久留米大学バイオ統計センター（2010 年 01 月 21 日）

Shrinkage estimators for large covariance matrices in multivariate normal distribution

日本女子大学理学部　　今野 良彦

January 19, 2010

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) 問題の背景・記号・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.

(5) 数値実験について

Real Wishart 分布（1）

$\mathbb{R}^p$-値確率ベクトル $X_1, X_2, \ldots, X_n$ は独立同一に多変量正規分布 $N_p(0, \boldsymbol{\Sigma})$ に従う：

$$P(dx) = \frac{1}{(2\pi)^{p/2}\mathrm{Det}(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}x'\boldsymbol{\Sigma}^{-1}x\right\}(dx)$$

ただし，

$$\boldsymbol{\Sigma} \in \mathrm{Sym}_p^+(\mathbb{R}) = \{\boldsymbol{S} \in \mathbb{R}^{p\times p} : x'\boldsymbol{S}x > 0 \quad \text{for } x \in \mathbb{R}^p,\, x \neq 0\},$$

dx は $\mathbb{R}^p$ 上のルベーグ測度，x' は縦ベクトル x の転置.

Real Wishart 分布（2）

Wishart Matrix

$$\boldsymbol{W} : p \times p = \sum_{j=1}^{n} X_i X_i'$$

とおく．$\boldsymbol{W}$ をウイシャート行列と呼ぶことにする．

定義より，$\boldsymbol{W} \in \overline{\mathrm{Sym}_p^+(\mathbb{R})}$（$\mathrm{Sym}_p^+(\mathbb{R})$ の閉包）であるが，

$$\mathbb{P}\{\boldsymbol{W} \in \mathrm{Sym}_p^+(\mathbb{R})\} = 1 \Longleftrightarrow n \geq p \quad \text{かつ} \quad \boldsymbol{\Sigma} \in \mathrm{Sym}_p^+(\mathbb{R})$$

となる．Eaton (1983, page 304) を参照.

Real Wishart 分布（3）

実または複素数値正規分布 $\overset{2\text{ 乗和}}{\Longrightarrow}$ ガンマ分布（χ^2 分布）

$\Downarrow$ 　　　　$\Downarrow$

多変量実正規分布 $\overset{2\text{ 乗和}}{\Longrightarrow}$ Real Wishart 分布

多変量複素正規分布 $\overset{2\text{ 乗和}}{\Longrightarrow}$ Complex Wishart 分布

Real Wishart 分布（4）

Real Wishart density ($\boldsymbol{\Sigma} \in \mathrm{Sym}_p^+(\mathbb{R})$，$n \geq p$ を仮定．)

$$W_p(\boldsymbol{\Sigma}, n)(dx) = \frac{1}{c_{p,n}(\mathrm{Det}\boldsymbol{\Sigma})^{n/2}}(\mathrm{Det}x)^{n/2}\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}x\right)\right) \times \mathbb{1}_{\mathrm{Sym}_p^+(\mathbb{R})}(x)\frac{(dx)}{(\mathrm{Det}\,x)^{(p+1)/2}}$$

この分布を $W_p(\Sigma, n)$ とかく．ただし，

$$c_{p,n} = 2^{np}\pi^{p(p-1)/4}\prod_{i=0}^{p-1}\Gamma((n-i)/2)$$

$$\mathbb{1}_{\mathrm{Sym}_p^+(\mathbb{R})}(x) = \begin{cases} 1 & (x \in \mathrm{Sym}_p^+(\mathbb{R})), \\ 0 & (\text{その他}) \end{cases}$$

Wishart 行列の極分解

極分解（$n \geq p$ の制限は必要なし）

$$\boldsymbol{W} = \boldsymbol{HLH}', \quad \boldsymbol{H} \in \mathrm{O}_p(\mathbb{R}), \quad \boldsymbol{L} = \mathrm{diag}(\ell_1, \ell_2, \ldots, \ell_p)$$

with $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_p \geq 0$

ただし，

$$\mathrm{O}_p(\mathbb{R}) = \{\boldsymbol{H} \in \mathbb{R}^{p\times p} : \boldsymbol{H}'\boldsymbol{H} = \boldsymbol{HH}' = \boldsymbol{I}_p\}$$

で $\boldsymbol{I}_p$ は $p \times p$ の単位行列とする．

Real Wishart 行列の固有根の分布（１）

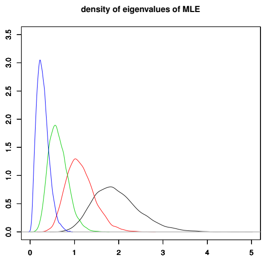

Figure 1: シミュレーションによる $W_4(\boldsymbol{I}_4, 10)$ の $(\ell_1, \ldots, \ell_4)/10$ の周辺確率密度関数$(p = 4,\ n = 10)$の作図．

Real Wishart 行列の固有根の分布（２）

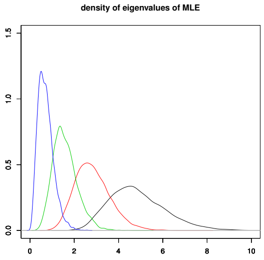

Figure 2: シミュレーションによる $W_{10}(\boldsymbol{I}_{10}, 4)$ の $(\ell_1, \ldots, \ell_{10})/4$ の周辺確率密度関数$(p = 10,\ n = 4)$の作図．

Real Wishart 行列の固有根の分布（３）

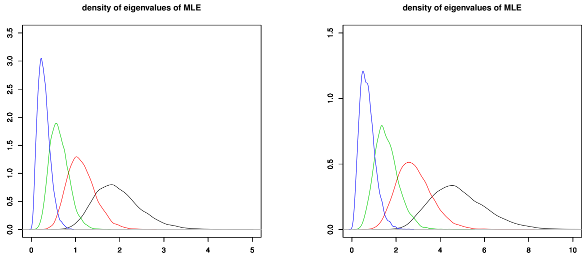

Figure 3: 左は $W_4(\boldsymbol{I}_4, 10)$；右は $W_{10}(\boldsymbol{I}_{10}, 4)$

Real Wishart 行列の固有根の分布（４）：from RMT view

Marčhenko and Pastur law[1]

$p = p(n)$ とし，$p/n \to \tau \leq 1(n \to \infty)$ のとき，

$$\frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{\lambda_i \leq x} \to \int_{-\infty}^{x} \frac{\sqrt{(b-t)(t-a)}}{2\pi t\tau} \mathbb{1}_{a \leq t \leq b}\, dt$$

(almost surely).

ただし，$a = (1 - \tau^{1/2})^2$, $b = (1 + \tau^{1/2})^2$.

[17] の Lecture 4 を参照．

Real Wishart 行列の固有根の分布（５）

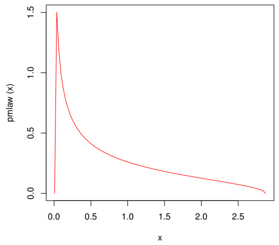

Figure 4: Density of Marčhenko-Pastur's quater-circle law ($\tau = 4/7 \fallingdotseq 0.57$, $a \fallingdotseq 0.06$, $b \fallingdotseq 3.086$).

Real Wishart 行列の固有根の周辺分布（６）

- $p \times p$ の正値対称行列 $\boldsymbol{W} \sim W_p(\boldsymbol{\Sigma}, n)\,(\Sigma > 0$ かつ $n \geq p)$ の固有根を $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_p$ とする．$\boldsymbol{W}/n$ は $\boldsymbol{\Sigma}$ の不偏推定量．
- しかし，$\boldsymbol{W}/n$ の固有根 $(\ell_1/n, \ldots, \ell_p/n)$ は $\boldsymbol{\Sigma}$ の固有根のバイアス推定量となる！

$$\Downarrow$$

Shrink eigenvalues toward a center!

- $n < p$ の場合はどうなるのか？

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) $\boldsymbol{W} := \sum_{k=1}^{n} X_i X_i'$．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．

⋆ $\boldsymbol{W}$ 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）

⋆ 平均を 0 としたことは本質的ではない．

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

⋆ Wishart 確率行列 $\boldsymbol{W}$ 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．

⋆ 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 $n^{-1}\boldsymbol{W}$ を評価できない．

先行研究について（１）

推定量 $n^{-1}\boldsymbol{W}$ の問題点

⋆ $\mathbb{E}[n^{-1}\boldsymbol{W}] = \boldsymbol{\Sigma}$ だが，$n^{-1}\boldsymbol{W}$ 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，$n^{-1}\boldsymbol{W}$ 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，$n^{-1}\boldsymbol{W}$ 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　$\rightarrow n < p$ の場合は？

⋆ 損失関数 L のもとでは，<u>Haff (1980) の結果</u>　$\rightarrow n < p$ 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，$n^{-1}\boldsymbol{W}$ 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”.

問題設定の復習

⋆ $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$．各 $\boldsymbol{X}_i\,(i = 1, 2, \ldots, n)$ 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) $\boldsymbol{W} := \sum_{k=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'$ に基づき，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;

⋆ $\boldsymbol{W}$ 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．

推定量のクラス

⋆ $\boldsymbol{W}=\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i'$　を分解する：$\ell_1\geq\cdots\geq\ell_n$ は $\boldsymbol{W}$ の固有値で,

$$\boldsymbol{W}=\boldsymbol{O}_1\boldsymbol{L}\boldsymbol{O}_1', \qquad \boldsymbol{L}=\mathrm{Diag}(\ell_1,\,\ldots,\,l_n);$$
$$\boldsymbol{O}_1 \text{ は } p\times n \text{ の半直交行列 s.t. } \boldsymbol{O}_1'\boldsymbol{O}_1=\boldsymbol{I}_n.$$

推定量のクラス
$$\widehat{\boldsymbol{\Sigma}}=\boldsymbol{O}_1\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{O}_1', \tag{1}$$
ただし，$\boldsymbol{\Psi}:=\boldsymbol{\Psi}(\boldsymbol{L})=\mathrm{Diag}(\psi_1,\,\psi_2,\,\ldots,\,\psi_n)$ で $\psi_k:=\psi_k(\boldsymbol{L})(k=1,\,2,\,\ldots,\,n)$ は $\mathbb{R}^n_{\geq}$ から $\mathbb{R}$ への可微分関数.

目標　$\boldsymbol{\Sigma}$ に依存するリスク $\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2]$ を評価したい！

推定量のクラスとリスクの評価の方針（ SURE 法)

⋆ リスク $\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2]$ の不偏推定量 $\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})$ ($\psi_1,\,\ldots,\,\psi_n$ と $\ell_1,\,\ldots,\,\ell_n$ を通してのみ $\boldsymbol{W}$ に依存)を導出：

$$\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2]=\mathbb{E}[\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})]$$

⋆ $\mathbb{E}[\mathrm{Tr}\,(n^{-1}\boldsymbol{W}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2]$ は定数リスクなので,

$$\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})\leq\mathbb{E}[\mathrm{Tr}\,(n^{-1}\boldsymbol{W}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2]$$

ならば,
$$\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2\leq\mathbb{E}[\mathrm{Tr}\,(n^{-1}\boldsymbol{W}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I}_p)^2]$$
がわかる.

SURE の導出　推定量の族 (1) に対して，リスクの不偏推定量 $\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})$ を導出する.

部分積分の公式と SURE 法（1）

⋆ $(x_{ij})_{i=1,\ldots,n;\,j=1,\ldots,p}:=[\boldsymbol{X}_1,\,\boldsymbol{X}_2,\,\ldots,\,\boldsymbol{X}_n]'\sim N_{n\times p}(\boldsymbol{0},\,\boldsymbol{I}_n\otimes\boldsymbol{\Sigma});$

⋆ $n\times p$ の行列作用素 $\nabla_{\boldsymbol{X}}$ を次で定める：

$$\nabla_{\boldsymbol{X}}=\left(\frac{\partial}{\partial x_{ij}}\right)_{\substack{i=1,\,2,\,\ldots,\,n\\ j=1,\,2,\,\ldots,\,p}};$$

⋆ 行列 $\nabla_{\boldsymbol{X}}\boldsymbol{A}$ の $(i,\,j)$ 成分を

$$(\nabla_{\boldsymbol{X}}\boldsymbol{A})_{ij}=\sum_{k=1}^p\frac{\partial a_{kj}}{\partial x_{ik}} \qquad \text{for } i=1,\,2,\,\ldots,\,n;\,j=1,\,2,\,\ldots,\,p.$$

部分積分の公式と SURE 法（2）

補題 1　$[\boldsymbol{X}_1,\,\boldsymbol{X}_2,\,\ldots,\,\boldsymbol{X}_n]'\sim N_{n\times p}(\boldsymbol{0},\,\boldsymbol{I}_n\otimes\boldsymbol{\Sigma})$ とし，$\boldsymbol{W}=\sum_{i=1}^n\boldsymbol{X}_i\boldsymbol{X}_i'$ とおいたとき，$p\times p$ 関数 $\boldsymbol{G}=\boldsymbol{G}(\boldsymbol{W})$ に対して,

$$\mathbb{E}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\boldsymbol{G}]=\mathbb{E}\,[n\boldsymbol{G}+(\boldsymbol{X}'\nabla_{\boldsymbol{X}})'\boldsymbol{G}].$$

特に,
$$\mathbb{E}\,[\mathrm{Tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\boldsymbol{G})]=\mathbb{E}\,[n\mathrm{Tr}\,(\boldsymbol{G})+\mathrm{Tr}\,(\boldsymbol{X}'\nabla_{\boldsymbol{X}}\boldsymbol{G}')].$$
ただし，$'$ は転置.

部分積分の公式と SURE 法（3）

⋆ 補題 1 において，$\boldsymbol{G}=\boldsymbol{O}_1\mathrm{Diag}(\ell_1^{-1}\psi_1,\,\ldots,\,\ell_n^{-1}\psi_n)\boldsymbol{O}_1'$ とおく：

補題 2　それぞれの期待値が存在するとき,

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

ただし，$\boldsymbol{\Psi}^{(1r)}=\mathrm{Diag}(\psi_1^{(1r)},\psi_2^{(1r)},\ldots,\,\psi_n^{(1r)})$ で
$$\psi_k^{(1r)}=\sum_{b\neq k}^n\frac{\psi_k-\psi_b}{\ell_k-\ell_b}+2\frac{\partial\psi_k}{\partial\ell_k}-\frac{\psi_k}{\ell_k}, \qquad k=1,\,2,\,\ldots,\,n.$$

特に，Kubokawa and Srivastava (2008)'s identity として,

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

部分積分の公式と SURE 法（4）

補題 3　$\widehat{\boldsymbol{\Sigma}}=\boldsymbol{O}_1\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{O}_1'$ に対して,

$$\mathbb{E}[\mathrm{Tr}\,\{\boldsymbol{\Sigma}^{-1}\boldsymbol{O}_1\boldsymbol{\Psi}\boldsymbol{O}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{O}_1\boldsymbol{\Psi}\boldsymbol{O}_1'\}]=\mathbb{E}[\mathrm{Tr}\,\{\boldsymbol{\Sigma}^{-1}\boldsymbol{O}_1\widetilde{\boldsymbol{\Psi}}^{(1)}\boldsymbol{O}_1'\}].$$

ただし，$\widetilde{\boldsymbol{\Psi}}^{(1)}=\mathrm{Diag}(\tilde{\psi}_1^{(1)},\,\tilde{\psi}_2^{(1)},\,\ldots,\,\tilde{\psi}_n^{(1)})$ で

$$\tilde{\psi}_k^{(1)}=(p-n-1)\frac{\psi_k^2}{\ell_k}+4\psi_k\cdot\frac{\partial\psi_k}{\partial\ell_k}+2\psi_k\cdot\sum_{b\neq k}^n\frac{\psi_k-\psi_b}{\ell_k-\ell_b}, \qquad k=1,\,2,\,\ldots,\,n.$$

部分積分の公式と SURE 法（5）

定理 4　$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{O}_1\boldsymbol{\Psi}(\boldsymbol{L})\boldsymbol{O}_1'$ に対して,

$$R(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathbb{E}\Big[\sum_{k=1}^{n}\Big\{(p-n-1)\big(\frac{\widetilde{\psi}_k^{(1)}}{\ell_k} - 2\frac{\psi_k}{\ell_k}\big) + 2\big(\frac{\partial\widetilde{\psi}_k^{(1)}}{\partial\ell_k} - 2\frac{\partial\psi_k}{\partial\ell_k}\big) + \sum_{b\neq k}^{n}\frac{(\widetilde{\psi}_k^{(1)} - 2\psi_k) - (\widetilde{\psi}_b^{(1)} - 2\psi_b)}{\ell_k - \ell_b}\Big\} + p\Big].$$

ただし，$(k = 1, 2, \ldots, n)$ で

$$\widetilde{\psi}_k^{(1)} = (p-n-1)\psi_k^2/\ell_k + 4\psi_k(\partial\psi_k/\partial\ell_k) + 2\psi_k\sum_{b\neq k}^{n}(\psi_k - \psi_b)/(\ell_k - \ell_b)$$

改良型推定量（1）

推定量の族　$n < p$ とする．つぎの推定量の族を考える：

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

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

結果　$\widehat{\boldsymbol{\Sigma}}_t$ 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．

$$0 < t < 2(n-1)(p-n-1)/\{(p-n+1)(p-n+2)\}$$

のとき，すべての $\boldsymbol{\Sigma}$ に対して，$R(\widehat{\boldsymbol{\Sigma}}_t, \boldsymbol{\Sigma}) \le R(n^{-1}\boldsymbol{W}, \boldsymbol{\Sigma})$ が成立する．

改良型推定量（2）

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

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

$$\widetilde{\boldsymbol{\Sigma}}_{\mathrm{HF}} = \frac{1}{p+n+1}\{\boldsymbol{W} + \frac{t_0}{\mathrm{Tr}\,\boldsymbol{W}}\boldsymbol{I}_p\}, \qquad t_0 = \frac{2(n-1)(p+n+1)}{(p-n+1)(p-n+3).}$$

⋆ 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！

数値実験の結果（1）

the percentage relative improvement in average loss of $\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}}$ over $\boldsymbol{W}/(n+p+1$

$$\mathsf{PRIAL}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}}) = \frac{\text{average loss of } \boldsymbol{W}/(n+p+1) - \text{average loss of } \widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}}}{\text{average loss of } \boldsymbol{W}/(n+p+1)}$$

parameter　$\alpha = \{p - (\mathrm{Tr}\,\boldsymbol{\Sigma}^{-1})^2/\mathrm{Tr}\,\boldsymbol{\Sigma}^{-2}\}/p.$

注意　$\boldsymbol{\Sigma} = \boldsymbol{I}_p \Longrightarrow \alpha = 0.$

数値実験の結果（2）

Table 1: Result of 1000 Monte-Carlo simulations for $p/n = 1/2$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_p$.

p	n	$\boldsymbol{W}/(n+p+1)$	$\widehat{\boldsymbol{\Sigma}}_{\mathrm{HF}}$	PRIAL
10	5	6.863 (0.013)	6.3465 (0.017)	7.5%
20	10	13.549 (0.014)	12.654 (0.017)	6.6%
40	20	26.891 (0.014)	25.647 (0.016)	4.6%
60	30	40.216 (0.014)	38.821 (0.016)	3.5%
80	40	53.570 (0.014)	52.095 (0.015)	2.8%
100	50	66.894 (0.014)	65.362 (0.015)	2.3%

The values in parentheses refer to the standard error on average loss.

数値実験の結果（3）

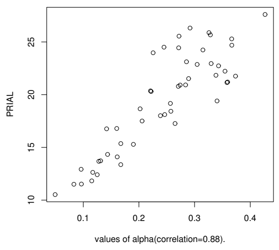

Figure 5: Effect of α for $p = 20$ and $n = 4$.

数値実験の結果（４）

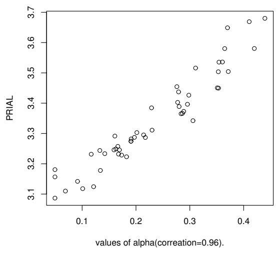

Figure 6: Effect of α for $p = 20$ and $n = 16$.

数値実験の結果（５）

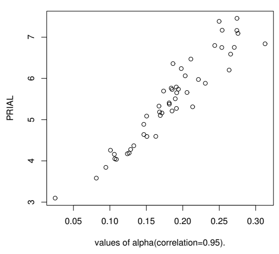

Figure 7: Effect of α for $p = 100$ and $n = 20$.

数値実験の結果（６）

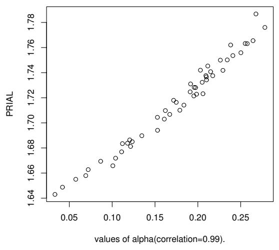

Figure 8: Effect of α for $p = 100$ and $n = 80$.

References

[1] Z.D. Bai, Methodologies in spectral analysis of large dimensional random matrices, a reivew. *Statistica Sinica* **9** (1999) 611-677.

[2] P.J. Bickel and E. Levina, Covariance regularization by thresholding. *Ann. Statist.* **36** (2008) 2577–2604.

[3] D.K. Dey and C. Srinivasan, Estimation of a covariance matrix under Stein's loss, *Ann. Statist.* **13** (1985) 1581-1591.

[4] M.L. Eaton, Group Invariance Application in Statistics, Regional conference series in Probability and Statistics Vol. 1, Institute of Mathematical Statistics (1989).

[5] R. Furrer and T. Bengtsson, Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants *J. Multivariate Analysis* **98** (2007) 227-255.

[6] L.R. Haff, Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8** (1980) 586–597.

[7] L.R. Haff, The variational form of certain Bayes estimators, *Ann. Statist.* **19** (1992) 1163-1190.

[8] W. James and C. Stein, Estimation with quadratic loss, in Proc. Fourth Berkeley Symp. Math. Statist. Prob. **1** 361-380, Univ. California Press (1961).

[9] Y. Konno, Estimation of a normal covariance matrix parametrized by irreducible symmetric cones under Stein's loss, *J. Multivairate Anal.* **98** (2007) 295-316.

[10] Y. Konno, Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invriant quadratic loss. *J. Multivariate Anal.* **100** (2009), 2237-2253.

[11] T. Kubokawa and M. Srivastava, Estimation of the precision matrix of s singular Wishart distribution and its application in high dimenstional data. *J. Multivariate Anal.* **99** (2008) 1906-1928.

[12] O. Ledoit and M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices. (English summary) *J. Multivariate Analysis* **88** (2004) 365–411.

[13] H. Massam and E. Neher, Estimation and testing for lattice conditional independence models of Euclidean Jordan algebras, *Ann. Statist.* **26** (1998) 1051-1082.

[14] R.J. Muirhead, Aspects of Multivariate Statistical Analysis, John Wiley & Sons, Inc (1982).

[15] J. Schäfer and K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Stat. Appl. Genet. Mol. Biol.* **4** (2005), Art. 32, 28 pp. (electronic).

[16] M. Srivastava, Singular Wishart and multivariate beta distribution. *Ann. Statist.* **31** (2003) 1537-1560.

[17] C. Stein, Lectures on the theory of estimation of many parameters, in Studies in the Statistical Theory of Estimation I(I. A. Ibragimov and M. S. Nikulin, eds.) (1977).

[18] W.B. Wu and M. Pourahmadi, Nonparametric estimation of large covariance matrices of longitudinal data *Biometrika* **90** (2003) 831-844