

令和 6 年 6 月 24 日現在

機関番号：32660

研究種目：若手研究

研究期間：2018～2023

課題番号：18K18015

研究課題名（和文）高次元データの統計的推測：スパイク性とスパース性

研究課題名（英文）Statistical inference for high-dimensional data: spiking and sparsity

研究代表者

石井 晶 (Ishii, Aki)

東京理科大学・創域理工学部情報計算科学科・講師

研究者番号：20801161

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：主に高次元独立標本に対する統計的推測に対する新しい理論や方法論の開発を行った。ゲノムデータなどの高次元小標本データは、遺伝子間の相関などによって高次元共分散行列の固有値が強く飛び抜けた（スパイクした）値をとるという高次元現象をもとに、高次元平均ベクトルの検定、高次元共分散行列の同等性検定、高次元共分散行列の構造検定、高次元判別分析に対して新たな手法の開発と、提案したものが理論的に高い精度保証を与えることを証明した。

研究成果の学術的意義や社会的意義

本研究は、高次元統計解析という、高次元データに対して高精度・低計算コスト・高い汎用性をもつ高次元データに対する新しい理論・方法論をもとに成果を挙げている。理論的に高い精度保証を与える手法を提案するだけでなく、実データをもとに応用例・解析例も多く与えた。上記から、理論研究・応用研究において課題となる点である、実データにおける汎用性や理論的な精度保証を解決する成果を挙げており、学術的意義がある。対象としているデータがゲノム等の高次元データであり、汎用性の高さから医療などの分野への応用も多く考えられ、社会的意義がある。

研究成果の概要（英文）：We have mainly developed new theories and methodologies for statistical inference for high-dimensional independent samples. Based on the high-dimensional phenomenon that eigenvalues of high-dimensional covariance matrices of high-dimension, low-sample-size data, such as genomic data, are strongly spiked due to correlation among genes, we developed new methods for testing high-dimensional mean vectors, equality tests for high-dimensional covariance matrices, tests for high-dimensional covariance structures, and high-dimensional discriminant analysis. We also theoretically showed that the proposed procedures guarantee high accuracy.

研究分野：高次元統計解析

キーワード：高次元データ 高次元小標本 強スパイク固有値 高次元統計的推測

様式 C-19、F-19-1 (共通)

1. 研究開始当初の背景

研究開始当初、ゲノム等の高次元小標本データは遺伝子間の相関等から、非常に強いノイズを含む強スパイクモデルをもつことが研究代表者の所属する研究グループの研究により知られてきていた。高次元データに対する種々の統計的推測において、その理論の中心を担うのは高次元中心極限定理であるが、非常に強いノイズを含む高次元データに対しては、そのままでは高次元中心極限定理が成り立たない。そこで、強スパイクモデルに対して強スパイクするノイズを除去するようなデータ変換をすることで高次元中心極限定理を成立させ、理論的な精度保証を与えるという研究が進められてきていた。しかしながら、データ変換をすることが適切でない場合もあり、その場合に対する新しい理論・方法論の構築は未開拓であった。

2. 研究の目的

非常に強いノイズを含む高次元データに対し、高次元中心極限定理が成り立たない場合における統計的推測の理論・方法論を構築する。その際、データ変換とは別の視点から理論の構築を目指し、強スパイクモデルを直接解析することで、新たな理論を構築することを主な研究目的とし、新たな統計的推測の手法を提案することを目的とした。

3. 研究の方法

強スパイクモデルに対する高次元統計的推測として、以下の四つを考え、それぞれに対してデータ変換技術が有用である場合にはその応用を考え、データ変換が適切でない場合は、統計量の漸近分布をカイ 2 乗分布に収束させることで新たな統計的推測を考えた。

- (1) 高次元平均ベクトルの統計的推測
- (2) 高次元共分散行列の同等性検定
- (3) 高次元共分散行列の構造検定
- (4) 高次元判別分析

- (1) 高次元平均ベクトルの統計的推測においては、データ変換技術が有用である。Aoshima and Yata (2018, Sinica)によって高次元平均ベクトルの 2 標本検定は与えられているため、本研究では多標本検定や他の平均ベクトルの推測まで含んだ研究成果を挙げることを目指す。さらに、データ変換技術を用いず、漸近論として次元数は無限大だが標本数は固定の枠組みで、非常に少ない標本数の高次元データにも適用できる高次元平均ベクトルの検定手法を与えることを目指す。
- (2) 高次元共分散行列に対してはデータ変換をすることが適切でない。そこで、強スパイクモデルを直接扱い、統計量の漸近分布をカイ 2 乗分布に収束させることで新たな検定手法を提案することを目指す。
- (3) 高次元共分散行列の構造検定では、強スパイクモデルをもつ共分散行列の構造に着目し、先行研究で仮定されているような、母集団の正規性なしに検定手法を構築する。
- (4) 高次元判別分析では、2 次判別分析に着目し、新たな判別手法を提案することを目指す。既存の高次元 2 次判別分析において、強スパイクモデルがどのような影響を与えるかを検証し、強スパイクモデルがもつ問題点を取り除くような判別手法を開発する。

4. 研究成果

(1) Ishii, Yata and Aoshima (2019, JJSJ)では、データ変換技術の応用をすることで高次元多標本検定まで含んだ、高次元平均ベクトルの推測に関する新しい手法を与えた。漸近論としてはデータ数も次元数を超えない程度に増加する場合を考えており、高次元中心極限定理の成立を証明している。多くのデータ解析例も与え、Springerlink 2019 highly downloaded の統計学 Top 10 にランクインした。石井・矢田・青嶋 (2020, 応用統計学)では、データ変換を用いることなく、統計量の漸近分布をカイ 2 乗分布に収束させることで、高次元平均ベクトルの検定について成果を挙げた。次元数は無限大だが、標本数は固定のもとで新たな検定手法を与えたことにより、非常に少ない標本数のデータにも適用可能である。先行研究よりも緩い仮定のもとで統計量を構築し、汎用性も高い。応用統計学会の奨励論文賞を受賞した。

- (2) Ishii, Yata and Aoshima (2019, JSPI)では、統計量の漸近分布をカイ 2 乗分布に収束

させることで、高次元共分散行列の同等性検定に対する新たな検定手法を与えた。高次元共分散行列の推測においては、強スパイクする部分が非常に大きな情報を含むため、その部分を除去するデータ変換技術を用いることは適切でない。本論文では、強スパイクする部分とそれ以外の部分で共分散行列を分割することで、高次元共分散行列のもつ多大な情報を余すことなく利用し、高い精度保証を与えた。

(3) Ishii, Yata and Aoshima (2021, AISM)では、高次元共分散行列の構造検定について、候補となる共分散行列の構造が強スパイクモデルとそうでないモデル(弱スパイクモデル)であっても統一的に検定することに成功した。それぞれのモデルにおいて高次元中心極限定理を導出し、高い精度保証を与えた。強スパイクモデルにおいては、固有ベクトルの検定も与えた。

(4) Ishii, Yata and Aoshima (2022, JMVA)では、高次元2次判別分析に対する新たな手法を提案した。データ変換技術を応用し、さらに、高次元データのもつ球面集中現象という幾何学的表現を利用することで、高次元で一致性を得られる判別手法を提案した。

5. 主な発表論文等

〔雑誌論文〕 計11件（うち査読付論文 10件 / うち国際共著 0件 / うちオープンアクセス 4件）

| | |
|--|-------------------------------|
| 1. 著者名 Ishii Aki, Yata Kazuyoshi, Aoshima Makoto | 4. 巻 188 |
| 2. 論文標題 Geometric classifiers for high-dimensional noisy data | 5. 発行年 2022年 |
| 3. 雑誌名 Journal of Multivariate Analysis | 6. 最初と最後の頁 104850 ~ 104850 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jmva.2021.104850 | 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 青嶋 誠、石井 晶、矢田和善 | 4. 巻 73 |
| 2. 論文標題 論説：高次元小標本における統計的仮説検定 | 5. 発行年 2021年 |
| 3. 雑誌名 日本数学会邦文誌「数学」 | 6. 最初と最後の頁 360 ~ 379 |
| 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 石井 晶、矢田 和善、青嶋 誠 | 4. 巻 49 |
| 2. 論文標題 単一強スパイク固有値モデルにおける高次元平均ベクトルの2標本検定 | 5. 発行年 2020年 |
| 3. 雑誌名 応用統計学 | 6. 最初と最後の頁 109-125 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.5023/jappstat.49.109 | 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 Ishii Aki, Yata Kazuyoshi, Aoshima Makoto | 4. 巻 2157 |
| 2. 論文標題 Tests for high-dimensional covariance structures under the non-strongly spiked eigenvalue model | 5. 発行年 2020年 |
| 3. 雑誌名 数理解析研究所講究録 | 6. 最初と最後の頁 21-30 |
| 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 無 |
| 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 A. ishii, K. Yata, M. Aoshima | 4. 巻 2 |
| 2. 論文標題 Inference on High-Dimensional Mean Vectors under the Strongly Spiked Eigenvalue Model | 5. 発行年 2019年 |
| 3. 雑誌名 Japanese Journal of Statistics and Data Science | 6. 最初と最後の頁 105-128 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-018-0029-z | 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 A. ishii | 4. 巻 49 |
| 2. 論文標題 A. ishii | 5. 発行年 2020年 |
| 3. 雑誌名 Communications in Statistics - Theory and Methods | 6. 最初と最後の頁 1561-1577 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/03610926.2018.1528365 | 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 A. ishii, K. Yata, M. Aoshima | 4. 巻 202 |
| 2. 論文標題 Equality tests of high-dimensional covariance matrices under the strongly spiked eigenvalue model | 5. 発行年 2019年 |
| 3. 雑誌名 Journal of Statistical Planning and Inference | 6. 最初と最後の頁 99-111 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jspi.2019.02.002 | 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 A. ishii, K. Yata, M. Aoshima | 4. 巻 - |
| 2. 論文標題 A quadratic classifier for high-dimension, low-sample-size data under the strongly spiked eigenvalue model | 5. 発行年 2019年 |
| 3. 雑誌名 Stochastic Models, Statistics and Their Applications, Proceedings of the 14th Workshop on Stochastic Models, Statistics and their Application | 6. 最初と最後の頁 131-142 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-28665-1_10 | 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 A. ishii, K. Yata, M. Aoshima | 4. 巻 印刷中 |
| 2. 論文標題 Inference on high-dimensional mean vectors under the strongly spiked eigenvalue model | 5. 発行年 2019年 |
| 3. 雑誌名 Japanese Journal of Statistics and Data Science | 6. 最初と最後の頁 印刷中 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-018-0029-z | 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 A. ishii, K. Yata, M. Aoshima | 4. 巻 202 |
| 2. 論文標題 Equality tests of high-dimensional covariance matrices under the strongly spiked eigenvalue model | 5. 発行年 2019年 |
| 3. 雑誌名 Journal of Statistical Planning and Inference | 6. 最初と最後の頁 99-111 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jspi.2019.02.002 | 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 A. ishii | 4. 巻 印刷中 |
| 2. 論文標題 A classifier under the strongly spiked eigenvalue model in High-dimension, low-sample-size context | 5. 発行年 2019年 |
| 3. 雑誌名 Communications in Statistics -Theory and Methods | 6. 最初と最後の頁 印刷中 |
| 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 - |

〔学会発表〕 計29件 (うち招待講演 13件 / うち国際学会 11件)

| |
|---|
| 1. 発表者名 Kazuyoshi Yata, Aki Ishii and Makoto Aoshima |
| 2. 発表標題 Estimation of eigenvectors for linear combinations of high-dimensional covariance matrices and its application |
| 3. 学会等名 The 5th International Conference on Econometrics and Statistics (招待講演) (国際学会) |
| 4. 発表年 2022年 |

| |
|-----------------------------------|
| 1. 発表者名 石井 晶 |
| 2. 発表標題 強スパイク固有値モデルにおける高次元統計解析 |
| 3. 学会等名 2022年度統計関連学会連合大会（招待講演） |
| 4. 発表年 2022年 |

| |
|--|
| 1. 発表者名 Kazuyoshi Yata, Aki Ishii and Makoto Aoshima |
| 2. 発表標題 Tests for covariance structures in high-dimensional data |
| 3. 学会等名 The 4th International Conference on Econometrics and Statistics（招待講演）（国際学会） |
| 4. 発表年 2021年 |

| |
|--|
| 1. 発表者名 Aki Ishii, Kazuyoshi Yata and Makoto Aoshima |
| 2. 発表標題 High-dimensional classifiers under the strongly spiked eigenvalue model |
| 3. 学会等名 The 4th International Conference on Econometrics and Statistics（招待講演）（国際学会） |
| 4. 発表年 2021年 |

| |
|--|
| 1. 発表者名 Aki Ishii, Kazuyoshi Yata and Makoto Aoshima |
| 2. 発表標題 High-dimensional quadratic classifiers under the strongly spiked eigenvalue model spiked eigenvalue model |
| 3. 学会等名 IISA 2021 Conference（招待講演）（国際学会） |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 石井 晶、矢田和善、青嶋 誠 |
| 2. 発表標題 単一強スパイク固有値モデルにおける高次元平均ベクトルの2標本検定 |
| 3. 学会等名 2021年度統計関連学会連合大会（招待講演） |
| 4. 発表年 2021年 |

| |
|--|
| 1. 発表者名 矢田和善、石井 晶、青嶋 誠 |
| 2. 発表標題 高次元データにおけるノイズ構造の高精度な解析に基づく統計的推測 |
| 3. 学会等名 2021年度統計関連学会連合大会（招待講演） |
| 4. 発表年 2021年 |

| |
|-----------------------------------|
| 1. 発表者名 石井 晶、矢田和善、青嶋 誠 |
| 2. 発表標題 強スパイク固有値モデルにおける高次元2次判別 |
| 3. 学会等名 応用統計学会年会2021年年会 |
| 4. 発表年 2021年 |

| |
|--|
| 1. 発表者名 石井 晶、矢田和善、青嶋 誠 |
| 2. 発表標題 Tests of high-dimensional correlation matrices under the strongly spiked eigenvalue model |
| 3. 学会等名 2020年度統計関連学会連合大会 |
| 4. 発表年 2020年 |

| |
|-------------------------------|
| 1. 発表者名 石井 晶, 矢田和善, 青嶋 誠 |
| 2. 発表標題 高次元固有ベクトルの検定について |
| 3. 学会等名 日本数学会2020年度秋季総合分科会 |
| 4. 発表年 2020年 |

| |
|--------------------------------|
| 1. 発表者名 石井 晶、矢田 和善、青嶋 誠 |
| 2. 発表標題 高次元データに対する共分散構造の検定 |
| 3. 学会等名 京都大学数理解析研究所RIMS研究集会 |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 石井 晶、矢田 和善、青嶋 誠 |
| 2. 発表標題 データ変換を用いた高次元次判別分析について |
| 3. 学会等名 日本学術振興会科学研究費によるシンポジウム「多様な高次元モデルにおける理論と方法論, 及び, 関連分野への応用」 |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 石井 晶、矢田 和善、青嶋 誠 |
| 2. 発表標題 単一強スパイク固有値モデルに対する高次元平均ベクトルの2標本検定 |
| 3. 学会等名 日本学術振興会科学研究費によるシンポジウム「統計的推測および確率解析に関する総合的研究」 |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 Aki Ishii, Kazuyoshi Yata and Makoto Aoshima |
| 2. 発表標題 Tests for high-dimensional covariance structures under the SSE model |
| 3. 学会等名 日本学術振興会科学研究費によるシンポジウム「International Symposium on Theories and Methodologies for Large Complex Data」 (招待講演) |
| 4. 発表年 2019年 |

| |
|--------------------------------------|
| 1. 発表者名 石井 晶、矢田 和善、青嶋 誠 |
| 2. 発表標題 単一強スパイク固有値モデルにおける高次元二標本検定 |
| 3. 学会等名 日本数学会2019年度秋季総合分科会 |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 Aki Ishii, Kazuyoshi Yata and Makoto Aoshima |
| 2. 発表標題 Tests for high-dimensional covariance structures based on eigenstructures |
| 3. 学会等名 2019年度統計関連学会連合大会 |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 矢田 和善、石井 晶、青嶋 誠 |
| 2. 発表標題 Geometrical quadratic discriminant analysis for high-dimension, strongly spiked eigenvalue models |
| 3. 学会等名 日本学術振興会科学研究費によるシンポジウム「高次元複雑データの統計モデリング」 |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 Aki Ishii, Kazuyoshi Yata and Makoto Aoshima |
| 2. 発表標題 Inference on mean vectors for high-dimensional data with the strongly spiked eigenstructure |
| 3. 学会等名 The 3rd International Conference on Econometrics and Statistics (招待講演) (国際学会) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 Aki Ishii, Kazuyoshi Yata and Makoto Aoshima |
| 2. 発表標題 Tests of high-dimensional correlation matrices on the basis of eigenstructures |
| 3. 学会等名 The 7th International Workshop in Sequential Methodologies (招待講演) (国際学会) |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 石井 晶 |
| 2. 発表標題 強スパイク固有値モデルにおける高次元統計的推測 (日本数学会特別講演) |
| 3. 学会等名 日本数学会2019年度年会 (招待講演) |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 石井 晶、矢田 和善、青嶋 誠 |
| 2. 発表標題 強スパイク固有値モデルにおける固有空間の推測と高次元平均ベクトルの検定 |
| 3. 学会等名 京都大学数理解析研究所RIMS研究集会 |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 Ishii, A., Yata, K., Aoshima, M. |
| 2. 発表標題 Tests of high-dimensional mean vectors under the SSE model |
| 3. 学会等名 Waseda International Symposium "Introduction of General Causality to Various Data & its Applications" (国際学会) |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 Ishii, A., Yata, K., Aoshima, M. |
| 2. 発表標題 Tests of high-dimensional mean vectors under the SSE model |
| 3. 学会等名 International Symposium on Statistical Theory and Methodology for Large Complex Data (国際学会) |
| 4. 発表年 2018年 |

| |
|--|
| 1. 発表者名 Ishii, A., Yata, K., Aoshima, M. |
| 2. 発表標題 A quadratic classifier for high-dimensional data under the strongly spiked eigenvalue model |
| 3. 学会等名 2018年度統計関連学会連合大会 |
| 4. 発表年 2018年 |

| |
|---|
| 1. 発表者名 石井 晶、矢田 和善、青嶋 誠 |
| 2. 発表標題 Strongly spiked eigenvalue モデルにおける高次元相関ベクトルの検定について |
| 3. 学会等名 日本数学会2018年度秋季総合分科会 |
| 4. 発表年 2018年 |

| |
|---|
| 1. 発表者名 Ishii, A., Yata, K., Aoshima, M. |
| 2. 発表標題 Equality tests for high-dimensional covariance matrices |
| 3. 学会等名 The 27th South Taiwan Statistical Conference (27th STSC) (招待講演) (国際学会) |
| 4. 発表年 2018年 |

| |
|--|
| 1. 発表者名 Ishii, A., Yata, K., Aoshima, M. |
| 2. 発表標題 Equality tests of high-dimensional covariance matrices with strongly spiked eigenstructures |
| 3. 学会等名 The 2nd International Conference on Econometrics and Statistics (国際学会) |
| 4. 発表年 2018年 |

| |
|--|
| 1. 発表者名 Yata, K., Aoshima, M., Ishii, A. |
| 2. 発表標題 Inference on High-Dimensional Mean Vectors Under the Strongly Spiked Eigenvalue Model |
| 3. 学会等名 9-th International Workshop on Applied Probability (招待講演) (国際学会) |
| 4. 発表年 2018年 |

| |
|---|
| 1. 発表者名 石井 晶、矢田 和善、青嶋 誠 |
| 2. 発表標題 Tests for high-dimensional covariance matrices and correlation matrices under the strongly spiked eigenvalue model |
| 3. 学会等名 日本学術振興会科学研究費によるシンポジウム「融合する統計科学」 |
| 4. 発表年 2018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 |
|--|---------------------------|-----------------------|----|
|--|---------------------------|-----------------------|----|

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 |
|---------|---------|
|---------|---------|