

氏名	石井 晶
学位の種類	博士(理学)
学位記番号	博甲第 8010 号
学位授与年月日	平成 29 年 3 月 24 日
学位授与の要件	学位規則第4条第1項該当
審査研究科	数理物質科学研究科
学位論文題目	Two-Sample Tests Based on Eigenstructures in the HDLSS Context (高次元小標本における固有空間の構造に基づく二標本検定)

主査	筑波大学教授	博士(理学)	青嶋 誠
副査	筑波大学准教授	博士(理学)	小池 健一
副査	筑波大学准教授	博士(理学)	矢田 和善
副査	筑波大学准教授	博士(数学)	塩谷 真弘

## 論 文 の 要 旨

本論文は、遺伝子発現データのような高次元小標本データに対して、固有空間の構造に基づく二標本検定の構築を考えている。高次元小標本データは、数万という膨大な次元数に対して、それを解析するための標本数が数十程度の小標本であるという特徴をもつ。このように次元数よりも標本数が小さい場合、従来の多変量統計解析は、理論と方法論が破綻する。新しい統計学が必要となり、本論文では、高次元統計解析を展開している。標本数が極端に少ないので大標本漸近理論は使えない。本論文は、標本数を固定した高次元漸近理論を駆使して、高次元空間における統計的推測を構築している。

本論文は、以下の内容を扱った4つの章から構成されている。

第1章 第一固有空間の推定と漸近分布の導出

第2章 寄与率の信頼区間の構築

第3章 二母集団の共分散行列に対する同等性検定の構築

第4章 固有空間の構造に基づいた二標本検定の構築

本論文の全体を通して、高次元データの固有値に、Aoshima and Yata (2016, Sinica)によって提唱された強スパイクモデルが仮定されている。この固有値モデルは、例えば高次元共分散行列の固有値に次元数の冪乗関数を考えたとき、最初の数個の固有値が0.5以上の冪で発散するようなモデルである。多くの遺伝子発現データは、強スパイクモデルに適合することが知られている。強スパイクモデルのもと、推測の要となるのは第一固有空間である。第一固有空間を如何に精確に推定できるかが、高次元データの推測の良し悪しを決めてしまう。第1章では、まず、固有値と固有ベクトルの推定として、双対標本共分散行列

の固有値と固有ベクトルによる推定法を紹介している。従来型の推定法は、次元数の増加とともにバイアスが発散し、例えば、固有ベクトルの推定には強不一致性が現れることを、高次元漸近理論を用いて明らかにしている。第2章は、高次元におけるノイズ空間の漸近的性質を解析し、ノイズ空間の幾何学的表現を導出することでノイズの大きさを見積もり、発散するバイアスを補正した固有値と固有ベクトルの新たな推定法を与えている。さらに、小標本における第一固有空間の高次元漸近分布を導出することにも成功し、それを用いて寄与率の信頼区間を構築している。第3章は、二母集団の共分散行列に対する同等性検定を考えている。従来の同等性検定は、母集団に正規分布を仮定した限定的なものであった。本論文は母集団分布の正規性の仮定を緩め、一般的な高次元データに対して、まず、第一固有空間の同等性検定を与え、それにノイズ空間の解析を加味して、共分散行列の同等性検定を構築することに成功している。第4章では、以上の準備の下で、高次元小標本データに対する二標本検定を考えている。高次元統計解析において、次元数と標本数がともに発散する場合の先行研究がある。しかし、本論文で対象とするような小標本の場合には、標本数が限定された状況でも精度を保証することができる新たな二標本検定が必要になる。本論文は、強スパイクモデルの性質を利用して検定統計量を新たに導出し、標本数を固定した高次元漸近理論で漸近分布を導出することに成功している。本論文で構築された二標本検定は、第一種の過誤と検出力の両方で、高次元における精度を保証することが証明されている。

## 審 査 の 要 旨

〔批評〕

本論文は、高次元データ空間における固有空間の構造に基づいて二標本検定を構築すべき、という発想に根ざしている。大胆な発想であり、大変に興味深い結果を幾つか導いている。理論の構築には、高次元漸近理論が用いられている。高次元漸近理論には、(1)次元数と標本数が独立して発散する場合、(2)次元数が標本数と関数関係を持ちながら発散する場合、(3)標本数を固定して次元数だけが発散する場合、の3つがある。本論文は、高次元小標本データの統計的推測がテーマである。標本数は小標本に限定されるため、本論文では、(3)の高次元漸近理論を扱っている。実は、統計量の高次元漸近分布の導出は、(3)の場合が最も複雑になる。本論文は、第一固有空間の高次元漸近分布、共分散行列の同等性検定の高次元漸近分布、そして、二標本検定の高次元漸近分布などを、(3)の高次元漸近理論を駆使して導出することに成功している。また、本論文で構築された二母集団の共分散行列に対する同等性検定は、高次元データ解析の方法論にとって大変に意義深いものである。高次元空間において共分散行列の同等性を仮定することは、母集団分布に正規性を仮定することと同様に、非現実的である。しかし、実際に同等性を検定する方法は、これまで限定的な状況でしか開発されておらず、未開拓なままであった。本論文の検定方式により、高次元共分散行列の同等性を小標本であっても検定することが可能となり、共分散行列の共通性と異質性が高い精度で明らかになった。実際に、遺伝子発現データに応用すると、比較する二群に第一固有空間の共通性が見つかり、本論文では、それらの二標本検定に、固有空間の構造を利用した推測方式が展開されている。本論文の貢献は、高次元データの理論と方法論の両面で大変に高く評価でき、また、高次元データの固有空間の構造に基づいて統計的推測を構築する発想は極めて斬新で、今後、高次元統計解析のコアな考え方になるものと予想される。

〔最終試験結果〕

平成29年2月13日、数理物質科学研究科学学位論文審査委員会において審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、審査委員全員によって、合格と判定された。

〔結論〕

上記の論文審査ならびに最終試験の結果に基づき、著者は博士(理学)の学位を受けるに十分な資格を有するものと認める。