高次元共分散行列の逆行列の推定問題とその応用

東京大・経済・院　伊藤 翼
東京大・経済　久保川 達也

多変量解析で登場するような統計的手法を実践する際に、共分散行列の逆行列（精度行列）の推定量が必要になることが多い。しかし、高次元データに対してその精度行列を求めようとすると、通常の標本共分散行列 Σ_p の逆行列は、推定精度が悪く、また次元 p がサンプルサイズ N を超える場合には S_p^{-1} は存在しないことは常識である。そこで以下では、$\mathbf{y}_1, \cdots, \mathbf{y}_N \sim i.i.d.(\mathbf{0}, \Sigma_p)$ のときに、$p, N \to \infty$ の枠組みで Σ_p^{-1} を推定することを考える。真の共分散行列 Σ_p の構造に関して事前の情報がないときに、この問題に対する有力なアプローチの一つは、S_p^{-1} を何らかの安定した統計量に縮小するというものである。具体的には、Σ_p^{-1} に対して、ある事前分布を想定することで導出される

$$\Omega_p^* = \alpha(S_p + \gamma I_p)^{-1} + \beta I_p$$

といった縮小推定量を考える。
パラメータは quadratic loss $L(\alpha, \beta, \gamma) = \frac{1}{p} Tr(\Omega_p^* \Sigma_p - I_p)^2$ を最小にするように推定すると

$$\begin{aligned}
\alpha_{opt}(\gamma) &= \frac{\frac{1}{p}Tr(\Sigma_p^2)\frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p] - \frac{1}{p}Tr(\Sigma_p)\frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p^2]}{\frac{1}{p}Tr(\Sigma_p^2)\frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p]^2 - \{\frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p^2]\}^2} \\
\beta_{opt}(\gamma) &= \frac{\frac{1}{p}Tr(\Sigma_p)\frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p]^2 - \frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p]\frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p^2]}{\frac{1}{p}Tr(\Sigma_p^2)\frac{1}{p}Tr[(S_p+\gamma I_p)^{-1}\Sigma_p]^2 - \{\frac{1}{p}Tr((S_p+\gamma I_p)^{-1}\Sigma_p^2)\}^2} \\
\gamma_{opt} &= argmin_{\gamma>0} L(\alpha_{opt}(\gamma), \beta_{opt}(\gamma), \gamma)
\end{aligned}$$

となる。ここで、行列の固有値の漸近的な挙動を調べる際に有用であるランダム行列理論の諸結果を適用することで、$\alpha_{opt}(\gamma), \beta_{opt}(\gamma), \gamma_{opt}$ の漸近的な値を求め、それに対する一致推定量を求めることで、α, β, γ の推定量を得ることができる。ランダム行列理論を適用する際には、特定の分布の仮定が必要でないため、この提案手法は、データの分布によらず一定のパフォーマンスを示し、また、既存の linear-type や ridge-type の推定量と比較した場合、特に、S_p の固有値のばらつきが大きい状況で、リスクを大幅に改善することが確認できる。

参考文献

[1] Bodnar, T., Gupta, A. K., and Parolya, N. (2014), Optimal linear shrinkage estimator for large dimensional precision matrix. *Journal of Multivariate Analysis* 132, 215-228.
[2] Wang, C., Pan, G., Tong, T., and Zhu, L. (2014), Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statistica Sinica* Preprint N0: SS-12-328R3