizumi, (2017). A Monte Carlo comparison of Jarque-Bera type tests and Henze-Zirkler test of multivariate normality.

Communications in Statistics - Simulation and Computation に掲載予定。査読有

DOI: 10.1080/03610918.2017.1315771

Ayaka Yagi and Takashi Seo, (2017). Tests for normal mean vectors with monotone incomplete data. *American Journal of Mathematical and Management Sciences*, **36**, 1-20. 査読有

DOI: 10.1080/01966324.2016.1236712

Miki Hosoya and Takashi Seo, (2016). On the likelihood ratio test for the equality of multivariate normal populations with two-step monotone missing data. *Journal of Statistical Theory and Practice*, **10**, 673-692. 査読有

DOI:10.1080/15598608.2016.1215942

Tamae Kawasaki and Takashi Seo, (2016). A test for subvector of mean vector with two-step monotone missing data. *SUT Journal of Mathematics*, **52**, 21-39. 査読有

<http://www.rs.tus.ac.jp/sutjmath/>

Tamae Kawasaki and Takashi Seo, (2016). Bias correction for T^2 type statistic with two-step monotone missing data. *STATISTICS*, **50**, 76-88. 査読有

DOI: 10.1080/02331888.2015.1060238

Masashi Hyodo, Tomohiro Mitani, Tetsuto Himeno and Takashi Seo, (2015). Approximate interval estimation for EPMC for improved linear discriminant rule under high dimensional framework. *SUT Journal of Mathematics*, **51**, 145-179. 査読有

<http://www.rs.tus.ac.jp/sutjmath/>

Miki Hosoya and Takashi Seo, (2015). Simultaneous testing of the mean vector and the covariance matrix with two-step monotone missing data. *SUT Journal of Mathematics*, **51**, 83-98. 査読有

<http://www.rs.tus.ac.jp/sutjmath/>

Hiroki Watanabe, Masashi Hyodo, Takashi Seo, Tatjana Pavlenko, (2015). Asymptotic properties of the misclassification rates for Euclidean distance discriminant rule in high-dimensional data. *Journal of Multivariate Analysis*, **140**, 234-244. 査読有

DOI: 10.1016/j.jmva.2015.05.008

Ayaka Yagi and Takashi Seo, (2015). Test for equality of mean vectors and simultaneous confidence intervals with two-step or three-step monotone missing data patterns. *American Journal of Mathematical and Management Sciences*, **34**, 213-233. 査読有

DOI:10.1080/01966324.2015.1020403

Tamae Kawasaki and Takashi Seo, (2015). A two sample test for mean vectors with unequal covariance matrices. *Communications in Statistics - Simulation and Computation*, **44**, 1850-1866. 査読有

DOI:10.1080/03610918.2013.824587

Ayaka Yagi and Takashi Seo, (2015). Tests for mean vectors with two-step and three-step monotone samples. *Josai Mathematical Monographs*, **8**, 49-71. 査読有

http://libir.josai.ac.jp/il/user_contents/02/G0000284repository/pdf/JOS-13447777-0806.pdf

Ayaka Yagi and Takashi Seo, (2014). A test for mean vector and simultaneous confidence intervals with three-step monotone missing data. *American Journal of Mathematical and Management Sciences*, **33**, 161-175. 査読有

DOI:10.1080/01966324.2014.911670

[学会発表](計 32 件)

八木文香, 瀬尾隆, On the Null Distribution of the Simplified T^2 Statistic with Monotone Missing Data, 第 16 回西東京統計研究会, 2017 年 3 月 16 日, 専修大学.

八木文香, 瀬尾隆, 単調欠測データの下での平均ベクトルに関する検定問題, 科研費研究集会「数理統計ひこね 2016」, 2016 年 12 月 2 日, 滋賀大学.

八木文香, 瀬尾隆, A Test for Zero Correlation with Monotone Missing Data, 日本計算機統計学会, 2016 年 11 月 25 日, プラサヴェルデ(沼津).

川崎玉恵, 瀬尾隆, Tests for a sub-mean vector with incomplete data, 統計数理研究所共同研究集会「環境・生態データと統計解析」, 2016 年 10 月 28 日, 統計数理研究所.

Ayaka Yagi, Takashi Seo, The null distribution of the LRT statistic for mean vectors with monotone missing data, ICOSDA2016 (International Conference on Statistical Distributions and Applications), 2016 年 10 月 16 日, Crowne Plaza, Niagara Falls Fallsview Hotel.

八木文香, 瀬尾隆, 単調欠測データの下での分散共分散行列の検定における尤度比検定統計量について, 日本数学会, 2016 年 9 月 18 日, 関西大学.

渡邊弘己, 山田雄紀, 兵頭昌, 瀬尾隆, ユークリッド距離型判別ルールにおける誤判別確率の推定量の数値的検証, 統計関連学会連合大会, 2016 年 9 月 5 日, 金沢大学.

八木文香, 瀬尾隆, 多標本問題における単調欠測データの下での平均ベクトルの尤度比検定統計量の漸近分布, 統計関連学会連合大会, 2016 年 9 月 5 日, 金沢大学.

篠崎絢, 橋口博樹, 瀬尾隆, 楕円ウィシャート行列における最大最小固有値の正確分布とその数値計算, 日本計算機統計学会, 2016 年 5 月 20 日, ハートピア京都.

八木文香, 瀬尾隆, 2 標本問題における単調欠測データの下での平均ベクトルの尤度比検定統計量の帰無分布, 日本計算機統計学会, 2016 年 5 月 19 日, ハートピア京都.

八木文香, 瀬尾隆, 単調欠測データをも

つ平均ベクトルの検定における尤度比検定統計量について, 日本数学会, 2016 年 3 月 19 日, 筑波大学.

川崎玉恵, 瀬尾隆, 2-step 単調欠測データのもとでの T^2 型統計量の帰無分布におけるカイ二乗近似, 応用統計学会, 2016 年 3 月 17 日, 統計数理研究所.

川崎玉恵, 首藤信通, 瀬尾隆, 2-step 単調欠測データにおける T^2 型検定統計量の帰無分布に対する漸近展開, 第 10 回日本統計学会春季集会, 2016 年 3 月 5 日, 東北大学.

細谷美貴, 瀬尾隆, Likelihood Ratio Test for Equality of Multivariate Normal Populations with Two-step Monotone Missing Data, 第 10 回日本統計学会春季集会, 2016 年 3 月 5 日, 東北大学.

細谷美貴, 瀬尾隆, 多標本問題における 2-step 単調欠測データの下での平均ベクトルと分散共分散行列の同時検定, 日本計算機統計学会, 2015 年 11 月 27 日, 釧路市生涯学習センター(まなぼっと幣舞).

Ayaka Yagi, Zofia Hanzus, Takashi Seo, LRT Statistics for Testing Equality of Two Mean Vectors with Monotone Missing Data, 2015 International Workshop for JSCS 30th Anniversary in Okinawa, 2015 年 10 月 31 日, 沖縄科学技術大学院大学.

Ryoko Yamaguchi, Ayaka Yagi, Takashi Seo, Simultaneous Testing of the Mean Vector and the Covariance Matrix with Monotone Missing Data, 2015 International Workshop for JSCS 30th Anniversary in Okinawa, 2015 年 10 月 31 日, 沖縄科学技術大学院大学.

八木文香, 瀬尾隆, 2 標本問題における単調欠測データの下での平均ベクトルの尤度比検定について, 日本数学会, 2015 年 9 月 15 日, 京都産業大学.

八木文香, 瀬尾隆, 単調欠測データの下での平均ベクトルに対する修正尤度比検定統計量について, 統計関連学会連合大会, 2015 年 9 月 8 日, 岡山大学.

川崎玉恵, 瀬尾隆, 2-step 単調欠測データの下での T^2 型検定統計量の漸近分布について, 統計関連学会連合大会, 2015 年 9 月 7 日, 岡山大学.

- ⑳ 八木文香, 瀬尾隆, 単調欠測データにおける平均ベクトルの尤度比検定について, 日本計算機統計学会, 2015年5月14日, 山梨県立図書館.
- ㉑ 川崎玉恵, 瀬尾隆, Chi-squared approximations to T-squared type statistic for two-step monotone missing data, Workshop on Distribution Theory and Applications (統計数理研究所共同研究集会「確率分布とその応用」), 2015年1月29日, 統計数理研究所.
- ㉒ 八木文香, 瀬尾隆, Tests for mean vectors with two-step and three-step monotone samples, ワークショップ「統計科学とその周辺」, 2014年12月7日, 城西大学.
- ㉓ 細谷美貴, 瀬尾隆, 1 標本問題および 2 標本問題における 2-step 単調欠測データの下での平均ベクトルと分散共分散行列の同時検定について, 科研費シンポジウム「統計的推測の理論的基礎とその応用」, 2014年12月2日, 筑波大学.
- ㉔ 細谷美貴, 瀬尾隆, 2-step 単調欠測データの下での 2 つの多変量正規母集団の同源性検定, 日本計算機統計学会, 2014年11月14日, 沖縄科学技術大学院大学.
- ㉕ 八木文香, 瀬尾隆, Testing the Equality of Two Mean Vectors Based on k-step Monotone Missing Data, 日本計算機統計学会, 2014年11月14日, 沖縄科学技術大学院大学.
- ㉖ 八木文香, 瀬尾隆, k-step 単調欠測データの下での平均ベクトルの検定, 日本数学会, 2014年9月28日, 広島大学.
- ㉗ 八木文香, 瀬尾隆, 2-step 単調欠測データの下で欠測の型が異なる平均ベクトルの検定, 統計関連学会連合大会, 2014年9月14日, 東京大学.
- ㉘ 川崎玉恵, 瀬尾隆, 2-step 単調欠測データのもとでの平均ベクトルに対する検定統計量のバイアス修正, 統計関連学会連合大会, 2014年9月14日, 東京大学.
- ㉙ 細谷美貴, 瀬尾隆, Likelihood Ratio Test for Simultaneous Testing of the Mean Vector and the Covariance Matrix with Two-step Monotone Missing Data, 統計関連学会連合大会, 2014年9月14日, 東京大学.
- ㉚ 八木文香, 瀬尾隆, 3-step 単調欠測データにおける平均ベクトル間の検定, 応用統計学会, 2014年5月22日, 統計数理研究所.

- ㉛ 細谷美貴, 瀬尾隆, 2-step 単調欠測データにおける分散共分散行列と平均ベクトルの同時検定に対する尤度比検定統計量について, 日本計算機統計学会, 2014年5月17日, 中央大学.

〔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) 研究代表者

瀬尾 隆 (SEO, Takashi)
東京理科大学・理学部・教授
研究者番号: 0 0 2 6 6 9 0 9

(2) 研究分担者

なし

(3) 連携研究者

西山 貴弘 (NISHIYAMA, Takahiro)
専修大学・経営学部・准教授
研究者番号: 3 0 5 1 6 4 7 2

小泉 和之 (KOIZUMI, Kazuyuki)
横浜市立大学・国際総合科学部・准教授
研究者番号: 7 0 5 4 8 1 4 8

首藤 信通 (SHUTOH, Nobumichi)
神戸大学・大学院海事科学研究科・講師
研究者番号: 0 0 6 3 4 0 9 9

兵頭 昌 (HYODO, Masashi)
大阪府立大学・工学(系)研究科(研究院)・准教授
研究者番号: 0 0 7 1 1 7 6 4

(4) 研究協力者

なし