Shrinkage estimation of the mean matrix of multivariate complex normal distributions

Yoshihiko KONNO

Japan Women's University, Faculty of Science

November 15, 2010

★ Z は標準複素正規分布 $\mathbb{C}N(0, 1)$ に従うとは

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

★ Z の確率密度関数 (w.r.t. Lebesgue measure on $\mathbb{C}$) は

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

ただし，$\bar{x}$ は x ($x \in \mathbb{C}$) の complex conjugate

★ $Z \sim \mathbb{C}N(0, 1)$, $\theta \in \mathbb{C}$, $\sigma \in \mathbb{R}_+$ に対して

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

★ X は $\mathbb{C}^p$ 値複素確率ベクトルとする．$\forall c \in \mathbb{C}^p$，$\theta \in \mathbb{C}^p$，$\Sigma \in \mathrm{Herm}(p, \mathbb{C})_+$ に対して，

$$c^* X \sim \mathbb{C}N(c^*\theta, c^*\Sigma c) \iff X \sim \mathbb{C}N_p(\theta, \Sigma).$$

ただし，c^* は c の transpose complex conjugate である．

★ $X \sim \mathbb{C}N_p(\theta, \Sigma)$ の確率密度関数 (w.r.t. Lebesgue measure on $\mathbb{C}^p$) は

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

★ $X \sim \mathbb{C}N_p(\theta, \Sigma)$ のとき，

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

ただし，$\mathrm{Re}\, \Sigma$, $\mathrm{Im}\, \Sigma$ は symmetric と skew-symmetric.

★ $\mathbb{C}^p$ 値複素確率ベクトル $X_1, X_2, \cdots, X_n$ は独立同一に $\mathbb{C}N_p(0, \Sigma)$ に従うとする．このとき，

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

は母数 Σ, p, n の複素 Wishart 分布に従うといい，$\mathbb{C}W_p(\Sigma, n)$ と書く．

★ $n \geq p$ のとき，$\mathbb{P}(S$ は正定値$) = 1$ で，S の確率密度関数 (w.r.t. Lebesgue measure on $\mathbf{Herm}(\mathbb{C}, p)$) は

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

ただし，$w \in \mathbf{Herm}^+(\mathbb{C}, p)$，$\Gamma(\cdot)$ は Euler's gamma function.

観測

- ★ X : $m \times p \sim \mathbb{C}N_{m\times p}(\Xi, I_m \otimes \Sigma)$
- ★ S : $p \times p \sim \mathbb{C}W_p(\Sigma, n)$
- ★ X と S は独立

  ただし，$n > p$ で，Σ は未知の $p \times p$ 正定値エルミート行列

損失関数

- ★ X と S に基づく未知の平均行列 Ξ の推定問題をつぎの損失関数のもとで考える：

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

  ただし，$\widehat{\Xi}$ は Ξ の推定量で，* は complex-conjugate transpose.

- ★ リスク関数：

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

方針

1. X^*XS^{-1} の固有値にのみ本質的に依存する修正項を X に加えた推定量の族 $\widehat{\Xi}$ を考える.
2. 推定量族 $\widehat{\Xi}$ のリスク $R(\widehat{\Xi}, (\Xi, \Sigma))$ は未知母数 (Ξ, Σ) に依存.
3. X のリスク $R(X, (\Xi, \Sigma)) = mp$ と比較するために，推定量の族のリスクの不偏推定量を求める.

$$R(\widehat{\Xi}, (\Xi, \Sigma)) = mp + \mathbb{E}\left[\widehat{\Delta}\right]$$

ただし，$\widehat{\Delta}$ は未知母数に依存せず，X^*XS^{-1} の固有値にのみ依存.

4. $\widehat{\Delta} \leq 0$ ならば，$\widehat{\Xi}$ は X を改良.
5. $\widehat{\Delta}$ の導出に部分積分 (Stein identity と Stein-Haff identity) の公式等を使う.

準備：記号

- ★ Let $\mathbb{R}$ and $\mathbb{C}$ denote the field of real and complex numbers.
- ★ $\mathrm{Re}\,c$ and $\mathrm{Im}\,c$: the real and imaginary parts of a complex number c.
- ★ $\mathbb{R}^p_> = \{(\ell_1, \ell_2, \dots, \ell_p) \in \mathbb{R}^p : \ell_1 > \ell_2 > \cdots > \ell_p > 0\}$
- ★ $\mathbb{R}^{m\times p}$ and $\mathbb{C}^{m\times p}$: the sets of all $m \times p$ matrices of real and complex entries.
- ★ $\mathbf{C}'$ and $\overline{\mathbf{C}}$: the transpose and the conjugate of $\mathbf{C} \in \mathbb{C}^{m\times p}$. $\mathbf{C}^* = \overline{\mathbf{C}'}$.
- ★ $\mathbb{C}^{p\times p}_+$: the set of $p \times p$ Hermitian positive definite matrices.

準備：記号 2

★ For $z = x + \sqrt{-1}\,y$ $(x, y \in \mathbb{R})$ and differentiable function $g(z) = u(z) + \sqrt{-1}\,v(z)$,

$$
\begin{aligned}
\frac{\partial}{\partial z} g &= \frac{1}{2}\left(\frac{\partial}{\partial x} - \sqrt{-1}\,\frac{\partial}{\partial y}\right) g \\
&= \frac{1}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{\sqrt{-1}}{2}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right), \\
\frac{\partial}{\partial \bar{z}} g &= \frac{1}{2}\left(\frac{\partial}{\partial x} + \sqrt{-1}\,\frac{\partial}{\partial y}\right) g \\
&= \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + \frac{\sqrt{-1}}{2}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right).
\end{aligned}
$$

準備：記号 3

★ $\mathbf{G} = (g_{ij})_{i=1,2,\dots,m,\,j=1,2,\dots,p}$ be an $m \times p$ matrix (g_{ij}'s are complex-valued differentiable functions on $\mathbb{C}^{m\times p}$).

★ For $\mathbf{x} = (x_{ij})_{i=1,2,\dots,m,\,j=1,2,\dots,p} \in \mathbb{C}^{m\times p}$, set

$$\nabla_{\mathbf{x}} = \left(\frac{\partial}{\partial x_{ij}}\right)_{i=1,2,\dots,m,\,j=1,2,\dots,p},$$

and define

$$\begin{aligned}\mathrm{Re}(\mathrm{Tr}\,(\nabla_x' \mathbf{G})) &= \mathrm{Tr}\,(\mathrm{Re}(\nabla_x' \mathbf{G})) \\ &= \sum_{j=1}^{p}\sum_{i=1}^{m}\left\{\frac{\partial(\mathrm{Re}\, g_{ij})}{\partial(\mathrm{Re}\, x_{ij})} + \frac{\partial(\mathrm{Im}\, g_{ij})}{\partial(\mathrm{Im}\, x_{ij})}\right\}.\end{aligned}$$

Lemma (Stein's identity)

Let X *be a* $p \times 1$ *complex random vector having* $\mathbb{C}N_p(\theta, \Sigma)$ *and let* $\mathbf{g} = (g_1, g_2, \ldots, g_p) : \mathbb{C}^p \to \mathbb{C}^p$ *be differentiable with*

$$\mathbb{E}\left|\frac{\partial\,(\mathrm{Re}\,g_i)}{\partial\,(\mathrm{Re}\,x_i)}\right|_{\mathbf{x}=X} < \infty, \quad \mathbb{E}\left|\frac{\partial\,(\mathrm{Im}\,g_i)}{\partial\,(\mathrm{Im}\,x_i)}\right|_{\mathbf{x}=X} < \infty, \quad i = 1, 2, \ldots, p,$$

and $\mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{C}^p$*. Then we have*

$$\begin{aligned}&\mathbb{E}[(X-\theta)^*\Sigma^{-1}\mathbf{g}(X) + \mathbf{g}^*(X)\Sigma^{-1}(X-\theta)]\\ &= \mathbb{E}\left[\sum_{i=1}^{p}\left\{\frac{\partial\,(\mathrm{Re}\,g_i)}{\partial\,(\mathrm{Re}\,x_i)} + \frac{\partial\,(\mathrm{Im}\,g_i)}{\partial\,(\mathrm{Im}\,x_i)}\right\}\Bigg|_{\mathbf{x}=X}\right].\end{aligned}$$

- ★ Assume that $\mathbf{S} = (s_{jk}) \sim \mathbb{C}W_p(\boldsymbol{\Sigma}, n)$ and let $\mathbf{G}(\mathbf{S}) = (g_{ij}(\mathbf{S}))$.
- ★ Let $\mathbf{D}_S = (\partial/\partial s_{jk})$ be a $p \times p$ operator matrix, and

$$\frac{\partial}{\partial s_{jk}} = \frac{1}{2}(1+\delta_{jk})\left\{\frac{\partial}{\partial(\mathrm{Re}\, s_{jk})} + \sqrt{-1}\frac{\partial}{\partial(\mathrm{Im}\, s_{jk})}\right\}, \tag{1}$$

for $j, k = 1, 2, \ldots, p$ and δ_{jk} is the Kronecker delta.
- ★

$$\begin{aligned}\{\mathbf{D}_S\mathbf{G}(\mathbf{S})\}_{jk} &= \sum_{l=1}^{p}\frac{\partial g_{lk}}{\partial s_{jl}}(\mathbf{S}) \\ &= \frac{1}{2}(1+\delta_{jl})\sum_{l=1}^{p}\left\{\frac{\partial g_{lk}}{\partial(\mathrm{Re}\, s_{jl})} + \sqrt{-1}\frac{\partial g_{lk}}{\partial(\mathrm{Im}\, s_{jl})}\right\}.\end{aligned}$$

Lemma (Stein-Haff identity)

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:*

$$\mathbb{E}[\mathrm{Tr}(\mathbf{G}(\mathbf{S})\boldsymbol{\Sigma}^{-1})] = \mathbb{E}[(n-p)\mathrm{Tr}(\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}) + \mathrm{Tr}(\mathbf{D_S}\mathbf{G}(\mathbf{S}))]. \quad (2)$$

推定量の族

- ★ Let $\mathbf{F} = \mathbf{Diag}\,(f_1, f_2, \ldots, f_{\min(m,p)})$ be the eigenvalues of $\mathbf{X}^*\mathbf{X}\mathbf{S}^{-1}$.
- ★ For $p > m$ decompose $\mathbf{X}\mathbf{S}^{-1}\mathbf{X}^* = \mathbf{U}\mathbf{F}\mathbf{U}^*$, 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

$$\widehat{\Xi}_H := \widehat{\Xi}_H(\mathbf{X}, \mathbf{S}) = \begin{cases} \mathbf{X}\{\mathbf{I}_p + \mathbf{A}\mathbf{H}(\mathbf{F})\mathbf{A}^{-1}\} & \text{if } m > p \\ \{\mathbf{I}_m + \mathbf{U}\mathbf{H}(\mathbf{F})\mathbf{U}^*\}\mathbf{X} & \text{if } p > m \end{cases}, \tag{3}$$

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

リスクの評価

- Stein's identity と Stein-Haff identity を用いて，未知母数を消去：

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

- $G = XAH(F)A^{-1}$ および $G = UH(F)U^*X$ はある種の不変性をもち, X^*XS^{-1} の固有値 $F = \mathrm{Diag}(f_1, f_2, \ldots, f_{\min(m,p)})$ にのみ本質的に依存することを利用する．

Theorem

Under the suitable conditions, we have

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

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

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

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

$$-2 f_k^2 h_{kk}(\mathrm{F}) h_k(\mathrm{F}) - 2 \sum_{b>k} \frac{f_k^2 h_k^2(\mathrm{F}) - f_b^2 h_b^2(\mathrm{F})}{f_k - f_b} \Big\}$$

Theorem

For $k = 1, 2, \ldots, \min(m, p)$, *let*

$$c_k^{(AS)} = \frac{m + p - 2k}{n - p + 2k},$$

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

Then the estimator

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

is minimax.

1. 実正規分布と Wishart 分布に対する平均行列の推定問題に関して得られた結果 (Konno (1991) and Kariya et al. (1999)) の複素バージョンが得られた.
2. 形式的にはほとんど変わらない！
3. 実正規分布の場合と比較して，縮小型推定量の存在条件が変わる：$m > p + 1 \longrightarrow m > p$ or $p > m + 1 \longrightarrow p > m$
4. 複素バージョンでしか得られない結果があればよいが，そのようなものはないのか？
5. 複素バージョンの場合の方が対称性があり，計算の見通しは楽.