

学位論文題名

主成分分析におけるノンパラメトリック検定に関する研究

学位論文内容の要旨

本論文は、主成分分析における分散共分散行列の固有値問題(固有値と固有ベクトルに関連した問題)に対して、母集団分布の条件に依存しない方法であるノンパラメトリック検定法を提案し、かつ、その有効性を検証したものである。

分散共分散行列は多変量解析のほとんどの手法の出発点となるものであり、したがって、あらゆる分析に先立って、その構造を明らかにすることはきわめて重要な作業である。とりわけ、固有値と固有ベクトルは直交変換に対して不変な特性値であり、その構造を把握する上で最も重要な指標である。

分散共分散行列の固有値問題に関する研究は歴史的なテーマであり、これまで多くの研究者によって研究されてきた。1939年は多変量解析にとって、いろいろな意味でbreakthroughの年であったとJacksonも指摘しているが、この年、Girshickは多変量正規分布の仮定のもとで分散共分散行列の固有値、固有ベクトルの最尤推定値に関する極限分布に関して重要な結果を与えている。その後、Andersonをはじめ、多変量正規分布のもとでさまざまな推定・検定問題に関する結果が導かれた。さらに、多変量正規分布が仮定できない場合についてもFujikoshiなどの漸近展開による方法やEfronの提唱したbootstrap法を応用する手順が開発されてきた。そして最近でも、Tsukadaらが固有ベクトルに関する新しい検定統計量を提案してその漸近的な性質を調べるなど、現在でもこのテーマに関する研究が続けられている。これまでの研究の特徴は、一部の精密な理論を除いて、多変量正規分布を仮定する場合も、またそうでない場合も、ほとんどは解析的に漸近分布を導出するなどの研究が中心であり、前提条件が厳しかったり、あるいは、漸近的な分布式の中に、再び未知の母数が多く含まれるなど、実用的な視点から見た場合、多くの困難な問題も抱えているように思われる。

理論的な研究は極めて難しい一方で、応用面での適用範囲は極めて広範に及んでいる。分散共分散行列の固有値問題を直接分析の対象とする手法は主成分分析であるが、その適用事例として、例えば、塩谷や杉山は、動物の骨の測定データの分析、単語に対する小学生の反応評価、成績データの分析、歯のこう耗度に関する分析、民族間の身体的特徴の比較、などを扱っているし、また、刈屋は計量経済分析における応用について議論している。さらに、柳井・前川は大学入試データの解析に応用している。これらは事例のほんの一部に過ぎず、実際のデータ解析の場面での応用事例は非常に多く、多変量解析の中でも、あるいは統計的手法全体の中でも最も応用される手法の一つと言える。

このような現状を考えると、あまり厳しい前提条件がなくても適用できる簡便な方法の開発が実際の場面では望まれるところである。

本論文は、以上のような背景のもとで、これまでの研究の経緯とは全く異なる観点から開発した検定手法を提案している。これまでの研究が母集団分布や母数に関していくつかの前提条件のもとで解析的な方法をベースとしているに対して、ここで提案している考え方は発見的かつ実用的な方法論である。手順はきわめて簡便であり、かつ、正規分布に関する知識

があるだけで応用できる。

本論文で扱っている固有値問題は、固有値の2標本問題と固有ベクトルの1標本、2標本問題であり、方法論としての特徴を挙げると以下のようなものである。

- (1) 固有値の問題も固有ベクトルの問題もともに共通の等分散性の検定問題として定式化される。このことは一つの発見である。
- (2) 等分散性のノンパラメトリック検定法が利用される。2つの方法が基本技法として用いられるが、そのうちの一つである Moses 検定は諸条件に対して極めて頑健な方法として再発見されたと言える。
- (3) 標準正規分布のパーセント点だけを用いて検定を行う。上の(2)の方法は、多変量解析が適用されるような多くのサンプルが利用できる場面では、正規近似の精度が実用的に十分である。

本論文の構成を以下に示す。

第1章では、本論文の背景と目的、及び章構成について述べている。

第2章では、分散共分散行列に関する固有値問題と主成分の関係について定義し、さらに、固有値や固有ベクトルの期待値や分散、あるいは仮説検定などの統計的推測法に関して、これまでに得られている主要な結果をまとめている。

第3章では、等分散性検定のための2種類のノンパラメトリック検定法について解説している。一つは Ansari-Bradley 検定、もう一つは Moses 検定である。これらの方法が4章以降の固有値問題で主要な役割を果たす。ここでは、手順や前提条件と適用上の留意事項などについて解説している。

第4章では、多変量正規分布を仮定した場合の固有値の2標本問題が扱っている。多変量正規分布を前提とする限りはこれまでの理論でも解決できる問題はあるが、ここでは主成分スコアにもとづくノンパラメトリック検定法の適用可能性について議論している。Ansari-Bradley 検定がここでは適用されている。主成分スコアを用いる場合の適用上の課題についても議論している。

第5章では、非正規性のもとでの固有値の2標本問題が扱っている。非正規分布として、elliptical 分布と lognormal 分布が取り上げられ、いろいろな分布条件のもとでシミュレーション実験によってノンパラメトリック検定の有効性が検証している。3章の方法に加えてここでは Moses 検定が導入され、分布の歪度や尖度と方法論の間の比較などに関して議論している。

第6章では、固有ベクトルの検定問題を扱っている。まず、固有値の2標本問題と共通の手順が適用できることが示されている。ここでは Moses 検定が主要な役割を果たす。主成分スコアを用いた4通りの検定法が提案され、lognormal 分布の母数をいろいろに変化させた形での実用性がシミュレーション実験によって検証している。さらに、2標本問題にも拡張している。

第7章では、まとめとして、本論文で提案したノンパラメトリック検定法の特徴と有効性について総括している。

学位論文審査の要旨

主 査 教 授 佐 藤 義 治
副 査 教 授 宮 腰 政 明
副 査 教 授 工 藤 峰 一
副 査 教 授 水 田 正 弘

学 位 論 文 題 名

主成分分析におけるノンパラメトリック検定に関する研究

主成分分析は、多次元データの縮約化・構造把握のための技法として、工学、経済学、心理学、医薬学など極めて広範な分野で応用される多変量解析の代表的な手法である。主成分分析の主要な統計量である固有値や固有ベクトルに関する研究は、歴史的なテーマであるものの、正規分布が前提になるなどの制約があり、実用的な観点からの研究は十分とは言えない段階にある。

本論文は、特に分散共分散行列の固有値問題(固有値と固有ベクトルに関連した問題)に対して、母集団分布の条件に依存しないノンパラメトリック検定法を提案し、かつ、その有効性を検証したものである。

主成分分析における固有値問題に関するこれまでの研究は、多変量正規分布を仮定する場合もまたそうでない場合も、解析的に漸近分布を導出することが中心であり、現実的視点からは前提条件が厳しく、また、漸近的な分布式の中に、再帰的に未知母数が含まれるなど、多くの困難な問題も抱えていた。以上のような背景の下で本論文は、従来の研究の経緯とは全く異なる観点から開発した検定手法、即ち、固有値問題に対して、等分散性のノンパラメトリック検定法を応用した発見的かつ実用的な方法論を開発したものである。

本論文は7章から構成され、各章における内容と成果はつぎの通りである。

第1章では、序論として、本論文の背景や目的、論文全体の構成を示し、第2章では、分散共分散行列に関する固有値問題と主成分の関係について定義し、さらに、固有値や固有ベクトルの期待値や分散、あるいは仮説検定などの統計的推測法に関して、従来得られている主要な結果をまとめている。これまでの研究成果から、固有値問題に影響を与える原因が、隣接する母固有値の比である点を指摘し、このことが本論文で提案する方法論の検証の論拠となっている。

第3章では等分散性検定のための2種類のノンパラメトリック検定法について解説している。一つは Ansari-Bradley 検定、もう一つは Moses 検定である。これらの方法が4章以降の固有値問題で主要な役割を果たす。ここでは、各方法の手順をまとめるだけにとどまらず、適用上の前提条件や留意事項について解説されており、4章以降の中では、それらの前提条件も議論しながら方法論が展開されている。

第4章では、固有値の2標本問題に対してノンパラメトリック検定法を提案し、さまざまな条件下で精度を検証している。方法論としての特徴は、固有値そのものを検定統計量とし

て用いるのではなく、主成分スコアをベースとする等分散検定の問題に帰着させるというアイデアである。本章では多変量正規分布が前提とされており、したがって、従来の研究結果と著者が開発した方法の比較が可能になるが、その結果として、固有値問題に対してノンパラメトリック検定法の適用可能性が確認できたという点で本章の研究意義は大きい。

第5章では、非正規性のもとでの固有値の2標本問題を扱っている。本来ノンパラメトリック法は母集団分布に関する条件に頑健であることが特徴であり、3章で用いた Ansari-Bradley 検定法に加えて、ここでは Moses 検定も導入され、母集団分布の歪度や尖度と方法論の比較に関して議論している。結果的に、F 検定は適用できないことを確認し、Moses 検定が最も頑健であることを明らかにしている。また、従来の漸近理論では母数が再起的に用いられる形でしか適用できなかった問題であり、その意味でも、まったく新しい方法論の開発としての意義は大きく、応用範囲を拡大することにつながるものと考えられる。

第6章では固有ベクトルの検定問題を扱っている。まず、固有値の2標本問題と共通の手順が適用できることを示し、Moses 検定を応用した4通りの検定法が提案されている。さらに、2標本問題にも拡張している。これらの方法論の考え方と手順は、極めて簡便であり、検定仮説に対する反応が直感的に把握し易い形になっている。固有ベクトルの検定問題に関しては、従来、正規性を仮定した Anderson(1963)の方法以外には有効な方法がなかったと言っても過言ではない。固有ベクトルの検定問題に対しても、固有値の検定問題と同じ方法論が利用できるというアイデアは、実際的な立場から大変優れたものであり、著者によって開発された方法は、極めて有用性が高いと考えられる。

7章ではまとめとして、本論文で提案したノンパラメトリック検定法の特徴と有効性について総括している。

これを要するに、著者は、非正規母集団の仮定の下で、従来の方法ではほとんど無力であった、主成分分析における固有値や固有ベクトルの検定問題に関して、等分散性のノンパラメトリック検定法を応用するという観点から新しい方法論を開発し、その有効性を示したものであり、情報学、計算機統計学の発展に寄与するところ大なるものがある。

よって、著者は北海道大学博士(工学)の学位を授与される資格あるものと認める。