Estimation of covariance matrix for Stein's loss on a two-step monotone incomplete sample

Shin-ichi TSUKADA

Abstract

For a two-step monotone incomplete sample, it is known that Anderson[1] derived the maximum likelihood estimator for the population mean vector and the population covariance matrix. Then Tsukada[8] derived the unbiased estimator for the covariance matrix.

On the other hand, Richards and Yamada[6] estimated the mean vector based on the loss function. In this article, we deal with the inference for the covariance matrix based on Stein's loss function. The estimator, which has the minimum risk, is proposed, and we compare the risks of the maximum likelihood estimator with the unbiased estimator, and with the proposed estimator, respectively.

Keywords. Stein's loss, Risk, Covariance matrix, Monotone incomplete sample

1 Introduction

We consider a monotone incomplete data, which was drawn from a multivariate normal population consisting of mutually independent observations of the following form;

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{Y}_1 \end{pmatrix}, \begin{pmatrix} \boldsymbol{X}_2 \\ \boldsymbol{Y}_2 \end{pmatrix}, \ldots, \begin{pmatrix} \boldsymbol{X}_n \\ \boldsymbol{Y}_n \end{pmatrix}, \begin{pmatrix} \boldsymbol{X}_{n+1} \\ * \end{pmatrix}, \ldots, \begin{pmatrix} \boldsymbol{X}_N \\ * \end{pmatrix}, \tag{1}$$

where each $\boldsymbol{X}_i \in \boldsymbol{R}^p$ and each $\boldsymbol{Y}_i \in \boldsymbol{R}^q$; $(\boldsymbol{X}_i, \boldsymbol{Y}_i)'$, $i = 1, \ldots, n$ are observations from $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and the incomplete data $\boldsymbol{X}_i$, $i = n+1, \ldots, N$, are observations of the first p elements of the same population.

To ensure that all means and variances are finite and that all integrals subsequently encountered are absolutely convergent, we also assume that $n > p+2$ and $N > n \geq p+q$ (Chang and Richards[3]). As explained by Yamada, et al.[9], we also assume that data are missing completely at random to derive the maximum likelihood estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$.

Anderson[1] and Anderson and Olkin[2] derived the maximum likelihood estimator (MLE) for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, while Kanda and Fujikoshi[5] investigated some of

*School of Education, Meisei University. 2-1-1 Hodokubo Hino City Tokyo 191-8506, JAPAN. E-mail:tsukada@ed.meisei-u.ac.jp.

their properties. Chang and Richards[3],[4] derived a stochastic representation for the exact distribution of the MLE $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ for a two-step monotone incomplete sample. They obtained an ellipsoidal confidence region for $\boldsymbol{\mu}$, and considered hypothesis testing for the covariance matrix. Richards and Yamada[6] studied the Stein phenomenon for a two-step monotone sample. They derived an improved estimator for $\boldsymbol{\mu}$. Recently, Tsukada[8] derived an unbiased estimator (UBE) for $\boldsymbol{\Sigma}$ and investigated its properties.

In this paper, we consider the inference for the covariance matrix $\boldsymbol{\Sigma}$ using Stein's loss function on a two-step monotone incomplete sample. We have obtained a new estimator and have discussed its properties in Section 2. In Section 3, we investigate the accuracy of the new estimator by performing numerical simulations.

2 Estimation of covariance matrix for Stein's loss

Let the missing ratio be $\tau = (N-n)/N$. We decompose $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as follows;

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are a p-element vector and q-element vector, respectively; $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'$, and $\boldsymbol{\Sigma}_{22}$ are of orders $p \times p$, $p \times q$, and $q \times q$, respectively. We also define the Schurz complement $\boldsymbol{\Sigma}_{22\cdot 1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Define the sample mean vectors

$$\bar{\boldsymbol{X}}_1 = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i,\ \bar{\boldsymbol{X}}_2 = \frac{1}{N-n}\sum_{i=n+1}^{N}\boldsymbol{X}_i,$$
$$\bar{\boldsymbol{Y}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Y}_i, \quad \bar{\boldsymbol{X}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i,$$

and the corresponding matrices of the sums of squares and products

$$\boldsymbol{A}_{11,n} = \sum_{i=1}^{n}\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_1\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_1\right)',\ \boldsymbol{A}_{12} = \boldsymbol{A}_{21}' = \sum_{i=1}^{n}\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_1\right)\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)',$$
$$\boldsymbol{A}_{22} = \sum_{i=1}^{n}\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)', \qquad \boldsymbol{A}_{11,N} = \sum_{i=1}^{N}\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)'.$$

The MLE is represented as follows:

$$\hat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{X}}, \quad \hat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{Y}} - (1-\tau)\boldsymbol{A}_{21}\boldsymbol{A}_{11,n}^{-1}\left(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2\right), \tag{2}$$

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{11} &= \frac{1}{N}\boldsymbol{A}_{11,N},\\ \hat{\boldsymbol{\Sigma}}_{12} &= \frac{1}{N}\boldsymbol{A}_{11,N}\boldsymbol{A}_{11,n}^{-1}\boldsymbol{A}_{12},\\ \hat{\boldsymbol{\Sigma}}_{22} &= \frac{1}{n}\boldsymbol{A}_{22\cdot 1,n} + \frac{1}{N}\boldsymbol{A}_{21}\boldsymbol{A}_{11,n}^{-1}\boldsymbol{A}_{11,N}\boldsymbol{A}_{11,n}^{-1}\boldsymbol{A}_{12}.\end{aligned} \tag{3}$$

As shown by Kanda and Fujikoshi[5], the expectation of $\hat{\boldsymbol{\Sigma}}$ is

$$E\left[\hat{\boldsymbol{\Sigma}}\right] = \frac{N-1}{N}\boldsymbol{\Sigma} + \frac{b_0}{N}\begin{pmatrix}\boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{\Sigma}_{22\cdot 1}\end{pmatrix},$$

where

$$b_0 = -\frac{(N-n)\{n-(p+1)(p+2)\}}{n(n-p-2)}.$$

The MLE is biased. Tsukada[8] proposes the UBE $\tilde{\boldsymbol{\Sigma}}$ as follows:

$$\tilde{\boldsymbol{\Sigma}}_{11} = \frac{N}{N-1}\hat{\boldsymbol{\Sigma}}_{11},\quad \tilde{\boldsymbol{\Sigma}}_{12} = \frac{N}{N-1}\hat{\boldsymbol{\Sigma}}_{12},\quad \tilde{\boldsymbol{\Sigma}}_{22} = \frac{N}{N-1}\hat{\boldsymbol{\Sigma}}_{22} - c_0\hat{\boldsymbol{\Sigma}}_{22\cdot 1}, \tag{4}$$

where

$$c_0 = \frac{(N-n)(p+1)(p+2) - n(N-n)}{(N-1)(n-p-2)(n-p-1)},$$

and shows that the risk of the UBE is smaller than the risk of MLE for Stein's loss. By expanding the MLE and the UBE, the asymptotic distribution of these estimators was derived.

Let $\boldsymbol{\Lambda}_{11}$ and $\boldsymbol{\Lambda}_{22}$ be $p\times p$ and $q\times q$ positive definite matrices, respectively. Let $\boldsymbol{\Lambda}_{21}$ be a $q\times p$ matrix and let $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ be $p\times 1$ and $q\times 1$ vectors, respectively. We define

$$\boldsymbol{\Lambda} = \begin{pmatrix}\boldsymbol{\Lambda}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{\Lambda}_{22}\end{pmatrix},\quad \boldsymbol{C} = \begin{pmatrix}\boldsymbol{I}_p & \boldsymbol{O}\\ \boldsymbol{\Lambda}_{21} & \boldsymbol{I}_q\end{pmatrix},\quad \boldsymbol{\nu} = \begin{pmatrix}\boldsymbol{\nu}_1\\ \boldsymbol{\nu}_2\end{pmatrix}, \tag{5}$$

and consider the set of affine transformations of the data in (1) to be of the form

$$\begin{aligned}\begin{pmatrix}\boldsymbol{X}_i^*\\ \boldsymbol{Y}_i^*\end{pmatrix} &= \boldsymbol{\Lambda}\boldsymbol{C}\begin{pmatrix}\boldsymbol{X}_i\\ \boldsymbol{Y}_i\end{pmatrix} + \boldsymbol{\nu},\ i = 1,\ldots,n,\\ \boldsymbol{X}_j^* &= \boldsymbol{\Lambda}_{11}\boldsymbol{X}_j + \boldsymbol{\nu}_1,\qquad j = n+1,\ldots,N.\end{aligned} \tag{6}$$

Romer and Richards[7] also considered the transformation in (6), and noted that as $\boldsymbol{\Lambda}_{11}$, $\boldsymbol{\Lambda}_{21}$, $\boldsymbol{\Lambda}_{22}$, and $\boldsymbol{\nu}$ vary over their respective parameter spaces. The set of all transformations in (6) forms a group; in particular, each such transformation is invertible.

We consider a class of estimators, which is of the form

$$\left\{\ddot{\boldsymbol{\Sigma}} \equiv d_1\hat{\boldsymbol{\Sigma}} + d_2\begin{pmatrix}\boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \hat{\boldsymbol{\Sigma}}_{22\cdot 1}\end{pmatrix}\middle|\, d_1, d_2 \in \boldsymbol{R}\right\}. \tag{7}$$

This class is invariant for the transformation in (6), and includes the MLE and UBE. We derive the estimator, which minimizes the risk for Stein's loss function

$$L(\boldsymbol{A}, \boldsymbol{\Sigma}) = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right) - \log \frac{|\boldsymbol{A}|}{|\boldsymbol{\Sigma}|} - (p+q), \tag{8}$$

where $\boldsymbol{A}$ is an estimator of $\boldsymbol{\Sigma}$. We calculate the risk of this class of estimators as follows:

$$\begin{aligned} R\left(\ddot{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\right) &= R\left(d_1\hat{\boldsymbol{\Sigma}} + d_2\begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \hat{\boldsymbol{\Sigma}}_{22\cdot 1}\end{pmatrix}, \boldsymbol{\Sigma}\right) \\ &= d_1 E\left[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}\right] + d_2 E\left[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \hat{\boldsymbol{\Sigma}}_{22\cdot 1}\end{pmatrix}\right] \\ &\quad - E\left[\log\begin{vmatrix} \ddot{\boldsymbol{\Sigma}}_{11} & \ddot{\boldsymbol{\Sigma}}_{12} \\ \ddot{\boldsymbol{\Sigma}}_{21} & \ddot{\boldsymbol{\Sigma}}_{22}\end{vmatrix} - \log|\boldsymbol{\Sigma}|\right] - (p+q). \end{aligned} \tag{9}$$

Since

$$\begin{aligned} E\left[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}\right] &= \frac{N-1}{N}(p+q) + \frac{b_0}{N}q, \\ E\left[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \hat{\boldsymbol{\Sigma}}_{22\cdot 1}\end{pmatrix}\right] &= \frac{n-p-1}{n}q, \\ E\left[\log\begin{vmatrix} \ddot{\boldsymbol{\Sigma}}_{11} & \ddot{\boldsymbol{\Sigma}}_{12} \\ \ddot{\boldsymbol{\Sigma}}_{21} & \ddot{\boldsymbol{\Sigma}}_{22}\end{vmatrix} - \log|\boldsymbol{\Sigma}|\right] &= E\left[\log\frac{\left|\ddot{\boldsymbol{\Sigma}}_{11}\right|}{|\boldsymbol{\Sigma}_{11}|} + \log\frac{\left|\ddot{\boldsymbol{\Sigma}}_{22\cdot 1}\right|}{|\boldsymbol{\Sigma}_{22\cdot 1}|}\right] \\ &= p\log|d_1| + M_{11} + q\log|d_1+d_2| + M_{22\cdot 1}, \end{aligned}$$

where

$$\begin{aligned} M_{11} &= E\left[\log\frac{\left|\hat{\boldsymbol{\Sigma}}_{11}\right|}{|\boldsymbol{\Sigma}_{11}|}\right] = -p\log\left(\frac{N}{2}\right) + \sum_{i=1}^{p}\frac{\Gamma'\left[(N-i)/2\right]}{\Gamma\left[(N-i)/2\right]}, \\ M_{22\cdot 1} &= E\left[\log\frac{\left|\hat{\boldsymbol{\Sigma}}_{22\cdot 1}\right|}{|\boldsymbol{\Sigma}_{22\cdot 1}|}\right] = -q\log\left(\frac{n}{2}\right) + \sum_{i=1}^{q}\frac{\Gamma'\left[(n-p-i)/2\right]}{\Gamma\left[(n-p-i)/2\right]}, \end{aligned}$$

the risk is

$$\begin{aligned} R(\ddot{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= d_1\left\{\frac{N-1}{N}(p+q) + \frac{b_0}{N}q\right\} + d_2\cdot\frac{n-p-1}{n}q \\ &\quad -p\log|d_1| - M_{11} - q\log|d_1+d_2| - M_{22\cdot 1} - (p+q). \end{aligned} \tag{10}$$

By differentiating the risk and assuming zero, we consider d_1 and d_2 in order to minimize the risk as follows:

$$
\begin{aligned}
d_1 &= \frac{N(n-p-2)}{N(n-p+q-2)-n-(p+2)(q-1)} \\
&= \frac{(1-\tau)N^2 - N(p+2)}{(1-\tau)N^2 - N(p-q+3-\tau)-(p+2)(q-1)}, \qquad (11) \\
d_2 &= \frac{N\left\{(p+1)(p+2)-n(p+q+1)\right\}+n(p+2)(q-1)+n^2}{N(n-p-1)(n-p+q-2)-(n-p-1)\left\{n+(p+2)(q-1)\right\}} \\
&= \frac{N^2(1-\tau)(p+q+\tau)-N(p+2)(p+q+\tau-\tau q)}{F(N)}, \qquad (12)
\end{aligned}
$$

where

$$
\begin{aligned}
F(N) &= N^3(1-\tau)^2 + N^2(2p-q+4-\tau) \\
&\quad + N\left\{p^2+p(5-2q+\tau q-2\tau)-3q+2\tau q-3\tau+5\right\} \\
&\quad + (p+1)(p+2)(q-1).
\end{aligned}
$$

Therefore, we obtain the following theorem.

Theorem 2.1. *We again denote the estimator as $\ddot{\boldsymbol{\Sigma}}$ and call the minimum risk estimator by this estimator, when d_1 is (11) and d_2 is (12). Since the minimum of the risk is*

$$
\begin{aligned}
R(\ddot{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}) &= -p\log\left|\frac{N(n-p-2)}{N(n-p+q-2)-n-(p+2)(q-1)}\right| \\
&\quad -q\log\left|\frac{n}{n-p-1}\right| - M_{11} - M_{22\cdot 1},
\end{aligned}
$$

the difference between the risk of the MLE and that of the minimum risk estimator is

$$
\begin{aligned}
& R(\hat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}) - R(\ddot{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}) \\
&= -\frac{p+q}{N} + \frac{b_0}{N}q \\
&\quad + p\log\left|\frac{N(n-p-2)}{N(n-p+q-2)-n-(p+2)(q-1)}\right| + q\log\left|\frac{n}{n-p-1}\right|.
\end{aligned}
$$

Proof. Because the risk of the MLE is

$$
R\left(\hat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}\right) = -\frac{p+q}{N} + \frac{b_0}{N}q - M_{11} - M_{22\cdot 1}, \qquad (13)
$$

from Tsukada[8], the difference between the risk of the MLE and that of the minimum risk estimator is obtained. □

Also, the expectation of this estimator is as follows:

Theorem 2.2. *The expectation of the risk minimum estimator is*

$$E\left[\ddot{\boldsymbol{\Sigma}}\right] = a_1\boldsymbol{\Sigma} + a_2\begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Sigma}_{22\cdot 1} \end{pmatrix}, \tag{14}$$

where

$$\begin{aligned} a_1 &= \frac{(N-1)(n-p-2)}{N(n-p+q-2)-n-(p+2)(q-1)}, \\ a_2 &= \frac{n^2-nN(p+q+1)+N(p+1)(p+2)+n(p+2)(q-1)}{n^2(N-1)-N(p-q+2)-n(p+2)(q-1)}. \end{aligned}$$

Proof. We can obtain this result using the following expectations:

$$\begin{aligned} E\left[\hat{\boldsymbol{\Sigma}}\right] &= \frac{N-1}{N}\boldsymbol{\Sigma} + \frac{b_0}{N}\begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Sigma}_{22\cdot 1} \end{pmatrix}, \\ E\left[\hat{\boldsymbol{\Sigma}}_{22\cdot 1}\right] &= \left\{\frac{N-1}{N} - \frac{p}{N}\cdot\frac{N-p-2}{n-p-2} - \frac{B_0}{N(n-p-2)}\right\}\boldsymbol{\Sigma}_{22} \\ &\quad + \left\{\frac{b_0}{N} + \frac{B_0}{N(n-p-2)}\right\}\boldsymbol{\Sigma}_{22\cdot 1}, \end{aligned}$$

where

$$B_0 = n^2 + n(N-n-2p-3) - (p+1)(2N-2n-p-2).$$

□

The asymptotic distribution of this estimator is derived in a manner that is similar to the case of the MLE and the UBE.

Theorem 2.3. *The estimator*

$$\sqrt{N}\left(vec\ \left(\ddot{\boldsymbol{\Sigma}}_{11}-\boldsymbol{\Sigma}_{11}\right), vec\ \left(\ddot{\boldsymbol{\Sigma}}_{12}-\boldsymbol{\Sigma}_{12}\right), vec\ \left(\ddot{\boldsymbol{\Sigma}}_{22}-\boldsymbol{\Sigma}_{22}\right)\right)'$$

is asymptotically distributed as a normal distribution with mean vector $\mathbf{0}$ *and a covariance matrix*

$$\Theta = \begin{pmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} \\ \Theta_{12}' & \Theta_{22} & \Theta_{23} \\ \Theta_{13}' & \Theta_{23}' & \Theta_{33} \end{pmatrix}, \tag{15}$$

where $\boldsymbol{K}_{ij}$ *is a commutation matrix and*

$$\begin{aligned} \Theta_{11} &= (\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Sigma}_{11}\otimes\boldsymbol{\Sigma}_{11}), \qquad \Theta_{12} = (\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Sigma}_{12}\otimes\boldsymbol{\Sigma}_{11}), \\ \Theta_{13} &= (\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Sigma}_{12}\otimes\boldsymbol{\Sigma}_{12}), \\ \Theta_{22} &= \frac{1}{1-\tau}(\boldsymbol{\Sigma}_{22}\otimes\boldsymbol{\Sigma}_{11}) - \frac{\tau}{1-\tau}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\otimes\boldsymbol{\Sigma}_{11}\right) + \boldsymbol{K}_{qp}(\boldsymbol{\Sigma}_{12}\otimes\boldsymbol{\Sigma}_{21}), \\ \Theta_{23} &= \frac{1}{1-\tau}(\boldsymbol{\Sigma}_{22}\otimes\boldsymbol{\Sigma}_{12})(\boldsymbol{I}_{q^2}+\boldsymbol{K}_{qq}) \\ &\quad -\frac{\tau}{1-\tau}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\otimes\boldsymbol{\Sigma}_{12}\right)(\boldsymbol{I}_{q^2}+\boldsymbol{K}_{qq}), \\ \Theta_{33} &= \frac{1}{1-\tau}(\boldsymbol{I}_{q^2}+\boldsymbol{K}_{qq})(\boldsymbol{\Sigma}_{22}\otimes\boldsymbol{\Sigma}_{22}) \\ &\quad -\frac{\tau}{1-\tau}(\boldsymbol{I}_{q^2}+\boldsymbol{K}_{qq})\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\otimes\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right). \end{aligned}$$

Proof. When the sample size N increases, the coefficient d_1 and d_2 including $\ddot{\boldsymbol{\Sigma}}$ converge to 1 and 0, respectively. Furthermore, according to Tsukada[8], the asymptotic distribution of the estimator $\ddot{\boldsymbol{\Sigma}}$ is the same as that of the MLE. □

The differences between the risks in Theorem 2.1, the expectation, and the convergence to the asymptotic distribution are evaluated in the next Section.

3 Numerical simulation

We perform numerical simulations to verify the expectation of the minimum risk estimator and the convergence to the asymptotic distribution of the estimator obtained in Section 2. Also, the risks were evaluated by numerical calculations.

3.1 Asymptotic distribution

We define the population distribution as follows. Write by

$$\boldsymbol{P} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \text{ and } \boldsymbol{\Lambda} = \text{diag}(\sigma^6, \sigma^5, \ldots, \sigma^2, \sigma, 1).$$

We assume that the population distribution is the 7-variate normal distribution with mean vector $\mathbf{0}$ and the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{P}\boldsymbol{\Lambda}$. Let $p = 4$ and $q = 3$. We set $\rho = 0.15$ and $\sigma = \sqrt{2}$. The total sample sizes N are 500 and 1000. The missing rates τ are 0.2, 0.4, 0.6, and 0.8. The number of simulations performed was ten thousand.

Let

$$\boldsymbol{v} = \sqrt{N}\left(\text{vec}\left(\ddot{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Sigma}_{11}\right), \text{vec}\left(\ddot{\boldsymbol{\Sigma}}_{12} - \boldsymbol{\Sigma}_{12}\right), \text{vec}\left(\ddot{\boldsymbol{\Sigma}}_{22} - \boldsymbol{\Sigma}_{22}\right)\right)'.$$

To investigate the convergence to the asymptotic distribution, we simulated the lower probability of $\boldsymbol{v}'\Theta^{-1}\boldsymbol{v}$ for the percentile of the chi-squared distribution with $(p+q)(p+q+1)/2 = 28$ degrees of freedom. Table 1 denotes the result in the case of $N = 500$, and Table 2 denotes the result in the case of $N = 1000$.

The convergence in the case of $N = 1000$ is better than that in the case of $N = 500$. As the missing rate τ increases, the convergence to the asymptotic distribution worsens, but we find that the distribution of the estimator almost converged to the asymptotic distribution for a range of more than 90%. Because most of the lower probability for $N = 500$ and $N = 1000$ converge when $\tau = 20\%$, it is believed that we can use the asymptotic distribution in the case of $\tau \leq 20\%$ and $N \geq 500$.

Table 1: The lower probability for $N = 500$.

τ	1%	5%	10%	50%	90%	95%	99%
20%	.0095	.0483	.0976	.5076	.9066	.9541	.9912
40%	.0089	.0458	.0950	.5029	.9076	.9548	.9913
60%	.0079	.0422	.0887	.4980	.9067	.9553	.9916
80%	.0048	.0291	.0666	.4560	.8979	.9505	.9903

Table 2: The lower probability for $N = 1000$.

τ	1%	5%	10%	50%	90%	95%	99%
20%	.0094	.0484	.0984	.5053	.9045	.9530	.9905
40%	.0095	.0482	.0978	.5027	.9041	.9520	.9904
60%	.0090	.0461	.0941	.4941	.9019	.9516	.9905
80%	.0068	.0387	.0827	.4767	.8996	.9505	.9899

3.2 Expectation of the minimum risk estimator

We assume that the population distribution is the same as mentioned in the above subsection and the number of simulations is ten thousand, and we investigated the expectation of the minimum risk estimator. The symbol S denotes the expectation of the minimum risk estimator obtained by simulation. To investigate the accuracy of equation (14), we simulated as follows. We estimate the expectation as

$$S = (s_{ij}) = \frac{1}{N_s}\sum_{k=1}^{N_s}\ddot{\mathbf{\Sigma}}_k,$$

where N_s is the number of simulations and $\ddot{\mathbf{\Sigma}}_k$ is the minimum risk estimator in each simulation, and we calculate the error

$$\begin{aligned} E &= \sum_{i=1}^{p+q}\sum_{j=i}^{p+q}(s_{ij} - E[\ddot{\sigma}_{ij}])^2 \\ &= \sum_{i=1}^{p}\sum_{j=i}^{p}(s_{ij} - E[\ddot{\sigma}_{ij}])^2 + \sum_{i=1}^{p}\sum_{j=p+1}^{p+q}(s_{ij} - E[\ddot{\sigma}_{ij}])^2 + \sum_{i=p+1}^{p+q}\sum_{j=i}^{p+q}(s_{ij} - E[\ddot{\sigma}_{ij}])^2 \\ &\equiv E_{11} + E_{12} + E_{22}. \end{aligned}$$

In each Table, the notation x^y indicates the value $x \times 10^y$.

Because each error is almost the same when $N = 200$ and $N = 500$, and the total error is considered to be a range of simulation errors: thus, the expectation (14) appears to be correct.

Table 3: Error between the expectation obtained by the simulation and expectation in (14)

	$N = 200$				$N = 500$			
τ	E_{11}	E_{12}	E_{22}	E	E_{11}	E_{12}	E_{22}	E
20%	7.342^{-4}	9.571^{-5}	8.516^{-5}	9.150^{-4}	1.618^{-4}	1.973^{-5}	9.370^{-6}	1.909^{-4}
40%	1.598^{-4}	2.810^{-5}	3.463^{-4}	5.342^{-4}	2.787^{-4}	2.761^{-5}	5.856^{-5}	3.649^{-4}
60%	6.616^{-4}	5.607^{-5}	1.541^{-3}	2.259^{-3}	2.094^{-4}	3.500^{-5}	2.378^{-4}	4.821^{-4}
80%	4.047^{-4}	4.105^{-5}	7.917^{-3}	8.363^{-3}	1.198^{-4}	6.271^{-5}	1.221^{-3}	1.403^{-3}

In the given scenario, we calculated the coefficients for $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_{22\cdot 1}$ in the case of $N = 200$ (Figure 1). The thin dashed line denotes the coefficient of the MLE and the dashed line denotes that of the minimum risk estimator. Similarly, the thin line denotes the coefficient for $\mathbf{\Sigma}_{22\cdot 1}$ of the MLE and the line does that of the minimum risk estimator. One sees that the coefficient for $\mathbf{\Sigma}$ converges to 1 and the coefficient for $\mathbf{\Sigma}_{22\cdot 1}$ converges to 0 as the missing rate τ decreases. By considering the speed of convergence of each coefficient, we have that the MLE is better than the minimum risk estimator from a biased perspective.

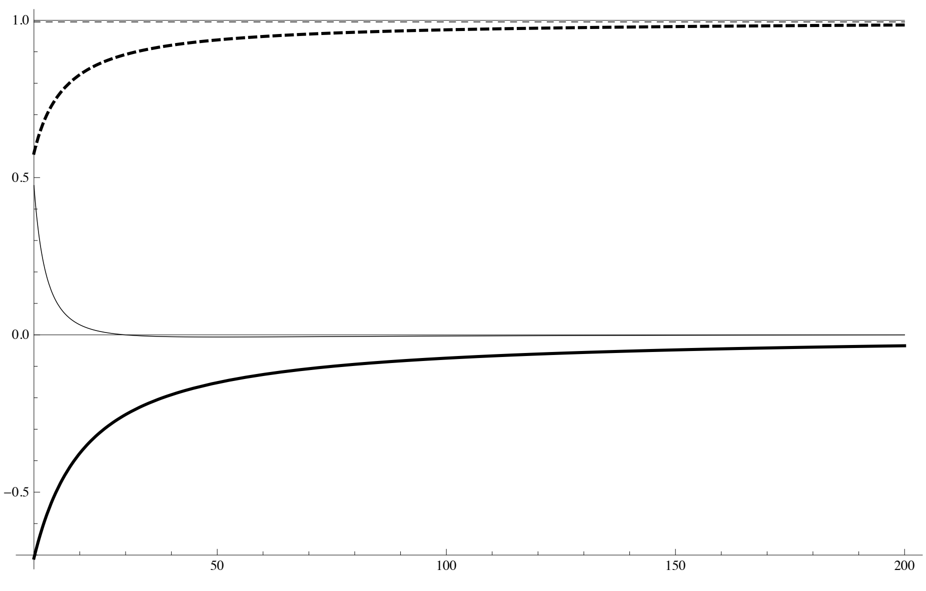

Figure 1: Coefficients in the case of $N = 200$

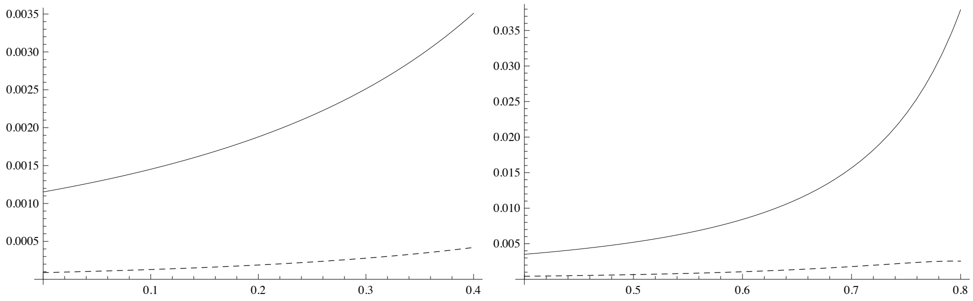

Figure 2: Difference in Risk for $N = 200$

3.3 Risk of estimators

We investigated the difference in the risk of the MLE and the UBE, and the difference in the risk of the MLE and that of the minimum risk estimator. Because the differences in the risks depend on the sample sizes N and n, and the dimensions p and q, we adopted the above setting as these parameters. In each figure, the dashed line denotes the difference in the risk of the MLE, and the thin line denotes the difference in the risk of the MLE and that of the minimum risk estimator.

The risks decrease when the sample size N increases. However, there are similar tendencies in both cases ($N = 200$ and $N = 500$). As the missing rate τ increases, the risks also increase. We have that the risk of the minimum risk estimator was the smallest, and the risk of the unbiased estimator was smaller than the risk of the MLE.

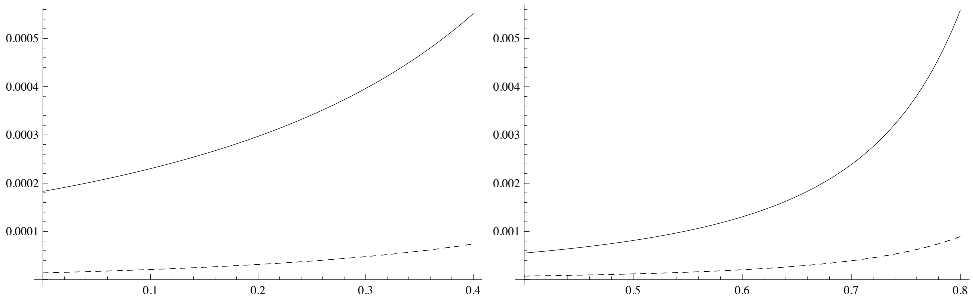

Figure 3: Difference in Risk for $N = 500$

4 Conclusions

We derived an estimator, which has the minimum risk for Stein's loss on a class of estimators, and investigated their properties. We obtained that its asymptotic distribution is the same as that of the MLE, but the estimator is biased. We are preparing to submit the result for the quadratic loss, and in future, we hope to study cases for other loss functions.

References

[1] T. W. Anderson, Maximum likelihood estimators for a multivariate normal distribution when some observations are missing, *J. Am. Stat. Assoc.*, **52** (1957), 200–203.

[2] T. W. Anderson and L. Olkin, Maximum likelihood estimation of the parameters of a multivariate normal distribution, *Linear Algebra Appl.*, **70** (1985), 147–171.

[3] W. -Y. Chang and D. St. P. Richards, Finite-sample inference with monotone incomplete multivariate normal data, I. *J. Multivariate Anal.*, **100** (2009), 1883-1899.

[4] W. -Y. Chang and D. St. P. Richards, Finite-sample inference with monotone incomplete multivariate normal data, II. *J. Multivariate Anal.*, **101** (2010), 603-620.

[5] T. Kanda and Y. Fujikoshi, Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data, *Am. J. Math.-S.*, **18** (1998), 161–190.

[6] D. St. P. Richards and T. Yamada, The Stein phenomenon for monotone incomplete multivariate normal data. *J. Multivariate Anal.*, **101** (2010), 657-678.

[7] M. M. Romer and D. St. P. Richards, Maximum likelihood estimation of the mean of a multivariate normal population with monotone incomplete data. *Statist. Probab. Letters*, **80** (2010), 1284—1288.

[8] S. Tsukada, Unbiased estimator for a covariance matrix under two-step monotone incomplete sample, *Commun. Stat.-Theor. M.*, accepted.

[9] T. Yamada, M. M. Romer and D. St. P. Richards, Kurtosis tests for multivariate normality with monotone incomplete data, Preprint, Penn State University (2012).