Shrinkage estimation of mean for complex multivariate normal distribution with unknown covariance when $p > n$

Yoshiko KONNO
collaborated with Satomi SEITA

Japan Women's University

Video Remote Presentation

Mathematical Methods of Modern Statistics 2

15–19 June 2020

(1) Let $\boldsymbol{n}, \boldsymbol{p} \in \mathbb{N}$ such that $\mathbf{min}(\boldsymbol{n}, \boldsymbol{p}) \geq \mathbf{2}$.
(2) For a matrix $\boldsymbol{A}$ with complex entries, $\boldsymbol{A}^*$ stands for a complex transpose conjugate of $\boldsymbol{A}$.
(3) A $\boldsymbol{p} \times \boldsymbol{p}$ matrix $\boldsymbol{C}$ is Hermitian if $\boldsymbol{C} = \boldsymbol{C}^*$. $\mathbf{Herm}^+(\boldsymbol{p}, \mathbb{C})$ stands for the set of all positive definite Hermitian matrices.
(4) Let $\boldsymbol{Z}$, a $\mathbb{C}^{\boldsymbol{p}}$-values random vector, follow a multivariate complex normal distritution with a mean vector $\boldsymbol{\theta} \in \mathbb{C}^{\boldsymbol{p}}$ and a covariance matrxi $\boldsymbol{\Sigma} \in \mathbf{Herm}^+(\boldsymbol{p}, \mathbb{C})$, i.e., $\boldsymbol{Z} \sim \mathbb{C}\boldsymbol{N}_{\boldsymbol{p}}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$.
(5) $\boldsymbol{Z}$ and $\boldsymbol{S}$ are independently distributed.
(6) A $\boldsymbol{p} \times \boldsymbol{p}$ semi-positive definite Hermitian matrix $\boldsymbol{S}$(not necessarily nonsingular) follow a complex Wishart distribution with the degrees of freedom $\boldsymbol{n}$ and a scale matirx $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{S} \sim \mathbb{C}\boldsymbol{W}_{\boldsymbol{p}}(\boldsymbol{n}, \boldsymbol{\Sigma})$.
(7) $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{Z}, \boldsymbol{S})$ is an estimator for $\boldsymbol{\theta}$.
(8) $\mathbb{E}[\,\cdot\,]$ stants for the expectation with respect to the joint distribution of $(\boldsymbol{Z}, \boldsymbol{S})$.

1. When the covariance matrix $\boldsymbol{\Sigma}$ is unknow and a sample size $\boldsymbol{n}$ is smaller than the dimension of the mean vextor, we consider a problem of estimating the unknown mean vector $\boldsymbol{\theta}$ based on observation $(\boldsymbol{Z}, \boldsymbol{S})$ under an invariant loss function.
2. This setup is a complex analogue of the problem of estimating mean vector of a multivariate real normal distribution, consider by Chételat and Wells (Ann. Statist., 2012).

An invariant loss function and its risk function

(1) A loss function is given by

$$L(\theta, \hat{\theta} | \Sigma) = (\hat{\theta} - \theta)^* \Sigma^{-1} (\hat{\theta} - \theta). \tag{1}$$

(2) The risk function is denoted by $R(\theta, \hat{\theta} | \Sigma) = \mathbb{E}[L(\theta, \hat{\theta} | \Sigma)]$.

Comparision of estimators

An estimator $\hat{\theta}_1$ is better than another estimator $\hat{\theta}_2$ if

$$R(\theta, \hat{\theta}_1 | \Sigma) \leq R(\theta, \hat{\theta}_1 | \Sigma) \quad \text{for } \forall (\theta, \Sigma) \in \mathbb{C}^p \times \mathrm{Herm}^+(p, \mathbb{C}),$$

and

$$R(\theta_0, \hat{\theta}_1 | \Sigma_0) < R(\theta_0, \hat{\theta}_1 | \Sigma_0) \quad \text{for } \exists (\theta_0, \Sigma_0) \in \mathbb{C}^p \times \mathrm{Herm}^+(p, \mathbb{C}).$$

Note that S is nonsingular if $n \geq p$ and singular if $p > n$. We focus on the situation of $p > n$, i.e., the case that S is singular. To derive a shrinkage estimator, we use the Moore-Penrose inverse of S.

Definition of the Moore-Penrose generalized inverse

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

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

First note that the maximum likelihood estimator of θ is $\hat{\theta}_0 = Z$ which is minimax with respect to the loss function (1).
Following the idea due to Chételat and Wells (Ann. Statist., 2012), we consider a class of estimators below. We consider the following class of estimators

Baranchik-like estimators

For bounded and differentiable functions $r : [0, \infty) \to (0, \infty)$, we define Baranchik-like estimators as

$$\begin{aligned} \hat{\theta}_r &= \left(I_p - \frac{r(F)}{F} SS^{\dagger}\right) Z \qquad (2) \\ &= P_{S^{\perp}} Z + \left(1 - \frac{r(F)}{F}\right) P_S Z, \end{aligned}$$

where I_p is a p-th identity matrix, $F = Z^* S^{\dagger} Z$, $P_S = SS^{\dagger}$ and $P_{S^{\perp}} = I_p - SS^{\dagger}$.

$\boxed{\text{Remark}}$ Since $\mathbf{\Sigma}$ is positive-definite, Note that $\boldsymbol{P(F > 0) = 1}$.

$\boxed{\text{Remark}}$ $\boldsymbol{P_S = SS^\dagger}$ and $\boldsymbol{P_{S^\perp} = I_p - SS^\dagger}$ are projections to the space spanned by the columns of $\boldsymbol{S}$ and the orthogonally complementant to its space, respectively.

Theorem 1

Let $\min(n, p) \geq 2$, $n \neq p$. If the function r in (2) satisfies the following conditions

(i)] $0 \leq r \leq \dfrac{2(\min(n, p) - 1)}{n + p - 2\min(n, p) + 2}$;

(ii) r is nondecreasing;

(iii) r', the derivative of r, is bounded,

the estimators (2) is minimax.

$\boxed{\text{Idea of Proof.}}$

We proved it almost in the same way as that in Chételat and Wells (Ann. Statist., 2012). There are three ingredients to prove the result:

1. Stein's identity for the multivariate complex normal,
2. Haff and Stein's identity for nonsingular complex Wishart distribution (see Konno(2009, JMVA)),
3. Derivative to the Moore-Penrose inverse.

Example: the James-Stein like estimators

Corollary 1. the James-Stein estimator

Let $p > n \geq 2$ and put $r = \dfrac{n-1}{p-n+2}$. Then the conditions (i)∼(iii) in the main theorem are satisfied.
Then the James-Stein-like estimator is given by

$$\begin{aligned} \hat{\theta}_{JSL} &= \left(I_p - \frac{p-1}{(n-p+2)F} SS^\dagger \right) Z \\ &= (I_p - SS^\dagger)Z + \left(1 - \frac{p-1}{(n-p+2)F} \right) SS^\dagger Z \end{aligned}$$

is better than $\hat{\theta}_0$.

Corollary 2. the James-Stein estimator

Let $n > p \geq 2$ and put $r = \dfrac{p-1}{n-p+2}$. Then the conditions (i)~(iii) in the main theorem are satisfied and the James-Stein estimator is given by

$$\hat{\theta}_{JSL} = \left(I_p - \frac{p-1}{(n-p+2)F} \right) Z; \quad F = Z^* S^{-1} Z, \tag{3}$$

is better than $\hat{\theta}_0$.

Further improvement over the Baranchik-like estimators $\hat{\theta}_r$

For a real number b, let $b_+ = \max(b, 0)$. Consider the estimator in the following:

$$\hat{\theta}_{r+} = (I_p - SS^\dagger)Z + \left(1 - \frac{r(F)}{F}\right)_+ SS^\dagger Z \tag{4}$$

Theorem 2

Let $\min(n, p) \geq 2$. If

$$R(\theta, \hat{\theta}_r | \Sigma) < \infty \text{ and } P(\hat{\theta}_{r+} \neq \hat{\theta}_r) > 0$$

for $\forall(\theta, \Sigma) \in \mathbb{C}^p \times \mathrm{Herm}^+(p)$, then the estimator $\hat{\theta}_{r+}$ is better than $\hat{\theta}_r$.

Two examples: An improvement over the Baranchik-like estimators $\hat{\theta}_r$

Corollary. 3

When $n > p \geq 2$, the positive-part estimator $\hat{\theta}_{JS+} = \left(1 - \dfrac{p-1}{(n-p+2)F}\right)_+ Z$ is better than the James-Stein estimator $\hat{\theta}_{JS}$.

Corollary. 4

When $p > n \geq 2$, the postive-part estimator $\hat{\theta}_{JSL+} = (I_p - SS^\dagger)Z + \left(1 - \dfrac{n-1}{(p-n+2)F}\right)_+ SS^\dagger Z$ is better than the James-Stein-like estimator $\hat{\theta}_{JSL}$.

Real case

Let $\min(n, p) \geq 3$. Let $X \sim \mathbb{R}N_p(\theta, \Sigma)$ and $S \sim \mathbb{R}W_p(n, \Sigma)$ independently. If the following three conditions **(i)** $\sim$ **(iii)** are satisfied, estimators $\hat{\theta}_r = (I_p - \frac{r(F)}{F} SS^{\dagger})X$ are better than $\hat{\theta}_0 = X$.
(i) $0 < r < \frac{2(\min(n,p)-2)}{n+p-2\min(n,p)+3}$; (ii) r is nondecreasing; (iii) r' is bounded,
where $r : [0, \infty) \to (0, \infty)$ and $F = X^T S^{\dagger} X > 0$.

Complex case

Let $\min(n, p) \geq 2$. Let $Z \sim \mathbb{C}N_p(\theta, \Sigma)$ and $S \sim \mathbb{C}W_p(n, \Sigma)$, independently. If the following three conditoins **(i)** $\sim$ **(iii)** are satisfied, estimators $\hat{\theta}_r = (I_p - \frac{r(F)}{F} SS^{\dagger})Z$ are better than $\hat{\theta}_0 = Z$.
(i) $0 < r < \frac{2(\min(n,p)-1)}{n+p-2\min(n,p)+2}$; (ii) r is nondecreasing; (iii) r' is bounded,
where $r : [0, \infty) \to (0, \infty)$ and $F = Z^* S^{\dagger} Z > 0$.

Summary of the talk

1. We proposed Baranchik-type shrinkage estimators for a complex mean vector of the multivariate complex normal distributions when the sample size n is smaller that the dimension of mean vector p.
2. Minimaxity is proved via using the integration-by-parts formulae, so-caleed Stein's identity for a complex normal distribution and Haff-Stein's identity for nonsingular complex Wishart distributions
3. We proved that the positive-part estimator works well.