Stochastic Models in Probability and Statistics
Vol. 2, No. 1, 2025, pp. 31–39. Research Article

On shrinkage estimation under divergence loss

Mohammad Arashi[†*], **Hidekazu Tanaka**[‡], **Morteza Amini**[§]

[†] *Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran*
[‡] *Graduate School of Science, Osaka Prefecture University, Sakai, Japan*
[§] *Department of Statistics, School of Mathematics, Statistics, and Computer Science, College of Science, University of Tehran, Tehran, Iran*

Email(s): arashi@um.ac.ir, tanaka@ms.osakafu-u.ac.jp, morteza.amini@ut.ac.ir

Abstract. In this paper, the superiority conditions for a general class of shrinkage estimators in the estimation problem of the normal mean are established under divergence loss. This approach is an extension of the work of Ghosh and Mergel (2009).

Keywords: Divergence loss; James-Stein estimate; Shrinkage estimate; Superharmonic function.

1 Introduction

In the line of seminal works of Stein (1956) and James and Stein (1961), there is growing interest in modifying and generalizing the latter work to bring a new estimator of shrinkage type in order to outperform the sample mean. Interesting studies may include in the couple of works done by Baranchik (1970), Efron and Morris (1973), Strawderman (1971), Faith (1978), Stein (1981), Brandwein and Strawderman (1980), Casella (1990), George (1991), Shao et al. (1994), Maruyama (2004), Srivastava and Kubokawa (2005), Ghosh et al. (2008), Wells and Zhou (2008), Ghosh and Mergel (2009) and Arashi and Tabatabey (2010) under different settings.

This work is arisen from the recent study due to Ghosh and Mergel (2009). They considerably investigated on the superiority conditions of Baranchik-type estimators over the sample mean in multivariate normal model, with divergence loss. In this paper, a minor extension is then carried out for another class of shrinkage estimators.

For the precise setup, first of all suppose that $\boldsymbol{X} \sim \mathscr{N}_p(\boldsymbol{\theta}, \sigma^2 I_p)$, where $\boldsymbol{\theta}$ and σ^2 are both unknown. Further, $S \sim (\sigma^2/(m+2))\chi^2_m$ is independent of $\boldsymbol{X}$. The aim of this work is to establish

*Corresponding author
Received: 01 November 2024 / Accepted: 15 March 2025
DOI: 10.22067/smps.2025.90474.1034

conditions in which the general class of shrinkage estimators given by

$$\boldsymbol{\delta}(\boldsymbol{X}) = \boldsymbol{X} + S\boldsymbol{g}(\boldsymbol{X}) \tag{1}$$

outperforms $\boldsymbol{X}$, where $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^p$ satisfies some regularity conditions which will be given later.

The outline of this paper is as follows: In Section 2, some preliminary results as well as some notations are given, while the main results are exhibited in Section 3. Some important remarks are also given in Section 4.

2 Preliminaries

Before revealing the main results, we express some useful notations. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$, let

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{j=1}^{p} x_j y_j, \quad \|\boldsymbol{x}\|^2 = \sum_{i=1}^{p} x_i^2.$$

Definition 1. *A function $h : \mathbb{R}^p \to \mathbb{R}$ is said to be almost differentiable if there exists a function $\blacktriangledown h : \mathbb{R}^p \to \mathbb{R}^p$ such that, for all $\boldsymbol{z} \in \mathbb{R}^p$,*

$$h(\boldsymbol{x}+\boldsymbol{z}) - h(\boldsymbol{x}) = \int_0^1 \boldsymbol{z} \cdot \blacktriangledown h(\boldsymbol{x}+t\boldsymbol{z}) dt$$

for (Lebesgue measure) almost all $\boldsymbol{x} \in \mathbb{R}^p$. A function $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^p$ is almost differentiable if all its coordinate functions are almost differentiable. Essentially, $\blacktriangledown$ is the vector differential operator of first partial derivatives with i^{th} coordinate $\triangledown_i = \partial/\partial x_i$.

Definition 2. *A function $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^p$ is said to be homogeneous of degree -1, if it satisfies*

$$\boldsymbol{g}(\lambda \boldsymbol{x}) = \frac{1}{\lambda} \boldsymbol{g}(\boldsymbol{x})$$

for all real $\lambda \neq 0$ and for all $\boldsymbol{x} \in \mathbb{R}^p$.

As an example satisfying the regularity condition in Definition 2, consider the reciprocal function $\boldsymbol{g}(\boldsymbol{x}) = (g(x_1), \ldots, g(x_p))'$ where $g(x) = x^{-1}$. Obviously, $g(\lambda x) = 1/(\lambda x) = \lambda^{-1} g(x)$.

In this paper, we employ the expectation of divergence loss, which includes Kullback-Leibler (KL for short) loss and Bhattacharyya-Hellinger (BH) loss, as a measurement. The divergence loss has been considered by many authors in other contexts. Among others, we refer to Amari (1982) and Cressie and Read (1984).

Definition 3. *Suppose that $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\theta}, \sigma^2 I_p)$ denotes the probability density function of $\mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 I_p)$. For an estimator $\boldsymbol{a}$ of $\boldsymbol{\theta}$, the divergence loss is defined by*

$$\begin{aligned} L_\beta(\boldsymbol{\theta}; \boldsymbol{a}) &= \frac{1}{\beta(1-\beta)} \left(1 - \int_{\mathbb{R}^p} \mathcal{N}^{1-\beta}(\boldsymbol{x}|\boldsymbol{\theta}, \sigma^2 I_p) \mathcal{N}^{\beta}(\boldsymbol{x}|\boldsymbol{a}, \sigma^2 I_p) d\boldsymbol{x} \right) \\ &= \frac{1}{b} \left(1 - \exp\left[-\frac{b}{2\sigma^2} \|\boldsymbol{a} - \boldsymbol{\theta}\|^2 \right] \right), \end{aligned} \tag{2}$$

where $b = \beta(1-\beta)$ and $\beta \in (0,1)$.

The second equality in (2) is a consequence of Lemma 2.2 of Ghosh et al. (2008). The KL loss and the BH loss occur as special cases of the divergence loss when $\beta \to 0$ or $\beta \to 1$, and $\beta = 1/2$, respectively. Some graphical displays are presented in Figure 1, illustrating the relative behavior of the loss function for $\sigma = 0.5$. In this figure, β varies within the range $(0,1)$, while $\|\boldsymbol{a}-\boldsymbol{\theta}\|^2$ changes from small to large values across different panes to demonstrate the effect of its magnitude. As shown, for moderate values of $\|\boldsymbol{a}-\boldsymbol{\theta}\|^2$, the shape of the loss function is clearly bath-shaped.

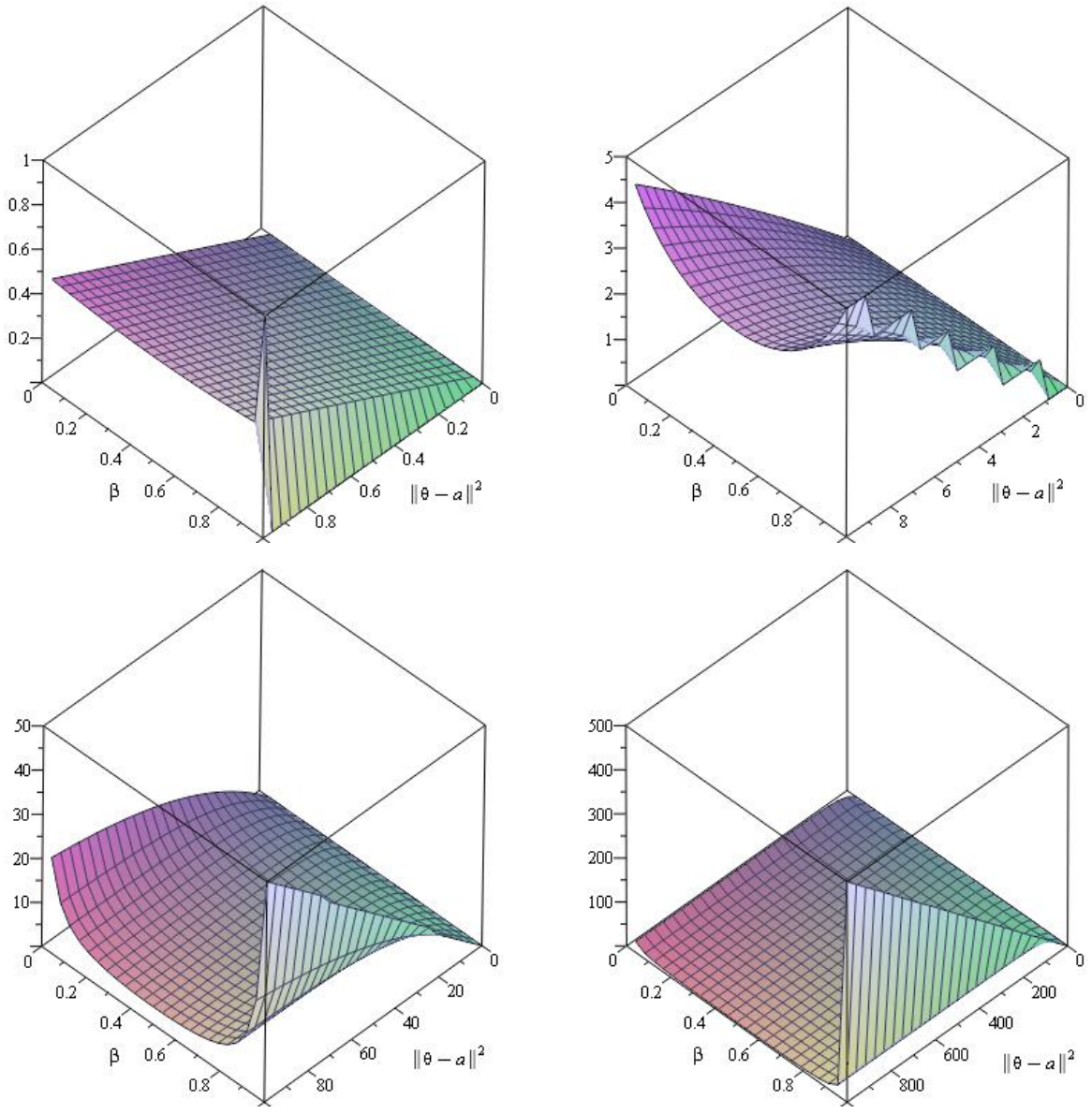

Figure 1: Behavior of the divergence loss relative to β and $\|\boldsymbol{a}-\boldsymbol{\theta}\|^2$.

To close this section, note that since $\|\boldsymbol{X}-\boldsymbol{\theta}\|^2 \sim \sigma^2\chi^2_p$, under the loss, it can be directly concluded that the risk of $\boldsymbol{X}$ is given by

$$R_\beta(\boldsymbol{\theta};\boldsymbol{X}) = \frac{1-(1+b)^{-p/2}}{b}. \tag{3}$$

3 Main results

In this section, we first give sufficient conditions in which the estimator $\boldsymbol{\delta}(\boldsymbol{X})$ given by (1) dominates $\boldsymbol{X}$ for the case of $\Sigma = \sigma^2 I_p$ (Case I). Also, for the case of the general framework Σ (Case II), we give sufficient conditions so that an estimator dominates $\boldsymbol{X}$.

Case I: ($\Sigma = \sigma^2 I_p$, σ^2 is unknown)

Theorem 1. *Assume $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 I_p)$ and $S \sim (\sigma^2/(m+2))\chi^2_m$ are independent, where $\boldsymbol{\theta}$ and σ^2 are both unknown. Further, assume that for a given estimator $\boldsymbol{\delta}(\boldsymbol{X})$ in (1), $\boldsymbol{g}$ is almost differentiable and homogenous function of degree -1 satisfying $E|(\partial/\partial W_i)g_i(\boldsymbol{W})| < \infty$ for all $\boldsymbol{\eta} \in \mathbb{R}^p$ and for $i = 1, \cdots, p$, where $\boldsymbol{W} \sim \mathcal{N}_p(\boldsymbol{\eta}, I_p/(b+1))$ is independent of S. Then, the estimator $\boldsymbol{\delta}(\boldsymbol{X})$ has smaller risk than $\boldsymbol{X}$, under the divergence loss (2), provided that*

$$\|\boldsymbol{g}(\boldsymbol{w})\|^2 + \frac{2}{b+1}\blacktriangledown \cdot \boldsymbol{g}(\boldsymbol{w}) \leq 0$$

for all $\boldsymbol{w} \in \mathbb{R}^p$.

Proof. Let $\boldsymbol{Y} = \boldsymbol{X}/\sigma$, $\boldsymbol{\eta} = \boldsymbol{\theta}/\sigma$, $S^* = S/\sigma^2$. Then, by the homogeneity of $\boldsymbol{g}$ and the inequality $e^x - e^y \geq e^y(x-y)$ $(x, y \in \mathbb{R})$, the risk difference between $\boldsymbol{X}$ and $\boldsymbol{\delta}(\boldsymbol{X})$ is given by

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{X}) - R_\beta(\boldsymbol{\theta};\boldsymbol{\delta}(\boldsymbol{X})) &= \frac{1}{b}E\left[\exp\left(-\frac{b}{2\sigma^2}\|\boldsymbol{X}+S\boldsymbol{g}(\boldsymbol{X})-\boldsymbol{\theta}\|^2\right) - \exp\left(-\frac{b}{2\sigma^2}\|\boldsymbol{X}-\boldsymbol{\theta}\|^2\right)\right] \\ &\geq -\frac{1}{2\sigma^2}E\left[\exp\left(-\frac{b}{2\sigma^2}\|\boldsymbol{X}-\boldsymbol{\theta}\|^2\right)\left(S^2\|\boldsymbol{g}(\boldsymbol{X})\|^2 + 2S(\boldsymbol{X}-\boldsymbol{\theta})\cdot\boldsymbol{g}(\boldsymbol{X})\right)\right] \\ &= -\frac{1}{2}E\left[\exp\left(-\frac{b}{2}\|\boldsymbol{Y}-\boldsymbol{\eta}\|^2\right)\left(S^{*2}\|\boldsymbol{g}(\boldsymbol{Y})\|^2 + 2S^*(\boldsymbol{Y}-\boldsymbol{\eta})\cdot\boldsymbol{g}(\boldsymbol{Y})\right)\right]. \end{aligned} \tag{4}$$

Since $\boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{\eta}, I_p)$ and $S^* \sim (m+2)^{-1}\chi^2_m$ are independent, and $E(S^*) = E\left(S^{*2}\right) = m/(m+2)$, the right hand side in (4) is rewritten by

$$\begin{aligned} &-\frac{m}{2(m+2)}E\left[\exp\left(-\frac{b}{2}\|\boldsymbol{Y}-\boldsymbol{\eta}\|^2\right)\left(\|\boldsymbol{g}(\boldsymbol{Y})\|^2 + 2(\boldsymbol{Y}-\boldsymbol{\eta})\cdot\boldsymbol{g}(\boldsymbol{Y})\right)\right] \\ &\quad = -\frac{m}{2(m+2)}(b+1)^{-p/2}E\left[\|\boldsymbol{g}(\boldsymbol{W})\|^2 + 2\boldsymbol{g}(\boldsymbol{W})\cdot(\boldsymbol{W}-\boldsymbol{\eta})\right]. \end{aligned} \tag{5}$$

By the Stein identity, we have

$$E\left[\boldsymbol{g}(\boldsymbol{W})\cdot(\boldsymbol{W}-\boldsymbol{\eta})\right] = \frac{1}{b+1}E\left[\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{W})\right]. \tag{6}$$

Substituting (5) and (6) in (4) concludes that

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{X}) - R_\beta(\boldsymbol{\theta};\boldsymbol{\delta}(\boldsymbol{X})) &\geq -\frac{m}{2(m+2)}(b+1)^{-p/2}E\left[\|\boldsymbol{g}(\boldsymbol{W})\|^2 + \frac{2}{b+1}\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{W})\right] \\ &\geq 0 \end{aligned} \tag{7}$$

for all $\boldsymbol{\theta}\in\mathbb{R}^p$ and for all $\sigma^2>0$. Here, we have to show that the inequality in (7) holds strictly for some $\boldsymbol{\theta}$ and σ^2. Since $P\{S\boldsymbol{g}(\boldsymbol{X})+2(\boldsymbol{X}-\boldsymbol{\theta})=0\}=0$, the equality in (4) holds only if $\boldsymbol{g}(\boldsymbol{X})=0$ (a.s.), that is, $\boldsymbol{\delta}(\boldsymbol{X})=\boldsymbol{X}$ (a.s.). This completes the proof. □

As a supplement, the earlier result can be proposed in a more general situation. In this regard, we have the following essential consequence.

Theorem 2. *Suppose that $\boldsymbol{X}\sim\mathcal{N}_p(\boldsymbol{\theta},\sigma^2 I_p)$ and $S\sim(\sigma^2/(m+2))\chi^2_m$ are independent, where $\boldsymbol{\theta}$ and σ^2 are both unknown. Consider the class of shrinkage estimators*

$$\boldsymbol{\delta}^*(\boldsymbol{X}) = \boldsymbol{X} + cS\boldsymbol{g}(\boldsymbol{X}),$$

where $\boldsymbol{g}$ is as in Theorem 1. Then, $\boldsymbol{\delta}^*(\boldsymbol{X})$ outperforms $\boldsymbol{X}$ under divergence loss provided that

(i) there exists a function $h(.)$ such that, for all $\boldsymbol{x}\in\mathbb{R}^p$

$$\frac{1}{2}\|\boldsymbol{g}(\boldsymbol{x})\|^2 \leq h(\boldsymbol{x}) \leq -\frac{1}{1+b}\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{x}),$$

(ii) $0<c\leq 1$.

Proof. By the same way as the proof of Theorem 1, the risk difference between $\boldsymbol{X}$ and $\boldsymbol{\delta}^*(\boldsymbol{X})$ is given by

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{X}) - R_\beta(\boldsymbol{\theta};\boldsymbol{\delta}^*(\boldsymbol{X})) &\geq -\frac{m}{2(m+2)}(b+1)^{-p/2}E\left[c^2\|\boldsymbol{g}(\boldsymbol{W})\|^2 + \frac{2c}{b+1}\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{W})\right] \\ &\geq -\frac{m}{m+2}(b+1)^{-p/2}c(c-1)E\left[h(\boldsymbol{W})\right] \\ &\geq 0 \end{aligned} \tag{8}$$

for all $c\in(0,1]$. By the same reason as Theorem 1, the equality in (8) holds for all $\boldsymbol{\theta}$ and for all σ^2 only if $\boldsymbol{\delta}(\boldsymbol{X})=\boldsymbol{X}$ (a.s.). □

Now, let $\mathscr{S}(p)$ be the set of all positive definite matrices of order p.

Case II: (Unknown $\Sigma\in\mathscr{S}(p)$)

In this case, first of all, let $n-1$ mutually independent random vectors $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_{n-1}\sim\mathcal{N}_p(\boldsymbol{0},\Sigma)$ be independent of $\boldsymbol{Z}_n\sim\mathcal{N}_p(\boldsymbol{\theta},\Sigma)$, in which $p\geq 3$. We deal with a more general type of improved shrinkage estimators, based on the sufficient statistic $(\boldsymbol{Z}_n,S)$, of the form

$$\boldsymbol{\gamma}(\boldsymbol{Z}_n,S) = \boldsymbol{Z}_n + \boldsymbol{F}(\boldsymbol{Z}_n,S), \tag{9}$$

where $S=\sum_{j=1}^{n-1}\boldsymbol{Z}_j\boldsymbol{Z}_j'$.

Definition 4. *Suppose that* $\mathscr{N}(\boldsymbol{x}|\boldsymbol{\theta},\Sigma)$ *denotes the probability density function of* $\mathscr{N}_p(\boldsymbol{\theta},\Sigma)$*. For an estimator* $\boldsymbol{a}$ *of* $\boldsymbol{\theta}$*, the generalized divergence loss (GDL) function is defined by*

$$\begin{aligned} L_\beta^\Sigma(\boldsymbol{\theta};\boldsymbol{a}) &= \frac{1}{\beta(1-\beta)}\left(1-\int_{\mathbb{R}^p}\mathscr{N}^{1-\beta}(\boldsymbol{x}|\boldsymbol{\theta},n^{-1}\Sigma)\mathscr{N}^{\beta}(\boldsymbol{x}|\boldsymbol{a},n^{-1}\Sigma)d\boldsymbol{x}\right) \\ &= \frac{1}{b}\left(1-\exp\left[-\frac{nb}{2}\|\boldsymbol{a}-\boldsymbol{\theta}\|_\Sigma^2\right]\right), \end{aligned} \tag{10}$$

where $\|\boldsymbol{a}-\boldsymbol{\theta}\|_\Sigma^2=(\boldsymbol{a}-\boldsymbol{\theta})'\Sigma^{-1}(\boldsymbol{a}-\boldsymbol{\theta})$*.*

Theorem 3. *Suppose that*

$$E\left[\boldsymbol{F}'(\boldsymbol{U},S)\boldsymbol{F}(\boldsymbol{U},S)\right]<\infty$$

for all $\boldsymbol{\theta}\in\mathbb{R}^p$ *and for all* $\Sigma\in\mathscr{S}(p)$*, where* $\boldsymbol{U}\sim\mathscr{N}_p(\boldsymbol{\theta},\Sigma/(nb+1))$ *is independent of* S*. Then, the estimator* $\boldsymbol{\gamma}(\boldsymbol{Z}_n,S)$ *given by* (9) *has smaller risk than* $\boldsymbol{Z}_n$*, under GDL function given by* (10) *provided that*

$$2\blacktriangledown_{\boldsymbol{u}}\cdot\boldsymbol{F}(\boldsymbol{u},S)+(n-p+2)(nb+1)\boldsymbol{F}'(\boldsymbol{u},S)S^{-1}\boldsymbol{F}(\boldsymbol{u},S)\le 0,$$

where $\blacktriangledown_{\boldsymbol{u}}$ *states getting derivative with respect to* $\boldsymbol{u}$*.*

Proof. By the same way as the proof of Theorem 1, the risk difference between $\boldsymbol{Z}_n$ and $\boldsymbol{\gamma}(\boldsymbol{Z}_n,S)$ is given by

$$R_\beta(\boldsymbol{\theta};\boldsymbol{Z}_n)-R_\beta(\boldsymbol{\theta};\boldsymbol{\gamma}(\boldsymbol{Z}_n,S))\ge-\frac{n}{2}E\left[\left(\|\boldsymbol{F}(\boldsymbol{Z}_n,S)\|_\Sigma^2+2\boldsymbol{F}'(\boldsymbol{Z}_n,S)\Sigma^{-1}(\boldsymbol{Z}_n-\boldsymbol{\theta})\right)e^{\left(-\frac{nb}{2}\|\boldsymbol{Z}_n-\boldsymbol{\theta}\|_\Sigma^2\right)}\right]. \tag{11}$$

Let $\boldsymbol{V}_i=\boldsymbol{Z}_i/\sqrt{nb+1}$ for $i=1,\cdots,n-1$, $T=\sum_{j=1}^{n-1}\boldsymbol{V}_j\boldsymbol{V}_j'$, $\boldsymbol{G}(\boldsymbol{u},T)=\boldsymbol{F}(\boldsymbol{u},(nb+1)T)$ and $\Lambda=\Sigma/(nb+1)$. Then, the right hand side in (11) is rewritten by

$$-\frac{n}{2}(nb+1)^{-1-p/2}E\left[\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}\boldsymbol{G}(\boldsymbol{U},T)+2\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}(\boldsymbol{U}-\boldsymbol{\theta})\right].$$

By applying Lemma 1 of Fourdrinier et al. (2003), we see that

$$\begin{aligned} &E\left[\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}\boldsymbol{G}(\boldsymbol{U},T)+2\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}(\boldsymbol{U}-\boldsymbol{\theta})\right] \\ &\quad\le E\left[2\blacktriangledown_{\boldsymbol{U}}\cdot\boldsymbol{G}(\boldsymbol{U},T)+(n-p+2)\boldsymbol{G}(\boldsymbol{U},T)'T^{-1}\boldsymbol{G}(\boldsymbol{U},T)\right]. \end{aligned} \tag{12}$$

Since $T=S/(nb+1)$, the right hand side in (12) is expressed as

$$E\left[2\blacktriangledown_{\boldsymbol{U}}\cdot\boldsymbol{F}(\boldsymbol{U},S)+(n-p+2)(nb+1)\boldsymbol{F}'(\boldsymbol{U},S)S^{-1}\boldsymbol{F}(\boldsymbol{U},S)\right]. \tag{13}$$

From (11) to (13), we have

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{Z}_n)-R_\beta(\boldsymbol{\theta};\boldsymbol{\gamma}(\boldsymbol{Z}_n,S)) &\ge -\frac{n}{2}(nb+1)^{-1-p/2} \\ &\quad\times E\left[2\blacktriangledown_{\boldsymbol{U}}\cdot\boldsymbol{F}(\boldsymbol{U},S)+(n-p+2)(nb+1)\boldsymbol{F}'(\boldsymbol{U},S)S^{-1}\boldsymbol{F}(\boldsymbol{U},S)\right] \\ &\ge 0 \end{aligned} \tag{14}$$

for all $\boldsymbol{\theta}\in\mathbb{R}^p$ and $\Sigma\in\mathscr{S}(p)$. By the same reason as Theorem 1, the equality in (14) holds only if $\boldsymbol{F}(\boldsymbol{Z}_n,S)=0$ (a.s.). □

4 Conclusions

In this paper, we establish the conditions under which a general class of shrinkage estimators outperforms the consistent estimator of a multivariate normal population mean when using a divergence loss function. The results obtained apply to both known and unknown covariance scenarios. Furthermore, our findings extend the earlier work of Ghosh and Mergel (2009) to a broader class of dominant estimators. The proofs of the proposed theorems are presented in a more straightforward manner than in the previous reference. This method can also be applied to exponential-type loss functions, such as LINEX, and reflected normal losses. Importantly, the conditions for superiority are robust concerning the squared error loss function, which further validates our derivations, as the divergence loss encompasses the square error loss as a special case.

4.1 Easy understanding

In conclusion, we provide a simple approach for making decisions regarding the superiority conditions based on divergence loss. As previously mentioned, the square error loss is a specific case of divergence loss. It is important to note that the graphs of both losses are parabolic. Additionally, the graph of the risk of $\mathbf{X}$ in Equation (3) also follows a parabolic form (see Figure 2). From a graphical perspective, for any estimator of $\boldsymbol{\theta}$ to be superior to $\mathbf{X}$, it must have a risk that is lower than that of the parabolic shape. Therefore, we can conclude that determining the superiority conditions for the proposed shrinkage estimators can be achieved by examining them under the square error loss. Interestingly, the conditions derived in this study align exactly with those found in the work of Ghosh and Mergel (2009), as they are the same under square error loss when $b \to 0$.

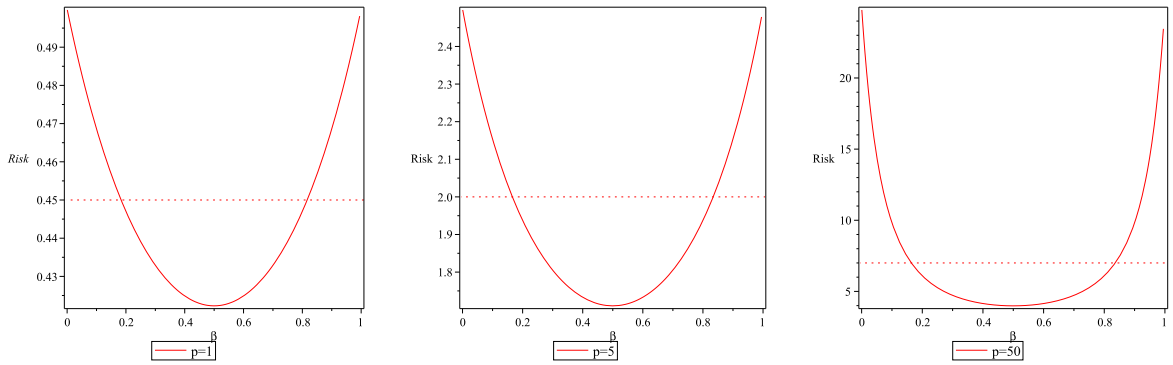

Figure 2: Risk of $\boldsymbol{X}$ under the divergence loss.

Acknowledgments

We would like to express our gratitude to the two anonymous reviewers for their constructive comments, which prompted us to include many additional details in the paper and significantly

improve its presentation. Mohammad Arashi's work is based on the research supported in part by the Iran National Science Foundation (INSF) grant No. 4015320.

References

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

Arashi, M. and Tabatabaey, S. M. M. (2010). A note on classical Stein-type estimators in elliptically contoured models. *Journal of Statistical Planning and Inference*, **140**, 1206–1213.

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *The Annals of Mathematical Statistics*, **41**, 642–645.

Brandwein, A. C. and Strawderman, W. E. (1980). Minimax estimators of location parameters for spherically symmetric distributions with concave loss. *The Annals of Statistics*, **8**, 279–284.

Casella, G. (1990). Estimators with nondecreasing risk: Application of a chi-square identity. *Statistics and Probability Letters*, **10**, 107–109.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **46**, 440–464.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors-An empirical Bayes approach. *Journal of the American Statistical Association*, **68**, 117–130.

Faith, R. E. (1978). Minimax Bayes estimators of a multivariate normal mean. *Journal of Multivariate Analysis*, **8**, 372–379.

Fourdrinier, D., Strawderman, W. E. and Wells, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. *Journal of Multivariate Analysis*, **85**, 24–39.

George, E. I. (1991). Shrinkage domination in a multivariate common mean problem. *The Annals of Statistics*, **19**, 952–960.

Ghosh, M., Mergel, V. and Datta, G. S. (2008). Estimation, prediction and the Stein phenomenon under divergence loss, *Journal of Multivariate Analysis*, **99**, 1941–1961.

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

James, W. and Stein, C. (1961). Estimation of quadratic loss. *In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 361–379.

Maruyama, Y. (2004). Stein's idea and minimax admissible estimation of a multivariate normal mean, *Journal of Multivariate Analysis*, **88**, 320–334.

Shao, P. Yi-Shi and Strawderman, W. E. (1994). Improving on the James-Stein positive-part estimator. *The Annals of Statistics*, **22**, 1517–1538.

Srivastava, M. S. and Kubokawa, T. (2005). Minimax multivariate empirical Bayes estimators under multicollinearity. *Journal of Multivariate Analysis*, **93**, 394–416.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, **9**, 1135–1151.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. *In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 197–206.

Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *The Annals of Mathematical Statistics*, **42**, 385–388.

Wells, M. T. and Zhou, G. (2008). Generalized Bayes minimax estimators of the mean of multivariate normal distribution with unknown variance. *Journal of Multivariate Analysis*, **99**, 2208–2220.