2008年度 統計関連学会連合大会（2008 年 9 月 08 日・慶應大学）

Shrinkage estimators for large covariance matrices in multivariate normal distributions

日本女子大学理学部　　今野 良彦

September 9, 2008

この講演の目的と構成

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 (Large Covariance matrix) 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．

本講演の構成

(1) 記号と問題設定.

(2) 先行研究について.

(3) リスクの不偏推定法 (SURE).

(4) 改良型推定量について.

記号と問題設定 (1)

⋆ $X_1, X_2, \ldots, X_n \sim N_p(0, \boldsymbol{\Sigma})$．各 $X_i\,(i = 1, 2, \ldots, n)$ は p 変量確率 (縦) ベクトルで，独立同一に p 変量正規分布に従う．ただし，$\boldsymbol{\Sigma}$ は $p \times p$ の正値対称行列で未知.

⋆ n は (標本数 - 1) で，p は変量の次元.

⋆ Wishart 確率行列 ($p \times p$ の行列) $\boldsymbol{S} := \sum_{k=1}^{n} X_i X_i'$．ただし，“$'$” はベクトルや行列の転置を示す.

⋆ 共分散行列 $\boldsymbol{\Sigma}$ の推定問題を損失関数

$$L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - I_p)^2$$

のもとで考える．ここで，$\widehat{\boldsymbol{\Sigma}}$ は $\boldsymbol{\Sigma}$ の推定量，$\boldsymbol{I}_p$ は $p \times p$ の単位行列，Tr は行列のトレースを表す.

⋆ $\boldsymbol{S}$ の分布に関する損失関数 L の期待値 $R(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) := \mathbb{E}[L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})]$ をリスクとよぶ．$\boldsymbol{\Sigma}$ に関して一様に推定量のリスクを比較したい.

記号と問題設定 (2)

$\star$ 平均を 0 としたことは本質的ではない.

$\star$ Wishart 確率行列 $\boldsymbol{S}$ は正定値 $\Longleftrightarrow n \geq p$.

$\star$ Wishart 確率行列 $\boldsymbol{S}$ の分布は常に存在するが，$p \times p$ の対称行列の空間上の Lebesgue 測度に関する確率密度関数は $n \geq p$ のとき存在.

$\star$ 変換 $\widehat{\boldsymbol{\Sigma}} \mapsto \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}\boldsymbol{A}'$; $\boldsymbol{\Sigma} \mapsto \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$($\boldsymbol{A}$ は $p \times p$ の正則行列) に関して不変な損失関数：

$$L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - I_p)^2;\ L_S(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log \mathrm{Det}(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - p.$$

ただし，Det は行列式. しかし，$n < p$ のとき，L_S（L_S の期待値) は $n^{-1}\boldsymbol{S}$ を評価できない.

先行研究について (1)

推定量 $n^{-1}\boldsymbol{S}$ の問題点

- ⋆ $\mathbb{E}[n^{-1}\boldsymbol{S}] = \boldsymbol{\Sigma}$ だが，$n^{-1}\boldsymbol{S}$ の固有根は，$\boldsymbol{\Sigma}$ の固有根よりも広がっている．
- ⋆ $n < p$ のとき，$\boldsymbol{\Sigma}$ は正定値であるにもかかわらず，$n^{-1}\boldsymbol{S}$ は正定値ではない．

$n \geq p$ の場合の先行研究

- ⋆ 損失関数 L_S のもとでは，$n^{-1}\boldsymbol{S}$ の固有根をShrinkage-expansion method を用いた改良型推定量．Stein (1977), Dey and Srivastava (1985), Haff (1991) 等を参照.
- ⋆ リスクを評価するために，SURE 法が有効 — 部分積分の公式と eigenvalue-caluculus　$\to n < p$ の場合は？
- ⋆ 損失関数 L のもとでは，Haff (1980) の結果　$\to n < p$ の場合は？

先行研究について (2)

$n < p$ の場合の先行研究

⋆ Ledoit and Wolf (2004): 損失関数 $\mathrm{Tr}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})^2$ のもとで，$n^{-1}\boldsymbol{S}$ と $\boldsymbol{I}_p$ の線形結合のなかで漸近的(n/p は有界)に最適なもの．積率の条件のみで分布に依存しない結果.

⋆ Wu and Pourahmadi (2003), Bickel and Levina (2008): banding approach. 漸近的に評価.

⋆ Furrer and Bengtsson (2007): “tapering”.

部分積分の公式と SURE 法(1)

⋆ 分解 $\boldsymbol{S}=\boldsymbol{O}_1\boldsymbol{L}\boldsymbol{O}_1'$．ここで，$\boldsymbol{L}=\mathbf{Diag}(\ell_1,\ell_2,\ldots,\ell_n)$，$\ell_1>\ell_2>\cdots>\ell_n>0$ は $\boldsymbol{S}$ の固有値，$\boldsymbol{O}_1$ は $p\times n$ の半直交行列，i.e., $\boldsymbol{O}_1'\boldsymbol{O}_1=\boldsymbol{I}_n$.

⋆ 推定量

$$\widehat{\boldsymbol{\Sigma}}=\boldsymbol{O}_1\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{O}_1',$$

を考える．ただし，

$$\begin{aligned}\boldsymbol{\Phi} &:= \boldsymbol{\Phi}(\boldsymbol{L})=\mathrm{Diag}(\varphi_1,\varphi_2,\ldots,\varphi_n),\\ \varphi_k &:= \varphi_k(\boldsymbol{L})\,(k=1,2,\ldots,n)\text{ は }\mathbb{R}^n_{>0}\text{ から }\mathbb{R}\text{ への部分可能な関数.}\end{aligned}$$

部分積分の公式と SURE 法(2)

$\boxed{\text{SURE}}$ $n+1<p$ とする．適当な条件のもと

$$
\begin{aligned}
R(\boldsymbol{O}_1\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{O}_1', \boldsymbol{\Sigma}) &= \mathbb{E}[\mathrm{Tr}\,\{\boldsymbol{\Sigma}^{-1}\boldsymbol{O}_1\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{O}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{O}_1\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{O}_1'\} \\
&\qquad -2\mathrm{Tr}\,\{\boldsymbol{\Sigma}^{-1}\boldsymbol{O}_1\boldsymbol{\Phi}(\boldsymbol{L})\boldsymbol{O}_1'\}+p] \\
&= \mathbb{E}\Bigg[\sum_{k=1}^{n}\Bigg\{(p-n-1)\Big(\frac{\varphi_k^{(1)}}{\ell_k}-2\frac{\varphi_k}{\ell_k}\Big)+2\Big(\frac{\partial\varphi_k^{(1)}}{\partial\ell_k}-2\frac{\partial\varphi_k}{\partial\ell_k}\Big) \\
&\qquad +\sum_{b\neq k}^{n}\frac{(\varphi_k^{(1)}-2\varphi_k)-(\varphi_b^{(1)}-2\varphi_b)}{\ell_k-\ell_b}\Bigg\}+p\Bigg],
\end{aligned}
$$

ただし，$\varphi_k^{(1)}=(p-n-1)\varphi_k^2/\ell_k+4\varphi_k(\partial\varphi_k/\partial\ell_k)+2\varphi_k\sum_{b\neq k}^{n}(\varphi_k-\varphi_b)/(\ell_k-\ell_b)$ $(k=1,\,2,\,\ldots,\,n)$.

改良型推定量(1)

推定量の族　$n+1<p$ とする．つぎの推定量の族を考える：

$$\widehat{\mathbf{\Sigma}}_t = \frac{1}{p+n+1}\Big(\boldsymbol{S} + \frac{t}{\mathrm{Tr}\,\boldsymbol{S}^+}\boldsymbol{O}_1\boldsymbol{O}_1'\Big).$$

ただし，$\boldsymbol{O}_1$ は $p\times n$ の半直交行列で, $\boldsymbol{S}$ の正の固有値に対応する固有ベクトルを並べたのもの，$\boldsymbol{S}^+$ は $\boldsymbol{S}$ の Moore-Penrose の逆行列，t は正の定数である．

結果　$\widehat{\mathbf{\Sigma}}_t$ のリスクの不偏推定量(SURE)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 $\mathbf{\Sigma}$ に対して，$R(\widehat{\mathbf{\Sigma}}_t, \mathbf{\Sigma}) \le R(n^{-1}\boldsymbol{S}, \mathbf{\Sigma})$ が成立する．

改良型推定量(2)

⋆ $\widehat{\boldsymbol{\Sigma}}_t$ は正定値ではない.

⋆ $\frac{1}{p+n+1}\Big(\boldsymbol{S} + \frac{t}{\mathrm{Tr}\,\boldsymbol{S}^+}\boldsymbol{O}_1\boldsymbol{O}_1'\Big)$ を修正したもの:

$$\widetilde{\boldsymbol{\Sigma}}_t = \frac{1}{p+n+1}\Big\{\boldsymbol{S} + \frac{t}{\mathrm{Tr}\,\boldsymbol{S}}\boldsymbol{I}_p\Big\}.$$

⋆ 残念なことに，推定量 $\widetilde{\boldsymbol{\Sigma}}_t$ のリスクを SURE を用いて評価できない！

⋆ 数値実験で調べた結果，$\widetilde{\boldsymbol{\Sigma}}_t$ は $n^{-1}\boldsymbol{S}$ と $(p+n+1)^{-1}\boldsymbol{S}$ を改良しているだろう.

⋆ $\boldsymbol{\Sigma}$ の固有根が散らばると改良率はあがる！

⋆ 複素 Wishart 行列に対しても同様な結果を得ることができる！

余談

⋆ Kubokawa and Srivastava (2008): $\mathbf{\Phi} = \mathrm{Diag}(\varphi_1, \varphi_2, \ldots, \varphi_n)$ に対し，

$$\mathbb{E}\left[\mathrm{Tr}\left\{\mathbf{\Sigma}^{-1}\boldsymbol{O}_1\mathbf{\Phi}\boldsymbol{O}_1'\right\}\right] = \mathbb{E}\left[\sum_{k=1}^{n}\Big\{(p-n-1)\frac{\phi_k}{\ell_k} + 2\frac{\partial\phi_k}{\partial\ell_k} + \sum_{b\neq k}^{n}\frac{\phi_k-\phi_b}{\ell_k-\ell_b}\Big\}\right].$$

⋆ $\mathbb{E}[\mathrm{Tr}\,(\mathbf{\Sigma}^{-1}\boldsymbol{O}_1\mathbf{\Phi}(\boldsymbol{L})\boldsymbol{O}_1'\mathbf{\Sigma}^{-1}\boldsymbol{O}_1\mathbf{\Phi}(\boldsymbol{L})\boldsymbol{O}_1')]$ を評価するために，Tr をはずしたい！

⋆ Matrix form の等式：

$$\mathbb{E}\left[\mathbf{\Sigma}^{-1}\boldsymbol{O}_1\mathbf{\Phi}\boldsymbol{O}_1'\right] = \mathbb{E}\left[\boldsymbol{O}_1\mathbf{\Psi}^{(1r)}\boldsymbol{O}_1' + \mathrm{Tr}\,(\boldsymbol{L}^{-1}\mathbf{\Phi})(\boldsymbol{I}_p - \boldsymbol{O}_1\boldsymbol{O}_1')\right].$$

ただし，$\mathbf{\Psi}^{(1r)} = \mathrm{Diag}(\phi_1^{(1r)}, \phi_2^{(1r)}, \ldots, \phi_n^{(1r)})$ ($k = 1, 2, \ldots, n$) で

$$\phi_k^{(1r)} = \sum_{b\neq k}^{n}(\phi_k - \phi_b)/(\ell_k - \ell_b) + 2(\partial\phi_k/\partial\ell_k) - \phi_k/\ell_k.$$