Shrinkage estimation of the mean matrix of multivariate complex normal distributions

Yoshihiko KONNO

Japan Women's University

CIMPA-MICINN-UNESCO RESEARCH SCHOOL
ANALYTICAL AND ALGEBRAIC TOOLS
IN STATISTICS AND GRAPHICAL MODELS
HAMMAMET, TUNISIA

5-17 September 2011

★ Good reference is Andersen et al. (1995, Lecture note in Statistics vol. 101).

★ We say that a random variable Z has a standard complex normal distributions $\mathbb{C}N(0, 1)$ if

$$[Z] = \begin{pmatrix} \mathrm{Re}Z \\ \mathrm{Im}Z \end{pmatrix} \sim N_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right].$$

★ Then the density (w.r.t. Lebesgue measure on $\mathbb{C}$) of Z is given as

$$f_Z(z) = \frac{1}{\sqrt{\pi}}\exp(-\bar{z}z)$$

where $\bar{z}$ is the conjugate of a complex x ($x \in \mathbb{C}$).

★ Assume that $Z \sim \mathbb{C}N(0, 1)$. For $\xi \in \mathbb{C}$, $\sigma \in \mathbb{R}_+$,

$$X := \xi + \sigma Z \sim \mathbb{C}N(\xi, \sigma^2).$$

★ Let X be $\mathbb{C}^p$-valued random vector. For $\forall c \in \mathbb{C}^p$, $\xi \in \mathbb{C}^p$, and $\Sigma \in \mathbb{C}_+^{p\times p}$

$$c^*X \sim \mathbb{C}N(c^*\xi, c^*\Sigma c) \Longleftrightarrow X \sim \mathbb{C}N_p(\xi, \Sigma),$$

where c^* is the transpose complex conjugate of a column vector c and $\mathbb{C}_+^{p\times p}$ is the set of all p.d. complex matrices.

★ Let $X \sim \mathbb{C}N_p(\xi, \Sigma)$. Then the density of X(w.r.t. Lebesgue measure on $\mathbb{C}^p$) is given as

$$f_X(x) = \frac{1}{\pi^p}(\mathrm{Det}\,\Sigma)^{-1}\exp\{-(x-\xi)^*\Sigma^{-1}(x-\xi)\}.$$

★ Let $X \sim \mathbb{C}N_p(\xi, \Sigma)$. Then the density of $[X]$(w.r.t Lebesgue measure on $\mathbb{R}^{2p}$) is given as

$$[X] := \begin{pmatrix} \mathrm{Re}\, X \\ \mathrm{Im}\, X \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \mathrm{Re}\, \xi \\ \mathrm{Im}\, \xi \end{pmatrix}, \frac{1}{2}\begin{pmatrix} \mathrm{Re}\, \Sigma & -\mathrm{Im}\, \Sigma \\ \mathrm{Im}\, \Sigma & \mathrm{Re}\, \Sigma \end{pmatrix}\right],$$

where $\mathrm{Re}\, \Sigma$ and $\mathrm{Im}\, \Sigma$ are symmetric and skew-symmetric.

★ Let $X_1, X_2, \ldots, X_n$ be mutually independent $\mathbb{C}N_p(0, \Sigma)$ random vectors. Then the distribution of

$$S := \sum_{i=1}^{n} X_i X_i^*$$

is called a complex Wishart distribution with parameters n, Σ (this is denoted by $\mathbb{C}W_p(n, \Sigma)$)

★ If $n \geq p$, then $\mathbb{P}(S \text{ is p.d.}) = 1$ and the density of S(w.r.t. Lebesgue measure on $\mathbb{C}_+^{p\times p}$) is given as

$$f_S(w) = \frac{\mathrm{Det}\,(w)^{n-p} \exp(-\mathrm{Tr}\,(w\Sigma^{-1}))}{\mathrm{Det}\,(\Sigma)^n \pi^{p(p-1)/2} \prod_{j=1}^{p} \Gamma(n+1-j)},$$

where $w \in \mathbb{C}_+^{p\times p}$, the set of all p.d. complex matrices, and $\Gamma(\cdot)$ is Euler's gamma function.

Observation for the problem of estimating mean matrix Ξ

★ Two random matrices $\mathbf{X} : m \times p$ and $\mathbf{S} : p \times p$ are independent and distributed as follows.

★ $\mathbf{X} : m \times p \sim \mathbb{C}N_{m\times p}(\Xi, \mathbf{I}_m \otimes \Sigma)$ and $\mathbf{S} : p \times p \sim \mathbb{C}W_p(n, \Sigma)$, where Ξ is unknown $m \times p$ matrix with complex entries and Σ is unknown Hermitian p.d. $p \times p$ matrix.
Write $\mathbf{X}' := (X_1, X_2, \ldots, X_m)$ and $\Xi' = (\xi_1, \xi_2, \ldots, \xi_m)$, the transpose of the matrix Ξ. Then X_i $(i = 1, 2, \ldots, m)$ are independently distributed as $\mathbb{C}N_p(\xi_i', \Sigma)$.

★ Assume that $n > p$.

Loss function

★ We consider the problem of estimating unknown mean matrix Ξ based on observation $\mathbf{X}$ and $\mathbf{S}$ under a loss function:

$$L((\Xi, \Sigma), \widehat{\Xi},) = \mathrm{Tr}\,\{\Sigma^{-1}(\widehat{\Xi} - \Xi)^*(\widehat{\Xi} - \Xi)\},$$

where $\widehat{\Xi}$ is an estimator for Ξ.

★ Risk function :

$$R((\Xi, \Sigma), \widehat{\Xi}) = \mathbb{E}[L((\Xi, \Sigma), \widehat{\Xi})],$$

Goal
Find estimators $\widehat{\Xi}$ s.t.

$$R((\Xi, \Sigma), \widehat{\Xi}) \leq R((\Xi, \Sigma), \mathbf{X}) = mp \qquad \text{for } \forall(\Xi, \Sigma).$$

Key ingredients

1. We consider a class of estimators $\widehat{\Xi}$ whose modification of the estimator $\mathbf{X}$ depends on the eigenvalues of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$.
2. Then $R((\Xi, \Sigma), \widehat{\Xi})$, the risk of estimators $\widehat{\Xi}$, depends on unknown parameters (Ξ, Σ).
3. To compare $R((\Xi, \Sigma), \widehat{\Xi})$ with $R((\Xi, \Sigma), \mathbf{X}) = mp$, we derive the unbiased estimate of the risk $R((\Xi, \Sigma), \widehat{\Xi})$.
$$R((\Xi, \Sigma), \widehat{\Xi}) = mp + \mathbb{E}\big[\widehat{\Delta}\big]$$
Here , $\widehat{\Delta}$ is a function of the eigenvalues of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$.
4. If $\widehat{\Delta} \le 0$, then $\widehat{\Xi}$ improves upon $\mathbf{X}$, i.e., $R((\Xi, \Sigma), \widehat{\Xi}) \le mp$ for $\forall(\Xi, \Sigma)$.
5. To derive $\widehat{\Delta}$, we use integration-by-parts formulae (Stein identity for complex normal distribution and Stein-Haff identity for complex Wishart distribution).

Notation 1

- ★ $\mathrm{Re}\, c$ and $\mathrm{Im}\, c$: the real and imaginary parts of a complex number c.
- ★ $\mathbb{R}^p_{\ge} = \{(\ell_1, \ell_2, \ldots, \ell_p) \in \mathbb{R}^p : \ell_1 \ge \ell_2 \ge \cdots \ge \ell_p > 0\}$
- ★ $\mathbb{C}^{m\times p}$: the sets of all $m \times p$ matrices of complex entries.
- ★ $\mathbf{C}'$ and $\overline{\mathbf{C}}$: the transpose and the conjugate of $\mathbf{C} \in \mathbb{C}^{m\times p}$. $\mathbf{C}^* = \overline{\mathbf{C}'}$.

Notation 2

- ★ For $z = x + \sqrt{-1}\,y$ $(x, y \in \mathbb{R})$ and differentiable function $g(z) = u(z) + \sqrt{-1}v(z)$ ($u(z)$ and $v(z)$ are real-valued differential functions),
$$\begin{aligned}\frac{\partial}{\partial z} g &= \frac{1}{2}\left(\frac{\partial}{\partial x} - \sqrt{-1}\frac{\partial}{\partial y}\right) g\\ &= \frac{1}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{\sqrt{-1}}{2}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right).\end{aligned}$$
Therefore
$$\mathrm{Re}\left(\frac{\partial g}{\partial z}\right) = \frac{1}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = \frac{1}{2}\left(\frac{\partial(\mathrm{Re}\, g)}{\partial(\mathrm{Re}\, z)} + \frac{\partial(\mathrm{Im}\, g)}{\partial(\mathrm{Im}\, z)}\right)$$

Notation 3

- ★ $\mathbf{G} = (g_{ij})_{i=1,2,\ldots,m,\, j=1,2,\ldots,p}$ be an $m \times p$ matrix (g_{ij}'s are complex-valued differentiable functions on $\mathbb{C}^{m\times p}$).
- ★ For $\mathbf{x} = (x_{ij})_{i=1,2,\ldots,m,\, j=1,2,\ldots,p} \in \mathbb{C}^{m\times p}$, set
$$\nabla_{\mathbf{x}} = \left(\frac{\partial}{\partial x_{ij}}\right)_{i=1,2,\ldots,m,\, j=1,2,\ldots,p},$$
and define
$$\begin{aligned}\mathrm{Re}(\mathrm{Tr}\,(\nabla'_x \mathbf{G})) &= \mathrm{Tr}\,(\mathrm{Re}(\nabla'_x \mathbf{G}))\\ &= \frac{1}{2}\sum_{j=1}^{p}\sum_{i=1}^{m}\left\{\frac{\partial(\mathrm{Re}\, g_{ij})}{\partial(\mathrm{Re}\, x_{ij})} + \frac{\partial(\mathrm{Im}\, g_{ij})}{\partial(\mathrm{Im}\, x_{ij})}\right\}.\end{aligned}$$

Lemma (Stein's identity for complex normal distribution)

Let X *be a* $m \times p$ *matrix having* $\mathbb{C}N_{m\times p}(\Xi, I_m \otimes \Sigma)$ *and let* $G = (g_{ij})_{i=1,2,\dots,m, j=1,2,\dots,p}$ *with*

$$\mathbb{E}\left|\frac{\partial(\mathrm{Re}\, g_{ij})}{\partial(\mathrm{Re}\, x_{ij})}\right|_{x=X} < \infty, \quad \mathbb{E}\left|\frac{\partial(\mathrm{Im}\, g_{ij})}{\partial(\mathrm{Im}\, x_{ij})}\right|_{x=X} < \infty,$$

and $x = (x_{ij})_{i=1,2,\dots,m, j=1,2,\dots,p}$ *in* $\mathbb{C}^{m\times p}$*. Then we have*

$$\mathbb{E}[\mathrm{Tr}\{\Sigma^{-1}(X-\Xi)^* G(X) + \mathrm{Tr}\{\Sigma^{-1}G^*(X)(X-\Xi)\}] = 2\mathbb{E}[\mathrm{Tr}(\mathrm{Re}(\nabla_x' G))|_{x=X}]$$

Sketch of proof

Recall that, for $X \sim \mathbb{C}N_p(\xi, \Sigma)$, we have

$$[X] := \begin{pmatrix} \mathrm{Re}\, X \\ \mathrm{Im}\, X \end{pmatrix} \sim N_{2p}\Bigg[\underbrace{\begin{pmatrix} \mathrm{Re}\, \xi \\ \mathrm{Im}\, \xi \end{pmatrix}}_{[\xi]:\, 2p\times 1}, \underbrace{\frac{1}{2}\begin{pmatrix} \mathrm{Re}\, \Sigma & -\mathrm{Im}\, \Sigma \\ \mathrm{Im}\, \Sigma & \mathrm{Re}\, \Sigma \end{pmatrix}}_{\{\Sigma\}:\, 2p\times 2p}\Bigg],$$

and note that

$$(X-\xi)^*\Sigma^{-1}g(X) + g^*(X)\Sigma^{-1}(X-\theta) = 2[X-\xi]'\{\Sigma\}^{-1}[g].$$

From Stein's identity for the multivariate real normal distribution, we have

$$\mathbb{E}[[X-\xi]'(\{\Sigma\}/2)^{-1}[g]] = \mathbb{E}\left[\sum_{i=1}^{p}\left(\frac{\partial(\mathrm{Re}\, g_i)}{\partial(\mathrm{Re}\, x_i)} + \frac{\partial(\mathrm{Im}\, g_i)}{\partial(\mathrm{Im}\, x_i)}\right)\Bigg|_{x=X}\right],$$

where $x' = (x_1, x_2, \dots, x_p)$ and $g' = (g_1, g_2, \dots, g_p)$.

We consider a real normal case for simplicity.
For $Z \sim N(\xi, 1)$ and a function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[g'(Z)] < \infty$, we have

$$\int_{-\infty}^{\infty}(z-\xi)g(z)\exp(-(z-\xi)^2/2)\,dz = \underbrace{\left[g(z)\exp(-(z-\xi)^2/2)\right]_{-\infty}^{\infty}}_{0} + \int_{-\infty}^{\infty} g'(z)\exp(-(z-\xi)^2/2)\,dz$$

because $\int_{-\infty}^{\infty} g(z)\exp(-(z-\xi)^2/2)\,dz < \infty$. This shows that

$$\mathbb{E}[(Z-\xi)g(Z)] = \mathbb{E}[g'(Z)].$$

Note that $\mathbb{E}[g'(Z)] < \infty$ implies $\mathbb{E}[Zg(Z)] < \infty$ and $\mathbb{E}[g(Z)] < \infty$.
See Sourav Chaterjee's lecture note(Stein's method and applications).

Notation

- ★ Let $s : p \times p = (s_{jk})$ and $G(s) = (g_{ij}(s))$.
- ★
$$\frac{\partial}{\partial s_{jk}} = \frac{1}{2}(1+\delta_{jk})\left\{\frac{\partial}{\partial(\mathrm{Re}\, s_{jk})} + (1-\delta_{jk})\sqrt{-1}\frac{\partial}{\partial(\mathrm{Im}\, s_{jk})}\right\},$$
for $j, k = 1, 2, \dots, p$ and δ_{jk} is the Kronecker delta.
- ★ Let $D_S = (\partial/\partial s_{jk})$ be a $p \times p$ operator matrix given by

$$\{D_S G(S)\}_{jk} = \sum_{l=1}^{p}\frac{\partial g_{lk}}{\partial s_{jl}}(S) = \frac{1+\delta_{jl}}{2}\sum_{l=1}^{p}\left\{\frac{\partial g_{lk}}{\partial(\mathrm{Re}\, s_{jl})} + (1-\delta_{j\ell})\sqrt{-1}\frac{\partial g_{lk}}{\partial(\mathrm{Im}\, s_{jl})}\right\}\Bigg|_{s=S}.$$

Lemma (Stein-Haff identity for complex Wishart distribution)

Assume that each entry of $\mathbf{G}(\mathbf{S})$ *is a partially differentiable function with respect to* $\mathrm{Re}\, s_{jk}$ *and* $\mathrm{Im}\, s_{jk}$, $j, k = 1, 2, \ldots, p$. *Under some regularity conditions on* $\mathbf{G}(\mathbf{S})$, *the following identity holds:*

$$\mathbb{E}[\mathrm{Tr}(\mathbf{G}(\mathbf{S})\boldsymbol{\Sigma}^{-1})] = \mathbb{E}[(n-p)\mathrm{Tr}(\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}) + \mathrm{Tr}(\mathbf{D}_S\mathbf{G}(\mathbf{S}))]. \quad (1)$$

Sketch of proof

For $x = (x_{ij})_{i=1,\ldots,n;\, j=1,\ldots,p}$, note that

$$\frac{\partial}{\partial x_{ij}} \exp(-\mathrm{Tr}\, x^* x) = -\bar{x}_{ij} \exp(-\mathrm{Tr}\, x^* x),$$

Assume that $X \sim \mathbb{C}N_{n\times p}(0, \mathbf{I}_n \otimes \mathbf{I}_p)$ and put $S = (s_{ij}) = X^*X \sim \mathbb{C}W(n, \mathbf{I}_p)$. Then we have

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^{p} s_{ij} g_{jl}\right] &= \mathbb{E}\left[\sum_{j=1}^{p}\sum_{k=1}^{n} \bar{x}_{ki} x_{kj} g_{jl}\right] = \mathbb{E}\left[\sum_{j=1}^{p}\sum_{k=1}^{n} \frac{\partial}{\partial x_{ki}}(x_{kj} g_{jl})\right] \\ &= \mathbb{E}\left[n g_{il} + \sum_{j=1}^{p}\sum_{k=1}^{n} x_{kj} \frac{\partial g_{jl}}{\partial x_{ki}}\right].\end{aligned}$$

In matrix form

$$\mathbb{E}[\mathbf{SG}] = \mathbb{E}[n\mathbf{G} + (\mathbf{X}'\nabla_{\mathbf{X}})'\mathbf{G}].$$

Transform $\mathbf{X} \to \mathbf{X}(\mathbf{P}^*)^{-1}$ ($\mathbf{S} \to \mathbf{P}^{-1}\mathbf{S}(\mathbf{P}^*)^{-1}$) and $\mathbf{G} \to \mathbf{P}^*\mathbf{G}(\mathbf{P}^*)^{-1}$, where $\boldsymbol{\Sigma} = \mathbf{PP}^*$ and $\mathbf{P}$ is a $p \times p$ nonsingular matrix. Then

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

Replacing $\mathbf{S}$ with $\mathbf{S}^{-1}\mathbf{G}$, taking the trace, and using the chain-rule of differentiation, we have

$$\mathbb{E}[\mathrm{Tr}\,\boldsymbol{\Sigma}^{-1}\mathbf{G}] = \mathbb{E}[(n-p)\mathrm{Tr}\,(\mathbf{G}\mathbf{S}^{-1}) + \mathrm{Tr}\,(D_S\mathbf{G})].$$

□

Class of estimators

- ★ Let $\mathbf{F} = \mathrm{Diag}\,(f_1, f_2, \ldots, f_{\min(m,p)})$ be the eigenvalues of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$.
- ★ For $p > m$ decompose $\mathbf{X}\mathbf{S}^{-1}\mathbf{X}^* = \mathbf{UFU}^*$, where $\mathbf{U}$ is an $m \times m$ unitary matrix.
- ★ For $m > p$ we decompose $\mathbf{S} = (\mathbf{A}^*)^{-1}\mathbf{A}^{-1}$ and $\mathbf{X}^*\mathbf{X} = (\mathbf{A}^*)^{-1}\mathbf{F}\mathbf{A}^{-1}$, where $\mathbf{A}$ is a $p \times p$ non-singular matrix.
- ★ Consider a class of estimators of the form

$$\widehat{\Xi}_H := \widehat{\Xi}_H(\mathbf{X}, \mathbf{S}) = \begin{cases} \mathbf{X}\{\mathbf{I}_p + \mathbf{AH}(\mathbf{F})\mathbf{A}^{-1}\} & \text{if } m > p \\ \{\mathbf{I}_m + \mathbf{UH}(\mathbf{F})\mathbf{U}^*\}\mathbf{X} & \text{if } m < p \end{cases}, \quad (2)$$

where $\mathbf{H} := \mathbf{H}(\mathbf{F}) = \mathrm{Diag}\,(h_1(\mathbf{F}), h_2(\mathbf{F}), \ldots, h_{\min(m,p)}(\mathbf{F}))$ whose i-th element $h_i := h_i(\mathbf{F})$, $i = 1, 2, \ldots, \min(m, p)$, is a real-valued function on $\mathbb{R}_{\geq}^{\min(m,p)}$.

Theorem

Under the suitable conditions, we have

$$R((\Xi,\Sigma),\widehat{\Xi}_H) = \begin{cases} mp + \mathbb{E}\Big[\widehat{\Delta}(n,m,p;\mathbf{H})\Big] & \text{if } m > p \\ mp + \mathbb{E}\Big[\widehat{\Delta}(n+m-p,p,m;\mathbf{H})\Big] & \text{if } m < p \end{cases}$$

where $h_{kk}(\mathbf{F}) = (\partial h_k/\partial f_k)(\mathbf{F})$ *and*

$$\widehat{\Delta}(n,m,p;\mathbf{H}) = \sum_{k=1}^{p}\Big\{2(m-p+1)h_k(\mathbf{F}) + 2f_k h_{kk}(\mathbf{F})$$

$$+4\sum_{b>k}\frac{f_k h_k(\mathbf{F}) - f_b h_b(\mathbf{F})}{f_k - f_b} + (n+p-2)f_k h_k^2(\mathbf{F})$$

$$-2f_k^2 h_{kk}(\mathbf{F})h_k(\mathbf{F}) - 2\sum_{b>k}\frac{f_k^2 h_k^2(\mathbf{F}) - f_b^2 h_b^2(\mathbf{F})}{f_k - f_b}\Big\}\mathbb{I}\{f_1 > \cdots > f_p > 0\}$$

Evaluation of risk

- ★ We have

$$R((\Xi,\Sigma),\widehat{\Xi}_H) = \mathbb{E}\Big[\mathrm{Tr}\,\{\Sigma^{-1}(\mathbf{X}+\mathbf{G}-\Xi)^*(\mathbf{X}+\mathbf{G}-\Xi)\}\Big]$$

$$= \underbrace{\mathbb{E}\Big[\mathrm{Tr}\,\{\Sigma^{-1}(\mathbf{X}-\Xi)^*(\mathbf{X}-\Xi)\}\Big]}_{mp}$$

$$+\underbrace{\mathbb{E}\Big[\mathrm{Tr}\,\{\Sigma^{-1}\mathbf{G}^*(\mathbf{X}-\Xi)\} + \mathrm{Tr}\,\{\Sigma^{-1}(\mathbf{X}-\Xi)^*\mathbf{G}\}\Big]}_{\text{Stein's identity}}$$

$$+ \underbrace{\mathbb{E}\Big[\mathrm{Tr}\,\{\Sigma^{-1}\mathbf{G}^*\mathbf{G}\}\Big]}_{\text{Stein and Haff identity}}$$

$$= mp + \mathbb{E}\Big[2\,\mathrm{Tr}\,\{\mathrm{Re}\,(\nabla'_X\mathbf{G})\}\Big]$$

$$+\mathbb{E}\Big[\mathrm{Tr}\,\{\mathbf{D}_S\mathbf{G}^*\mathbf{G}\} + (n-p)\mathrm{Tr}\,\{\mathbf{S}^{-1}\mathbf{G}^*\mathbf{G}\}\Big].$$

- ★ When $m > p$, recall that $\widehat{\Xi}_H = \mathbf{X}+\mathbf{G} = \mathbf{X}+\mathbf{XAH(F)A}^{-1}$, where $\mathbf{A}$ is a $p\times p$ nonsingular matrix s.t. $\mathbf{S} = (\mathbf{A}^*)^{-1}\mathbf{A}^{-1}$ and $\mathbf{A}^*\mathbf{X}^*\mathbf{XA} = \mathrm{F} = \mathrm{Diag}\,(f_1,\ldots,f_p)$, and that the modification $\mathbf{H}$ depends only on the eigenvalues of $\mathbf{X}^*\mathbf{XS}^{-1}$, .
- ★

$$R((\Xi,\Sigma),\widehat{\Xi}_H) = mp + \mathbb{E}\Big[2\,\mathrm{Tr}\,\{\mathrm{Re}\,(\nabla'_X\mathbf{G})\}\Big]$$

$$+\mathbb{E}\Big[\mathrm{Tr}\,\{\mathbf{D}_S\mathbf{G}^*\mathbf{G}\} + (n-p)\mathrm{Tr}\,\{\mathbf{S}^{-1}\mathbf{G}^*\mathbf{G}\}\Big].$$

$$= mp + \mathbb{E}\Big[2\,\mathrm{Tr}\,\{\mathrm{Re}\,(\nabla'_X\mathbf{XAHA}^{-1})\}\Big]$$

$$+\mathbb{E}\Big[\mathrm{Tr}\,\{\mathbf{D}_S\,\underbrace{(\mathbf{XAHA}^{-1})^*(\mathbf{XAHA}^{-1})}_{(\mathbf{A}^*)^{-1}\mathbf{HFH(F)A}^{-1}}\}$$

$$+(n-p)\,\underbrace{\mathrm{Tr}\,\{\mathbf{S}^{-1}(\mathbf{XAHA}^{-1})^*(\mathbf{XAHA}^{-1})\}}_{\mathrm{Tr}\,\{\mathbf{FH}^2)\}}\Big].$$

- ★ The terms inside the first and second expectations are written as

$$\mathrm{Tr}\,\{\mathrm{Re}\,(\nabla'_X\mathbf{XAHA}^{-1})\}$$

$$= \sum_{k=1}^{p}\Big\{f_k h_{kk} + (m-p+1)h_{kk} + 2\sum_{b>k}\frac{f_k h_k - f_b h_b}{f_k - f_b}\Big\}$$

$$\mathrm{Tr}\,\{\mathbf{D}_S((\mathbf{A}^*)^{-1}\mathbf{HFHA}^{-1})\}$$

$$= \sum_{k=1}^{p}\Big\{2(p-1)f_k h_k^2 - 2f_k h_{kk}h_k - 2\sum_{b>k}\frac{f_k^2h_k^2 - f_b^2h_b^2}{f_k - f_b}\Big\},$$

where $\mathbf{H} = \mathrm{Diag}(h_1,\ldots,h_p)$ and $h_{kk} = \partial h_k/\partial f_k$, and $\mathbf{A}$ is a $p\times p$ nonsingular matrix such $\mathbf{S} = (\mathbf{A}^*)^{-1}\mathbf{A}^{-1}$, and that the modification $\mathbf{H}$ depends only on the eigenvalues of $\mathbf{X}^*\mathbf{XS}^{-1}$, $\mathbf{F} = \mathrm{Diag}(f_1, f_2, \ldots, f_p)$.

- ★ When $m < p$, recall that $\widehat{\Xi}_H := \mathbf{X} + \mathbf{G}$ where $\mathbf{G} = \mathbf{UHU^*X}$, where $\mathbf{U}$ is an $m \times m$ unitary matrix such that $\mathbf{XS^{-1}X^*} = \mathbf{UFU^*}$, and recall that an $m \times m$ diagonal matrix $\mathbf{H}$ depends only on $\mathbf{F} = \mathrm{Diag}(f_1, f_2, \dots, f_m)$.
- ★ For $m < p$, we have

$$\begin{aligned} R((\Xi, \Sigma), \widehat{\Xi}_H) &= mp + \mathbb{E}\Big[2\,\mathrm{Tr}\,\{\mathrm{Re}\,(\nabla'_X \mathbf{G})\}\Big] \\ &\quad + \mathbb{E}\Big[\mathrm{Tr}\,\{\mathbf{D_S G^* G} + (n-p)\mathrm{Tr}\,\{\mathbf{S^{-1}G^*G}\}\Big]. \\ &= mp + \mathbb{E}\Big[2\,\mathrm{Tr}\,\{\mathrm{Re}\,(\mathbf{UH(F)U^*X})\}\Big] \\ &\quad + \mathbb{E}\Big[\mathrm{Tr}\,\{\mathbf{D_S} \underbrace{(\mathbf{UH(F)U^*X})^*(\mathbf{UH(F)U^*X})}_{\mathbf{X^*UH^2U^*X}}\} \\ &\quad + (n-p)\underbrace{\mathrm{Tr}\,\{\mathbf{S^{-1}}(\mathbf{UH(F)U^*X})^*(\mathbf{UH(F)U^*X})\}}_{\mathrm{Tr}\,\{\mathbf{FH^2}\}}\Big]. \end{aligned}$$

- ★ The terms inside the first and second expectations are written as

$$\begin{aligned} &\mathrm{Tr}\,\{\mathrm{Re}\,(\nabla'_X \mathbf{UHU^*X})\} \\ &\quad = \sum_{k=1}^{m}\left\{f_k h_{kk} + (p-m+1)h_{kk} + 2\sum_{b>k}\frac{f_k h_k - f_b h_b}{f_k - f_b}\right\} \\ &\mathrm{Tr}\,\{\mathbf{D_S}(\mathbf{X^*UH^2UX})\} \\ &\quad = \sum_{k=1}^{m}\left\{2(m-1)f_k h_k^2 - 2f_k^2 h_{kk} h_k - 2\sum_{b>k}\frac{f_k^2 h_k - f_b^2 h_b}{f_k - f_b}\right\}, \end{aligned}$$

where $\mathbf{H} = \mathrm{Diag}(h_1, \dots, h_m)$ and $h_{kk} = \partial h_k / \partial f_k$, and $\mathbf{U}$ is an $m \times m$ unitary matrix s.t. $\mathbf{XS^{-1}X^*} = \mathbf{U}\mathrm{Diag}(f_1, f_2, \dots, f_m)\mathbf{U^*}$. □

Theorem (Efron-Morris type estimator)

Let

$$c = \begin{cases} \frac{m-p}{n+p} & \text{if } m > p \\ \frac{p-m}{n+2m-p} & \text{if } m < p \end{cases}$$

and put

$$\mathbf{H(F)} = -\mathrm{Diag}\left(\frac{c}{f_1}, \frac{c}{f_2}, \dots, \frac{c}{f_{\min(m,p)}}\right)$$

Then the estimator

$$\widehat{\Xi}_H = \begin{cases} \mathbf{X}\{\mathbf{I}_p - c(\mathbf{X^*X})^{-1}\mathbf{S}\} & \text{if } m > p \\ \{\mathbf{I}_m - c(\mathbf{XS^{-1}X^*})^{-1}\}\mathbf{X} & \text{if } m < p \end{cases}$$

improves upon $\mathbf{X}$, *i.e.*, $R((\Xi, \Sigma), \widehat{\Xi}_H) \le R((\Xi, \Sigma), \mathbf{X})$ *for* $\forall(\Xi, \Sigma)$.

Theorem

Let $c_k = \frac{m+p-2k}{n-p+2k}$ $(k = 1, 2, \dots, \min(m, p))$ *and put*

$$\mathbf{H(F)} = -\mathrm{Diag}\left(\frac{c_1}{f_1}, \frac{c_2}{f_2}, \dots, \frac{c_{\min(m,p)}}{f_{\min(m,p)}}\right).$$

Then the estimator

$$\widehat{\Xi} = \begin{cases} \mathbf{X}\{\mathbf{I}_p + \mathbf{AH(F)A^{-1}}\} & \text{if } m > p \\ \{\mathbf{I}_m + \mathbf{UH(F)U^*}\}\mathbf{X} & \text{if } m > p \end{cases}$$

improves upon $\mathbf{X}$, *i.e.*, $R((\Xi, \Sigma), \widehat{\Xi}_H) \le R((\Xi, \Sigma), \mathbf{X})$ *for* $\forall(\Xi, \Sigma)$.

Recall that $\mathbf{U}$ is an $m \times m$ unitary matrix s.t. $\mathbf{XS^{-1}X^*} = \mathbf{UFU^*}$ and that $\mathbf{A}$ is a $p \times p$ nonsingular matrix s.t. $\mathbf{S} = (\mathbf{A^*})^{-1}\mathbf{A}^{-1}$.

1. The results are complex version of those for estimation problem for the mean matrix of the real multivariate normal distribution (see Konno (JMVA, 1991) and Kariya et al. (JMVA, 1999)).
2. However the results are parallel to real case.
3. Note that the substitution rule to get the second part from the first part is $(p, m, n) \to (m, p, n+m-p)$, i.e.,

$$R((\Xi, \Sigma), \widehat{\Xi}_H) = \begin{cases} mp + \mathbb{E}\left[\widehat{\Delta}(n, m, p; \mathrm{H})\right] & \text{if } m > p \\ mp + \mathbb{E}\left[\widehat{\Delta}(n+m-p, p, m; \mathrm{H})\right] & \text{if } m < p \end{cases}$$

This is the same as the rule substitution for the joint distribution of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$.

When $\Xi = \mathbf{0}$, the joint distribution of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$ is, apparting from normalizing constants,

$$\Pi_{k=1}^{p} \frac{f_k^{m-p}}{(1+f_k)^{n+m}} \Pi_{k=1}^{p-1} \Pi_{j=k+1}^{p} (f_k - f_j)^2 \Pi_{k=1}^{p} \mathrm{d} f_k$$

if $m > p$ while it is

$$\Pi_{k=1}^{m} \frac{f_k^{p-m}}{(1+f_k)^{n+m}} \Pi_{k=1}^{m-1} \Pi_{j=k+1}^{m} (f_k - f_j)^2 \Pi_{k=1}^{m} \mathrm{d} f_k,$$

if $p > m$. The second one is obtained by the substitution rule $(p, m, n) \to (m, p, n+m-p)$. Good reference is Khatri(1965, Ann. ,Math. Statist).