

パス解析モデルのスパース同定に関する研究開発

稲岡 雄介

パス解析は観測変数間の因果関係を調べる分析法であり、重回帰分析よりも複雑なモデルを考えられるモデルである。パス解析は構造方程式モデリング(SEM)の枠組みの中で発展してきたが、確認的手法でありソフトウェアで実行する際は分析者自身がどの部分にパスがあるか指定したうえでパス係数や誤差分散を推定していた。一方統計学・機械学習の分野において推定する係数や行列の一部の要素を0と推定するスパース推定に関する研究が盛んになり、SEM の下位モデルである因子分析(Factor Analysis)でもスパース因子分析 (Adachi & Trendafilov, 2014; Hirose & Yamamoto, 2014)が開発された。因子分析とは違い、パス解析には探索的モデルは主流ではなく事前にパスの位置を指定する手間があるため、本研究ではスパースの理論を適用しその手間を省けるモデル・アルゴリズムを提案した。

大きさ p のデータベクトルを $z=[x',y']'$ とし、 x は大きさ q の説明変数ベクトル、 y は大きさ r の従属変数とし、分析者は事前に説明変数と従属変数に変数を分ける必要がある。またパス係数行列を A 、誤差ベクトルを e として、提案手法のモデルは

$$y = Az + e$$

となる。また識別性のための制約と、説明変数と従属変数に多変量正規分布を仮定すると、このモデルの対数尤度関数は

$$L(\Theta) \propto -n/2 \{ \log|\Psi| + \text{tr}(S_{YY} - 2S_{YZ}A' + AS_{ZZ}A') \Psi^{-1} + \log|\Phi| + \text{tr}S_{xx}\Phi^{-1} \}$$

と書くことができる。また本研究の目的はパス係数行列 A のスパース推定であるため $L(\Theta)$ にペナルティを加えた式を目的関数とする。ペナルティには MCP (Minimax Convex Penalty) (Zhang, 2010) を用いる。ペナルティに L_1 ノルムを用いる Lasso (Tibshirani, 1996) は値を縮小推定してしまう欠点があるが、MCP はペナルティの強さを示すパラメータ λ に加え、軟閾値作用素と硬閾値作用素のどちらに近づくかを決めるパラメータ γ があり分析者自身が決定できる。推定するパラメータ Θ は A 、 Ω 、 Ψ の 3 つであるが A 、 Ψ は解析的に解が得られないため交互に反復を行う。 A の更新は目的関数を要素ごとに展開し、座標降下法 (coordinate descent method) (Friedman et al. 2010) を用いて更新を行う。

数値シミュレーションを行い、データ数 n が大きくなれば真値と推定値の二乗誤差が小さくなり、非ゼロ位置の再現も $n=500, 1000$ の場合においては 100%に近い値を示した。また実データ解析を 3 つのデータに対して行い、従来のパス解析よりも高速に解釈のしやすい解が得られ、提案手法の有用性が示された。

今後の課題としては依然分析者が決める説明変数と従属変数の因果関係を間違えたときの挙動や得られた解の性質、分析者が決めるペナルティのパラメータ λ と γ の選択基準を考えることなどがあげられる。また識別可能な別の制約を用いたモデルや正規分布以外の分布を用いる、ペナルティ関数を変えるという拡張が考えられる。(行動統計科学)