IMS Collections
Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman
Vol. 8 (2012) 24–41

DOI: 10.1214/11-IMSCOLL802

Minimax estimation of linear combinations of restricted location parameters

Tatsuya Kubokawa

University of Tokyo

Abstract: The estimation of a linear combination of several restricted location parameters is addressed from a decision-theoretic point of view. A benchmark estimator of the linear combination is an unbiased estimator, which is minimax, but inadmissible relative to the mean squared error. An interesting issue is what is a prior distribution which results in the generalized Bayes and minimax estimator. Although it seems plausible that the generalized Bayes estimator against the uniform prior over the restricted space should be minimax, it is shown to be not minimax when the number of the location parameters, k, is more than or equal to three, while it is minimax for $k = 1$. In the case of $k = 2$, a necessary and sufficient condition for the minimaxity is given, namely, the minimaxity depends on signs of coefficients of the linear combination. When the underlying distributions are normal, we can obtain a prior distribution which results in the generalized Bayes estimator satisfying minimaxity and admissibility. Finally, it is demonstrated that the estimation of ratio of normal variances converges to the estimation of difference of the normal positive means, which gives a motivation of the issue studied here.

Contents

1. Introduction

The point estimation of restricted parameters has been studied from a decision-theoretic point of view since Katz (1961) showed that the generalized Bayes estimator of a restricted parameter is minimax and admissible in a one-parameter

Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan e-mail: tatsuya@e.u-tokyo.ac.jp

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: admissibility, decision theory, generalized Bayes estimator, minimaxity, restricted parameters, Stein estimation

exponential family. Farrell (1964) established the minimaxity and admissibility in the general location family. This classical problem was revisited by Marchand and Strawderman (2004, 2005) and Kubokawa (1990, 2004). Hartigan (2004) considered the simultaneous estimation of a mean vector restricted to a convex cone in a k-variate normal distribution and used the Gauss divergence theorem to show that the generalized Bayes estimator against the uniform prior dominates the unbiased estimator. Tsukuma and Kubokawa (2008) established the minimaxity of the generalized Bayes estimator and proved that it is admissible for $k = 1, 2$ and inadmissible for $k \geq 3$, which is an extension of the Stein result. For a good survey of admissibility, see Rukhin (1995).

In this paper, we consider the estimation of the linear combination of the several location parameters where each location parameter is restricted to be positive. More specifically, we consider the following simple model: Let $X_1, \ldots, X_k$ be mutually independent random variables where X_i has probability density function $f_i(x_i - \mu_i)$ with location parameter μ_i such that $\mu_i > 0$ for $i = 1, \ldots, k$. It is assumed that $E[X_i^2] < \infty$. In matrix notation, let $\boldsymbol{X} = (X_1, \ldots, X_k)^t$, $\boldsymbol{x} = (x_1, \ldots, x_k)^t$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^t$ where $\boldsymbol{X}^t$ denotes the transpose of $\boldsymbol{X}$. Then, the joint density of $\boldsymbol{X}$ is denoted by

$$f(\boldsymbol{x} - \boldsymbol{\mu}) = \prod_{i=1}^{k} f_i(x_i - \mu_i), \tag{1.1}$$

and $\boldsymbol{\mu}$ is restricted on the space,

$$D = \{\boldsymbol{\mu} \mid \mu_i > 0,\ i = 1, \ldots, k\}.$$

For real constants a_i's and $\boldsymbol{a} = (a_1, \ldots, a_k)^t$, consider a linear combination of $\boldsymbol{\mu}$ given by

$$\theta = \sum_{i=1}^{k} a_i \mu_i = \boldsymbol{a}^t \boldsymbol{\mu}.$$

We study the estimation of θ in a decision-theoretic framework, where an estimator $\hat{\theta}$ of θ is evaluated by the mean squared error $R(\boldsymbol{\mu}, \hat{\theta}) = E[(\hat{\theta} - \theta)^2]$.

An unbiased estimator of θ is given by

$$\hat{\theta}^U = \sum_{i=1}^{k} a_i \widehat{\mu}_i^U, \tag{1.2}$$

where $\widehat{\mu}_i^U$ is the unbiased estimator of μ_i given by

$$\widehat{\mu}_i^U = X_i - c_i, \quad \text{for} \quad c_i = E[X_i - \mu_i].$$

As shown in Section 2, $\hat{\theta}^U$ is minimax, but inadmissible because of the restriction on the parameter $\boldsymbol{\mu}$. Thus, it is of interest to obtain the admissible and minimax estimator of θ. To this end, consider the uniform prior

$$\pi(\boldsymbol{\mu}) \mathrm{d}\boldsymbol{\mu} = \mathrm{d}\boldsymbol{\mu} I(\boldsymbol{\mu} \in D), \tag{1.3}$$

where $\mathrm{d}\boldsymbol{\mu} = \prod_{i=1}^{k} \mathrm{d}\mu_i$ and $I(\boldsymbol{\mu} \in D)$ is the indicator function such that $I(\boldsymbol{\mu} \in D) = 1$ if $\boldsymbol{\mu} \in D$, $I(\boldsymbol{\mu} \in D) = 0$ otherwise. The resulting generalized Bayes estimator of θ is

$$\begin{aligned} \hat{\theta}^{GB} &= \int_D \boldsymbol{a}^t \boldsymbol{\mu} f(\boldsymbol{X} - \boldsymbol{\mu}) \mathrm{d}\boldsymbol{\mu} \Big/ \int_D f(\boldsymbol{X} - \boldsymbol{\mu}) \mathrm{d}\boldsymbol{\mu} \\ &= \sum_{i=1}^{k} a_i \int_0^\infty \mu_i f_i(X_i - \mu_i) \mathrm{d}\mu_i \Big/ \int_0^\infty f_i(X_i - \mu_i) \mathrm{d}\mu_i, \end{aligned} \tag{1.4}$$

and our first concern is whether $\hat{\theta}^{GB}$ is minimax or not. We investigate this problem in Section 3 and show that $\hat{\theta}^{GB}$ is not minimax for $k \geq 3$, but minimax for $k = 1$. The minimaxity in the case of $k = 2$ depends on the signs of the coefficients a_1 and a_2, and a necessary and sufficient condition for the minimaxity of $\hat{\theta}^{GB}$ is that $a_1 a_2 \leq 0$. This means that, for example, the generalized Bayes estimator $\hat{\theta}^{GB}$ is not minimax in the estimation of the sum $\mu_1 + \mu_2$, but minimax in the estimation of the difference $\mu_1 - \mu_2$.

Concerning minimaxity of the generalized Bayes estimator against the uniform prior, it is interesting to note that we have different answers for the simultaneous estimation of $\boldsymbol{\mu}$ and the estimation of the linear combination $\boldsymbol{a}^t\boldsymbol{\mu}$. Namely, $\widehat{\boldsymbol{\mu}}^{GB} = \int_D \boldsymbol{\mu} f(\boldsymbol{X} - \boldsymbol{\mu}) \mathrm{d}\boldsymbol{\mu} / \int_D f(\boldsymbol{X} - \boldsymbol{\mu}) \mathrm{d}\boldsymbol{\mu}$ is always minimax for the simultaneous estimation of $\boldsymbol{\mu}$ under a quadratic loss, while $\hat{\theta}^{GB}$ is not necessarily minimax and its minimaxity depends on the dimension of $\boldsymbol{\mu}$.

In Section 4, we focus on the normal distributions, and suggest a specific prior distribution such that the resulting generalized Bayes estimator is minimax and admissible. In Section 5, we use the arguments as in Rukhin (1992) to show that the estimation of ratio of normal variances asymptotically reduces to the estimation of difference of positive normal means, which gives a motivation of the estimation problem studied here.

2. Minimaxity and inadmissibility of the unbiased estimator

In this section, we show that the unbiased estimator $\hat{\theta}^U$ given in (1.2) is minimax, but inadmissible under the assumption that $E[X_i^2] < \infty$ for $i = 1, \ldots, k$. The minimaxity of $\hat{\theta}^U$ can be verified by using similar arguments as in Girshick and Savage (1951).

Proposition 2.1 (minimaxity of the unbiased estimator). *The unbiased estimator $\hat{\theta}^U$ of $\theta = \sum_{i-1}^k a_i \mu_i$ is minimax in the estimation of the restricted parameters on D, and the risk function $R_0 = R(\boldsymbol{\mu}, \hat{\theta}^U)$ is a constant.*

Proof. Let $D_m = \{\boldsymbol{\mu} |\ 0 < \mu_i < m,\ i = 1, \ldots, k\}$ for $m = 1, 2, \ldots$, and consider the sequence of prior distributions given by

$$\pi_m(\boldsymbol{\mu}) = \begin{cases} m^{-k} & \text{if } \boldsymbol{\mu} \in D_m \\ 0 & \text{otherwise,} \end{cases}$$

which yields the Bayes estimators

$$\hat{\theta}_m^\pi = \hat{\theta}_m^\pi(\boldsymbol{X}) = \int_{D_m} \boldsymbol{a}^t \boldsymbol{u} f(\boldsymbol{X} - \boldsymbol{u}) \mathrm{d}\boldsymbol{u} \int_{D_m} f(\boldsymbol{X} - \boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

with the Bayes risk function

$$r_m(\pi_m, \hat{\theta}_m^\pi) = \frac{1}{m^k} \int_{D_m} \int \left\{ \hat{\theta}_m^\pi(\boldsymbol{x}) - \boldsymbol{a}^t \boldsymbol{\mu} \right\}^2 f(\boldsymbol{x} - \boldsymbol{\mu}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{\mu}. \tag{2.1}$$

Since $r_m(\pi_m, \hat{\theta}_m^\pi) \leq r_m(\pi_m, \hat{\theta}^U) = R_0$, it is sufficient to show that $\liminf_{m \to \infty} r_m(\pi_m, \hat{\theta}_m^\pi) \geq R_0$. Making the transformations $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{\mu}$ and $\boldsymbol{t} = \boldsymbol{u} - \boldsymbol{\mu}$

with $\mathrm{d}\boldsymbol{z} = \mathrm{d}\boldsymbol{x}$ and $\mathrm{d}\boldsymbol{t} = \mathrm{d}\boldsymbol{u}$ gives that

$$\begin{aligned}\hat{\theta}_m^\pi(\boldsymbol{x}) - \boldsymbol{a}^t\boldsymbol{\mu} &= \hat{\theta}_m^\pi(\boldsymbol{z}+\boldsymbol{\mu}) - \boldsymbol{a}^t\boldsymbol{\mu} \\ &= \int_{D_m} \boldsymbol{a}^t(\boldsymbol{u}-\boldsymbol{\mu}) f(\boldsymbol{z}+\boldsymbol{\mu}-\boldsymbol{u})\mathrm{d}\boldsymbol{u} \Big/ \int_{D_m} f(\boldsymbol{z}+\boldsymbol{\mu}-\boldsymbol{u})\mathrm{d}\boldsymbol{u} \\ &= \int_{\boldsymbol{t}+\boldsymbol{\mu}\in D_m} \boldsymbol{a}^t\boldsymbol{t} f(\boldsymbol{z}-\boldsymbol{t})\mathrm{d}\boldsymbol{t} \Big/ \int_{\boldsymbol{t}+\boldsymbol{\mu}\in D_m} f(\boldsymbol{z}-\boldsymbol{t})\mathrm{d}\boldsymbol{t}. \end{aligned} \tag{2.2}$$

Making the transformation $\xi_i = (2/m)(\mu_i - m/2)$ with $\mathrm{d}\boldsymbol{\xi} = (2/m)^k \mathrm{d}\boldsymbol{\mu}$ for $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)^t$, we can rewrite the condition $0 < \mu_i < m$ as $|\xi_i| < 1$. Also the condition that $0 < t_i + \mu_i < m$ for $\boldsymbol{t} = (t_1, \dots, t_k)^t$ is expressed by the inequality $-(m/2)(\xi_i + 1) < t_i < (m/2)(1-\xi_i)$. Let $D_m^* = \{\boldsymbol{t} | -(m/2)(\xi_i+1) < t_i < (m/2)(1-\xi_i)\}$. Then the transformations are used in (2.2) and (2.1) to obtain that

$$\hat{\theta}_m^\pi(\boldsymbol{x}) - \boldsymbol{a}^t\boldsymbol{\mu} = \int_{D_m^*} \boldsymbol{a}^t\boldsymbol{t} f(\boldsymbol{z}-\boldsymbol{t})\mathrm{d}\boldsymbol{t} \Big/ \int_{D_m^*} f(\boldsymbol{z}-\boldsymbol{t})\mathrm{d}\boldsymbol{t} \equiv \hat{\theta}_m^*(\boldsymbol{z}|\boldsymbol{\xi}), \tag{2.3}$$

and

$$r_m(\pi_m, \hat{\theta}_m^\pi) = \frac{1}{2^k}\int_{|\xi_i|<1, i=1,\dots,k} \int \left\{\hat{\theta}_m^*(\boldsymbol{z}|\boldsymbol{\xi})\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{\xi}.$$

For a small $\varepsilon > 0$, it is observed that

$$r_m(\pi_m, \hat{\theta}_m^\pi) \ge \frac{1}{2^k}\int_{|\xi_i|<1-\varepsilon, i=1,\dots,k} \int \left\{\hat{\theta}_m^*(\boldsymbol{z}|\boldsymbol{\xi})\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{\xi}.$$

The range of $\boldsymbol{t}$ in the integrals in $\hat{\theta}_m^*(\boldsymbol{z}|\boldsymbol{\xi})$ given by (2.3) is $D_m^* = \{\boldsymbol{t}| -(m/2)(\xi_i+1) < t_i < (m/2)(1-\xi_i)\}$. Since $|\xi_i| < 1-\varepsilon$, it is noted that $1-\xi_i > 1-(1-\varepsilon) = \varepsilon > 0$ and $1+\xi_i > 1 + (-1+\varepsilon) = \varepsilon > 0$. These inequalities imply that the end point $(m/2)(1-\xi_i)$ tends to infinity as $m \to \infty$. Thus $\hat{\theta}_m^*(\boldsymbol{z}|\boldsymbol{\xi})$ converges to $\hat{\theta}^U(\boldsymbol{z})$. Using the Fatou lemma, we obtain that

$$\begin{aligned}\liminf_{m\to\infty} r_m(\pi_m, \hat{\theta}_m^\pi) &\ge \liminf_{m\to\infty} \frac{1}{2^k}\int_{|\xi_i|<1-\varepsilon, i=1,\dots,k} \int \left\{\hat{\theta}_m^*(\boldsymbol{z}|\boldsymbol{\xi})\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{\xi} \\ &\ge \frac{1}{2^k}\int_{|\xi_i|<1-\varepsilon, i=1,\dots,k} \int \left\{\liminf_{m\to\infty} \hat{\theta}_m^*(\boldsymbol{z}|\boldsymbol{\xi})\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{\xi} \\ &= \frac{1}{2^k}\int_{|\xi_i|<1-\varepsilon, i=1,\dots,k} \mathrm{d}\boldsymbol{\xi} \int \left\{\hat{\theta}^U(\boldsymbol{z})\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \\ &= (1-\varepsilon)^k R(\mu, \hat{\theta}^U) = (1-\varepsilon)^k R_0 \end{aligned}$$

From the arbitrariness of $\varepsilon > 0$, it follows that $\liminf_{m\to\infty} r_m(\pi_m, \hat{\theta}_m^\pi) \ge R_0$, completing the proof of Proposition 2.1. □

Proposition 2.1 is an extension of the results of Marchand and Strawderman (2005) and Kubokawa (2004) who treated the case of $k = 1$.

Since the unbiased estimator $\widehat{\mu}_i^U = X_i - c_i$ of the positive parameter μ_i takes a negative value with a positive probability for $i = 1, \dots, k$, it is plausible that $\hat{\theta}^U = \sum_{i=1}^k a_i \widehat{\mu}_i^U$ can be improved on by a truncated procedure. Let Λ_+ and Λ_- be subsets of $\{1, \dots, k\}$ such that

$$a_i > 0 \quad \text{if} \quad i \in \Lambda_+, \quad \text{and} \quad a_j < 0 \quad \text{if} \quad j \in \Lambda_-. \tag{2.4}$$

Then θ and $\hat{\theta}^U$ are decomposed as

$$
\begin{aligned}
\theta &= \theta_+ - \theta_- \quad \text{for} \quad \theta_+ = \sum_{i\in\Lambda_+} a_i\mu_i \quad \text{and} \quad \theta_- = -\sum_{i\in\Lambda_-} a_i\mu_i, \\
\hat{\theta}^U &= \hat{\theta}^U_+ - \hat{\theta}^U_- \quad \text{for} \quad \hat{\theta}^U_+ = \sum_{i\in\Lambda_+} a_i\widehat{\mu}^U_i \quad \text{and} \quad \hat{\theta}^U_- = -\sum_{i\in\Lambda_-} a_i\widehat{\mu}^U_i.
\end{aligned} \tag{2.5}
$$

Since θ_+ and θ_- are positive, it is reasonable to truncate $\hat{\theta}^U_+$ and $\hat{\theta}^U_-$ at zero, namely, $\hat{\theta}^{TR}_+ = \max\{\hat{\theta}^U_+, 0\}$ and $\hat{\theta}^{TR}_- = \max\{\hat{\theta}^U_-, 0\}$, which results in the truncated estimator

$$
\hat{\theta}^{TR} = \hat{\theta}^{TR}_+ - \hat{\theta}^{TR}_-.
$$

Proposition 2.2 (inadmissibility of the unbiased estimator). *The truncated estimator $\hat{\theta}^{TR}$ dominates the unbiased estimator $\hat{\theta}^U$, namely $\hat{\theta}^{TR}$ is minimax.*

Proof. Noting that $\hat{\theta}^U_+$ and $\hat{\theta}^U_-$ are mutually independent, we can write the risk difference $\Delta_k = E[(\hat{\theta}^U - \theta)^2] - E[(\hat{\theta}^{TR} - \theta)^2]$ as

$$
\begin{aligned}
\Delta_k = {} & E[(\hat{\theta}^U_+ - \theta_+)^2 - (\hat{\theta}^{TR}_+ - \theta_+)^2] + E[(\hat{\theta}^U_- - \theta_-)^2 - (\hat{\theta}^{TR}_- - \theta_-)^2] \\
& + 2E[\hat{\theta}^{TR}_+ - \theta_+]E[\hat{\theta}^{TR}_- - \theta_-].
\end{aligned}
$$

It can be seen that $(\hat{\theta}^U_+ - \theta_+)^2 - (\hat{\theta}^{TR}_+ - \theta_+)^2 = \hat{\theta}^U_+(\hat{\theta}^U_+ - 2\theta_+)I(\hat{\theta}^U_+ < 0) > 0$ where $I(A)$ is the indicator function such that $I(A) = 1$ if A is true, $I(A) = 0$ otherwise. Also, $E[\hat{\theta}^{TR}_+ - \theta_+] = E[\max\{\hat{\theta}^U_+, 0\} - \theta_+] = E[\hat{\theta}^U_+ - \theta_+ + \max\{0, -\hat{\theta}^U_+\}] = E[\max\{0, -\hat{\theta}^U_+\}] \geq 0$. These observations show that $\Delta_k > 0$ for any $\boldsymbol{\mu} \in D$. □

3. Is the uniform prior Bayes estimator minimax?

We now investigate whether the generalized Bayes estimator $\hat{\theta}^{GB}$ for the uniform prior over D is minimax or not. As shown below, the minimaxity depends on the dimension k of the location vector $\boldsymbol{\mu}$.

3.1. Minimaxity in the case of $k = 1$

Let X be a random variable whose density function is given by $f(x - \mu)$ where μ is a location parameter restricted on the space $\{\mu \in \boldsymbol{R} | \mu > 0\}$. The unbiased estimator of μ is $\widehat{\mu}^U = X - c_0$ for $c_0 = E[X - \mu] = \int uf(u)\mathrm{d}u$, which is minimax. We first consider a class of estimators of the form

$$
\widehat{\mu}(\phi) = X - \phi(X)
$$

for an absolutely continuous function $\phi(\cdot)$, and derive sufficient conditions on $\phi(\cdot)$ for the minimaxity. From Kubokawa (1994a, 1999, 2004), we can obtain an integral expression for the risk difference of two estimators $\widehat{\mu}^U$ and $\widehat{\mu}(\phi)$.

Lemma 3.1. *Assume that $\phi(\cdot)$ is an absolutely continuous function such that $\lim_{w\to\infty}\phi(w) = c_0$. Then, the difference of the risk functions of $\widehat{\mu}^U$ and $\widehat{\mu}(\phi)$ can be written as*

$$
\begin{aligned}
\Delta &\equiv R(\mu, \widehat{\mu}^U) - R(\mu, \widehat{\mu}(\phi)) \\
&= -2\int \Big\{ \int_{-\infty}^{w} uf(u)\mathrm{d}u - \phi(w+\mu)\int_{-\infty}^{w} f(u)\mathrm{d}u \Big\} \phi'(w+\mu)\mathrm{d}w.
\end{aligned} \tag{3.1}
$$

Proof. Since $\lim_{w\to\infty}\phi(w)=c_0$, it can be seen that

$$\Delta = E[[(X-\phi(X+t)-\mu)^2]_{t=0}^{\infty}] = E[\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}(X-\phi(X+t)-\mu)^2\mathrm{d}t],$$

which can be rewritten as

$$\Delta = -2\int\int_0^\infty \{x-\phi(x+t)-\mu\}\phi'(x+t)\mathrm{d}tf(x-\mu)\mathrm{d}x.$$

Making the transformations $w=x+t-\mu$ and $u=w-t$ with $\mathrm{d}w=\mathrm{d}x$ and $\mathrm{d}u=-\mathrm{d}t$ in turn gives

$$\begin{aligned}\Delta =& -2\int\int_0^\infty \{w-t-\phi(w+\mu)\}\phi'(w+\mu)f(w-t)\mathrm{d}t\mathrm{d}w\\ =& -2\int\int_{-\infty}^{w}\{u-\phi(w+\mu)\}\phi'(w+\mu)f(u)\mathrm{d}u\mathrm{d}w,\end{aligned}$$

which yields (3.1). □

Lemma 3.1 provides a class of estimators improving on $\widehat{\mu}^U$.

Proposition 3.1. *Assume that $\phi(\cdot)$ is an absolutely continuous function such that (a) $\phi(w)$ is nondecreasing in w, $\lim_{w\to\infty}\phi(w)=c_0$, and (b) $\phi(w)\geq\phi^{GB}(w)$, where*

$$\phi^{GB}(w)=\int_{-\infty}^{w}uf(u)\mathrm{d}u/\int_{-\infty}^{w}f(u)\mathrm{d}u.$$

Then the estimator $\widehat{\mu}(\phi)$ dominates $\widehat{\mu}^U$, namely $\widehat{\mu}(\phi)$ is minimax.

It is easy to see that the function $\phi^{GB}(w)$ is nondecreasing and $\lim_{w\to\infty}\phi^{GB}(w)=c_0$. Since $\phi^{GB}(w)\leq w$, it is also seen that $\phi^{GB}(w)\leq\phi^{TR}(w)=\min\{w,c_0\}$. Thus, $\phi^{GB}(w)$ and $\phi^{TR}(w)$ satisfy the conditions in Proposition 3.1, and we get the improved estimators

$$\begin{aligned}&\widehat{\mu}^{GB}=X-\phi^{GB}(X)=\int_0^\infty \mu f(X-\mu)\mathrm{d}\mu/\int_0^\infty f(X-\mu)\mathrm{d}\mu,\\ &\widehat{\mu}^{TR}=X-\phi^{TR}(X)=\max\{X,0\}.\end{aligned}$$

Note that $\widehat{\mu}^{GB}$ is the generalized Bayes estimator of μ against the uniform prior $\mathrm{d}\mu$ over the space of $\mu>0$, and $\widehat{\mu}^{TR}$ is the maximum likelihood estimator of μ.

It can be easily seen that $\lim_{\mu\to\infty}R(\mu,\widehat{\mu}^{GB})=R_0=R(\mu,\widehat{\mu}^U)$. Also from Lemma 3.1, we get the following risk property for the generalized Bayes estimator $\widehat{\mu}^{GB}$.

Proposition 3.2. *Both estimators $\widehat{\mu}^{GB}$ and $\widehat{\mu}^U$ have the same risk at $\mu=0$, namely, $R_0=R(0,\widehat{\mu}^U)=R(0,\widehat{\mu}^{GB})$. Also, $R(\mu,\widehat{\mu}^{GB})$ converges to R_0 as $\mu\to\infty$.*

3.2. Minimaxity and non-minimaxity in the case of $k=2$

Let X_1 and X_2 be two mutually independent random variables whose densities are $f_1(x_1-\mu_1)$ and $f_2(x_2-\mu_2)$, respectively, where μ_1 and μ_2 are unknown location parameters, $\mu_1>0$ and $\mu_2>0$. Let us consider the problem of estimating the linear combination of μ_1 and μ_2, namely,

$$\theta=\theta_{a_1,a_2}=a_1\mu_1+a_2\mu_2,$$

where a_1 and a_2 are known real constants. From the results in the previous subsection, it can be guessed that the generalized Bayes estimator $\hat{\theta}^{GB}$ of θ against the uniform prior $\mathrm{d}\mu_1\mathrm{d}\mu_2$ over the space of $\mu_1 > 0$ and $\mu_2 > 0$ improves on the unbiased estimator $\hat{\theta}^U = a_1\widehat{\mu}_i^U + a_2\widehat{\mu}_2^U$ in terms of the mean squares error, $R(\mu_1, \mu_2, \hat{\theta}^U) = E[(\hat{\theta}^U - \theta)^2]$. Here $\widehat{\mu}_i^U = X_i - c_i$ and $c_i = E[X_i - \mu_i]$ for $i = 1, 2$. However, this conjecture is not true. As shown below, the condition for the minimaxity of $\hat{\theta}^{GB}$ depends on signs of a_1 and a_2.

In general, let us consider a class of estimators of the form $\hat{\theta}(\phi_1, \phi_2) = a_1\widehat{\mu}_1(\phi_1) + a_2\widehat{\mu}_2(\phi_2)$, where $\widehat{\mu}_i(\phi_i) = X_i - \phi_i(X_i)$ for $i = 1, 2$ and $\phi_i(\cdot)$ is an absolutely continuous function.

Lemma 3.2. *The risk difference of the estimators $\hat{\theta}^U$ and $\hat{\theta}(\phi_1, \phi_2)$ is written as*

$$
\begin{aligned}
R(\mu_1,&\mu_2, \hat{\theta}^U) - R(\mu_1, \mu_2, \hat{\theta}(\phi_1, \phi_2)) \\
&= a_1^2\{R(\mu_1, \widehat{\mu}_1^U) - R(\mu_1, \widehat{\mu}_1(\phi_1))\} + a_2^2\{R(\mu_2, \widehat{\mu}_2^U) - R(\mu_2, \widehat{\mu}_2(\phi_2))\} \\
&\quad - 2a_1a_2E[\widehat{\mu}_1(\phi_1) - \mu_1]E[\widehat{\mu}_2(\phi_2) - \mu_2].
\end{aligned}
$$

It is noted that $E[\widehat{\mu}_i(\phi_i) - \mu_i] = E[X_i - \mu_i - \phi_i(X_i)] = c_i - E[\phi_i(X_i)]$. If $\phi_i(w)$ is a nondecreasing function with $\lim_{w\to\infty}\phi_i(w) = c_i$, then it can be seen that $E[\widehat{\mu}_i(\phi_i) - \mu_i] \geq 0$. Hence from Proposition 3.1 and Lemma 3.2, we get the following proposition.

Proposition 3.3. *For $i = 1, 2$, assume that $\phi_i(\cdot)$ is an absolutely continuous function such that (a) $\phi_i(w)$ is nondecreasing in w and $\lim_{w\to\infty}\phi_i(w) = c_0$, and (b) $\phi_i(w) \geq \phi_i^{GB}(w)$, where*

$$
\phi_i^{GB}(w) = \int_{-\infty}^{w} uf_i(u)\mathrm{d}u \Big/ \int_{-\infty}^{w} f_i(u)\mathrm{d}u.
$$

If $a_1a_2 \leq 0$, then the estimator $\hat{\theta}(\phi_1, \phi_2)$ is minimax.

It is interesting to note that the condition $a_1a_2 \leq 0$ is necessary and sufficient for the minimaxity of the generalized Bayes estimator against the uniform prior over the restricted space, which is expressed as $\hat{\theta}^{GB} = a_1\widehat{\mu}_1^{GB} + a_2\widehat{\mu}_2^{GB}$ for $\widehat{\mu}_i^{GB} = X_i - \phi_i^{GB}(X_i)$.

Proposition 3.4. *The generalized Bayes estimator $\hat{\theta}^{GB} = a_1\widehat{\mu}_1^{GB} + a_2\widehat{\mu}_2^{GB}$ against the uniform prior $\mathrm{d}\mu_1\mathrm{d}\mu_2$, $\mu_1 > 0$ and $\mu_2 > 0$, is minimax relative to the squared error loss if and only if $a_1a_2 \leq 0$.*

Proof. From Proposition 3.1, it follows that $R(\mu_i, \widehat{\mu}_i^U) - R(\mu_i, \widehat{\mu}_i^{GB}) \geq 0$ for $i = 1, 2$. Since ϕ_i^{GB} satisfies condition (a) of Proposition 3.3,

$$
E[\widehat{\mu}_i(\phi_i^{GB}) - \mu_i] = c_i - E[\phi_i^{GB}(X_i)] > 0.
$$

If $a_1a_2 \leq 0$, it is seen that $-2a_1a_2E[\widehat{\mu}_1(\phi_i)\widehat{\mu}_2(\phi_2)] \geq 0$. Thus, the dominance of $\hat{\theta}^{GB}$ over $\hat{\theta}^U$ is proved.

Reversely, suppose that $\hat{\theta}^{GB}$ dominates $\hat{\theta}^U$. We show that supposing the inequality $a_1a_2 > 0$ yields a contradiction. From Lemma 3.2, it is seen that at $(\mu_1, \mu_2) = (0, 0)$,

$$
\begin{aligned}
R(0,&0, \hat{\theta}^U) - R(0, 0, \hat{\theta}^{GB}) \\
&= a_1^2\{R(0, \widehat{\mu}_1^U) - R(0, \widehat{\mu}_1^{GB})\} + a_2^2\{R(0, \widehat{\mu}_2^U) - R(0, \widehat{\mu}_2^{GB})\} \\
&\quad - 2a_1a_2E_0[\widehat{\mu}_1^{GB}]E_0[\widehat{\mu}_2^{GB}],
\end{aligned}
$$

which is equal to $-2a_1a_2E_0[\widehat{\mu}_1^{GB}]E_0[\widehat{\mu}_2^{GB}]$ from Proposition 3.2. Under the condition $a_1a_2 > 0$, it is clear that $-2a_1a_2E_0[\widehat{\mu}_1^{GB}]E_0[\widehat{\mu}_2^{GB}] < 0$ at $(\mu_1, \mu_2) = (0, 0)$. This contradicts to the fact that $\hat{\theta}^{GB}$ dominates $\hat{\theta}^{U}$. Hence, if $\hat{\theta}^{GB}$ dominates $\hat{\theta}^{U}$, then $a_1a_2 \le 0$. □

3.3. Non-minimaxity for $k \ge 3$

We treat here the case of $k \ge 3$ where the random variables $X_1, \ldots, X_k$ are as in (1.1). Although it may be guessed that the generalized Bayes estimator against the uniform prior over the parameter D is minimax, the following proposition shows that this conjecture is not correct.

Proposition 3.5. *The generalized Bayes estimator $\hat{\theta}^{GB}$, given in (1.4), against the uniform prior over D is not minimax if $k \ge 3$.*

Proof. Corresponding to the decompositions given in (2.5), we can write $\hat{\theta}^{GB}$ as $\hat{\theta}^{GB} = \hat{\theta}_+^{GB} - \hat{\theta}_-^{GB}$ for $\hat{\theta}_+^{GB} = \sum_{i\in\Lambda_+} a_i\widehat{\mu}_i^{GB}$ and $\hat{\theta}_-^{GB} = -\sum_{i\in\Lambda_-} a_i\widehat{\mu}_i^{GB}$. Since $k \ge 3$, either Λ_+ or Λ_- includes more than two elements. We suppose here that Λ_+ has more than two elements without any loss of generality. The risk difference of the two estimators $\hat{\theta}^{U}$ and $\hat{\theta}^{GB}$ is expressed as

$$\begin{aligned}\Delta(\boldsymbol{\mu}) &= R(\boldsymbol{\mu}, \hat{\theta}^{U}) - R(\boldsymbol{\mu}, \hat{\theta}^{GB})\\ &= E[(\hat{\theta}_+^{U} - \theta_+)^2 - (\hat{\theta}_+^{GB} - \theta_+)^2] + E[(\hat{\theta}_-^{U} - \theta_-)^2 - (\hat{\theta}_-^{GB} - \theta_-)^2]\\ &\quad + 2E[\hat{\theta}_+^{GB} - \theta_+]E[\hat{\theta}_-^{GB} - \theta_-]\\ &= \Delta_+(\boldsymbol{\mu}) + \Delta_-(\boldsymbol{\mu}) + 2B_+(\boldsymbol{\mu})B_-(\boldsymbol{\mu}), \quad \text{(say)}\end{aligned}$$

for $B_+(\boldsymbol{\mu}) = E[\hat{\theta}_+^{GB} - \theta_+]$ and $B_-(\boldsymbol{\mu}) = E[\hat{\theta}_-^{GB} - \theta_-]$. Note that $B_-(\boldsymbol{\mu}) = -\sum_{i\in\Lambda_-} a_iB_i(\mu_i)$ for $B_i(\mu_i) = E_{\mu_i}[X_i - \mu_i - \phi_i^{GB}(X_i)]$ and that $B_i(\mu_i) = c_i - E_0[\phi_i^{GB}(X_i + \mu_i)]$. Since $B_i(\mu_i) \to 0$ as $\mu_i \to \infty$, it is seen that $B_-(\boldsymbol{\mu}) \to 0$ as $\mu_i \to \infty$ for all $i \in \Lambda_-$. Since

$$\Delta_-(\boldsymbol{\mu}) = \sum_{i\in\Lambda_-} a_i^2\{E[(\widehat{\mu}_i^{U} - \mu_i)^2] - E[(\widehat{\mu}_i^{GB} - \mu_i)^2]\} - 2\sum_{i\in\Lambda_-}\sum_{j\ne i, j\in\Lambda_-} a_ia_jB_i(\mu_i)B_j(\mu_j),$$

from Proposition 3.2, it can be seen that $\Delta_-(\boldsymbol{\mu}) \to 0$ as $\mu_i \to \infty$ for all $i \in \Lambda_-$. Thus,

$$\lim_{\mu_i\to\infty, i\in\Lambda_-} \Delta(\boldsymbol{\mu}) = \Delta_+(\boldsymbol{\mu}).$$

Similarly, $\Delta_+(\boldsymbol{\mu})$ is written as

$$\Delta_+(\boldsymbol{\mu}) = \sum_{i\in\Lambda_+} a_i^2\{E[(\widehat{\mu}_i^{U} - \mu_i)^2] - E[(\widehat{\mu}_i^{GB} - \mu_i)^2]\} - 2\sum_{i\in\Lambda_+}\sum_{j\ne i, j\in\Lambda_+} a_ia_jB_i(\mu_i)B_j(\mu_j),$$

and from Proposition 3.1, it follows that the first term in the r.h.s. is equal to zero when $\mu_i = 0$ for all $i \in \Lambda_+$. Since $B_i(0) = c_i - E_0[\phi_i^{GB}(X_i)] > 0$ and $a_ia_j > 0$ for any $i, j \in \Lambda_+$, it is concluded that

$$\lim_{\mu_i\to 0, i\in\Lambda_+}\ \lim_{\mu_i\to\infty, i\in\Lambda_-} \Delta(\boldsymbol{\mu}) = \lim_{\mu_i\to 0, i\in\Lambda_+} \Delta_+(\boldsymbol{\mu}) = -2\sum_{i\in\Lambda_+}\sum_{j\ne i, j\in\Lambda_+} a_ia_jB_i(0)B_j(0),$$

which is negative. That is, $R(\boldsymbol{\mu}, \hat{\theta}^{U}) < R(\boldsymbol{\mu}, \hat{\theta}^{GB})$ for a $\boldsymbol{\mu} \in D$, which means that $\hat{\theta}^{GB}$ is not minimax. □

4. Admissible and minimax estimation in normal distributions

The generalized Bayes estimator against the uniform prior over D is not necessarily minimax as shown in the previous section. An interesting query is what is a prior distribution which results in the minimax and Bayes estimator. Although it may be hard to answer this query for the general location family, we can find an affirmative solution in a setup where the underlying distributions are normal.

Let $X_1, \ldots, X_k$ be mutually independent random variables where X_i has a normal distribution with mean μ_i and unit variance, $\mathcal{N}(\mu_i, 1)$ for $\mu_i > 0$. We use the same notations D, $\boldsymbol{\mu}$, θ, $\boldsymbol{a}$ as defined in (1.1). For the prior distribution considered here, denoted by $\pi^*(\boldsymbol{\mu})$, with probability one,

$$\mu_i = \alpha_i \xi_+ \quad \text{for} \quad i \in \Lambda_+, \quad \text{and} \quad \mu_j = \beta_j \xi_- \quad \text{for} \quad j \in \Lambda_-,$$

where $\alpha_i = a_i \sum_{j\in\Lambda_+} a_j / \sum_{j\in\Lambda_+} a_j^2$, $\beta_j = a_j \sum_{i\in\Lambda_-} a_i / \sum_{i\in\Lambda_-} a_i^2$, and ξ_+ and ξ_- are distributed uniformly over the set $\{(\xi_+, \xi_-) | \xi_+ > 0, \xi_- > 0\}$. For notational simplicity, let $A_1 = \sum_{i\in\Lambda_+} a_i$, $A_2 = \sum_{i\in\Lambda_+} \alpha_i^2$, $B_1 = -\sum_{i\in\Lambda_-} a_i$ and $B_2 = \sum_{i\in\Lambda_-} a_i^2$. Then it is noted that $\sum_{i\in\Lambda_+} \alpha_i^2 = A_1^2/A_2$, $\sum_{i\in\Lambda_+} \alpha_i x_i / \sum_{i\in\Lambda_+} \alpha_i^2 = \hat{\theta}_+^U / A_1$, $\sum_{i\in\Lambda_+} a_i \alpha_i = A_1$, and similar equalities are satisfied for β_j. The joint density function of $(\boldsymbol{X}, \boldsymbol{\mu})$ is

$$(2\pi)^{-k/2} \exp\{-\frac{1}{2}\sum_{i\in\Lambda_+}(x_i - \alpha_i\xi_+)^2 - \frac{1}{2}\sum_{j\in\Lambda_+}(x_j - \beta_j\xi_-)^2\} \mathrm{d}\boldsymbol{x}\mathrm{d}\xi_+\mathrm{d}\xi_-$$

$$= \exp\{-\frac{A_1^2}{2A_2}(\xi_+ - \frac{\hat{\theta}_+^U}{A_1})^2 - \frac{B_1^2}{2B_2}(\xi_- - \frac{\hat{\theta}_-^U}{B_1})^2\} p(S_1, S_2) \mathrm{d}\boldsymbol{x}\mathrm{d}\xi_+\mathrm{d}\xi_-, \tag{4.1}$$

where $p(S_1, S_2) = (2\pi)^{-k/2} \exp\{-(S_+ + S_-)/2\}$ for

$$S_+ = \sum_{i\in\Lambda_+} x_i^2 - (\sum_{i\in\Lambda_+} \alpha_i x_i)^2 / \sum_{i\in\Lambda_+} \alpha_i^2,$$
$$S_- = \sum_{j\in\Lambda_-} x_j^2 - (\sum_{j\in\Lambda_-} \beta_j x_j)^2 / \sum_{j\in\Lambda_-} \beta_i^2.$$

To simplify the notation, let

$$\theta_1 = \theta_+/\sqrt{A_2}, \quad \theta_2 = \theta_-/\sqrt{B_2}, \quad z_1 = \hat{\theta}_+^U/\sqrt{A_2}, \quad z_2 = \hat{\theta}_-^U/\sqrt{B_2}.$$

Then, z_1 and z_2 are mutually independently distributed as $\mathcal{N}(\theta_1, 1)$ and $\mathcal{N}(\theta_2, 1)$, respectively, and

$$\theta = \sqrt{A_2}\theta_1 - \sqrt{B_2}\theta_2. \tag{4.2}$$

Making the transformation $\xi_1 = A_1 A_2^{-1/2} \xi_+$ and $\xi_2 = B_1 B_2^{-1/2} \xi_-$, we can rewrite the joint density function of $(\boldsymbol{X}, \boldsymbol{\mu})$ given in (4.1) as

$$\exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\} \frac{\sqrt{A_2 B_2}}{A_1 B_1} p(S_1, S_2) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{\xi}, \tag{4.3}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2)^t$ and $\boldsymbol{z} = (z_1, z_2)^t$. Since $\sum_{i\in\Lambda_+} a_i \alpha_i \xi_+ + \sum_{j\in\Lambda_-} a_j \beta_j \xi_- = \sqrt{A_2}\xi_1 - \sqrt{B_2}\xi_2$, the generalized Bayes estimator of θ against the prior π^* can be written as

$$\hat{\theta}^{GB*} = \frac{\int_D (\sqrt{A_2}\xi_1 - \sqrt{B_2}\xi_2)\exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\mathrm{d}\boldsymbol{\xi}}{\int_D \exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\mathrm{d}\boldsymbol{\xi}} = \sqrt{A_2}\{z_1 - \phi^{GB*}(z_1)\} - \sqrt{B_2}\{z_2 - \phi^{GB*}(z_2)\}, \tag{4.4}$$

where $\boldsymbol{a}^t\boldsymbol{X} = \hat{\theta}_+^U - \hat{\theta}_-^U = \sqrt{A_2}z_1 - \sqrt{B_2}z_2$ and

$$\phi^{GB*}(w) = \int_{-\infty}^{w} u\exp\{-u^2/2\}\mathrm{d}u \Big/ \int_{-\infty}^{w} \exp\{-u^2/2\}\mathrm{d}u,$$

Minimaxity and admissibility of $\hat{\theta}^{GB*}$ can be established in the following proposition.

Proposition 4.1. *The generalized Bayes estimator $\hat{\theta}^{GB*}$ of θ against the prior π^* is admissible and minimax.*

Proof. The minimaxity of $\hat{\theta}^{GB*}$ follows from Proposition 3.3. In fact, the arguments given in (4.2) and (4.4) mean that the generalized Bayes estimator of $\theta = \sqrt{A_2}\theta_1 - \sqrt{B_2}\theta_2$ is based on $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{I}_2)$, where $\boldsymbol{\theta} = (\theta_1, \theta_2)^t$ for $\theta_1 > 0$ and $\theta_2 > 0$. Thus, it can be seen that $\phi^{GB*}(w)$ satisfies the conditions (a) and (b) of Proposition 3.3, so that $\hat{\theta}^{GB*}$ is minimax.

We next prove the admissibility of $\hat{\theta}^{GB*}$ using the method of Brown and Hwang (1982). Consider a sequence of the prior distributions $\pi_n^*(\boldsymbol{\mu})$ such that with probability one,

$$\mu_i = \alpha_i\xi_+ \quad \text{for} \quad i \in \Lambda_+, \quad \text{and} \quad \mu_j = \beta_j\xi_- \quad \text{for} \quad j \in \Lambda_-,$$

where (ξ_+, ξ_-) is distributed as $\{h_n(A_1A_2^{-1/2}\xi_+ + B_1B_2^{-1/2}\xi_-)\}^2$ for

$$h_n(t) = \begin{cases} 1, & \text{if } 0 \le t < 1 \\ 1 - \log t/\log n, & \text{if } 1 \le t \le n \\ 0, & \text{if } n < t. \end{cases}$$

Similarly to (4.3), we can write the joint density function of $(\boldsymbol{X}, \boldsymbol{\mu})$ given in (4.1) as

$$\exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\{h_n(|\boldsymbol{\xi}|)\}^2\frac{\sqrt{A_2B_2}}{A_1B_1}p(S_1, S_2)\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{\xi}, \tag{4.5}$$

where $|\boldsymbol{\xi}|$ denotes $|\boldsymbol{\xi}| = \xi_1 + \xi_2$. The generalized Bayes estimator δ_n against the prior $\pi_n^*(\boldsymbol{\mu})$ can be expressed as

$$\delta_n = \frac{\int_D (\sqrt{A_2}\xi_1 - \sqrt{B_2}\xi_2)\{h_n(|\boldsymbol{\xi}|)\}^2\exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\mathrm{d}\boldsymbol{\xi}}{\int_D \{h_n(|\boldsymbol{\xi}|)\}^2\exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\mathrm{d}\boldsymbol{\xi}}$$

and the generalized Bayes estimator $\hat{\theta}^{GB*}$ corresponds to the case of $h_n(|\boldsymbol{\xi}|) = 1$, where $D = \{\boldsymbol{\xi} | \xi_1 > 0, \xi_2 > 0\}$. From (4.3), the difference of the Bayes risk functions of two estimators $\hat{\theta}^{GB*}$ and δ_n is written by

$$\begin{aligned}\Delta_n &= \int_D \int \Big\{(\hat{\theta}^{GB*} - \theta)^2 - (\delta_n - \theta)^2\Big\} \\ &\qquad \times \exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\{h_n(|\boldsymbol{\xi}|)\}^2\frac{\sqrt{A_2B_2}}{A_1B_1}p(S_1, S_2)\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{\xi} \\ &= \int (\hat{\theta}^{GB*} - \delta_n)^2 \int_D \exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\{h_n(|\boldsymbol{\xi}|)\}^2\frac{\sqrt{A_2B_2}}{A_1B_1}p(S_1, S_2)\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{x},\end{aligned}$$

where $\theta = \sqrt{A_2}\xi_1 - \sqrt{B_2}\xi_2$ in the above bracket. Noting that z_1, z_2 and (S_+, S_-) are mutually independent, we can evaluate Δ_n as

$$
\begin{aligned}
\Delta_n &= C\int(\hat{\theta}^{GB*}-\delta_n)^2\int_D \exp\{-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2\}\{h_n(|\boldsymbol{\xi}|)\}^2\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{z}\\
&\le 2CA_2\int\Big\{\int_D \xi_1 f_2(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}-\int_D \xi_1 f_1(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}\Big\}^2\\
&\qquad\qquad \times\int_D \exp\{-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2\}\{h_n(|\boldsymbol{\xi}|)\}^2\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{z}\\
&\quad + 2CB_2\int\Big\{\int_D \xi_2 f_2(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}-\int_D \xi_2 f_1(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}\Big\}^2\\
&\qquad\qquad \times\int_D \exp\{-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2\}\{h_n(|\boldsymbol{\xi}|)\}^2\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{z}\\
&= \Delta_+ + \Delta_-, \qquad \text{(say)}
\end{aligned}
\tag{4.6}
$$

where C is an appropriate positive constant, and

$$
\begin{aligned}
f_1(\boldsymbol{\xi}) &= \frac{\{h_n(|\boldsymbol{\xi}|)\}^2\exp(-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2)}{\int_D\{h_n(|\boldsymbol{\xi}|)\}^2\exp(-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}},\\
f_2(\boldsymbol{\xi}) &= \frac{\exp(-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2)}{\int_D\exp(-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}}.
\end{aligned}
$$

We now show that $\Delta_+\to 0$ and $\Delta_-\to 0$ as $n\to\infty$ by using the same arguments as in Tsukuma and Kubokawa (2008). Let $\boldsymbol{\theta}\vee\boldsymbol{\eta} = (\max(\theta_1,\eta_1),\max(\theta_2,\eta_2))^t$ and $\boldsymbol{\theta}\wedge\boldsymbol{\eta} = (\min(\theta_1,\eta_1),\min(\theta_2,\eta_2))^t$ for $\boldsymbol{\theta} = (\theta_1,\theta_2)^t$ and $\boldsymbol{\eta} = (\eta_1,\eta_2)^t$. Since $\{h_n(t)\}^2$ is nonincreasing in t, it is noted that $\{h_n(|\boldsymbol{\theta}\wedge\boldsymbol{\eta}|)\}^2 \ge \{h_n(|\boldsymbol{\xi}|)\}^2$, which implies that $f_1(\boldsymbol{\theta})f_2(\boldsymbol{\eta}) \le f_2(\boldsymbol{\theta}\vee\boldsymbol{\eta})f_1(\boldsymbol{\theta}\wedge\boldsymbol{\eta})$. Hence it follows from Karlin and Rinott (1980) that

$$
\int_D \xi_i f_1(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} \le \int_D \xi_i f_2(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}, \qquad i = 1,\, 2. \tag{4.7}
$$

Using the integration by parts, we can see that

$$
\begin{aligned}
&\int_D(\xi_1-z_1)\{h_n(|\boldsymbol{\xi}|)\}^2\exp(-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}\\
&\quad = \exp(-z_1^2/2)\int_0^n\{h_n(\xi_2)\}^2\exp(-(\xi_2-x_2)^2/2)\mathrm{d}\xi_2\\
&\qquad\quad -\int_{\boldsymbol{\xi}\in D,1\le|\boldsymbol{\xi}|\le n} h_n(|\boldsymbol{\xi}|)\{1/(|\boldsymbol{\xi}|\log n)\}\exp(-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi},
\end{aligned}
\tag{4.8}
$$

$$
\begin{aligned}
&\int_D(\xi_1-z_1)\exp(-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}\\
&\quad = \exp(-z_1^2/2)\int_0^\infty\exp(-(\xi_2-z_2)^2/2)\mathrm{d}\xi_2.
\end{aligned}
\tag{4.9}
$$

Combining (4.7), (4.8) and (4.9) gives

$$
\begin{aligned}
0 &\le \int_D (\xi_1 - z_1) f_2(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} - \int_D (\xi_1 - z_1) f_1(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} \\
&= \frac{\exp(-z_1^2/2)}{\int_0^\infty \exp(-(\xi_1 - z_1)^2/2)\mathrm{d}\xi_1} \\
&\quad - \frac{\exp(-z_1^2/2)\int_0^n \{h_n(\xi_2)\}^2 \exp(-(\xi_2 - z_2)^2/2)\mathrm{d}\xi_2}{\int_D \{h_n(|\boldsymbol{\xi}|)\}^2 \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}} \\
&\quad + \frac{\int_{\boldsymbol{\xi}\in D, 1\le|\boldsymbol{\xi}|\le n} (|\boldsymbol{\xi}|\log n)^{-1} h_n(|\boldsymbol{\xi}|)\exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}}{\int_D \{h_n(|\boldsymbol{\xi}|)\}^2 \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}}.
\end{aligned} \tag{4.10}
$$

Since $h_n(|\boldsymbol{\xi}|)I(\boldsymbol{\xi} \in D,\, 0 \le |\boldsymbol{\xi}| \le n) \le h_n(\xi_2)I(0 \le \xi_1 \le n,\, 0 \le \xi_2 \le n)$, we observe that

$$
\begin{aligned}
&\int_D \{h_n(|\boldsymbol{\xi}|)\}^2 \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi} \\
&\le \int_{0\le\xi_1\le n,\, 0\le\xi_2\le n} \{h_n(\xi_2)\}^2 \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi} \\
&\le \int_0^\infty \exp(-(\xi_1 - z_1)^2/2)\mathrm{d}\xi_1 \int_0^n \{h_n(\xi_2)\}^2 \exp(-(\xi_2 - z_2)^2/2)\mathrm{d}\xi_2,
\end{aligned}
$$

which is used to evaluate the second term in the r.h.s. of the equation (4.10). Hence from (4.10),

$$
\begin{aligned}
0 &\le \int_D (\xi_1 - z_1) f_2(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} - \int_D (\xi_1 - z_1) f_1(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} \\
&\le \frac{\int_{\boldsymbol{\xi}\in D, 1\le|\boldsymbol{\xi}|\le n} (|\boldsymbol{\xi}|\log n)^{-1} h_n(|\boldsymbol{\xi}|)\exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}}{\int_D \{h_n(|\boldsymbol{\xi}|)\}^2 \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}}.
\end{aligned}
$$

Using the Cauchy-Schwarz inequality, we can see that

$$
\begin{aligned}
&\left\{\int_D \xi_1 f_2(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} - \int_D \xi_1 f_1(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}\right\}^2 \\
&\le \frac{\{\int_{\boldsymbol{\xi}\in D,\, 1\le|\boldsymbol{\xi}|\le n} (|\boldsymbol{\xi}|\log n)^{-1} h_n(|\boldsymbol{\xi}|)\exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}\}^2}{\{\int_D \{h_n(|\boldsymbol{\xi}|)\}^2 \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}\}^2} \\
&\le \frac{\int_{\boldsymbol{\xi}\in D,\, 1\le|\boldsymbol{\xi}|\le n} (|\boldsymbol{\xi}|\log n)^{-2} \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}}{\int_D \{h_n(|\boldsymbol{\xi}|)\}^2 \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}},
\end{aligned}
$$

which implies

$$
\begin{aligned}
\Delta_+ &= 2CA_2 \int \left\{\int_D \xi_1 f_2(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi} - \int_D \xi_1 f_1(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}\right\}^2 \\
&\qquad\qquad \times \int_D \exp\{-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2\}\{h_n(|\boldsymbol{\xi}|)\}^2 \mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{z} \\
&\le 2CA_2 \int \int_{\boldsymbol{\xi}\in D,\, 1\le|\boldsymbol{\xi}|\le n} (|\boldsymbol{\xi}|\log n)^{-2} \exp(-\|\boldsymbol{\xi} - \boldsymbol{z}\|^2/2)\mathrm{d}\boldsymbol{\xi}\mathrm{d}\boldsymbol{z} \\
&= 2CA_2(2\pi) \int_{\boldsymbol{\xi}\in D,\, 1\le\xi_1+\xi_2\le n} (\xi_1 + \xi_2)^{-2} \mathrm{d}\boldsymbol{\xi} (\log n)^{-2}.
\end{aligned}
$$

Making the transformations $u = \xi_1 + \xi_2$ and $w = \xi_1/(\xi_1 + \xi_2)$, we see that

$$\int_{\boldsymbol{\xi}\in D,\, 1\le\xi_1+\xi_2\le n} (\xi_1 + \xi_2)^{-2} \mathrm{d}\boldsymbol{\xi} = \int_0^1 \mathrm{d}w \int_1^n u^{-1}\mathrm{d}u = \log n,$$

so that

$$\Delta_+ \le 2CA_2(2\pi)(\log n)^{-1},$$

which goes to zero as $n \to \infty$. Similarly, we have $\Delta_- \le 2CB_2(2\pi)(\log n)^{-1}$. Therefore, the admissibility of $\hat{\theta}^{GB*}$ is established. □

Finally, we give an expression of the risk function of $\hat{\theta}^{GB*}$. As seen from (4.3) and (4.4), the estimator $\hat{\theta}^{GB*}$ corresponds to the case $k = 2$ in the generalized Bayes estimator $\hat{\theta}^{GB}$ against the uniform prior over D given in (1.3). Thus, we begin with the estimator $\hat{\theta}^{GB}$. First, the generalized Bayes estimator of the mean vector $\boldsymbol{\mu}$ against the uniform prior is given by $\widehat{\boldsymbol{\mu}}^{GB} = \boldsymbol{X} - \boldsymbol{\phi}^{GB}(\boldsymbol{X})$, where

$$\boldsymbol{\phi}^{GB}(\boldsymbol{X}) = \frac{\int_D (\boldsymbol{X} - \boldsymbol{\xi}) \exp\{-\|\boldsymbol{X} - \boldsymbol{\xi}\|^2/2\}\mathrm{d}\boldsymbol{\xi}}{\int_D \exp\{-\|\boldsymbol{X} - \boldsymbol{\xi}\|^2/2\}\mathrm{d}\boldsymbol{\xi}}$$

for $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)^t$. It can be seen that $\boldsymbol{\phi}^{GB}(\boldsymbol{X}) = (\phi_1^{GB}(X_1), \ldots, \phi_k^{GB}(X_k))^t$ where $\phi_i^{GB}(X_i) = \int_0^\infty (X_i - \xi_i) \exp\{-(X_i - \xi_i)^2/2\}\mathrm{d}\xi_i / \int_0^\infty \exp\{-(X_i - \xi_i)^2/2\}\mathrm{d}\xi_i$. The function $\phi_i^{GB}(X_i)$ can be further rewritten as

$$\phi_i^{GB}(X_i) = \phi_i^{GB} = \frac{\int_{-\infty}^{X_i} u \exp\{-u^2/2\}\mathrm{d}u}{\int_{-\infty}^{X_i} \exp\{-u^2/2\}\mathrm{d}u} = -\frac{\exp\{-X_i^2/2\}}{\int_{-\infty}^{X_i} \exp\{-u^2/2\}\mathrm{d}u}, \tag{4.11}$$

which is negative. In the context of the simultaneous estimation of $\boldsymbol{\mu}$, Hartigan (2004) derived an expression of the risk function of $\widehat{\boldsymbol{\mu}}^{GB}$, which is given by

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^{GB}) = E[\|\widehat{\boldsymbol{\mu}}^{GB} - \boldsymbol{\mu}\|^2] = k + E_{\boldsymbol{\mu}}\Big[\sum_{i=1}^k \mu_i \phi_i^{GB}(X_i)\Big].$$

This demonstrates that $\widehat{\boldsymbol{\mu}}^{GB}$ dominates $\boldsymbol{X}$, namely, $\widehat{\boldsymbol{\mu}}^{GB}$ is minimax for any dimension k. In contrast, the dominance results obtained in Section 3 mean that the generalized Bayes estimator $\hat{\theta}^{GB}$ is not necessarily minimax. Using the same arguments as in Hartigan (2004), we can get a similar expression of the risk function. Using the same notation as in (4.11), we can express the generalized Bayes estimator $\hat{\theta}^{GB}$ of $\theta = \boldsymbol{a}^t\boldsymbol{\mu}$ as

$$\hat{\theta}^{GB} = \boldsymbol{a}^t\widehat{\boldsymbol{\mu}}^{GB} = \boldsymbol{a}^t\boldsymbol{X} - \boldsymbol{a}^t\boldsymbol{\phi}^{GB}(\boldsymbol{X}),$$

whose risk function is given in the following proposition.

Proposition 4.2. *The risk function $R(\boldsymbol{\mu}, \hat{\theta}^{GB}) = E[(\hat{\theta}^{GB} - \theta)^2]$ has the form*

$$R(\boldsymbol{\mu}, \hat{\theta}^{GB}) = \boldsymbol{a}^t\boldsymbol{a} + E_{\boldsymbol{\mu}}\Big[\sum_{i=1}^k a_i^2 \mu_i \phi_i^{GB}(X_i) + 2\sum_{i=1}^k \sum_{j>i} a_i a_j \phi_i^{GB}(X_i)\phi_j^{GB}(X_j)\Big]. \tag{4.12}$$

Proof. For notational simplicity, let $\phi_i = \phi_i^{GB}(X_i)$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)^t$. Let $\Delta = R(\boldsymbol{\mu}, \hat{\theta}^{GB}) - R(\boldsymbol{\mu}, \boldsymbol{a}^t\boldsymbol{X})$. Since $R(\boldsymbol{\mu}, \boldsymbol{a}^t\boldsymbol{X}) = \boldsymbol{a}^t\boldsymbol{a}$, it is easy to see that

$$\Delta = \boldsymbol{a}^t E[-(\boldsymbol{X} - \boldsymbol{\mu})\boldsymbol{\phi}^t - \boldsymbol{\phi}(\boldsymbol{X} - \boldsymbol{\mu})^t + \boldsymbol{\phi}\boldsymbol{\phi}^t]\boldsymbol{a}.$$

Applying the Stein identity to a cross product term gives

$$\begin{aligned}
E[(\boldsymbol{X}-\boldsymbol{\mu})\boldsymbol{\phi}^t] &= E[\mathrm{diag}_i\big((X_i-\mu_i)\phi_i\big)] \\
&= E[\mathrm{diag}_i\Big(1-\frac{\int_0^\infty (X_i-\xi_i)^2\exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i}{\int_0^\infty \exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i}+\phi_i^2\Big)],
\end{aligned}$$

where $\mathrm{diag}_i(d_i)$ denotes $\mathrm{diag}(d_1,\ldots,d_k)$. Since $(X_i-\mu_i)(X_i-\xi_i)=(X_i-\xi_i)^2+(\xi_i-\mu_i)(X_i-\xi_i)$, the other cross product term can be written as

$$\begin{aligned}
E[\boldsymbol{\phi}(\boldsymbol{X}-\boldsymbol{\mu})^t] = E[&\mathrm{diag}_i\Big(\frac{\int_0^\infty (X_i-\xi_i)^2\exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i}{\int_0^\infty \exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i}\Big) \\
&+\mathrm{diag}_i\Big(\frac{\int_0^\infty (\xi_i-\mu_i)(X_i-\xi_i)\exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i}{\int_0^\infty \exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i}\Big)].
\end{aligned}$$

From integration by parts, it is observed that

$$\begin{aligned}
&\int_0^\infty (\xi_i-\mu_i)(X_i-\xi_i)\exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i \\
&\qquad = \mu_i\exp\{-X_i^2/2\}-\int_0^\infty \exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i,
\end{aligned}$$

From (4.11), it follows that

$$\frac{\int_0^\infty (\xi_i-\mu_i)(X_i-\xi_i)\exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i}{\int_0^\infty \exp\{-(X_i-\xi_i)^2/2\}\mathrm{d}\xi_i} = -\mu_i\phi_i-1.$$

Combining the above observations gives

$$\Delta = \boldsymbol{a}^t E[-\mathrm{diag}_i(\{\phi_i\}^2-\mu_i\phi_i)+\boldsymbol{\phi}\boldsymbol{\phi}^t]\boldsymbol{a},$$

which yields expression (4.12). □

When μ_i is zero, it is seen that $E_0[\phi_i(X_i)]=E_0[\phi_1(X_1)]$ for $i=2,\ldots,k$. According to Proposition 4.2,

$$R(\boldsymbol{0},\hat{\theta}^{GB}) = \boldsymbol{a}^t\boldsymbol{a}+2\sum_{i=1}^k\sum_{j>i} a_ia_j\{E_0[\phi_1^{GB}(X_1)]\}^2,$$

which implies that a necessary condition for the minimaxity is $\sum_{i=1}^k\sum_{j>i}a_ia_j\le 0$ or $k=1$. As seen from Propositions 3.1 and 3.3, this is a sufficient condition as well. However, Proposition 3.5 shows that it is not sufficient in the case of $k\ge 3$.

As in the case of $k=2$ in Proposition 4.2, we can provide an expression of the risk of the generalized Bayes and minimax estimator $\hat{\theta}^{GB*}$ given in (4.4),

$$\begin{aligned}
R(\boldsymbol{\mu},\hat{\theta}^{GB*}) = A_2^2+B_2^2 &+ E_{\boldsymbol{\mu}}\big[A_2\theta_1\phi^{GBs}(z_1)+B_2\theta_2\phi^{GBs}(z_2)\big] \\
&- 2E_{\boldsymbol{\mu}}\big[\sqrt{A_2B_2}\phi^{GB*}(z_1)\phi^{GB*}(z_2)\big].
\end{aligned}$$

5. A relation to the Stein problem in variance estimation

In this section, we explain that the estimation of the restricted mean in a normal distribution is related to the Stein problem in the estimation of variance. This fact was established by Rukhin (1992) in a canonical form of a normal distributional model. We here use the same arguments to clarify the conditions on the parameters under which the Stein estimator of variance in a linear regression model asymptotically reduces to the truncated estimator of the restricted normal mean. We also show that the Stein problem in estimation of ratio of variances converges to the estimation of the difference of two restricted normal means.

Let us consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{5.1}$$

where $\boldsymbol{y}$ and $\boldsymbol{X}$ are $n \times 1$ and $n \times p$ matrices, respectively, and $\boldsymbol{\beta}$ is a p-vector of the regression parameters and $\boldsymbol{\epsilon}$ is an n-vector of errors having a distribution $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$. It is assumed that $\boldsymbol{X}$ is of full rank. Let $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$ and $S = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^t(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$, which are distributed as $\mathcal{N}(\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_p)$ and $\sigma^2\chi^2_m$ for $m = n - p$. Stein (1964) showed that the best scale estimator based on S is inadmissible and is improved on by using information contained in $\widehat{\boldsymbol{\beta}}$. For instance, the unbiased estimator $\hat{\sigma}^{2U} = S/m$ is improved on by the truncated estimator

$$\hat{\sigma}^{2S} = \{S/m,\ (S + \widehat{\boldsymbol{\beta}}^t\boldsymbol{X}^t\boldsymbol{X}\widehat{\boldsymbol{\beta}})/n\}$$

relative to an entropy loss function. Rukhin (1992) showed that this dominance result can be viewed as the estimation of a positive mean in a normal distribution.

Consider the asymptotic approximation under the following setup which is a slightly different framework from that of Rukhin (1992):

(A1) The dimension p behaves as $p = n - d_n$ where $d_n > 0$ and $d_n = O(n^\delta)$ for $0 \le \delta < 1$.

(A2) $\boldsymbol{X}^t\boldsymbol{X}/n$ converges to a positive definite matrix, and there is a positive constant θ such that
$$\lim_{n\to\infty} \sqrt{m}\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}/(n\sigma^2) = \sqrt{2}\theta.$$

Under (A1), it is easy to see that $m = O(n^\delta)$ and $m \to \infty$ as $n \to \infty$. Let $Z = (S - m\sigma^2)/(\sqrt{2m}\sigma^2)$ and $\boldsymbol{U} = (\boldsymbol{X}^t\boldsymbol{X})^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma$. Then $\boldsymbol{U}$ have $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I})$. Since $E[S] = m\sigma^2$ and $Var[S] = 2m\sigma^4$, it is seen that Z converges to $\mathcal{N}(0, 1)$ as $m \to \infty$. Thus,

$$\sqrt{m}(\hat{\sigma}^{2U} - \sigma^2)/\sigma^2 = \sqrt{2}Z = -\sqrt{2}(Y - \theta), \tag{5.2}$$

where $Y = -Z + \theta$ and it converges to $\mathcal{N}(\theta, 1)$. Since $\widehat{\boldsymbol{\beta}}^t\boldsymbol{X}^t\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \sigma^2\boldsymbol{U}^t\boldsymbol{U} + 2\sigma\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X})^{1/2}\boldsymbol{U} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}$, it can be seen that

$$\begin{aligned}
&\sqrt{m}(\hat{\sigma}^{2S} - \sigma^2)/\sigma^2 \\
&\quad = \sqrt{m}(\hat{\sigma}^{2U} - \sigma^2)/\sigma^2 - \sqrt{m}\max\{0, \frac{p}{nm}S - \frac{\widehat{\boldsymbol{\beta}}^t\boldsymbol{X}^t\boldsymbol{X}\widehat{\boldsymbol{\beta}}}{n}\} \\
&\quad = \sqrt{2}Z - \max\{0, \frac{p}{n}(\sqrt{2}Z + \sqrt{m}) - \frac{\sqrt{m}}{n}\boldsymbol{U}^t\boldsymbol{U} \\
&\qquad\qquad - 2\frac{\sqrt{m}}{\sigma\sqrt{n}}\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)^{1/2}\boldsymbol{U} - \frac{\sqrt{m}\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}}{n\sigma^2}\} \\
&\quad = \sqrt{2}Z - \max\{0, \frac{p}{n}\sqrt{2}Z + \frac{\sqrt{mp}}{n}\sqrt{p}(\boldsymbol{U}^t\boldsymbol{U}/p - 1) \\
&\qquad\qquad - 2\frac{\sqrt{m}}{\sigma\sqrt{n}}\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)^{1/2}\boldsymbol{U} - \frac{\sqrt{m}\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}}{n\sigma^2}\}.
\end{aligned}$$

From the assumptions, it is observed that $p/n \to 1$, $\sqrt{mp}/n = O(n^{(\delta-1)/2}) \to 0$ and $\sqrt{p}(\boldsymbol{U}^t\boldsymbol{U}/p-1) = O_p(1)$, so that $(\sqrt{mp}/n)\sqrt{p}(\boldsymbol{U}^t\boldsymbol{U}/p-1) \to 0$. Note that

$$\frac{\sqrt{m}}{\sigma\sqrt{n}}\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)^{1/2}\boldsymbol{U} = \sqrt{\sqrt{m}/n}\sqrt{\sqrt{m}\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}/n}\frac{\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)^{1/2}\boldsymbol{U}}{\sqrt{\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)\boldsymbol{\beta}}}.$$

Since $\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)^{1/2}\boldsymbol{U}/\sqrt{\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)\boldsymbol{\beta}} \sim \mathcal{N}(0,1)$ and $\sqrt{m}/n = O(n^{\delta/2-1})$, it is seen that $\{\sqrt{m}/\sigma\sqrt{n}\}\boldsymbol{\beta}^t(\boldsymbol{X}^t\boldsymbol{X}/n)^{1/2}\boldsymbol{U} \to 0$. Thus,

$$\begin{aligned}\sqrt{m}(\hat{\sigma}^{2S}-\sigma^2)/\sigma^2 &\to \sqrt{2}Z - \max\{0, \sqrt{2}Z-\sqrt{2}\theta\}\\ &= -\sqrt{2}(\max\{Y,0\}-\theta),\end{aligned} \tag{5.3}$$

where $Y = -Z+\theta$ converges to $\mathcal{N}(\theta,1)$ for $\theta > 0$. This shows that Stein's truncated estimator of σ^2 converges to the nonnegative estimator $\max(Y,0)$ of θ where $Y \sim \mathcal{N}(\theta,1)$ for $\theta > 0$.

We next consider the estimation of ratio of variances in two linear models, given by $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$, $i = 1,2$, where $\boldsymbol{\epsilon}_i \sim \mathcal{N}_n(\boldsymbol{0}, \sigma_i^2\boldsymbol{I}_n)$, $\boldsymbol{\beta}_i$ is a $p\times 1$ vector and the other variables are defined similarly to (5.1). Let $\widehat{\boldsymbol{\beta}}_i$ and S_i be defined as similar statistics as in model (5.1). Kubokawa (1994b), Kubokawa and Srivastava (1996) and Iliopoulos and Kourouklis (1999) showed that the best multiple by the ratio S_1/S_1 can be improved on by using information on $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ in the estimation of the ratio $\rho = \sigma_2^2/\sigma_1^2$. Let $\hat{\sigma}_i^{2U} = S_i/m$ and $\hat{\sigma}_i^{2S} = \min\{\hat{\sigma}_i^{2U}, (S_i + \widehat{\boldsymbol{\beta}}_i^t\boldsymbol{X}_i^t\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_i)/n\}$ for $i = 1,2$ and $m = n-p$. For instance, the ratio of the unbiased estimators $\hat{\sigma}_2^{2U}/\hat{\sigma}_1^{2U}$ should be improved on by the ratio of the truncated estimators $\hat{\sigma}_2^{2S}/\hat{\sigma}_1^{2S}$. To derive the asymptotic approximations of these ratio estimators, it is noted that for two estimators $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$,

$$\begin{aligned}\sqrt{m}(\hat{\sigma}_2^2/\hat{\sigma}_1^2 - \sigma_2^2/\sigma_1^2) &= \frac{\sqrt{m}(\hat{\sigma}_2^2-\sigma_2^2)/\sigma_2^2 + \sqrt{m}}{(\hat{\sigma}_1^2-\sigma_1^2)/\sigma_1^2+1}\frac{\sigma_2^2}{\sigma_1^2} - \sqrt{m}\frac{\sigma_2^2}{\sigma_1^2}\\ &= \frac{\sigma_2^2}{\sigma_1^2}\Big(\frac{\sqrt{m}(\hat{\sigma}_2^2-\sigma_2^2)/\sigma_2^2}{(\hat{\sigma}_1^2-\sigma_1^2)/\sigma_1^2+1} - \frac{\sqrt{m}(\hat{\sigma}_1^2-\sigma_1^2)/\sigma_1^2}{(\hat{\sigma}_1^2-\sigma_1^2)/\sigma_1^2+1}\Big).\end{aligned} \tag{5.4}$$

Assume the condition (A1) and

(A2′) For $i = 1,2$, $\boldsymbol{X}_i^t\boldsymbol{X}_i/n$ converges to a positive definite matrix, and there is positive constant θ_i such that

$$\lim_{n\to\infty}\sqrt{m}\boldsymbol{\beta}_i^t\boldsymbol{X}_i^t\boldsymbol{X}_i\boldsymbol{\beta}_i/(n\sigma_i^2) = \sqrt{2}\theta_i.$$

For $i = 1,2$, let $Z_i = (S_i - m\sigma_i^2)/(\sqrt{2m}\sigma_i^2)$ and $Y_i = -Z_i + \theta_i$, which has $\mathcal{N}(\theta_i, 1)$. Hence from (5.2) and (5.4), it is observed that

$$\sqrt{m}(\hat{\sigma}_2^{2U}/\hat{\sigma}_1^{2U} - \sigma_2^2/\sigma_1^2) \to (\sigma_2^2/\sigma_1^2)\{(Y_1-Y_2)-(\theta_1-\theta_2)\}.$$

Also from (5.3) and (5.4),

$$\sqrt{m}(\hat{\sigma}_2^{2S}/\hat{\sigma}_1^{2S} - \sigma_2^2/\sigma_1^2) \to (\sigma_2^2/\sigma_1^2)\{(\max(Y_1,0)-\max(Y_2,0))-(\theta_1-\theta_2)\}.$$

This shows that the estimation of the ratio of the variances can be approximated by the estimation of the difference of the positive means of normal distributions. Thus, the estimation of the mean difference can be motivated from the estimation of ratio of variances.

Acknowledgments

The author is grateful to Professor Andrew Rukhin for his valuable comments. This research was supported in part by Grant-in-Aid for Scientific Research Nos. 19200020 and 21540114 from Japan Society for the Promotion of Science.

References

Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics* III (S. S. Gupta, J. Berger, eds.), 205–230. Academic Press, New York.

Farrell, R. H. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.*, **35**, 949–998.

Girshick, M. A. and Savage, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. In *Proc. Second Berkeley Symp. Math. Statist. Probab.*, **1**, 53–74. University of California Press, Berkeley.

Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statist. Prob. Letters*, **67**, 285–288.

Iliopoulos, G. and Kourouklis, S. (1999). Improving on the best affine equivariant estimator of the ratio of the generalized variances. *J. Multivariate Analysis*, **68**, 176–192.

Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.*, **10**, 467–498.

Katz, M. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Ann. Math. Statist.*, **32**, 136–142.

Kubokawa, T. (1994a). A unified approach to improving equivariant estimators. *Ann. Statist.*, **22**, 290–299.

Kubokawa, T. (1994b). Double shrinkage estimation of ratio of scale parameters. *Ann. Inst. Statist. Math.*, **46**, 95–119.

Kubokawa, T. (1999). Shrinkage and modification techniques in estimation of variance and the related problems: A review. *Commun. Statist.–Theory and Methods*, **28**, 613–650.

Kubokawa, T. (2004). Minimaxity in estimation of restricted parameters. *J. Japan Statist. Soc.*, **34**, 229–253.

Kubokawa, T. and Saleh, A. K. Md. E. (1998). Estimation of location and scale parameters under order restrictions. *J. Statistical Research*, **28**, 41–51.

Kubokawa, T. and Srivastava, M. S. (1996). Double shrinkage estimators of ratio of variances. *Proceedings of the Sixth Lukacs Symposium*, 139–154.

Marchand, E. and Strawderman, W. E. (2004). Estimation in restricted parameter spaces: A review. *Festschrift for Herman Rubin, IMS Lecture Notes-Monograph Series*, **45**, 21–44.

Marchand, E. and Strawderman, W. E. (2005). Improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Ann. Inst. Statist. Math.*, **57**, 129–143.

Rukhin, A. (1992). Asymptotic risk behavior of mean vector and variance estimators and the problem of positive normal mean. *Ann. Inst. Statist. Math.*, **44**, 299–311.

Rukhin, A. (1995). Admissibility: Survey of a concept in progress. *International Statistical Review*, **63**, 95–115.

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, **16**, 155–160.

TSUKUMA, H. and KUBOKAWA, T. (2008). Stein's phenomenon in estimation of means restricted to a polyhedral convex cone. *J. Multivariate Analysis*, **99**, 141–164.