柔軟で反復計算を減らした精度行列の縮小推定法の提案とその特徴

中京大学 国際教養学部 永井 勇

各個体が p 次元のデータからなる n 個の個体の多変量データ $\boldsymbol{Y}$ の分析を考える．ここで, 個体間のデータは無相関で, 各個体の分散共分散行列を $\boldsymbol{\Sigma}$ (未知の正定値行列) とする, つまり $\mathrm{Cov}[\mathrm{vec}(\boldsymbol{Y})] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$ とする ($\mathrm{vec}(\cdot)$ は vec 作用素, $\otimes$ はクロネッカー積である)．このとき, よく使われる $\boldsymbol{\Sigma}$ の推定量は

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{Y}' \left\{\boldsymbol{I}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\right\}\boldsymbol{Y},$$

である (例えば Schott (2017) 参照)．ここで $\mathbf{1}_n$ は全ての要素が 1 の n 次元ベクトルである．また, $\boldsymbol{S}$ の固有値を $d_1, \ldots, d_p$ とすると, $\boldsymbol{S}$ の求め方から $d_i \geq 0$ $(i = 1, \ldots, p)$ である．

判別分析やモデル選択などの $\boldsymbol{Y}$ の様々な分析においては, 非常に多くの場面で $\boldsymbol{S}^{-1}$ (精度行列) が必要となる．しかし, $\mathrm{rank}(\boldsymbol{S}) < p$ の場合, いくつかの i で d_i が 0 となるため, $\boldsymbol{S}^{-1}$ は存在しない．また, $\mathrm{rank}(\boldsymbol{S}) = p$ であっても, $\boldsymbol{Y}$ の列に相関が強い組がある場合, d_i が 0 に近い値となるため, $\boldsymbol{S}^{-1}$ が不安定になってしまう．

そこで, $\boldsymbol{S}^{-1}$ の代わりとして以下の様々な形での逆行列を用いる手法が提案されている;

- Moore-Penrose 型一般逆行列
- $\boldsymbol{S}$ の非対角成分を全て 0 にした行列の逆行列
- $\boldsymbol{S}$ にリッジ型の罰則を付けた行列の逆行列

Moore-Penrose 型一般逆行列については Schott (2017) Sec. 5 などを参照, 他はそれぞれ Srivastava ら (2013), Wang ら (2015) で提案されている．本講演では, これらの手法の中でリッジ型の罰則を付けた行列の逆行列を $\boldsymbol{S}^{-1}$ の代わりに用いる手法に着目する．この手法は, 二つの正のパラメータ λ と α を用いて, $\lambda(\boldsymbol{S} + \alpha\boldsymbol{I}_p)^{-1}$ を $\boldsymbol{S}^{-1}$ の代わりに用いる手法である．この手法は固有値を通じて見ると, α (> 0) を $d_1, \ldots, d_p$ に加えることで $(\boldsymbol{S} + \alpha\boldsymbol{I}_p)^{-1}$ の存在を保証し, λ によりスケール調整を行う手法である．

この手法には二つの問題点がある．一つ目は, p 個全ての固有値に対して一つの正のパラメータ α だけを用いて一様に調整を行っている点である．二つ目は, λ と α の同時最適化のための反復計算が必要となる点である．本講演では, 一般化リッジ回帰による推定法 (Hoerl & Kennard, 1970) を用いて, これらの問題点を回避する手法を提案する．

具体的には, $\boldsymbol{Q}$ を $\boldsymbol{S}$ を対角化する直交行列として, $\boldsymbol{S}^{-1}(\lambda, \boldsymbol{\theta}) = \lambda(\boldsymbol{S} + \boldsymbol{Q}\mathrm{diag}(\boldsymbol{\theta})\boldsymbol{Q}')^{-1}$ を $\boldsymbol{S}^{-1}$ の代わりに用いる手法を提案する．ここで, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ $(\theta_i \geq 0)$, $\mathrm{diag}(\boldsymbol{\theta})$ は $\theta_1, \ldots, \theta_p$ を対角に並べた行列である．この $\boldsymbol{S}^{-1}(\lambda, \boldsymbol{\theta})$ は, 各 i で d_i に θ_i を加えてから逆行列を取り, λ でスケール調整を行う推定量である．パラメータ数が $(p+1)$ 個になってしまう一方で, 各パラメータ θ_i を各固有値 d_i の大きさに合わせて調整することができるので, リッジ型の罰則を付ける手法に比べて非常に柔軟な調整が可能な形となっている．本講演ではさらに, 提案する推定量での λ や $\boldsymbol{\theta}$ の最適化について考える．例えば, Wang ら (2015) でも用いられている次のロス関数の最小化に基づく最適化を考える;

$$L^{[2]}(\lambda, \boldsymbol{\theta}) = \mathrm{tr}\left\{\left(\boldsymbol{S}^{-1}(\lambda, \boldsymbol{\theta})\boldsymbol{\Sigma} - \boldsymbol{I}_p\right)^2\right\}.$$

本講演では, λ を固定した下で $L^{[2]}(\lambda, \boldsymbol{\theta})$ を最小にする $\boldsymbol{\theta}$ が陽に求まることを示す．また, 得られた $\boldsymbol{\theta}$ を用いた場合の $\boldsymbol{S}^{-1}(\lambda, \boldsymbol{\theta})$ の特徴についても述べる．さらに, λ の最適化についても述べる．

また本講演ではその他の様々なロス関数についても触れ, 数値実験による比較などについて報告する．

引用文献:

[1] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics*, **12**, 69–82.

[2] Schott, J. R. (2017). *Matrix Analysis for Statistics* (Third Edition).

[3] Srivastava, M. S., Katayama, S. and Kano, Y. (2013). A two sample test in high dimensional data. *J. Multivariate Anal.*, **114**, 349–358.

[4] Wang, C., Pan, G., Tong, T., and Zhu, L. (2015). Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statistica Sinica*, **25**, 993–1008.