

## 母集団間で分散共分散行列が異なる場合における 統計的手法の開発

### 研究者の所属・氏名等

フリガナ 氏名	カワサキ タマエ 川崎 玉恵
所属等	東京理科大学 理学部第一部 応用数学科 嘱託助教
プロフィール	2007年3月 私立多摩大学目黒高等学校 卒業 2011年3月 東京理科大学理学部第一部数理情報科学科卒業 2013年3月 東京理科大学大学院理学研究科数理情報科学専攻 修士課程修了 2016年3月 東京理科大学大学院理学研究科数理情報科学専攻 博士後期課程修了(博士(理学)) 2016年4月 東京理科大学理学部第一部数理情報科学科 嘱託助教 2017年4月 応用数学科(数理情報科学科から名称変更) 嘱託助教 現在に至る

### 1. 研究の概要

近年盛んに発展しているデータサイエンスや機械学習、人工知能の分野は、統計学の一分野である数理統計学が基礎理論となって構築されている。数理統計学は統計学の様々な分野においても基盤となっている重要な学問であり、その中でも平均における仮説検定問題は、実験や観測を行った際にグループ間で差があるのかを調べるなどにおいて有用であり、品質管理や臨床医学をはじめとしたあらゆる応用分野で利用されている。さらに、数理統計学における理論を、複数の変数を同時に扱うことができるようにと発展して生まれた多変量解析は、世の中の多くのデータに適応できることから、様々な分野で活用されている。

本研究では、多変量解析において難しい問題としてよく知られる母集団分布の分散共分散行列が異なる場合における平均ベクトルの仮説検定問題について議論を行った。

### 2. 研究の動機、目的

申請者は多変量解析を専門としているが、その中で多群における平均ベクトルの仮説検定問題は、母集団分布の分散共分散行列が母集団間で「等しい」か「等しくない」かによって検定手法としての扱いやすさが変わってくることを知った。分散共分散行列が母集団間で「等しい」場合には、検定統計量として Hotelling の  $T^2$  統計量（1変量の場合における  $t$  統計量に対応するもの）がよく知られており、この検定統計量は正確に  $F$  分布に従うため仮説検定を容易に行うことができる。一方で、分散共分散行列が母集団間で「等しくない」場合には、様々な検定統計量が議論されているが、その正確な分布の導出は難しく、パーセント点の値が容易には求められないことがため「(多変量) Behrens-Fisher 問題」と呼ばれ、難しい問題として知られている。この問題に対する検定統計量の近似分布に関しては、Cornish-Fisher 展開や Bartlett 補正を用いて  $\chi^2$  分布に近似させる方法や、 $\chi^2$  分布の自由度を調整して与える方法等が先行研究でなされている。しかしこれらの提案手法は、母集団分布の分散共分散行列の構造による影響が大きく、さらに次元数に対してサンプル数が少ない場合や次元数とサンプル数が近い場合における近似が良くない。

そこで申請者は解決策のひとつとして、検定統計量を摂動展開することにより  $F$  分布の自

由度を調整して近似解を与えた Yanagihara and Yuan (2005, *Comm. Statist., Simulation Comput.*) に着目した。しかし、この結果では上記のような近似が悪い場合での改善ができていないため、申請者は Kawasaki and Seo (2015, *Comm. Statist., Simulation Comput.*) で検定統計量の摂動展開のオーダーを先行研究よりも高次まで与えることで、近似精度の改良を行った。申請者による提案によって、先行研究で提案されてきた結果における近似の悪い部分に対する改良はなされたが、新たな問題としては式が複雑であり、応用場面では使いにくいという問題が生じた。そこで本研究課題は、近似の悪さを改良しつつも、より簡便な近似解の提案を目指すものである。

### 3. 研究の結果

まずは応用分野への適応を目指し、より簡便な近似解の提案を目指した。具体的には、Kawasaki and Seo (2015) で与えた近似分布として用いた F 分布の自由度を調整する方法について、推定するパラメータを 1 つ増やすことにより高次までの摂動展開をすることなく、近似精度の改良と近似解の簡便化を考えた。

得られた結果としては、まずパラメータが増えたことにより、検定統計量の摂動展開を行い、その後期待値を求める際に必要となる期待値計算の定理が必要となる。この定理となる期待値の計算公式は、多変量解析であるため複数の変数に対して分布を仮定し、それらの期待値を個別に計算していく必要があるため、パソコンを用いてプログラムによる計算を行い、本研究のみならず他の研究にも有用となる期待値の定理を複数与えることができた。さらに書籍により、申請者がこれまでの研究では行ってこなかった分散共分散行列に構造を与える手法についても学び、本研究の仮定として用いることを現在検討している。分散共分散分析に構造を仮定することは、実際の問題に適応させる場合に有用となることも多く、これまで数多くの先行研究が行われている。

### 4. これからの展望

例えば環境問題における水質や土壌の調査において、観測地点で得られるデータは、どんなに同じ条件・状況を目指していてもある程度の観測誤差が生じてしまう。さらに、採取する環境や条件を統計的に差のない状況にして行えるかどうか大きな問題となり、この場合の分散や変数間の共分散は母集団間で差が生じやすくなる。このように、観測・測定する計画とは別の要因で、どうしても同等な調査が行えないというケースは環境問題に関わらず、多く存在している。また、このようなケースで得られたデータには欠測値を含んだものが多く存在する。以上の観点から、本研究は理論的な貢献だけではなく、様々な分野への発展が期待できると確信している。さらに、得られたデータから分散共分散行列の構造などのモデル式を作ることにより、その採取環境等に即した (多変量) Behrens-Fisher 問題への当てはまりの良い近似解の提案が行える。未解決な問題である分、多くの発展性が見込める研究課題であるため、理論的かつ応用面も視野に入れた包括的な研究を行っていきたい。

### 5. 社会に対するメッセージ

本研究課題は、まずは理論的な部分における発展を目指しており、様々な現象や状態を数学的に一般化した形での提案となる。しかし、(多変量) Behrens-Fisher 問題は多くの応用分野が直面している問題であるにも関わらず、使いやすく有効である決定的な手法の提案には至っていない現状がある。そこで本研究課題は、理論的な立場ではあるが、応用面で直面している問題について挑み、統計学における理論的な発展と、データサイエンスが進む社会に対する貢献の 2 つを目指している。

理論的な基礎研究の積み重ねが、今日の応用分野の発展につながっている一方で、応用面には直接的に結び付かない結果があるのも事実である。貴事業団からのご支援により、これまで改めて向き合うことのなかった構造に関する研究など、多くのことを学ぶ機会となり、今後の研究への大きな基盤づくりができた 1 年であった。このような理論的な基礎研究に対し深いご理解を示していただき、ご支援くださったことに対して深く感謝申し上げます。