STATISTICA, anno LXXVII, n. 4, 2017

MINIMAX ESTIMATION OF THE MEAN MATRIX OF THE MATRIX VARIATE NORMAL DISTRIBUTION UNDER THE DIVERGENCE LOSS FUNCTION

Shokofeh Zinodiny
Amirkabir University of Technology, Tehran, Iran
Sadegh Rezaei
Amirkabir University of Technology, Tehran, Iran
Saralees Nadarajah [1]
University of Manchester, Manchester M13 9PL, UK

1. INTRODUCTION

Let $\mathbf{X} = \left(x_{i,j}\right)$ be a $p \times m$ random matrix having a matrix variate normal distribution with mean matrix $\mathbf{\Theta} = \left(\theta_{i,j}\right)$ and covariance matrix $\mathbf{V} \otimes \mathbf{I}_m$, where $\mathbf{V}$ is known or unknown, $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\otimes$ denotes the Kronecker product. This note considers estimation of $\mathbf{\Theta}$ relative to the divergence loss function

$$L(\mathbf{a};\mathbf{\Theta}) = \frac{1}{\beta(1-\beta)}\left[1 - \int_{\mathbf{R}^{p\times m}} f^{\beta}(\mathbf{X}|\mathbf{a}) f^{1-\beta}(\mathbf{X}|\mathbf{\Theta})\, d\mathbf{x}\right], \tag{1}$$

where $0 < \beta < 1$, $f(\mathbf{X}|\mathbf{\Theta})$ is the conditional density function of $\mathbf{X}$ given $\mathbf{\Theta}$ and $\mathbf{a}$ is an estimator of $\mathbf{\Theta}$.

Estimators for the mean matrix have been proposed under different loss functions. Efron and Morris (1972) proposed an empirical Bayes estimator outperforming the maximum likelihood estimator, X, for the case $m > p + 1$. Stein (1973), Zhang (1986a), Zhang (1986b), Bilodeau and Kariya (1989), Ghosh and Shieh (1991), Konno (1991), Tsukuma (2008), Tsukuma (2010) and others found minimax estimators better than the maximum likelihood estimator for quadratic loss and general quadratic loss functions.

The most recent minimax estimator for $\mathbf{\Theta}$ is that due to Zinodiny *et al.* (2017). They estimated $\mathbf{\Theta}$ under the balanced loss function. But Zinodiny *et al.* (2017) considered only the case $\mathbf{V} = \sigma^2 \mathbf{I}_p$, where σ^2 is unknown. In this note, we consider three cases

[1] Corresponding author. E-mail: mbbsssn2@manchester.ac.uk

as outlined in the abstract. Furthermore, the divergence loss function in (1) has attractive properties not shared by other loss functions, including the balanced loss function. Firstly, it is invariant to one-to-one coordinate transformation of $\mathbf{a}$. Secondly, the Fisher information function, used commonly to measure the sensitivity of estimates, is proportional to (1). Thirdly, many loss functions including the balanced loss function are symmetric, but (1) is not symmetric. If β is closer to 1 more weight is given to $\mathbf{a}$. If β is closer to 0 more weight is given to $\boldsymbol{\Theta}$. Other attractive properties of (1) can be found in Kashyap (1974).

The divergence loss function is directly comparable to the densities $f(\mathbf{X}|\boldsymbol{\Theta})$ and $f(\mathbf{X}|\mathbf{a})$. Robert (2001) refers to it as an intrinsic loss function, see also Amari (1982) and Cressie and Read (1984). The divergence loss function has been applied in many areas. Some examples include: discrimination between stationary Gaussian processes (Corduas, 1985); the Vocal Joystick engine, a real-time software library which can be used to map non-linguistic vocalizations into realizable continuous control signals (Malkin *et al.*, 2011); risk-aware modeling framework for speech summarization (Chen and Lin, 2012); models for web image retrieval (Yang *et al.*, 2012); models for coclustering the mouse brain atlas (Ji *et al.*, 2013); properties of record values (Paul and Thomas, 2016).

The aim of this note is to estimate the mean matrix of the matrix variate normal distribution under the divergence loss function. The contents are organized as follows: Section 2 shows that the empirical Bayes estimators dominate the maximum likelihood estimator under (1) for $m > p+1$ and hence the latter is inadmissible for the case $\mathbf{V}$ is known. We find a general class of minimax estimators using a technique due to Stein (1973) when $\mathbf{V} = \mathbf{I}_p$. Section 3 shows that X is inadmissible for $m > p+1$ when $\mathbf{V} = \sigma^2\mathbf{I}_p$, where σ^2 is unknown. Section 4 shows that X is inadmissible for the case $\mathbf{V}$ is unknown. These sections in fact extend the results of Ghosh *et al.* (2008) and Ghosh and Mergel (2009) to the multivariate case. Section 5 performs a simulation study to compare one of the derived estimators with that in Zinodiny *et al.* (2017). Section 6 concludes the note.

2. Estimation of the mean matrix when $\mathbf{V}$ is known

Here, we suppose

$$\mathbf{X} \sim N_{p\times m}(\boldsymbol{\Theta}, \mathbf{V}\otimes\mathbf{I}_m),$$

where $N_{p\times m}(\boldsymbol{\Theta}, \mathbf{V}\otimes\mathbf{I}_m)$ denotes the matrix variate normal distribution with mean matrix $\boldsymbol{\Theta}$ and covariance matrix $\mathbf{V}\otimes\mathbf{I}_m$, where $\mathbf{V}$ is assumed known. We consider the problem of estimating the mean matrix $\boldsymbol{\Theta}$ under the loss function (1), namely

$$\begin{aligned} L(\mathbf{a};\boldsymbol{\Theta}) &= \frac{1}{\beta(1-\beta)}\left[1-\int_{\mathbf{R}^{p\times m}} f^{\beta}(\mathbf{X}|\mathbf{a}) f^{1-\beta}(\mathbf{X}|\boldsymbol{\Theta})\, d\mathbf{x}\right] \\ &= \frac{1}{\beta(1-\beta)}\left[1-e^{-\frac{\beta(1-\beta)\mathrm{tr}\left((\mathbf{a}-\boldsymbol{\Theta})'\mathbf{V}^{-1}(\mathbf{a}-\boldsymbol{\Theta})\right)}{2}}\right], \end{aligned} \tag{2}$$

where $\mathrm{tr}(\mathbf{A})$ and $\mathbf{A}'$ denote, respectively, the trace and the transpose of a matrix $\mathbf{A}$, and $f(\mathbf{X}|\boldsymbol{\Theta})$ denotes the matrix variate normal density function. The usual estimator for $\boldsymbol{\Theta}$ is the maximum likelihood estimator, X.

Lemma 1 shows that the maximum likelihood estimator is minimax under the loss function (2).

LEMMA 1. *The estimator X is minimax under the loss function (2).*

PROOF. Suppose the proper prior sequence $\pi_n(\boldsymbol{\Theta})$ is distributed according to a matrix variate normal distribution with mean matrix zero and covariance matrix $n\mathbf{C}\otimes\mathbf{I}_m$, i.e. $\boldsymbol{\Theta} \sim N_{p\times m}\left(\mathbf{0}_{p\times m}, n\mathbf{C}\otimes\mathbf{I}_m\right)$, where $\mathbf{C}$ is a $p \times p$ known positive definite matrix. Then, the posterior distribution is

$$\boldsymbol{\Theta}|\mathbf{X} \sim N_{p\times m}\left(\left(\mathbf{I}_p - \mathbf{V}(n\mathbf{C}+\mathbf{V})^{-1}\right)\mathbf{X}, \left(\mathbf{V}^{-1}+(n\mathbf{C})^{-1}\right)^{-1}\otimes\mathbf{I}_m\right).$$

Thus, the Bayes estimator is

$$\boldsymbol{\delta}_{\pi_n}(\mathbf{X}) = E[\boldsymbol{\Theta}|\mathbf{X}] = \left[\mathbf{I}_p - \mathbf{V}(n\mathbf{C}+\mathbf{V})^{-1}\right]\mathbf{X}.$$

Moreover, the risk function of $\mathbf{X}$ and Bayes risk function of $\boldsymbol{\delta}_{\pi_n}(\mathbf{X})$ are

$$R(\mathbf{X};\boldsymbol{\Theta}) = \frac{1-[1+\beta(1-\beta)]^{-\frac{pm}{2}}}{\beta(1-\beta)}$$

and

$$r_n = r\left(\pi_n, \boldsymbol{\delta}_{\pi_n}(\mathbf{X})\right) = \frac{1-\left|\mathbf{V}^{-1}+(n\mathbf{C})^{-1}\right|^{\frac{m}{2}}\left|[1+\beta(1-\beta)]\mathbf{V}^{-1}+(n\mathbf{C})^{-1}\right|^{-\frac{m}{2}}}{\beta(1-\beta)},$$

respectively, where $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$. Since the determinant of a matrix is a continuous function, we have

$$\lim_{n\to\infty} r_n = \frac{1-[1+\beta(1-\beta)]^{-\frac{pm}{2}}}{\beta(1-\beta)} = \sup_{\boldsymbol{\Theta}} R(\mathbf{X};\boldsymbol{\Theta}).$$

Hence, using Theorem 1.12 on page 613 of Lehmann and Casella (1998) and results of Blyth (1951), X is minimax under the loss function (2). □

Now we construct a class of empirical Bayes estimators better than X. Assume we have some additional information about $\boldsymbol{\Theta}$ that can be written as

$$\boldsymbol{\Theta} \sim N_{p\times m}\left(\mathbf{0}_{p\times m}, \mathbf{A}\otimes\mathbf{I}_m\right).$$

The conditional distribution of $\boldsymbol{\Theta}$ given $\mathbf{X}$ is

$$N_{p\times m}\Big(\big(\mathbf{I}_p-\mathbf{V}(\mathbf{V}+\mathbf{A})^{-1}\big)\mathbf{X},\big(\mathbf{V}^{-1}+\mathbf{A}^{-1}\big)^{-1}\otimes\mathbf{I}_m\Big).$$

So, the Bayes estimator of $\boldsymbol{\Theta}$ under (2) is

$$\widehat{\boldsymbol{\Theta}}_B=E[\boldsymbol{\Theta}|\mathbf{X}]=\big(\mathbf{I}_p-\mathbf{V}\boldsymbol{\Sigma}^{-1}\big)\mathbf{X},$$

where $\boldsymbol{\Sigma}=\mathbf{A}+\mathbf{V}$. In an empirical Bayes scenario, $\boldsymbol{\Sigma}$ is unknown, and is estimated from the marginal distribution of $\mathbf{X}$. The marginal distribution of $\mathbf{X}$ is $N_{p\times m}\big(\mathbf{0}_{p\times m},\boldsymbol{\Sigma}\otimes\mathbf{I}_m\big)$. So, $\mathbf{S}=\mathbf{X}\mathbf{X}'$ is completely sufficient for $\boldsymbol{\Sigma}$ and an empirical Bayes estimator is

$$\widehat{\boldsymbol{\Theta}}_{EB}=E[\boldsymbol{\Theta}|\mathbf{X}]=\big[\mathbf{I}_p-\mathbf{V}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\big]\mathbf{X},$$

where $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ is an estimator for $\boldsymbol{\Sigma}^{-1}$ and $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ depends on $\mathbf{X}$ only through $\mathbf{S}$. According to Ghosh and Shieh (1991), a natural candidate for $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ is $(m-p-1)\mathbf{S}^{-1}$. Ghosh and Shieh (1991) obtained this empirical Bayes estimator when the loss function was

$$L_1(\boldsymbol{\delta};\boldsymbol{\Theta})=\operatorname{tr}\big((\boldsymbol{\delta}-\boldsymbol{\Theta})'\mathbf{Q}(\boldsymbol{\delta}-\boldsymbol{\Theta})\big),\tag{3}$$

where $\mathbf{Q}$ is a $p\times p$ known positive definite matrix.

To continue, we use the following notations borrowed from Ghosh and Shieh (1991): let $E_{\boldsymbol{\Theta}}$ denote the expectation conditional on $\boldsymbol{\Theta}$; $\widetilde{E}$ denotes the expectation over the marginal distribution of $\boldsymbol{\Theta}$; E denotes the expectation over the joint distribution of $\mathbf{X}$ and $\boldsymbol{\Theta}$. Let $R_i(\boldsymbol{\delta};\boldsymbol{\Theta})=E_{\boldsymbol{\Theta}}[L_i(\boldsymbol{\delta};\boldsymbol{\Theta})]$, $i=1,2,3$ and $R(\boldsymbol{\delta};\boldsymbol{\Theta})=E_{\boldsymbol{\Theta}}[L(\boldsymbol{\delta};\boldsymbol{\Theta})]$. We use the notation $\mathbf{D}=\big(d_{i,j}\big)$, where $d_{i,j}=\Big(\frac{1+\delta_{i,j}}{2}\Big)\frac{\partial}{\partial s_{i,j}}$ (see the last line of page 308 in Ghosh and Shieh (1991)), $\delta_{i,j}$ being the Kronecker deltas and $s_{i,j}$ the (i,j)th element of $\mathbf{S}$. Note that $\mathbf{D}$ is a differential operator in the form of a matrix. If we assume $f(\mathbf{A})$ is a real-valued function of a $p\times p$ matrix $\mathbf{A}$, then $\mathbf{D}f(\mathbf{A})$ is a $p\times p$ matrix with (i,j)th element $\frac{1+\delta_{i,j}}{2}\frac{\partial f(\mathbf{A})}{\partial s_{i,j}}$. For example, if $p=2$ and $f(\mathbf{A})=s_{1,1}^2+s_{2,2}^2$ then $\mathbf{D}f(\mathbf{A})=\begin{pmatrix}2s_{1,1}&0\\0&2s_{2,2}\end{pmatrix}$. Also for any $p\times p$ matrix $\mathbf{T}$, $\mathbf{D}(\mathbf{T})$ is a $p\times p$ matrix with (i,j)th element $\sum_{l=1}^{p}d_{i,l}t_{l,j}$. For example, if $p=2$ then $\mathbf{D}(\mathbf{S})=\begin{pmatrix}\frac{3}{2}&0\\0&\frac{3}{2}\end{pmatrix}$.

THEOREM 2. *The empirical Bayes estimator in (3) is minimax under (2) if it is minimax under the quadratic loss function (3) when* $\mathbf{Q}=\mathbf{V}^{-1}$.

PROOF. The difference of the risks of $\widehat{\boldsymbol{\Theta}}_{EB}$ and $\mathbf{X}$ is

$$R\big(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}\big)-R(\mathbf{X};\boldsymbol{\Theta})=\frac{1}{\beta(1-\beta)}\Big[g(\mathbf{X};\boldsymbol{\Theta})-g\big(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}\big)\Big],$$

where

$$g(\boldsymbol{\delta};\boldsymbol{\Theta}) = E\left[e^{-\beta(1-\beta)\operatorname{tr}\left((\boldsymbol{\delta}-\boldsymbol{\Theta})'\mathbf{V}^{-1}(\boldsymbol{\delta}-\boldsymbol{\Theta})\right)/2}\right].$$

Using the fact that $e^{-x} \geq 1-x$, we see that

$$g(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}) \geq g(\mathbf{X};\boldsymbol{\Theta}) - \frac{\beta(1-\beta)}{2[1+\beta(1-\beta)]^{\frac{pm}{2}}} E_{\boldsymbol{\Theta}}[\operatorname{tr}(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{V}'\mathbf{V}^{-1}\mathbf{V}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} - 2\operatorname{tr}(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{V}'\mathbf{V}^{-1}(\mathbf{X}-\boldsymbol{\Theta}))],$$

where $E_{\boldsymbol{\Theta}}$ is taken with respect to $N_{p\times m}\left(\boldsymbol{\Theta}, \frac{\mathbf{V}}{1+\beta(1-\beta)} \otimes \mathbf{I}_m\right)$. Using Theorem 2 in Ghosh and Shieh (1991) and the notation $R_1(\boldsymbol{\delta};\boldsymbol{\Theta}) = E_{\boldsymbol{\Theta}}[L_1(\boldsymbol{\delta};\boldsymbol{\Theta})]$, we can write

$$g\left(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}\right) - g(\mathbf{X};\boldsymbol{\Theta}) \geq -\frac{\beta(1-\beta)}{2}[1+\beta(1-\beta)]^{-\frac{pm}{2}}\left[R_1\left(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}\right) - R_1(\mathbf{X};\boldsymbol{\Theta})\right]. \quad (4)$$

Hence, if $R_1\left(\widehat{\boldsymbol{\Theta}}_{EB},\boldsymbol{\Theta}\right) \leq R_1(\mathbf{X};\boldsymbol{\Theta})$, then $R\left(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}\right) \leq R(\mathbf{X};\boldsymbol{\Theta})$. □

Similar to Ghosh and Shieh (1991), we can write (4) as

$$g\left(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}\right) - g(\mathbf{X};\boldsymbol{\Theta}) \geq -\frac{\beta(1-\beta)}{2}[1+\beta(1-\beta)]^{-\frac{pm}{2}} E\left[\operatorname{tr}\left(\mathbf{V}^{\frac{1}{2}}\mathbf{U}_1\mathbf{V}^{\frac{1}{2}}\right)\right]$$

and

$$\mathbf{U}_1 = \widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{S}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S}) - 4\mathbf{D}\left(\mathbf{S}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\right) - 2(m-p-1)\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S}),$$

and $R_1\left(\widehat{\boldsymbol{\Theta}}_{EB};\boldsymbol{\Theta}\right) \leq R_1(\mathbf{X};\boldsymbol{\Theta})$ if $\mathbf{U}_1 \leq 0$ has a positive probability for some $\boldsymbol{\Theta}$. If $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ is chosen in such a way that $\widehat{\boldsymbol{\Theta}}_{EB}$ improves on X under the quadratic loss function (3), then it improves on X also under (2).

Let $\mathbf{O}_p$ be the set of orthogonal matrices of order p and let $\mathbf{V}_{m,p}$ be the Stiefel manifold, namely, $\mathbf{V}_{m,p} = \left\{\mathbf{V} \in \mathbf{R}^{m\times p}, \mathbf{V}'\mathbf{V} = \mathbf{I}_p\right\}$. Write the singular value decomposition of $\mathbf{X}$ as $\mathbf{ULV}'$, where $\mathbf{U} \in \mathbf{O}_p$, $\mathbf{V} \in \mathbf{V}_{m,p}$ and $\mathbf{L} = \operatorname{diag}\left(l_1, l_2, \ldots, l_p\right)$ with $l_1 > l_2 > \cdots > l_p > 0$. Ghosh and Shieh (1991) showed that if $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S}) = \mathbf{UL}^{-1}\boldsymbol{\Psi}^*(\mathbf{L})\mathbf{U}'$, where $\boldsymbol{\Psi}^*(\mathbf{L})$ is a diagonal matrix with diagonal elements equal to $\psi_i^*(\mathbf{L}), i = 1,\ldots,p$, then $\widehat{\boldsymbol{\Theta}}_{EB}$ is minimax under the following conditions when the loss function is (3) with $\mathbf{Q} = \mathbf{V}^{-1}$.

THEOREM 3. *Suppose that* $\psi_i^*(\mathbf{L})$, $i = 1,\ldots,p$ *satisfy*

I. $0 < \psi_i^*(\mathbf{L}) < 2(m-p-1)$, $i = 1,\ldots,p$,

II. *For every* $i = 1, \ldots, p$, $\frac{\partial \psi_i^*(\mathbf{L})}{\partial l_i} \geq 0$,

III. $\psi_i^*(\mathbf{L})$ *are similarly ordered to* l_i, *that is* $\left[\psi_i^*(\mathbf{L}) - \psi_t^*(\mathbf{L})\right](l_i - l_t) \geq 0$ *for every* $1 \leq i, t \leq p$.

Then, $\widehat{\mathbf{\Theta}}_{EB} = \left[\mathbf{I}_p - \mathbf{VUL}^{-1}\mathbf{\Psi}^*(\mathbf{L})\mathbf{U}'\right]\mathbf{X}$ *improves on* X *for* $m > p+1$ *under the loss function (3).*

PROOF. See Ghosh and Shieh (1991). □

Using Theorem 2, we can see that $\widehat{\mathbf{\Theta}}_{EB} = \left[\mathbf{I}_p - \mathbf{VUL}^{-1}\mathbf{\Psi}^*(\mathbf{L})\mathbf{U}'\right]\mathbf{X}$ is minimax for $m > p+1$ under the conditions of Theorem 3 when the loss function is (2). This means, X is inadmissible for $m > p+1$ under the loss function (2). Examples (1) and (2) in Ghosh and Shieh (1991) illustrated the above result. Of course, one may obtain estimators of the form (3) which dominate X when the conditions of Theorem 3 are not met. Some examples are given in Ghosh and Shieh (1991).

In the rest of this section, we consider the problem of estimating $\mathbf{\Theta}$ under the loss function (2) for the case $\mathbf{V} = \mathbf{I}_p$. Let $\boldsymbol{\delta} = \mathbf{X} + \mathbf{G}$, where $\mathbf{G} = \mathbf{G}(\mathbf{X})$ is a $p \times m$ matrix valued function of $\mathbf{X}$. Also let ∇ be a $p \times m$ matrix with (i,j) element equal to the differential operator $\frac{\partial}{\partial x_{i,j}}$. We obtain a general condition for minimaxity.

THEOREM 4. *Suppose*

$$\boldsymbol{\delta} = \mathbf{X} + \mathbf{G}$$

is minimax under (3). Then it is also minimax under (2).

PROOF. The difference of the risks of $\boldsymbol{\delta}$ and $\mathbf{X}$ is

$$R(\boldsymbol{\delta}; \mathbf{\Theta}) - R(\mathbf{X}; \mathbf{\Theta}) = g(\mathbf{X}; \mathbf{\Theta}) - g(\boldsymbol{\delta}; \mathbf{\Theta}),$$

where

$$g(\boldsymbol{\delta}; \mathbf{\Theta}) = E_{\mathbf{\Theta}}\left[e^{-\frac{\beta(1-\beta)\mathrm{tr}\left((\boldsymbol{\delta}-\mathbf{\Theta})'(\boldsymbol{\delta}-\mathbf{\Theta})\right)}{2}}\right].$$

Using $e^{-x} \geq 1 - x$, we see that

$$g(\boldsymbol{\delta}; \mathbf{\Theta}) \geq g(\mathbf{X}; \mathbf{\Theta}) - \frac{\beta(1-\beta)}{2}\left[1 + \beta(1-\beta)\right]^{-\frac{pm}{2}} E\left[\mathrm{tr}\left(\mathbf{G}'\mathbf{G}\right) + 2\,\mathrm{tr}\left(\mathbf{G}'(\mathbf{X} - \mathbf{\Theta})\right)\right],$$

where $E_{\mathbf{\Theta}}$ is taken with respect to $N_{p\times m}\left(\mathbf{\Theta}, \frac{\mathbf{I}_p}{1+\beta(1-\beta)} \otimes \mathbf{I}_m\right)$. Using Theorem 2 in Ghosh and Shieh (1991), we can write

$$g(\boldsymbol{\delta}; \mathbf{\Theta}) - g(\mathbf{X}; \mathbf{\Theta}) \geq -\frac{\beta(1-\beta)}{2}\left[1 + \beta(1-\beta)\right]^{-\frac{pm}{2}}\left[R_1(\boldsymbol{\delta}; \mathbf{\Theta}) - R_1(\mathbf{X}; \mathbf{\Theta})\right].$$

Hence, if $R_1(\boldsymbol{\delta}; \mathbf{\Theta}) \leq R_1(\mathbf{X}; \mathbf{\Theta})$, then $R(\boldsymbol{\delta}; \mathbf{\Theta}) \leq R(\mathbf{X}; \mathbf{\Theta})$. □

By Theorem 4, if $E_{\boldsymbol{\Theta}}[\operatorname{tr}(\mathbf{G}'\mathbf{G})+2\operatorname{tr}(\mathbf{G}'(\mathbf{X}-\boldsymbol{\Theta}))]\leq 0$ then $R(\boldsymbol{\delta};\boldsymbol{\Theta})\leq R(\mathbf{X};\boldsymbol{\Theta})$. Using Stein (1973)'s identities, $E_{\boldsymbol{\Theta}}\left[\frac{\operatorname{tr}(\mathbf{G}'(\mathbf{X}-\boldsymbol{\Theta}))}{1+\beta(1-\beta)}\right]=E_{\boldsymbol{\Theta}}[\operatorname{tr}(\nabla\mathbf{G}')]$, so we can write

$$E_{\boldsymbol{\Theta}}\left[\operatorname{tr}\left(\mathbf{G}'\mathbf{G}\right)+2\operatorname{tr}\left(\mathbf{G}'(\mathbf{X}-\boldsymbol{\Theta})\right)\right]=E\left\{\operatorname{tr}\left(\mathbf{G}'\mathbf{G}\right)+2[1+\beta(1-\beta)]\operatorname{tr}\left(\nabla\mathbf{G}'\right)\right\}.$$

So, if $\operatorname{tr}(\mathbf{G}'\mathbf{G})+2[1+\beta(1-\beta)]\operatorname{tr}(\nabla\mathbf{G}')\leq 0$ with a positive probability for some $\boldsymbol{\Theta}$, then $\boldsymbol{\delta}$ is minimax under the loss function (2).

Now, assume the class of shrinkage estimators

$$\boldsymbol{\delta}=\left[\mathbf{I}_p-\mathbf{U}\mathbf{F}^{-1}\boldsymbol{\Psi}(\mathbf{F})\mathbf{U}'\right]\mathbf{X}, \tag{5}$$

where $\boldsymbol{\Psi}(\mathbf{F})=\operatorname{diag}\left(\psi_1,\ldots,\psi_p\right)$ is a diagonal matrix with ψ_i being functions of $\mathbf{F}=\mathbf{L}^2$. We obtain the following by applying Theorem 4.

COROLLARY 5. *Suppose ψ_i satisfy*

I. For fixed $f_1,\ldots,f_{i-1},f_{i+1},\ldots,f_p$, $\frac{\partial\psi_i}{\partial f_i}\geq 0$, where $i=1,\ldots,p$,

II. $0\leq\psi_p\leq\cdots\leq\psi_1\leq 2(m-p-1)$.

Then, $\boldsymbol{\delta}$ in (5) is minimax under the loss function (2).

PROOF. With the notation $\boldsymbol{\Phi}(\mathbf{F})=(\mathbf{F}')^{-1}\boldsymbol{\Psi}(\mathbf{F})$, Stein (1973) showed that $\boldsymbol{\delta}=\left[\mathbf{I}_p-\mathbf{U}\boldsymbol{\Phi}(\mathbf{F})\mathbf{U}'\right]\mathbf{X}$ is minimax under conditions (I) and (II) when the loss function is (3). Using Theorem 4, we can see that $\boldsymbol{\delta}=\mathbf{X}+\mathbf{G}$ is minimax under (2), where $\mathbf{G}=-\mathbf{U}\boldsymbol{\Phi}(\mathbf{F})\mathbf{U}'\mathbf{X}$. □

Tsukuma (2008) showed that the proper Bayes estimators with respect to the following prior

$$\boldsymbol{\Theta}\sim N_{p\times m}\left(0_{p\times m},\boldsymbol{\Lambda}^{-1}\left(\mathbf{I}_p-\boldsymbol{\Lambda}\right)\otimes\mathbf{I}_m\right)$$

are of the form $\boldsymbol{\delta}=\left[\mathbf{I}_p-\mathbf{U}\mathbf{F}^{-1}\boldsymbol{\Psi}(\mathbf{F})\mathbf{U}'\right]\mathbf{X}$, where $\boldsymbol{\Lambda}$ is a $p\times m$ random matrix with the density function

$$\pi(\boldsymbol{\Lambda})\propto|\boldsymbol{\Lambda}|^{\frac{a}{2}-1}\mathbf{I}\left(0_{p\times p}<\boldsymbol{\Lambda}<\mathbf{I}_p\right).$$

Under certain a and loss function (3), these estimators are minimax, thus we can see from Corollary 5 that $\boldsymbol{\delta}$ is minimax under (2).

3. Estimation of the mean matrix when $\mathbf{V}=\sigma^2\mathbf{I}_p$ and σ^2 is unknown

Here, we assume that

$$\mathbf{X}\sim N_{p\times m}\left(\mathbf{\Theta},\sigma^2\mathbf{I}_p\otimes\mathbf{I}_m\right),$$

where σ^2 is unknown. So, S is independent of X and

$$S\sim\sigma^2\chi^2_n,$$

where χ^2_n denotes a chi-square random variable with n degrees of freedom. We consider estimation of $\mathbf{\Theta}$ under

$$\begin{aligned} L(\mathbf{a};\mathbf{\Theta}) &= \frac{1}{\beta(1-\beta)}\left[1-\int_{\mathbf{R}^{p\times m}} f^{\beta}(\mathbf{X}|\mathbf{a})f^{1-\beta}(\mathbf{X}|\mathbf{\Theta})\,d\mathbf{x}\right] \\ &= \frac{1}{\beta(1-\beta)}\left[1-e^{-\frac{\beta(1-\beta)\mathrm{tr}\left((\mathbf{a}-\mathbf{\Theta})'(\mathbf{a}-\mathbf{\Theta})\right)}{2\sigma^2}}\right]. \end{aligned} \tag{6}$$

Similar to Lemma 1, one can show that the maximum likelihood estimator, X, is minimax under the loss function (6). Tsukuma (2010) obtained general conditions for minimaxity of estimators having the form $\boldsymbol{\delta}_1=\mathbf{X}+\mathbf{G}(\mathbf{X},S)$, where $\mathbf{G}(\mathbf{X},S)$ is a $p\times m$ matrix valued function of $\mathbf{X}$ and S under the quadratic loss function

$$L_2(\mathbf{a};\mathbf{\Theta})=\frac{\mathrm{tr}((\mathbf{a}-\mathbf{\Theta})'(\mathbf{a}-\mathbf{\Theta}))}{\sigma^2}. \tag{7}$$

Theorem 6. *The estimator $\boldsymbol{\delta}_1=\mathbf{X}+\mathbf{G}(\mathbf{X},S)$ is minimax under the quadratic loss function (6) if it is minimax under (7) for $m>p+1$.*

Proof. We have

$$R(\boldsymbol{\delta}_1;\mathbf{\Theta})-R(\mathbf{X};\mathbf{\Theta})=\frac{1}{\beta(1-\beta)}\left[g(\mathbf{X};\mathbf{\Theta})-g(\boldsymbol{\delta}_1;\mathbf{\Theta})\right].$$

Using $e^{-x}\geq 1-x$,

$$g(\boldsymbol{\delta}_1;\mathbf{\Theta})\geq g(\mathbf{X};\mathbf{\Theta})-\frac{\beta(1-\beta)}{2}\left[1+\beta(1-\beta)\right]^{-\frac{pm}{2}}E_{\mathbf{\Theta}}\left[\frac{\mathrm{tr}(\mathbf{G}'\mathbf{G})+2\,\mathrm{tr}(\mathbf{G}'(\mathbf{X}-\mathbf{\Theta}))}{\sigma^2}\right]$$

and

$$g(\boldsymbol{\delta}_1;\mathbf{\Theta})-g(\mathbf{X};\mathbf{\Theta})\geq-\frac{\beta(1-\beta)}{2}\left[1+\beta(1-\beta)\right]^{-\frac{pm}{2}}\left[R_2(\boldsymbol{\delta}_1;\mathbf{\Theta})-R_2(\mathbf{X};\mathbf{\Theta})\right],$$

where

$$R_2(\boldsymbol{\delta};\mathbf{\Theta})=E_{\mathbf{\Theta}}\left[L_2(\boldsymbol{\delta};\mathbf{\Theta})\right].$$

So, if

$$R_2(\boldsymbol{\delta}_1;\boldsymbol{\Theta}) \leq R_2(\mathbf{X};\boldsymbol{\Theta}),$$

then

$$R(\boldsymbol{\delta}_1;\boldsymbol{\Theta}) \leq R(\mathbf{X};\boldsymbol{\Theta}).$$

The proof is complete. □

Tsukuma (2010) showed that if

$$\frac{(n-2)\operatorname{tr}(\mathbf{G}'\mathbf{G})}{s} + 2\operatorname{tr}\left(\nabla\mathbf{G}'\right) + \frac{\partial \operatorname{tr}(\mathbf{G}'\mathbf{G})}{\partial s} \leq 0$$

then $\boldsymbol{\delta}_1 = \mathbf{X} + \mathbf{G}(\mathbf{X}, S)$ is minimax under the loss function (7). So, using Theorem 6, we see that $\boldsymbol{\delta}$ is minimax under the loss function (6).

THEOREM 7. *Consider the estimator*

$$\boldsymbol{\delta}_2(\mathbf{X}, S) = \left[\mathbf{I}_p - \mathbf{U}\mathbf{F}^{-1}\boldsymbol{\Psi}(\mathbf{F}, S)\mathbf{U}'\right]\mathbf{X},$$

where $\mathbf{F} = \frac{\mathbf{L}}{s} = \operatorname{diag}\left(f_1,\ldots,f_p\right)$ *and* $\boldsymbol{\Psi}(\mathbf{F},s) = \operatorname{diag}\left(\psi_1,\ldots,\psi_p\right)$ *are diagonal matrices with* ψ_i *being functions of* $\mathbf{F}$ *and* s. *If the following conditions hold, then* $\boldsymbol{\delta}_2$ *is minimax under the loss function (7)*

I. For fixed $\mathbf{F}$, $\frac{\partial \psi_i}{\partial s} \leq 0$,

II. For fixed $f_1,\ldots,f_{j-1},f_{j+1},\ldots,f_p$, $\frac{\partial \psi_i}{\partial f_j} \geq 0$, *where* $i, j = 1,\ldots,p$,

III. $0 \leq \psi_p \leq \cdots \leq \psi_1 \leq \frac{2(m-p-1)}{n+2}$.

PROOF. Tsukuma (2010) showed that $\boldsymbol{\delta}_2$ is minimax under conditions I-III when the loss function is (7). So, using Theorem 6, we see that $\boldsymbol{\delta}_2$ is minimax under the loss function (6) if conditions I-III hold. □

4. ESTIMATION OF THE MEAN MATRIX WHEN $\mathbf{V}$ IS UNKNOWN

Here, we assume

$$\mathbf{X} \sim N_{p\times m}(\boldsymbol{\Theta}, \mathbf{V}\otimes\mathbf{I}_m),$$

where $\mathbf{V}$ is unknown. So, $\widehat{\mathbf{V}}$ is independent of X and

$$\widehat{\mathbf{V}} \sim W_p(\mathbf{V}, n),$$

where $W_p(\mathbf{V}, n)$ denotes a Wishart random vector with n degrees of freedom and mean $n\mathbf{V}$. We consider estimating $\boldsymbol{\Theta}$ under the loss function

$$\begin{aligned} L(\mathbf{a};\boldsymbol{\Theta}) &= \frac{1}{\beta(1-\beta)}\left[1-\int_{\mathbf{R}^{p\times m}} f^{\beta}(\mathbf{X}|\mathbf{a}) f^{1-\beta}(\mathbf{X}|\boldsymbol{\Theta})\, d\mathbf{x}\right] \\ &= \frac{1}{\beta(1-\beta)}\left[1-e^{-\frac{\beta(1-\beta)\operatorname{tr}\left((\mathbf{a}-\boldsymbol{\Theta})'\mathbf{V}^{-1}(\mathbf{a}-\boldsymbol{\Theta})\right)}{2}}\right]. \end{aligned} \tag{8}$$

Similar to Lemma 1, one can show that the maximum likelihood estimator, X, is minimax under the loss function (8).

We now consider the relationship between this loss function and the following quadratic loss function

$$L_3(\mathbf{a};\boldsymbol{\Theta},\mathbf{V}) = \operatorname{tr}\left((\mathbf{a}-\boldsymbol{\Theta})'\mathbf{Q}^*(\mathbf{a}-\boldsymbol{\Theta})\right), \tag{9}$$

where $\mathbf{Q}^* = \mathbf{V}^{-\frac{1}{2}}\mathbf{Q}\mathbf{V}^{-\frac{1}{2}}$. Assume $\boldsymbol{\delta}_3 = \mathbf{X} + \mathbf{G}\left(\mathbf{X}, \widehat{\mathbf{V}}\right)$, where $\mathbf{G}\left(\mathbf{X}, \widehat{\mathbf{V}}\right)$ is a $p \times m$ matrix valued function of $\mathbf{X}$ and $\widehat{\mathbf{V}}$.

THEOREM 8. *The estimator $\boldsymbol{\delta}_3 = \mathbf{X} + \mathbf{G}\left(\mathbf{X}, \widehat{\mathbf{V}}\right)$ is minimax under the loss function (8) if it is minimax under (9).*

PROOF. We have

$$R(\boldsymbol{\delta}_3;\boldsymbol{\Theta}) - R(\mathbf{X};\boldsymbol{\Theta}) = \frac{1}{\beta(1-\beta)}\left[g(\mathbf{X};\boldsymbol{\Theta}) - g(\boldsymbol{\delta}_3;\boldsymbol{\Theta})\right],$$

where

$$g(\boldsymbol{\delta}_3;\boldsymbol{\Theta}) = E\left[e^{-\frac{\beta(1-\beta)\operatorname{tr}\left((\boldsymbol{\delta}_3-\boldsymbol{\Theta})'\mathbf{V}^{-1}(\boldsymbol{\delta}_3-\boldsymbol{\Theta})\right)}{2}}\right].$$

So,

$$\begin{aligned} g(\boldsymbol{\delta}_3;\boldsymbol{\Theta}) \geq g(\mathbf{X};\boldsymbol{\Theta}) - \frac{\beta(1-\beta)}{2}&[1+\beta(1-\beta)]^{-\frac{pm}{2}} \\ &E_{\boldsymbol{\Theta}}[\operatorname{tr}(\mathbf{G}'\mathbf{Q}^*\mathbf{G}) + 2\operatorname{tr}(\mathbf{G}'\mathbf{Q}^*(\mathbf{X}-\boldsymbol{\Theta}))], \end{aligned}$$

where $E_{\boldsymbol{\Theta}}$ is taken with respect to $N_{p\times m}\left(\boldsymbol{\Theta}, \frac{\mathbf{V}}{1+\beta(1-\beta)} \otimes \mathbf{I}_m\right)$. Using

$$R_3(\boldsymbol{\delta};\boldsymbol{\Theta}) = E_{\boldsymbol{\Theta}}[L_3(\boldsymbol{\delta};\boldsymbol{\Theta})],$$

we can write

$$g(\boldsymbol{\delta}_3;\boldsymbol{\Theta}) - g(\mathbf{X};\boldsymbol{\Theta}) \geq -\frac{\beta(1-\beta)}{2}\left[1+\beta(1-\beta)\right]^{-\frac{pm}{2}}\left[R_3(\boldsymbol{\delta}_3;\boldsymbol{\Theta}) - R_3(\mathbf{X};\boldsymbol{\Theta})\right].$$

Hence, if $R_3(\boldsymbol{\delta}_3;\boldsymbol{\Theta}) \leq R_3(\mathbf{X};\boldsymbol{\Theta})$, then $R(\boldsymbol{\delta}_3;\boldsymbol{\Theta}) \leq R(\mathbf{X};\boldsymbol{\Theta})$. □

Shieh (1993) considered empirical Bayes estimators of the form

$$\widehat{\boldsymbol{\Theta}}_{1EB} = \left[\mathbf{I}_p - \widehat{\mathbf{V}}\mathbf{S}^{-1}\tau\left(\widehat{\mathbf{V}},\mathbf{S}\right)\right]\mathbf{X},$$

where $\mathbf{S} = \mathbf{X}\mathbf{X}'$ and $\tau\left(\widehat{\mathbf{V}},\mathbf{S}\right)$ is a symmetric matrix. Shieh (1993) showed that $\widehat{\boldsymbol{\Theta}}_{1EB}$ is better than the maximum likelihood estimator, X, under (8). For example, he showed that if $\tau\left(\widehat{\mathbf{V}},\mathbf{S}\right) = (m-p-1)\mathbf{I}_p$, then $\widehat{\boldsymbol{\Theta}}_{EB}$ improves on X for $m > p+1$.

For the case $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$, Konno (1990), Konno (1991), Konno (1992) considered the estimators

$$\boldsymbol{\delta}_4 = \begin{cases} \left[\mathbf{I}_p - \mathbf{R}\mathbf{F}^{-1}\boldsymbol{\Phi}(\mathbf{F})\mathbf{R}'\right]\mathbf{X}, & \text{if } m < p, \\ \left[\mathbf{I}_p - \mathbf{Q}\mathbf{F}^{-1}\boldsymbol{\Phi}(\mathbf{F})\mathbf{R}\mathbf{Q}^{-1}\right]\mathbf{X}, & \text{if } m \geq p, \end{cases}$$

where, for $p < m$, $\mathbf{X}\widehat{\mathbf{V}}^{-1}\mathbf{X}' = \mathbf{R}\mathbf{F}\mathbf{R}'$, $\mathbf{R} \in \mathbf{O}_p$ and, for $p \geq m$, $\mathbf{Q}\widehat{\mathbf{V}}^{-1}\mathbf{Q}' = \mathbf{I}_p$, $\mathbf{Q}'\mathbf{X}\mathbf{X}'\mathbf{Q} = \mathbf{F} = \operatorname{diag}\left(f_1,\ldots,f_{m\wedge p}\right)$, $f_1 \geq \cdots \geq f_{m\wedge p} > 0$. Under the following conditions, Konno (1990), Konno (1991), Konno (1992) showed that $\boldsymbol{\delta}_4$ improves on X when the loss function is (8):

i) For $i = 1,\ldots,p$, $\phi_i(\mathbf{F})$ is nondecreasing in f_i;

ii) $0 \leq \phi_{m\wedge p}(\mathbf{F}) \leq \phi_{m\wedge p-1}(\mathbf{F}) \leq \cdots \leq \phi_1(\mathbf{F}) \leq \frac{2(m\vee p - m\wedge p - 1)}{n+(2m-p)\wedge p+1}$.

So, using Theorem 8, we see that these estimators are minimax under (8).

5. Simulation study

As mentioned in Section 1, Zinodiny *et al.* (2017) estimated $\boldsymbol{\Theta}$ under the balanced loss function when $\mathbf{V} = \sigma^2\mathbf{I}_p$, where σ^2 is unknown. Here, we perform a simulation study to compare this estimator with the estimator under the divergence loss function, see Theorem 8. The simulation study was performed as follows:

1. simulate 10000 random samples each of size n from a matrix variate normal distribution with zero means and $\mathbf{V} = \sigma^2\mathbf{I}_p$;

2. compute Zinodiny *et al.* (2017)'s estimator and the estimator in Theorem 8 for each of the samples, say $\left\{\widehat{\boldsymbol{\Theta}}_1, \widehat{\boldsymbol{\Theta}}_2, \ldots, \widehat{\boldsymbol{\Theta}}_{10000}\right\}$ and $\left\{\widetilde{\boldsymbol{\Theta}}_1, \widetilde{\boldsymbol{\Theta}}_2, \ldots, \widetilde{\boldsymbol{\Theta}}_{10000}\right\}$;

3. compute the biases of Zinodiny *et al.* (2017)'s estimator and the estimator in Theorem 8 as

$$\frac{1}{10000mp}\sum_{i=1}^{10000}\sum_{j=1}^{p}\sum_{k=1}^{m}\widehat{\Theta}_{i,j,k}$$

and

$$\frac{1}{10000mp}\sum_{i=1}^{10000}\sum_{j=1}^{p}\sum_{k=1}^{m}\widetilde{\Theta}_{i,j,k};$$

4. compute the mean squared errors of Zinodiny *et al.* (2017)'s estimator and the estimator in Theorem 8 as

$$\frac{1}{10000mp}\sum_{i=1}^{10000}\sum_{j=1}^{p}\sum_{k=1}^{m}\widehat{\Theta}^2_{i,j,k}$$

and

$$\frac{1}{10000mp}\sum_{i=1}^{10000}\sum_{j=1}^{p}\sum_{k=1}^{m}\widetilde{\Theta}^2_{i,j,k}.$$

We repeated this procedure for $n = 10, 12, \ldots, 109$. We chose $\sigma = 1$, $m = 10$ and $p = 5$.

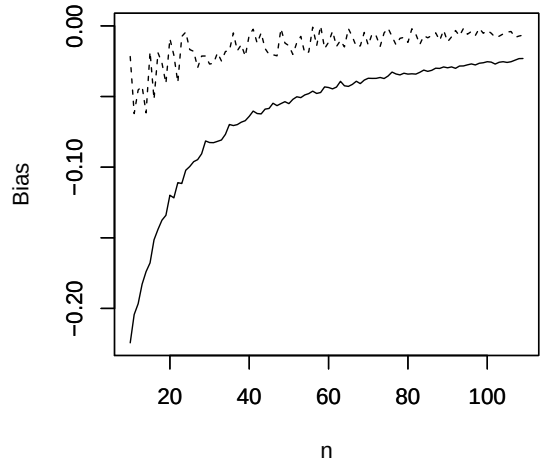

Figure 1 – Biases of Zinodiny *et al.* (2017)'s estimator (solid curve) and the estimator in Theorem 8 (broken curve) versus n.

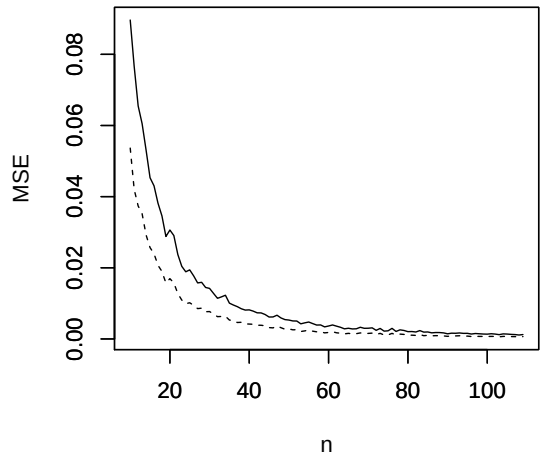

Figure 2 – Mean squared errors of Zinodiny *et al.* (2017)'s estimator (solid curve) and the estimator in Theorem 8 (broken curve) versus n.

The biases versus n are drawn in Figure 1. The mean squared errors versus n are drawn in Figure 2. We see that the estimator in Theorem 8 has consistently smaller bias and consistently smaller mean squared error for every n. We took $\sigma = 1$, $m = 10$ and $p = 5$ in the simulations. But the same observation was noted for a wide range of other values of σ, m and p. Hence, the estimator in Theorem 8 is a better estimator with respect to both bias and mean squared error.

6. Conclusions

We have considered the problem of estimating the mean of a matrix variate normal distribution. We have provided minimax estimators under the divergence loss function. The divergence loss function has several attractive features compared to other loss functions considered in the literature.

We have given minimax estimators for the mean considering different forms for the covariance matrix. The forms considered include the cases that the covariance matrix is known and the covariance matrix is completely unknown.

We have performed a simulation study to compare one of our estimators with the most recently proposed estimator. The simulation shows that our estimator has consistently smaller bias and consistently smaller mean squared error.

ACKNOWLEDGEMENTS

The authors would like to thank the referee and the Editor for careful reading and comments which greatly improved the paper.

REFERENCES

S. AMARI (1982). *Differential geometry of curved exponential families-curvatures and information loss*. Annals of Statistics, 10, pp. 357–387.

M. BILODEAU, T. KARIYA (1989). *Minimax estimators in the normal manova model*. Journal of Multivariate Analysis, 28, pp. 260–270.

C. R. BLYTH (1951). *On minimax statistical decision procedures and their admissibility*. Annals of Mathematical Statistics, 22, pp. 22–42.

B. CHEN, S. H. LIN (2012). *A risk-aware modeling framework for speech summarization*. IEEE Transactions on Audio, Speech, and Language Processing, 20, pp. 211–222.

M. CORDUAS (1985). *On the divergence between linear processes*. Statistica, 45, pp. 393–401.

N. CRESSIE, T. R. C. READ (1984). *Multinomial goodness-of-fit tests*. Journal of the Royal Statistical Society Series B, 46, pp. 440–464.

B. EFRON, C. MORRIS (1972). *Empirical Bayes on vector observations: An extension of Stein's method*. Biometrika, 59, pp. 335–347.

M. GHOSH, V. MERGEL (2009). *On the stein phenomenon under divergence loss and an unknown variance-covariance matrix*. Journal of Multivariate Analysis, 100, pp. 2331–2336.

M. GHOSH, V. MERGEL, G. S. DATTA (2008). *Estimation, prediction and the stein phenomenon under divergence loss*. Journal of Multivariate Analysis, 99, pp. 1941–1961.

M. GHOSH, G. SHIEH (1991). *Empirical bayes minimax estimators of matrix normal means*. Journal of Multivariate Analysis, 38, pp. 306–318.

S. JI, W. ZHANG, R. LI (2013). *A probabilistic latent semantic analysis model for coclustering the mouse brain atlas*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 10, pp. 1460–1468.

R. L. KASHYAP (1974). *Minimax estimation with divergence loss function*. Information Sciences, 7, pp. 341–364.

Y. KONNO (1990). *Families of minimax estimators of matrix of normal means with unknown covariance matrix*. Journal of the Japan Statistical Society, 20, pp. 191–201.

Y. KONNO (1991). *On estimation of a matrix of normal means with unknown covariance matrix*. Journal of Multivariate Analysis, 36, pp. 44–55.

Y. KONNO (1992). *Improved estimation of matrix of normal mean and eigenvalues in the multivariate F-distribution*. Ph.D. thesis, Institute of Mathematics, University of Tsukuba.

E. L. LEHMANN, G. CASELLA (1998). *Theory of Point Estimation*. Springer Verlag, New York, 2 ed.

J. MALKIN, X. LI, S. HARADA, J. LANDAY, J. BILMES (2011). *The vocal joystick engine v1.0*. Computer Speech and Language, 25, pp. 535–555.

J. PAUL, P. Y. THOMAS (2016). *Sharma-mittal entropy properties on record values*. Statistica, 76, pp. 273–287.

C. P. ROBERT (2001). *The Bayesian Choice, second edition*. Springer Verlag, New York.

G. SHIEH (1993). *Empirical bayes minimax estimators of matrix normal means for arbitrary quadratic loss and unknown covariance matrix*. Statistics and Decisions, 11, pp. 317–341.

C. STEIN (1973). *Estimation of the mean of a multivariate normal distribution*. In J. HÁJEK (ed.), *Proceedings of the Prague Symposium on Asymptotic statistics*. Universita Karlova, Prague, pp. 345–381.

H. TSUKUMA (2008). *Admissibility and minimaxity of Bayes estimators for a normal mean matrix*. Journal of Multivariate Analysis, 99, pp. 2251–2264.

H. TSUKUMA (2010). *Proper Bayes minimax estimators of the normal mean matrix with common unknown variances*. Journal of Statistical Planning and Inference, 140, pp. 2596–2606.

L. YANG, B. GENG, A. HANJALIC, X. S. HUA (2012). *A unified context model for web image retrieval*. ACM Transactions on Multimedia Computing, Communications, and Applications, 8. Article No. 28.

Z. ZHANG (1986a). *On estimation of matrix of normal mean*. Journal of Multivariate Analysis, 18, pp. 70–82.

Z. ZHANG (1986b). *Selecting a minimax estimator doing well, at a point*. Journal of Multivariate Analysis, 19, pp. 14–23.

S. ZINODINY, S. REZAEI, S. NADARAJAH (2017). *Bayes minimax estimation of the mean matrix of matrix-variate normal distribution under balanced loss function*. Statistics and Probability Letters, 125, pp. 110–120.

SUMMARY

The problem of estimating the mean matrix of a matrix-variate normal distribution with a covariance matrix is considered under two loss functions. We construct a class of empirical Bayes estimators which are better than the maximum likelihood estimator under the first loss function and hence show that the maximum likelihood estimator is inadmissible. We find a general class of minimax estimators. Also we give a class of estimators that improve on the maximum likelihood estimator under the second loss function and hence show that the maximum likelihood estimator is inadmissible.

Keywords: Empirical Bayes estimation; Matrix variate normal distribution; Mean matrix; Minimax estimation.