

クラスタリングを伴う因子分析法の研究開発

宇野 光平

因子分析は人の知能を説明するための数理モデルとして提案された経緯があり、心理統計学において数理的発展を遂げてきた。また心理学以外の分野でも、因子分析による説明は直感的かつ有効であることから応用可能な分野は数多く存在する。因子分析の基本的な考え方はデータの分割であり、データを共通部分と独自部分の2つの側面から説明する。各個人が共通部分や独自部分でどのような影響下にあるかは、それぞれ共通因子と独自因子得点によって知ることができる。これら各個人の因子得点の扱いによって因子分析を3種類のモデル、変量モデル・母数モデル・行列モデルに分類することができる。変量モデルは、因子得点を確率に基づいて変動する確率変数とみなすもので因子得点をパラメータとして最適推定することはできない。母数モデルは独自因子得点を確率変数として扱うものの、共通因子得点はパラメータと扱うため、その最適値の推定は可能である。しかしながら最尤法による推定ができないという問題がある。行列モデルは、共通因子得点と独自因子得点それぞれをパラメータとみなすモデルであり、最適な因子得点を数式で表現できるが、解が一意に定まらないという問題がある。本学位論文では上記の3つのモデルのうちの母数モデルと行列モデルに着目し、クラスター制約、つまり各個人の共通因子得点はグルーピング可能であるという制約を導入することで、問題の解決を目指す。

本論文は主に2つの新手法の提案で構成されるが、1つめの提案手法では母数モデルにクラスター制約を導入した。この結果、パラメータである共通因子得点が自由に取れる値の範囲が狭まり最尤推定が可能となった。この提案手法は、いままで最尤推定が不可能であった母数モデル因子分析を、制約を加えることにより最尤推定可能とした新たな手法である。一方で、クラスター分析の文脈でこの提案手法を考えると、共通因子得点によるクラスター分析と捉えることができる。このとき、共通因子得点の推定とクラスター所属の推定を同一のプロセス、つまり尤度最大化を行う中で共通因子得点の推定とクラスター所属の推定を同時に達成している点が重要である。これは、たとえば変量モデル因子分析を行ったあと何らかの因子得点の推定をし、その因子得点を用いてクラスター分析を行う、タンデム分析とは明らかに意味合いが異なる。なぜならば、タンデム分析における因子分析では個人のクラスターはパラメータである因子負荷量と独自分散の推定には全く影響を及ぼさないのに対し、提案手法は個人のクラスターを考慮したうえでパラメータの推定を行なっているからである。この結果、提案手法はタンデム分析では抽出できない因子構造を抽出することができる。実際に、タンデム分析と比べた同時的分析の有用性は、クラスタリングを伴う主成分分析でも確認されている。また、実データを用いた分析例により、クラスタリングを伴う主成分分析よりも提案手法であるクラスタリングを伴う母数モデル因子分析の分類精度のほうが、優れていることが経験的に示された。この結果は、独自分散に強い仮定を加えることで因子分析と主成分分析が一致し、より柔軟にパラメータ推定を行える因子分析のほうがより良い推定が可能である事実を示唆している。事実、個人の分類には寄与しない変数の独自分散は大きな値をとる傾向にあり、逆に個人の分類に重要である変数は小さな値をとる傾向にあった。本論文では、独自分散の値を変数ごとに自由に推定可能であることが、クラスタリングを伴う主成分分析よりも提案手法が優位である理由と推察した。

もう1つの提案手法は、行列モデルに関する手法である。行列モデル因子分析では、共通因子得点と独自因子得点それぞれがパラメータとして扱われるため、個人ごとの因子得点が最適推定可能である。しかしながら、この行列モデルには変量モデルと同様に2つの不定性が存在する。回転の不定性と因子得点の不定性である。後者の因子得点の不定性は、因子得点は解が一意に定まらないことを意味している。

このため、個人ごとに得られる因子得点は解釈として用いることは難しいという問題があった。そこで、この因子得点を一意に定める新たな手法を提案した。この手法は、回転の不定性を利用して解釈しやすい因子負荷量を得る手法である回転法に着想を得ており、因子得点の不定性を利用して解釈しやすい因子得点を得ることができる。ただし何らかの基準を作り、その基準を最大化あるいは最小化することが必要となる。提案手法では、解釈しやすい因子得点をターゲットパラメータとして設定し、そのターゲットパラメータと実際の因子得点との差が最小となるようにした。解釈しやすい因子得点を今回はクラスター構造を持つ共通因子得点とした。シミュレーションや実データへの解析例によって有用性を確認し、さらには独自因子得点も一意に定めることができることから独自因子得点の解釈にも言及した。

前述の通り、因子分析において回転の不定性は解釈を行ううえで避けることのできない問題である。しかし、これは因子負荷量への制約を加えたうえでの推定によって回避可能である。こうした研究はスパース推定と呼ばれるもので、因子分析も多くの多変量解析法と同様にスパース推定可能であることが近年の研究で明らかになった。本論文では、このスパース因子分析の他にも因子負荷量に制約を加えた手法を紹介し実データへの適用例を示すことで、今後スパース因子分析が今まで別々に考えられてきた探索的因子分析と検証的因子分析の橋渡しの手法になりうることを考察した。(行動統計科学)