

群サイズを固定した個体判別法の研究開発

斉藤 光平

1. 研究背景と目的

多変量データを取り扱う際、所属する群がわかっている対象個体が、どのような基準で群に分けられているかを判別し、また、その判別の基準に従って新規個体をいずれかの群に分類するために、判別分析と呼ばれる多変量解析の手法が用いられる。さらに、対象データの個体判別・分類を目的とする場面は多く、例えば、心理学や教育学、工学や医学といった幅広い分野で判別分析法が活用されている。

しかし、従来の判別分析法では、100 人を 10 人と 90 人の群に分けるといったような、各群のサイズを固定した状態で判別基準を求めることは困難であった。そこで本研究では、2 群の判別において、各群のサイズを固定したまま、最適な判別基準を求めるためのアルゴリズムを考案し、それをサイズ固定判別分析(Size Fixed Discriminant Analysis;SFDA)と名付けた。さらに、SFDAを用いて3つの実データを分析し、それに基づき、最後に考察を行った。

2. 提案手法

SFDA の目的関数は、データ行列の右から重みベクトルをかけたものと、置換行列・メンバシップ行列・重心ベクトルを、それぞれ順にかけたものの差を最小化する形で定式化される。このとき、推定すべきパラメータは重みベクトル、置換行列、重心ベクトルの 3 つである。一般的な判別分析法では、メンバシップ行列もパラメータとして扱い、推定の対象となるが、SFDA では、メンバシップ行列の各列に含まれる 1 の個数を固定し、置換行列を用いてメンバシップ行列の行を置換することによって、群サイズを固定した状態でパラメータの推定が可能となるため、メンバシップ行列はパラメータとして扱わない。

これらのパラメータは、交互最小二乗法と呼ばれるアルゴリズムから推定される。交互最小二乗法とは、複数のパラメータを推定する際、最小二乗法を用いて 1 つ 1 つのパラメータを交互に推定する計算アルゴリズムのことを意味する。

また、提案手法において、置換行列は最小二乗置換(足立, 2011)と呼ばれる計算法を用いることで推定される。さらに、重みベクトルは回帰問題(ten Berge, 1993)、及び、重心ベクトルは固有値問題(ten Berge, 1983, 1993; 足立・村上, 2011)に帰着させて解くことで、推定値を導出することができる。これらの計算を、任意の収束基準に到達するまで反復し、収束基準に到達した際に得られた値を、各パラメータの解とした。

3. 実データ解析

3 種類のデータ(Fisher, 1936; Hartigan, 1977; MacFarland, n.d.)を用いて、SFDA の有用性を確認した。Fisher(1936)のアイリスデータのうち、setosa 種と versicolor 種の 2 群だけを用いて、SFDA の判別基準の正確さと、新規個体の分類の様子を確認した。その結果、群サイズがデータにおける各群のサイズと一致している場合、設けられた判別基準は全ての個体を正しく判別できており、新規個体も正しい分類を得ることができた。

さらに、Harigan(1977)の合衆国最高裁判事のデータと、MacFarland(n.d.)の SAT & College GPA データを SFDA に適応させた。その結果、合衆国最高裁判事のデータにおいては、群サイズの変動によらず、いずれの分け方であっても、サイズがより少数となる群の個体を選ぶ際に重要となる変数があることがわかった。また、SAT & College GPA データにおいては、群サイズの変動によって、サイズがより少数となる群の個体を選ぶ際に重要となる変数が異なることがわかった。

4. 総合考察

本研究で提案した SFDA は、群サイズを固定した状態で個体の所属群を判別するための手法であった。また、その上で 2 群に判別できるため、提案手法は複数科目の試験による選抜などで有用となると考えられる。なぜならば、試験では受験者の人数に関わらず、試験に合格できる人数は予め決まっているからである。このように、決まった人数を選抜する際、群サイズを固定した状態で判別を行うことのできる SFDA が有用であることがわかる。さらに、既に得られた判別基準を用いて、同じレベルの選抜を行うことができるという利点も挙げられる。以上のことから、SFDA は、特に人数制限のある選抜を行う場面や、その基準を用いた新しい受験者の合否などを考える際に有用となり、実用性の高い分析手法であると言える。

ただし、本研究において群に関する制約として課しているのは、あくまでも各群のサイズ数の条件だけである。すなわち、サイズ数が少数である群が相対的に優秀な群、つまり、上位群であるという保証は得られない。そのため、上述のように、試験において上位の何名かを選抜するためには、サイズ数がより少数の群が、より優秀な個体を選抜するような制約を SFDA に課した分析手法の開発が、本研究の今後の課題である。(行動統計科学)