Improved mean-matrix estimation of multivariate complex normal distributions by a singular value shrinkage

Yoshihiko KONNO(今野良彦)

Japan Women's University（日本女子大学）

OCAMI 研究集会
統計的推測理論への幾何学的アプローチ

11–12 December 2021

Matrix Estimation Problem of $\boldsymbol{\Theta}$ in MANOVA [1]

$$\begin{array}{cc} & p \\ \begin{matrix} m \\ n \end{matrix} & \begin{pmatrix} Z \\ U \end{pmatrix} \end{array} = \begin{pmatrix} \Theta \\ 0 \end{pmatrix} + E; \qquad E = \begin{array}{c} p \\ \begin{pmatrix} \longleftrightarrow \\ \longleftrightarrow \\ \vdots \\ \longleftrightarrow \\ \longleftrightarrow \end{pmatrix} \end{array}$$

where $\begin{bmatrix} Z \\ U \end{bmatrix}$ is observation and Θ is an $m \times p$ non-random matrix(unknown), E is an $(m+n) \times p$ error matrix(unobservable) whose rows are identically distributed as $N_p(0, \Sigma)$ $(\mathbb{C}N_p(0, \Sigma))$. Here Σ is a $p \times p$ positive-definite and unknown matrix(Hermitian).

---

[1] =Multivariate ANalysis Of VAriance

Our approach

1. $n < p$ case(singular Wishart matrix case) is included; high dimensional data case.
    1. Chételat and Wells (Ann. Statist., 2012) ; they projected Z onto the space spaned by columns of a Wishart matrix $U^* U =: S$ and shrink the projected factor $ZS^{\dagger}Z^* =: F$, where $S^{\dagger}$ and Z^* are the Poore-Menrose inverse of S and the complex-conjugate transpose of Z respectively.
    2. Tsukuma and Kubokawa (JMVA, 2015); they proposed a general class of estimators, shrink singular values of F matrix and took positive parts of shrinked singular values.
2. A method of evaluation of estimators: a classical SURE(Stein's unbiased risk estimator).

Notation

(1) Let $m, n, p \in \mathbb{N}$ such that $\max(m, p) \geq 2$, $n \geq 2$ and $n \neq p$.

(2) $\mathbf{M}_{m,p}(\mathbb{C})$ denotes the set of $m \times p$ matrices over $\mathbb{C}$. For a matrix $C \in \mathbf{M}_{m,p}(\mathbb{C})$, C^* stands for the complex-conjugate transpose of C.

(3) A $p \times p$ matrix $C \in \mathbf{M}_{p,p}(\mathbb{C})$ is Hermitian if $C = C^*$. $\mathbf{Herm}^{++}(p, \mathbb{C})$ stands for the set of all $p \times p$ positive-definite Hermitian matrices.

(4) For $m \geq p$, $\mathbf{V}_{m,p}(\mathbb{C}) = \{V \in \mathbf{M}_{m,p}(\mathbb{C});\ V^* V = I_p\}$

(5) Let $\mathbf{D}_p^+ = \{\mathrm{Diag}(d_1, d_2, \ldots, d_p);\ d_1 > d_2 > \cdots > d_p > 0\}$.

Model: A canonical form

(6) Let

$$\begin{aligned} \boldsymbol{Z} : \boldsymbol{m} \times \boldsymbol{p} &= (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_m)^\top \\ \boldsymbol{U} : \boldsymbol{n} \times \boldsymbol{p} &= (\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_n)^\top, \end{aligned}$$

where $\boldsymbol{Z}_i' \boldsymbol{s}$ and $\boldsymbol{U}_i' \boldsymbol{s}$ are independently distributed as

$$\begin{aligned} \boldsymbol{Z}_i &\sim \mathbb{C}\boldsymbol{N}_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma})\ (\boldsymbol{i} = 1, 2, \ldots, \boldsymbol{m}) \\ \boldsymbol{U}_i &\sim \mathbb{C}\boldsymbol{N}_p(0, \boldsymbol{\Sigma})\ (\boldsymbol{i} = 1, 2, \ldots, \boldsymbol{n}) \end{aligned}$$

with $\boldsymbol{\theta}_i \in \mathbb{C}^p$ $(\boldsymbol{i} = 1, 2, \ldots, \boldsymbol{m})$ and $\boldsymbol{\Sigma} \in \mathbf{Herm}^{++}(\mathbb{C})$ being unknown.

(7) Put

$$\boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{U}_i \boldsymbol{U}_i^* \sim \boldsymbol{W}_p(\boldsymbol{n}, \boldsymbol{\Sigma}).$$

Note that $\boldsymbol{S}$ is singular if $2 \leq \boldsymbol{n} < \boldsymbol{p}$.

Set-up of problem

(8) Consider the problem of estimating $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)^\top$ under an invariant loss function

$$\mathbf{L}(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Theta} | \boldsymbol{\Sigma}) = \mathrm{tr}\{(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^*\}$$

where $\widehat{\boldsymbol{\Theta}}$: $\boldsymbol{m} \times \boldsymbol{p}$ is an estimator of $\boldsymbol{\Theta}$ based on $(\boldsymbol{Z}, \boldsymbol{S})$.

(9) The risk of $\widehat{\boldsymbol{\Theta}}$ is denote by

$$\mathbf{R}(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Theta} | \boldsymbol{\Sigma}) = \mathbb{E}[\mathbf{L}(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Theta} | \boldsymbol{\Sigma})]$$

where the expectation is taken with respect to the joint distribution of $(\boldsymbol{Z}, \boldsymbol{S})$.

The Moore-Penrose generalized inverse

To derive a shrinkage estimator, we use the Moore-Penrose inverse of $\boldsymbol{S}$.

Definition of the Moore-Penrose generalized inverse

For an $m \times n$ complex matrix $\boldsymbol{A}$, $n \times m$ complex matirx $\boldsymbol{A}^{\dagger}$ is the Moore-Pensrose generalized inverse of $\boldsymbol{A}$ if following conditions (i)~(iv) are satisfied:
(i) $\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{A}$;
(ii) $\boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{\dagger}$ (reflective condition);
(iii) $(\boldsymbol{A}\boldsymbol{A}^{\dagger})^{*} = \boldsymbol{A}\boldsymbol{A}^{\dagger}$ (minimum least squared condition);
(iv) $(\boldsymbol{A}^{\dagger}\boldsymbol{A})^{*} = \boldsymbol{A}^{\dagger}\boldsymbol{A}$ (minimum norm condition).

Remark For any $m \times n$ matrices $\boldsymbol{A}$, the Moore-Penrose generalized inverse of $\boldsymbol{A}$ exists uniquely.

A general class of estimators

Following the idea due to Chételat and Wells (Ann. Statist., 2012) and Tsukuma and Kubokawa (JMVA, 2015), we consider a class of estimators below.

Idea of estimator

(1) Project each row of $\boldsymbol{Z}$ into the space spanned by the columns of $\boldsymbol{S}$.

$$\mathbb{C}^p = \underbrace{\mathbf{span}(\boldsymbol{S})}_{\text{Shrink}} \oplus \mathbf{span}(\boldsymbol{S})^\perp$$

(2) Shrink the first factor above by using singular values of $\boldsymbol{Z}\boldsymbol{S}^\dagger\boldsymbol{Z}^*$.

A general class of estimators

$$\widehat{\Theta}^{SH} = \underbrace{Z(I_p - SS^{\dagger})}_{\mathrm{span}(S)^{\perp}} + R \underbrace{(I_\ell - \Phi(F))}_{\text{Shrinkage factor}} R^* \underbrace{ZSS^{\dagger}}_{\mathrm{span}(S)} .$$

(1) $m = 1,\ n < p\ (\ell = 1)$: Chételat and Wells (AOS, 2012) .
$\ell \geq 2$: Tsukuma and Kubokawa (JMVA, 2015).

(2) Define $ZS^{\dagger}Z^* = RFR^*$ where $R \in \mathrm{V}_{m,\ell}(\mathbb{C})$ and $F = \mathrm{Diag}(f_1, \ldots, f_\ell) \in \mathrm{D}_\ell^+$.

- Decompose $S = HLH^*$ where $L \in \mathrm{D}^+_{n\wedge p}$ and $H \in \mathrm{V}_{p,n\wedge p}(\mathbb{C})$.
- Define by $ZHL^{-1/2} = RF^{1/2}V^*$ and $V \in \mathrm{V}_{n\wedge p,\ell}(\mathbb{C})$.

(3) $\Phi(F) = \mathrm{Diag}(\phi_1(F), \ldots, \phi_\ell(F))$ where $\phi_i : \mathrm{D}_\ell^+ \longrightarrow [0, \infty)$ is a weakly differentiable $(i = 1, 2, \ldots, \ell)$.

Positive-part

$$\widehat{\Theta}^{SH+} = \underbrace{Z(I_p - SS^{\dagger})}_{\text{span}(S)^{\perp}} + R \underbrace{(I_\ell - \Phi(F))_+}_{\text{Shrinkage factor}} R^* \underbrace{ZSS^{\dagger}}_{\text{span}(S)}$$

where

$$(I_\ell - \Phi(F))_+ = \text{Diag}(\max(1-\phi_1, 0), \ldots, \max(1-\phi_\ell, 0)).$$

Stein's Unbiased Risk Estimator

Under certain regularity conditions, Stein's unbiased risk estimator for δ^{SH} is given by

$$\mathbf{R}(\widehat{\Theta}^{SH}, \Theta | \Sigma) - \mathbf{R}(Z, \Theta | \Sigma) = \mathbb{E}[\,\mathbf{SURE}(\Phi)],$$

$$\mathbf{SURE}(\Phi) = \sum_{i=1}^{\ell} \Bigg\{ a f_i \phi_i^2 - 2 f_i^2 \phi_i \frac{\partial \phi_i}{\partial f_i} - 2 \sum_{j>i} \frac{f_i^2 \phi_i^2 - f_j^2 \phi_j^2}{f_i - f_j}$$

$$-2 b \phi_i - 2 f_i \frac{\partial \phi_i}{\partial f_i} - 4 \sum_{j>i} \frac{f_i \phi_i - f_j \phi_j}{f_i - f_j} \Bigg\}$$

where $a = \{|n-p| + 2m\} \wedge (n+p) - 2$ and $b = |m - (n \wedge p)| + 1$.

Complex case $\longrightarrow$ Real case: $2 \longrightarrow 3$ and $2 \longrightarrow 4$.

Improvement

1. Find $\boldsymbol{\phi_i}$'s to satisfy $\mathbf{SURE(\Phi) \le 0}$. Then

$$\mathbf{R(\widehat{\Theta}^{SH}, \Theta | \Sigma) \le R(Z, \Theta | \Sigma)} \text{ for } \forall \mathbf{(\Theta, \Sigma)}.$$

2. If $\mathbb{P}(\widehat{\Theta}^{SH} \neq \widehat{\Theta}^{SH+}) > 0$, then

$$\mathbf{R(\widehat{\Theta}^{SH+}, \Theta | \Sigma) < R(\widehat{\Theta}^{SH}, \Theta | \Sigma)} \quad \text{for } \forall \mathbf{(\Theta, \Sigma)}.$$

3. Example

$$\phi_i = \max\left\{1 - \frac{c_i}{f_i} - \frac{d}{\sum_{i=1}^{\ell} f_i}, 0\right\}; \quad c_i = \frac{b - 1 + (\ell - i)}{a + 2 - 2(\ell - i)}$$

d is a positive constant depending on c_i and m, n, p.

Efective degrees of freedom

Assume that $n \neq p$.

(1) Observation: $\begin{bmatrix} Z \\ U \end{bmatrix} \sim N_{(m+n)\times p}\left(\begin{bmatrix} \Theta \\ 0 \end{bmatrix}, I_{m+n} \otimes \Sigma\right)$.

(2) Estimator: $\widehat{\Theta} = Z(I_p - SS^{\dagger}) + R(I_{\ell} - \Phi)R^{*}ZSS^{\dagger}$

(3) Future observation: $Z^{0} \sim N_{m\times p}(\Theta, I_m \otimes \Sigma)$.

(4) Prediction error

$$\mathrm{Err} = \|Z^{0} - \widehat{\Theta}\|^{2}_{F,\Sigma^{-1}} = \mathrm{Tr}\left[(Z^{0} - \widehat{\Theta})\Sigma^{-1}(Z^{0} - \widehat{\Theta})^{*}\right].$$

(5) Naive error:

$$\begin{aligned} \mathrm{err} &= |n-p| \times \|Z - \widehat{\Theta}\|_{F,S^{\dagger}} \\ &= |n-p| \times \mathrm{Tr}\left[(Z - \widehat{\Theta})S^{\dagger}(Z - \widehat{\Theta})^{*}\right]. \end{aligned}$$

(6)

$$\begin{aligned} \mathbb{E}[\,\mathrm{Err}\,] &= mp + \mathrm{R}(\widehat{\Theta}, \Theta, |\,\Sigma) = 2mp + \mathbb{E}[\mathrm{SURE}(\Phi)], \\ \mathbb{E}[\,\mathrm{err}\,] &= mp + \mathbb{E}\left[|n-p| \sum_{i=1}^{\ell} f_i \phi_i^2\right], \end{aligned}$$

where the expectation in the left hand side of the first equation is taken with respect to (Z^0, Z, S) and the others are taken with respect to $Z, S)$.

(7) Effective degrees of freedom

$$
\begin{aligned}
\widehat{\mathrm{Err}} &= mp + \mathrm{err} + \mathbf{SURE}(\boldsymbol{\Phi}) - |n-p| \sum_{i=1}^{\ell} f_i \phi_i^2 \\
&= \mathrm{err} + \mathrm{penalty} \\
\mathrm{penalty} &= mp + \sum_{i=1}^{\ell} \Big\{ \tilde{a} f_i \phi_i^2 - 2 f_i^2 \phi_i \frac{\partial \phi_i}{\partial f_i} - 2 \sum_{j>i} \frac{f_i^2 \phi_i^2 - f_j^2 \phi_j^2}{f_i - f_j} \\
&\qquad -2b\phi_i - 2 f_i \frac{\partial \phi_i}{\partial f_i} - 4 \sum_{j>i} \frac{f_i \phi_i - f_j \phi_j}{f_i - f_j} \Big\} \\
\tilde{a} &= 2(m \wedge n \wedge p) - 2; \quad b = |m - (n \wedge p)| + 1.
\end{aligned}
$$

Covariance penalty

(8) For $\boldsymbol{n > p}$,

$$\begin{aligned} \mathbf{R}(\widehat{\Theta}, \Theta | \Sigma) &= \mathbb{E}[\,\mathrm{err}\,] + mp \\ &+ (n-p)\mathrm{Tr}\left[\mathbb{COV}(S^{-1}, (Z-\widehat{\Theta})(Z-\widehat{\Theta})^*)\right] \\ &+ 2\mathrm{Tr}\left[(I_m \otimes \Sigma^{-1/2})\mathbb{COV}(Z, \widehat{\Theta})(I_m \otimes \Sigma^{-1/2})\right], \end{aligned}$$

where

$$\begin{aligned} \mathbb{COV}(A, B) &= \mathbb{E}[(\mathrm{vec}(A) - \mathbb{E}[\mathrm{vec}(A)])(\mathrm{vec}(A) - \mathbb{E}[\mathrm{vec}(A)])^*] \\ \mathrm{vec}(A) &= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}; \quad A = \begin{pmatrix} a_1^\top \\ a_2^\top \\ \vdots \\ a_m^\top \end{pmatrix}; \quad a_1 : p \times 1. \end{aligned}$$

References

1. Candés, E.J., Sing-Long, C.A., and Trzasko, J.D. (2013): IEEE on Signal Processing **61** 4643–4657.
2. Chételat, D. and Wells, M.T. (2012): AOS **40** 3137–3160.
3. Efron, B. (2004): JASA **99** 619–642.
4. Hansen, N.R. (2018): SPL **135** 76–88.
5. Tsukuma, H. and Kubokawa, T. (2015): JMVA **139** 312–328.