

Shrinkage estimators for large covariance matrices in multivariate normal distributions

今野 良彦 (日本女子大学 理学部)

はじめに 近年, データ数よりも変量の次元が高いデータ (高次元データ) の解析のための多変量推測理論の構築が注目を集めている. 本講演では, 高次元データの設定のもとで多変量正規分布の共分散行列の推定問題を統計的決定理論の枠組みで考察した結果を報告する.

問題設定 n をデータ数, p を変量の次元とし, $n < p$ とする. X を $n \times p$ の確率行列とし, 各行は多変量正規分布 $N_p(\mathbf{0}_p, \Sigma)$ に独立同一に従うとする. ただし, Σ は $p \times p$ の正値対称行列である. 正則ではない Wishart 行列 $S := X'X$ ('は転置を表す) に基づき, Σ の推定問題を損失関数 $L(\hat{\Sigma}, \Sigma) = \text{Tr}(\hat{\Sigma}\Sigma^{-1} - I_p)^2$ のもとで考える. ここで, $\hat{\Sigma}$ は Σ の推定量, I_p は $p \times p$ の単位行列, Tr は行列のトレースを表す. S の分布に関する L の期待値 $R(\hat{\Sigma}, \Sigma) := \mathbb{E}[L(\hat{\Sigma}, \Sigma)]$ をリスクとよぶ.

推定量 $n^{-1}S$ の問題点 (1) $n < p$ のとき, Σ は正定値であるにもかかわらず, $n^{-1}S$ は正定値ではない. (2) $n^{-1}S$ の固有根は, Σ の固有根よりも広がっている.

問題点 (2) は $n < p$ の場合のみならず, 小標本の場合や n と p が近いときにも存在する問題点である. $n > p+1$ のときは, S の固有根に対して shrinkage-expansion technique を適用することにより改良型推定量 (Dey and Srivastava (1985) を参照) が得られていることが知られている.

推定量の族 $n+1 < p$ とする. つぎの推定量の族を考える:

$$\hat{\Sigma}_t = \frac{1}{p+n+1} \left(S + \frac{t}{\text{Tr } S^+} O_1 O_1' \right).$$

ただし, O_1 は $p \times n$ の半直交行列で, S の正の固有値に対応する固有ベクトルを並べたもの, S^+ は S の Moore-Penrose の逆行列, t は正の定数である.

結果 正則でない Wishart 行列に対する部分積分の公式 (the Stein-Haff identity のアナロジー) を用いて, $\hat{\Sigma}_t$ のリスクの不偏推定量を導出し, リスクを評価することにより次の結果を得る.

$0 < t < 2(n-1)(p-n-1)/\{(p-n+1)(p-n+3)\}$ のとき, すべての Σ に対して, $R(\hat{\Sigma}_t, \Sigma) \leq R(n^{-1}S, \Sigma)$ が成立する.

$\hat{\Sigma}_t$ は正定値ではない. これを修正したもの $\tilde{\Sigma}_t = (p+n+1)^{-1} \{ S + t(\text{Tr } S^+)^{-1} I_p \}$ についてはリスクの評価が難しいので, 数値実験でリスクを評価した結果を報告する.

参考文献 Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. ANN. STATIST. **13** 1581-1591.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. J. MULTIVARIATE ANAL. **88** 365-411.

Konno, Y., Shrinkage Estimators for Large Covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss, submitted. April 2008.