CIRJE-F-570

Integral Inequality for Minimaxity in the Stein Problem

Tatsuya Kubokawa
University of Tokyo

June 2008; Revised in August 2009

J. Japan Statist. Soc.
Vol. 3? No. ? 200? 1–1

INTEGRAL INEQUALITY FOR MINIMAXITY IN THE STEIN PROBLEM

Tatsuya Kubokawa*

In the estimation of a multivariate normal mean, it is shown that the problem of deriving shrinkage estimators improving on the maximum likelihood estimator can be reduced to that of solving an integral inequality. The integral inequality not only provides a more general condition than a conventional differential inequality studied in the literature, but also handles non-differentiable or discontinuous estimators. The paper also gives general conditions on prior distributions such that the resulting generalized Bayes estimators are minimax. Finally, a simple proof for constructing a class of estimators improving on the James-Stein estimator is given based on the integral expression of the risk.

Key words and phrases: Decision theory, differential inequality, estimation, inadmissibility, integral inequality, James-Stein estimator, linear regression model, normal distribution, regression coefficients, risk function, uniform domination.

1. Introduction

One of the most attractive topics in theoretical statistics is the Stein problem in the estimation of a mean vector of a multivariate normal distribution. A considerable amount of studies have been developed since Stein (1956) and James and Stein (1961) discovered the inadmissibility of the maximum likelihood estimator (MLE) when the dimension of the mean vector is larger than or equal to three. This phenomenon of the admissibility and the inadmissibility has been studied by Brown (1971), Johnstone (1984) and Eaton (2004) in the relation with the recurrence of a Markov chain and a diffusion process. Since the MLE is a minimax estimator with a constant risk, the problem of finding estimators improving on the MLE is equivalent to that of deriving minimax estimators, and Stein (1973, 81) showed that this problem can be reduced to solving a differential inequality. A powerful tool used there is the so-called Stein identity, namely the integration by parts in a normal distribution. The identity has been extended to discrete and continuous exponential families by Hudson (1978) as well as to the Wishart distribution by Stein (1977) and Haff (1979). Those identities provided differential inequalities for finding improved estimators in various estimation problems. Differential inequalities are quite useful, but require absolute continuity of shrinkage functions. Instead of the differential inequality, in this paper, we propose an integral inequality for deriving the minimax estimators, which can not only provide more general conditions for the minimaxity, but also eliminate the continuity condition for the shrinkage function. Also it can be used to obtain general conditions on prior distributions such that the resulting

Received ??. Revised ??. Accepted ??, ??.
*Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN.

generalized Bayes estimators are minimax.

To explain the outlines of the paper, we here describe the model and the estimation problem. Let $\boldsymbol{X} = (X_1, \ldots, X_p)^t$ and S be mutually independent random variables distributed as

$$\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I}_p) \quad \text{and} \quad S/\sigma^2 \sim \chi^2_n,$$

where $\mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I}_p)$ denotes a p-variate normal distribution with mean $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^t$ and covariance matrix $\sigma^2 \boldsymbol{I}_p$ for the $p \times p$ identity matrix $\boldsymbol{I}_p$, and χ^2_n denotes a chi-square distribution with n degrees of freedom. This is a canonical form of a linear regression model. The problem of estimating the mean vector $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}$ is considered relative to the quadratic loss

$$L(\omega, \widehat{\boldsymbol{\theta}}) = \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2/\sigma^2 = (\hat{\theta} - \boldsymbol{\theta})^t(\hat{\theta} - \boldsymbol{\theta})/\sigma^2,$$

for $\omega = (\boldsymbol{\theta}, \sigma^2)$, unknown parameters. Estimator $\widehat{\boldsymbol{\theta}}$ is evaluated in terms of the risk function $R(\omega, \widehat{\boldsymbol{\theta}}) = E_\omega[L(\omega, \widehat{\boldsymbol{\theta}})]$ where $E_\omega[\cdot]$ is the expectation with respect to $\boldsymbol{X}$ and S. It is noted that $(\boldsymbol{X}, S)$ is the minimal sufficient statistic for ω, so that the mean vector $\boldsymbol{\theta}$ is estimated based on $(\boldsymbol{X}, S)$.

The maximum likelihood estimator of $\boldsymbol{\theta}$ is $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{X}$. Since it is minimax with the constant risk $R(\omega, \widehat{\boldsymbol{\theta}}_0) = p$, improving on $\widehat{\boldsymbol{\theta}}_0$ is equivalent to deriving minimax estimators but $\widehat{\boldsymbol{\theta}}_0$. To find a minimax estimator, Stein (1956) considered a class of the estimators

$$\widehat{\boldsymbol{\theta}}_\psi = (1 - \psi(W)/W)\, \boldsymbol{X}, \quad \text{for} \quad W = \|\boldsymbol{X}\|^2/S,$$

where $\psi(w)$ is a nonnegative function. In fact, out of the class, James and Stein (1961) found the estimator

$$\widehat{\boldsymbol{\theta}}^{JS} = (1 - a_0/W)\, \boldsymbol{X}, \quad \text{for} \quad a_0 = (p-2)/(n+2), \tag{1.1}$$

and established that if $p \geq 3$, then the James-Stein estimator $\widehat{\boldsymbol{\theta}}^{JS}$ dominates $\widehat{\boldsymbol{\theta}}_0$ for $p \geq 3$, namely, $R(\omega, \widehat{\boldsymbol{\theta}}^{JS}) \leq R(\omega, \widehat{\boldsymbol{\theta}}_0)$ for any ω and the strict inequality holds for some ω. Since $\widehat{\boldsymbol{\theta}}_0$ is minimax, the James-Stein estimator $\widehat{\boldsymbol{\theta}}^{JS}$ is minimax for $p \geq 3$. To characterize minimax shrinkage estimators, Efron and Morris (1976) used the Stein identity and the chi-square identity to derive an unbiased estimator $\widehat{R}(\widehat{\boldsymbol{\theta}}_\psi)$ of the risk function $R(\omega, \widehat{\boldsymbol{\theta}}_\psi)$, namely $R(\omega, \widehat{\boldsymbol{\theta}}_\psi) = E_\omega[\widehat{R}(\widehat{\boldsymbol{\theta}}_\psi)]$, where

$$\widehat{R}(\widehat{\boldsymbol{\theta}}_\psi) = p + \{(n+2)\psi(W) - 2(p-2)\}\, \psi(W)/W - 4\psi'(W) - 4\psi(W)\psi'(W).$$

This implies that the estimator $\widehat{\boldsymbol{\theta}}_\psi$ is minimax if the function $\psi(w)$ satisfies the differential inequality:

$$\mathcal{D}(w) \equiv (n+2)\, \{\psi(w) - 2a_0\}\, \psi(w)/w - 4\psi'(w) - 4\psi(w)\psi'(w) \leq 0. \tag{1.2}$$

Such an approach to finding improved estimators is generally called *unbiased-estimator-of-risk method*. Some solutions of the differential inequality (1.2) have been provided by Baranchick (1970), Alam (1973) and Efron and Morris (1976).

In this paper, we obtain an expression of the risk function $R(\omega, \widehat{\boldsymbol{\theta}}_\psi)$ based on an integral and derive an integral inequality for improvement on $\widehat{\boldsymbol{\theta}}_0$. As derived in Section 2., it is shown that $\widehat{\boldsymbol{\theta}}_\psi$ is minimax if $\psi(w)$ satisfies the integral inequality

$$\mathcal{I}(w) \equiv \psi^2(w) + 2\psi(w) - (n+p)\int_0^1 z^{n/2}\psi(w/z)\mathrm{d}z \le 0,$$

which does not assume the continuity nor the differentiability of $\psi(w)$. Section 2. presents some examples of non-differentiable or discontinuous minimax estimators. Several sufficient conditions for the integral inequality $\mathcal{I}(w) \le 0$ are given, and the relation between the integral and differential inequalities is clarified, namely, the differential inequality $\mathcal{D}(w) \le 0$ implies the integral inequality $\mathcal{I}(w) \le 0$. In particular, the integral inequality gives a useful condition for the minimaxity when $w^c\psi(w)$ is non-decreasing in w for $c \ge 0$.

In Section 3., we investigate the minimaxity of the generalized Bayes estimators in similar prior distributions as in Wells and Zhou (2008), who derived the general and nice conditions on the priors. Their interesting point is that for the shrinkage function $\psi_\pi(w)$ of the generalized Bayes estimator, their arguments can handle the case that $\psi_\pi(w)$ is not increasing, but $w^c\psi_\pi(w)$ is increasing for some $c > 0$. In this paper, we reexamine the minimaxity of the generalized Bayes estimators under the similar setup as in Wells and Zhou (2008), and obtain more general and slightly better conditions.

In Section 4., we give another simple proof for constructing a class of equivariant estimators $\widehat{\boldsymbol{\theta}}_\psi$ improving on the James-Stein estimator. This class was given by Kubokawa (1994) without a proof, because the proof is complicated. Using the integaral expression given in Section 2., we can prove it more simply,

2. Integral Inequality for the Minimaxity

2.1. Derivation of the integral inequality

We shall derive a condition based on the integral inequality under which the shrinkage estimator $\widehat{\boldsymbol{\theta}}_\psi$ is minimax in terms of the risk $R(\omega, \widehat{\boldsymbol{\theta}}_\psi)$. To this end, define $\mathcal{I}(w)$ by

$$\mathcal{I}(w) = \psi^2(w) + 2\psi(w) - (n+p)\int_0^1 z^{n/2}\psi(w/z)\mathrm{d}z. \tag{2.1}$$

Then the risk function can be expressed based on the function $\mathcal{I}(w)$.

Theorem 1. *Assume that $\psi(w)$ is a nonnegative and bounded function and $p \ge 3$. Then the risk function of the estimator $\widehat{\boldsymbol{\theta}}_\psi$ is expressed as*

$$R(\omega, \widehat{\boldsymbol{\theta}}_\psi) = p + E\left[\frac{S}{\sigma^2 W}\mathcal{I}(W)\right]. \tag{2.2}$$

Proof. The risk function of the estimator $\widehat{\boldsymbol{\theta}}_\psi$ is written as

$$R(\omega, \widehat{\boldsymbol{\theta}}_\psi) = p - 2E\left[\frac{\psi(W)}{\sigma^2 W}\boldsymbol{X}^t(\boldsymbol{X}-\boldsymbol{\theta})\right] + E\left[\frac{S}{\sigma^2}\frac{\psi^2(W)}{W}\right]. \tag{2.3}$$

To evaluate the second term in the right hand side of this equality, we use the Stein identity given by Stein (1973) and the chi-square identity given by Efron and Morris (1976), which are, respectively, given by

$$E\left[(X_i-\theta_i)h(\boldsymbol{X})\right] = \sigma^2 E\left[\frac{\partial h(\boldsymbol{X})}{\partial X_i}\right], \tag{2.4}$$

$$E[\varphi(S)S] = \sigma^2 E[n\varphi(S) + 2S\varphi'(S)], \tag{2.5}$$

where $h(\cdot)$ and $\varphi(\cdot)$ are absolutely continuous functions and all the expectations are assumed to be finite.

Define a function $\Psi(W)$ by

$$\Psi(W) = \frac{1}{2W}\int_0^1 z^{n/2}\psi(W/z)\mathrm{d}z = \frac{W^{n/2}}{2}\int_W^\infty \frac{\psi(t)}{t^{n/2+2}}\mathrm{d}t,$$

where the transformation $t = W/z$ is made at the second equation. Using the chi-square identity (2.5), we obtain the equations given by

$$E^{S|\boldsymbol{X}}[\Psi(W)S] = \sigma^2 E^{S|\boldsymbol{X}}\left[n\Psi(W) + 2S\frac{\partial}{\partial S}\Psi(W)\right] = \sigma^2 E^{S|\boldsymbol{X}}\left[\frac{\psi(W)}{W}\right],$$

where $E^{S|\boldsymbol{X}}[\cdot]$ denotes the conditional expectation with respect to S given $\boldsymbol{X}$, and all the expectations are finite since $\psi(w)$ is bounded. It is interesting to note that $\Psi(W)$ is a solution of the differential equation $n\varphi(S)+2S\varphi'(S) = \psi(W)/W$ with respect to $\varphi(S)$ as a function of S.

Using the equation in the previous display, we can rewrite the cross product term in (2.3) as

$$\begin{aligned} E\left[\frac{\psi(W)}{W}\frac{\boldsymbol{X}^t(\boldsymbol{X}-\boldsymbol{\theta})}{\sigma^2}\right] =& E\left[\frac{S}{\sigma^2}\Psi(W)\frac{\boldsymbol{X}^t(\boldsymbol{X}-\boldsymbol{\theta})}{\sigma^2}\right] \\ =& E\left[\frac{S}{\sigma^2}\frac{\boldsymbol{X}^t(\boldsymbol{X}-\boldsymbol{\theta})}{\sigma^2}\frac{W^{n/2}}{2}\int_W^\infty \frac{\psi(t)}{t^{n/2+2}}\mathrm{d}t\right]. \end{aligned}$$

Then, the Stein identity (2.4) is applied to this expectation and we observe that

$$\begin{aligned} & E\left[\frac{S}{\sigma^2}\Psi(W)\frac{\boldsymbol{X}^t(\boldsymbol{X}-\boldsymbol{\theta})}{\sigma^2}\right] \\ &= E\left[\frac{S}{\sigma^2}\left\{p\Psi(W) + 2W\Psi'(W)\right\}\right] \\ &= E\left[\frac{S}{\sigma^2}\left\{p\Psi(W) + \frac{n}{2}W^{n/2}\int_W^\infty \frac{\psi(t)}{t^{n/2+2}}\mathrm{d}t - W^{n/2+1}\frac{\psi(W)}{W^{n/2+2}}\right\}\right] \\ &= E\left[\frac{S}{\sigma^2}\left\{(p+n)\Psi(W) - \frac{\psi(W)}{W}\right\}\right]. \end{aligned} \tag{2.6}$$

Substituting (2.6) into (2.3) and recalling the definition of $\Psi(W)$, we obtain the expression (2.2) in Theorem 1. ■

The condition for the minimaxity of the estimator $\widehat{\boldsymbol{\theta}}_\psi$ can be directly provided from Theorem 1.

Proposition 1. *Assume that $\psi(w)$ is a nonnegative and bounded function and $p \geq 3$. Then the estimator $\widehat{\boldsymbol{\theta}}_\psi$ is minimax if $\psi(w)$ satisfies the integral inequality*

$$\mathcal{I}(w) \leq 0 \quad \text{for almost all } w > 0, \tag{2.7}$$

with respect to the Lebesgue measure.

An interesting point in Theorem 1 and Proposition 1 is that the continuity of the function $\psi(w)$ is not assumed. Since the function $\psi(w)$ is differentiable for the generalized Bayes estimators of $\boldsymbol{\theta}$, this result may not be important. However, a class of minimax estimators can be clearly extended.

The function $\mathcal{I}(w)$ has several variants. For instance, from (2.1.) it is written as

$$\mathcal{I}(w) = \psi^2(w) + 2\psi(w) - (n+p)w^{n/2+1} \int_w^\infty \frac{\psi(t)}{t^{n/2+2}} \mathrm{d}t. \tag{2.8}$$

It is also expressed as

$$\mathcal{I}(w) = \psi^2(w) - 2a_0\psi(w) - (n+p) \int_0^1 z^{n/2} \{\psi(w/z) - \psi(w)\} \mathrm{d}z,$$

for $a_0 = (p-2)/(n+2)$. If $\psi(w)$ is absolutely continuous, then it is further rewritten as

$$\mathcal{I}(w) = \psi^2(w) - 2a_0\psi(w) - 2\frac{n+p}{n+2} w \int_0^1 z^{n/2-1}\psi'(w/z)\mathrm{d}z, \tag{2.9}$$

since

$$\begin{aligned}\int_0^1 z^{n/2-1}\psi'(w/z)\mathrm{d}z &= \int_0^1 \left\{-\frac{z^{n/2+1}}{w}\right\} \frac{\mathrm{d}}{\mathrm{d}z}\psi(w/z)\mathrm{d}z \\ &= -\frac{1}{w}\psi(w) + \frac{n+2}{2w}\int_0^1 z^{n/2}\psi(w/z)\mathrm{d}z,\end{aligned}$$

where $\int_0^1 z^{n/2-1}|\psi'(w/z)|\mathrm{d}z$ is assumed to be finite.

2.2. Relations with the differential inequality

In the previous subsection, we have obtained the condition based on the integral inequality (2.7) for the minimaxity of the shrinkage estimator $\widehat{\boldsymbol{\theta}}_\psi$. We here provide general sufficient conditions for the integral inequality and clarify the relation between the integral inequality and the differential inequality (1.2) for the minimaxity.

Proposition 2. *Assume that $\psi(w)$ is a nonnegative and bounded function and $p \geq 3$. Then $\mathcal{I}(w)$ is expressed as*

$$\mathcal{I}(w) = (n+p)\int_0^1 z^{n/2}\left\{F(w) - F(w/z)\right\}\frac{\psi(w/z)}{F(w/z)}\mathrm{d}z, \tag{2.10}$$

where

$$\begin{aligned} F(w) =& \frac{\psi(w)\{\psi(w)+2\}}{w^{n/2+1}}\exp\left\{\int_1^w \frac{n+p}{t\{\psi(t)+2\}}\mathrm{d}t\right\} \\ =& \psi(w)\{\psi(w)+2\}\exp\left\{\frac{n+2}{2}\int_1^w \frac{2a_0-\psi(t)}{t\{\psi(t)+2\}}\mathrm{d}t\right\}. \end{aligned} \tag{2.11}$$

If $F(w)$ is nondecreasing for almost all w, then $\mathcal{I}(w) \leq 0$ for almost all w.

Proof. From (2.1), $\mathcal{I}(w)$ can be rewritten as

$$\begin{aligned} \mathcal{I}(w) =& F(w)\left\{\{\psi(w)+2\}\frac{\psi(w)}{F(w)} - (n+p)\int_0^1 z^{n/2}\frac{\psi(w/z)}{F(w/z)}\mathrm{d}z\right\} \\ &+ (n+p)\int_0^1 z^{n/2}\left\{F(w)-F(w/z)\right\}\frac{\psi(w/z)}{F(w/z)}\mathrm{d}z \\ =& I_1 + I_2, \quad \text{(say)} \end{aligned}$$

so that it is sufficient to show that $I_1 = 0$. Making the transformation $t = w/z$ and noting the definition of $F(w)$, we observe that

$$\begin{aligned} \int_0^1 z^{n/2}\frac{\psi(w/z)}{F(w/z)}\mathrm{d}z =& w^{n/2+1}\int_w^\infty \frac{1}{t^{n/2+2}}\frac{\psi(t)}{F(t)}\mathrm{d}t \\ =& w^{n/2+1}\int_w^\infty \frac{1}{t\{\psi(t)+2\}}\exp\left\{-\int_1^t \frac{n+p}{s\{\psi(s)+2\}}\mathrm{d}s\right\}\mathrm{d}t, \end{aligned}$$

which is equal to

$$\begin{aligned} &-\frac{w^{n/2+1}}{n+p}\int_w^\infty \frac{\mathrm{d}}{\mathrm{d}t}\exp\left\{-\int_1^t \frac{n+p}{s\{\psi(s)+2\}}\mathrm{d}s\right\}\mathrm{d}t \\ &= -\frac{w^{n/2+1}}{n+p}\left[\exp\left\{-\int_1^t \frac{n+p}{s\{\psi(s)+2\}}\mathrm{d}s\right\}\mathrm{d}t\right]_{t=w}^{\infty}. \end{aligned} \tag{2.12}$$

Since $\psi(w)$ is bounded and nonnegative, it is noted that $\lim_{t\to\infty}\int_1^t [s\{\psi(s)+2\}]^{-1}\mathrm{d}s = \infty$. Hence, (2.12) is equal to

$$\frac{w^{n/2+1}}{n+p}\exp\left\{-\int_1^w \frac{n+p}{s\{\psi(s)+2\}}\mathrm{d}s\right\} = \frac{\psi(w)+2}{n+p}\frac{\psi(w)}{F(w)},$$

which means that $I_1 = 0$. We thus obtain the expression (2.10). The second equality in (2.11) can be shown by noting that

$$\begin{aligned} \frac{n+2}{2}\int_1^w \frac{2a_0-\psi(t)}{t\{\psi(t)+2\}}\mathrm{d}t =& \int_1^w \frac{p-2-(n+2)(\psi(t)+2-2)/2}{t\{\psi(t)+2\}}\mathrm{d}t \\ =& \int_1^w \frac{n+p}{t\{\psi(t)+2\}}\mathrm{d}t - \frac{n+2}{2}\int_1^w \frac{1}{t}\mathrm{d}t. \end{aligned}$$

From this expression, it follows that $\mathcal{I}(w) \leq 0$ if $F(w)$ is nondecreasing, and Proposition 2 is proved. ■

We here assume that $\psi(w)$ is absolutely continuous. Under this assumption, the relationship between the differential inequality (1.2) and the integral inequality (2.7) can be clarified.

Proposition 3. *Assume that $\psi(w)$ is a nonnegative, bounded and absolutely continuous function and $p \geq 3$. Then the function* (2.1) *is rewritten as*

$$\mathcal{I}(w) = \frac{w^{n/2+1}}{2} \int_w^\infty \frac{1}{t^{n/2+1}} \mathcal{D}(t) \mathrm{d}t, \tag{2.13}$$

where $\mathcal{D}(w)$ is given by (1.2). *For $F(w)$ defined in* (2.11), *the derivative of $F(w)$ is expressed as*

$$F'(w) = -\frac{F(w)}{2\psi(w)\{\psi(w)+2\}} \mathcal{D}(w). \tag{2.14}$$

If the inequality $\mathcal{D}(w) \leq 0$ is satisfied, then $F(w)$ is nondecreasing, so that the inequality $\mathcal{I}(w) \leq 0$ holds.

Proof. It is noted that

$$\begin{aligned}
\psi^2(w) + 2\psi(w) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}z} \Big\{ z^{n/2+1} \left[\psi^2(w/z) + 2\psi(w/z) \right] \Big\} \mathrm{d}z \\
&= \int_0^1 \frac{z^{n/2}}{2} \big\{ (n+2) \left[\psi^2(w/z) + 2\psi(w/z) \right] \\
&\qquad - 4 \left[\psi(w/z) + 1 \right] \psi'(w/z) w/z \big\} \mathrm{d}z.
\end{aligned} \tag{2.15}$$

Combining (2.1) and (2.15) gives that

$$\begin{aligned}
\mathcal{I}(w) = \int_0^1 \frac{z^{n/2}}{2} \big\{ &(n+2)\psi^2(w/z) - 2(p-2)\psi(w/z) \\
&- 4[\psi(w/z)+1]\psi'(w/z)w/z \big\} \mathrm{d}z,
\end{aligned}$$

which is equal to (2.13). The expression (2.14) of $F'(w)$ can be derived by differentiating $\log F(w)$ with respect to w where the function $F(w)$ is given by the r.h.s. of the second equality in (2.11). Hence, Proposition 3 is proved. ■

When $\psi(w)$ is an absolutely continuous function satisfying that $0 < \psi(w) < 2a_0$, Efron and Morris (1976) showed that $\mathcal{D}(w) \leq 0$ if and only if $M(w)$ is nondecreasing, where

$$M(w) = w^{p/2-1} \psi(w) / \{2a_0 - \psi(w)\}^{1+a_0}.$$

Combining this result and Proposition 3, we can clarify the relations between the sufficient conditions for the minimaxity illuatrated below, where $\psi(w) \uparrow$ means that $\psi(w)$ is nondecreasing.

$$\begin{array}{ccccccc} \psi(w)\uparrow & \Longrightarrow & \mathcal{D}(w)\le 0 & \Longrightarrow & F(w)\uparrow & \Longrightarrow & \mathcal{I}(w)\le 0 \\ & & \Updownarrow & & & & \Downarrow \\ & & M(w)\uparrow & & & & \text{minimax} \end{array}$$

We conclude this section with notes on the differential inequality. Assuming that $\psi(w)$ is absolutely continuous, we can apply the chi-square identity (2.5) to the expectation (2.2) with the expression (2.8) and get

$$\begin{aligned} R(\omega,\widehat{\boldsymbol{\theta}}_\psi)-p =& E\left[\frac{S}{\sigma^2 W}\mathcal{I}(W)\right] \\ =& E\left[\frac{n}{W}\mathcal{I}(W)+2S\left(-\frac{\|\boldsymbol{X}\|^2}{S^2}\right)\frac{\mathrm{d}}{\mathrm{d}W}\left\{W^{-1}\mathcal{I}(W)\right\}\right]=E\left[\mathcal{D}(W)\right], \end{aligned}$$

which is the expression derived by Efron and Morris (1976). This means that $p+\mathcal{D}(W)$ is an unbiased estimator of the risk function $R(\omega,\widehat{\boldsymbol{\theta}}_\psi)$. Remark 4 and Example 3 in Efron and Morris (1976) treated the discontinuous case for the function $\psi(w)$ and stated that the condition (1.2) can be extended to the discontinuous case by using a delta function at points of discontinuity of $\psi(w)$. Theorem 1 and Proposition 1 provide the general conditions which can cover the discontinuous case.

2.3. Useful conditions for the minimaxity and simple examples

We now provide useful conditions for the minimaxity which can be derived from the integral inequality $\mathcal{I}(w)\le 0$. One of them can be obtained under the assumption that $w^c\psi(w)$ is nondecreasing in w for a nonnegative constant c. Then, it is observed that

$$\begin{aligned} \int_0^1 z^{n/2}\psi(w/z)\mathrm{d}z =& \int_0^1 z^{n/2}(z/w)^c\{(w/z)^c\psi(w/z)\}\mathrm{d}z \\ \ge& \int_0^1 z^{n/2}(z/w)^c\{w^c\psi(w)\}\mathrm{d}z \\ =& \int_0^1 z^{n/2+c}\mathrm{d}z\psi(w)=\frac{2}{n+2+2c}\psi(w), \end{aligned}$$

since $(w/z)^c\psi(w/z)\ge w^c\psi(w)$ for $0<z<1$. From (2.1), we have

$$\begin{aligned} \mathcal{I}(w) \le& \psi^2(w)+2\psi(w)-2\frac{n+p}{n+2+2c}\psi(w) \\ =& \psi^2(w)-2\frac{p-2-2c}{n+2+2c}\psi(w), \end{aligned}$$

which implies the following proposition.

Proposition 4. *Assume that for $p\ge 3$ and a constant $c\ge 0$, the function $\psi(w)$ satisfies the conditions for any $w>0$:*

(a) *$w^c\psi(w)$ is nondecreasing in w,*

(b) *$0\le\psi(w)\le 2(p-2-2c)/(n+2+2c)$.*

Then, the estimator $\widehat{\boldsymbol{\theta}}_\psi$ is minimax.

This is a variant of Alam (1973) for known variance case and very useful for checking the minimaxity of the generalized Bayes estimators in the next section. Wells and Zhou (2008) derived another variant where their condition (b) can be written as

(WZ-b) $0 \le \psi(w) \le 2(p-2-2c)/(n+2+4c)$,

from Lemma 4.1 of Wells and Zhou (2008). This shows that our condition (b) is slightly better than (WZ-b) for $c > 0$. When $c = 0$, the conditions (a) and (b) correspond to those of Baranchik (1970).

Example 1 (Discontinuous case) Let us partition the positive real line into k intervals as $r_0 = 0 < r_1 < \cdots < r_{k-1} < r_k = \infty$. For a nondecreasing function $d(r)$, consider a step function of the form

$$\psi_d(w) = \sum_{i=0}^{k-1} d(r_i) I(r_i \le w < r_{i+1}),$$

where $I(r_i \le w < r_{i+1})$ is the indicator function such that $I(r_i \le w < r_{i+1}) = 1$ on the interval $[r_i, r_{i+1})$ and $I(r_i \le w < r_{i+1}) = 0$ otherwise. As examples of $d(r)$, we can consider the two functions

$$d_1(r) = \min\{r, a_0\},$$
$$d_2(r) = a_0 - \frac{2}{n+2}\left[\int_0^1 \frac{(1+r)^{(p+n)/2}}{(1+rz)^{(p+n)/2+1}} z^{p/2-2} \mathrm{d}z\right]^{-1},$$

for $a_0 = (p-2)/(n+2)$. For $d_2(r)$, $\psi_{d_2}(w)$ is identical to the function $\psi_0(w)$ given in (4.1), and the estimator with the discontinuous function $\psi_{d_2}(w)$ was treated in Kubokawa (1994) in the process of deriving the Brewster-Zidek type estimator. Since $\psi_{d_i}(w)$ satisfies the conditions (a) and (b) of Proposition 4 for $c = 0$, the corresponding discontinuous estimators are minimax. ■

The integral inequality can provide another simple sufficient condition for the minimaxity described below.

Proposition 5. *Assume that the function $\psi(w)$ is essentially bounded. Define the essential infimum of $\psi(w)$ on the set $\{x|w < x\}$ by $\mathrm{ess\,inf}_{x>w}\psi(x)$. If $\psi(w)$ satisfies the inequality*

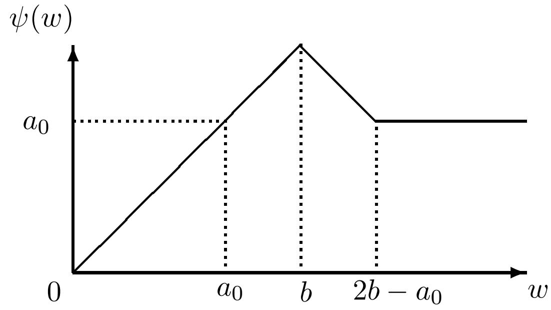
$$\mathcal{I}_*(w) = \psi^2(w) + 2\psi(w) - 2(1+a_0)\ \mathrm{ess\,inf}_{x>w}\psi(x) \le 0, \tag{2.16}$$

then the estimator $\widehat{\boldsymbol{\theta}}_\psi$ is minimax.

Example 2 (Partly decreasing case) Let us consider a non-differentiable function of the form

$$\psi(w) = \begin{cases} w & \text{if} \quad w \le b, \\ -w + 2b & \text{if} \quad b < w \le 2b - a_0, \\ a_0 & \text{if} \quad 2b - a_0 < w, \end{cases}$$

for a positive constant $b(\geq a_0)$. The function $\psi(w)$ is partly decreasing as illustrated in the following figure:



In the case of $w \leq a_0$, the function $\mathcal{I}_*(w)$ given by (2.16) is written by

$$\mathcal{I}_*(w) = \psi^2(w) + 2\psi(w) - 2(1 + a_0)\psi(w) = \psi^2(w) - 2a_0\psi(w),$$

which is not positive. It is also seen that $\mathcal{I}_*(w) = -a_0^2 < 0$ for $w > 2b - a_0$. In the case of $a_0 < w \leq 2b - a_0$, it is noted that $a_0 \leq \psi(w) \leq b$, so that

$$\mathcal{I}_*(w) = \psi^2(w) + 2\psi(w) - 2(1 + a_0)a_0 \leq b^2 + 2b - 2(1 + a_0)a_0,$$

which is not positive if b satisfies the inequality

$$0 < b \leq -1 + (1 + a_0)\sqrt{1 + a_0^2/(1 + a_0)^2}.$$

Hence, the estimator $\widehat{\boldsymbol{\theta}}_\psi$ is minimax under this condition on b. ■

3. Minimaxity of the Generalized Bayes Estimators

We now derive conditions for minimaxity of the generalized Bayes estimators. Wells and Zhou (2008), hereafter abbreviated by W&Z, recently developed nice results for the minimaxity, and we use their arguments and Proposition 4 to obtain slightly improved conditions for the minimaxity.

For the model and the prior assumption, a similar setup as in W&Z is used here where their notation m is n in ours notation: The model is expressed as $\boldsymbol{X}|(\boldsymbol{\theta}, \eta) \sim \mathcal{N}_p(\boldsymbol{\theta}, \eta^{-2}\boldsymbol{I}_p)$ and $S|\eta \sim S^{n/2-1}\eta^n \exp\{-S\eta^2\}$, and the prior distribution $\pi(\boldsymbol{\theta}, \eta)$ is given by

$$\boldsymbol{\theta}|(\nu, \eta) \sim \mathcal{N}_p(\mathbf{0}, \nu\eta^{-2}\boldsymbol{I}_p), \quad \nu \sim h(\nu)I(\nu \geq \nu_0), \quad \eta \sim \eta^{-K},$$

where $I(\nu \geq \nu_0)$ is the indicator function for a nonnegative and known constant ν_0. It is assumed that K and $h(\nu)$ satisfy the following conditions:

(C1) $K > 0$ and $A \equiv -K + (p + n + 3)/2 > 1$,
(C2) $\int_0^{\lambda_0} \lambda^{p/2-2}h((1 - \lambda)/\lambda)\mathrm{d}\lambda < \infty$ for $\lambda = 1/(1 + \nu)$ and $\lambda_0 = 1/(1 + \nu_0)$,
(C3) $\lim_{\lambda\to 0} \lambda^{p/2-1}h((1 - \lambda)/\lambda) = \lim_{\nu\to\infty} h(\nu)/(1 + \nu)^{p/2-1} = 0$, and
(C4) $h(\nu)$ is differentiable.

Then the generalized Bayes estimator of $\boldsymbol{\theta}$ is given by

$$\widehat{\boldsymbol{\theta}}_{\pi}^{GB}(\boldsymbol{X}, S) = (1 - \psi_{\pi}(W)/W)\boldsymbol{X},$$

where $W = \|\boldsymbol{X}\|^2/S$ and

$$\begin{aligned}\psi_{\pi}(w) =& w\frac{\int_0^{\lambda_0} \lambda^{p/2-1}(1+\lambda W)^{-A}h((1-\lambda)/\lambda)\mathrm{d}\lambda}{\int_0^{\lambda_0} \lambda^{p/2-2}(1+\lambda W)^{-A}h((1-\lambda)/\lambda)\mathrm{d}\lambda}\\ =& w\frac{\int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda W)^{-A}H(\lambda)\mathrm{d}\lambda}{\int_0^{\lambda_0} \lambda^{\alpha}(1+\lambda W)^{-A}H(\lambda)\mathrm{d}\lambda},\end{aligned}$$

for $\alpha = p/2 - 2$ and $H(\lambda) = h((1-\lambda)/\lambda)$.

To derive the conditions for the minimaxity of $\widehat{\boldsymbol{\theta}}_{\pi}^{GB}(\boldsymbol{X}, S)$, we prepares two lemmas which were implicitly given in W&Z. Let

$$C(w) = \frac{\int_0^{\lambda_0} \lambda^{\alpha+2}(1+\lambda w)^{-A}H'(\lambda)\mathrm{d}\lambda}{\int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda} - \frac{\int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda w)^{-A}H'(\lambda)\mathrm{d}\lambda}{\int_0^{\lambda_0} \lambda^{\alpha}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda}. \tag{3.1}$$

Then we can get an inequality used for showing the monotonicity of $\psi_{\pi}(w)$.

Lemma 1. *Assume the conditions (C1)-(C4). For a nonnegative constant c, the following inequality holds:*

$$\frac{\mathrm{d}}{\mathrm{d}w}\left\{w^c\psi_{\pi}(w)\right\} \geq w^{c-1}\psi_{\pi}(w)\left\{c - C(w)\right\}. \tag{3.2}$$

Proof. Differentiating $w^c\psi_{\pi}(w)$ with respect to w gives that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}w}\left\{w^c\psi_{\pi}(w)\right\} =& w^{c-1}\psi_{\pi}(w)\Big\{c + 1 - Aw\frac{\int_0^{\lambda_0} \lambda^{\alpha+2}(1+\lambda w)^{-A-1}H(\lambda)\mathrm{d}\lambda}{\int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda}\\ &+ Aw\frac{\int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda w)^{-A-1}H(\lambda)\mathrm{d}\lambda}{\int_0^{\lambda_0} \lambda^{\alpha}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda}\Big\}.\end{aligned}$$

By the integration by part, it is observed that

$$\begin{aligned}Aw\int_0^{\lambda_0} \lambda^{\alpha+2}(1+\lambda w)^{-A-1}H(\lambda)\mathrm{d}\lambda =& -\left[\lambda^{\alpha+2}(1+\lambda w)^{-A}H(\lambda)\right]_{\lambda=0}^{\lambda_0}\\ + (\alpha+2)\int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda &+ \int_0^{\lambda_0} \lambda^{\alpha+2}(1+\lambda w)^{-A}H'(\lambda)\mathrm{d}\lambda.\end{aligned} \tag{3.3}$$

We can get a similar equality about $Aw\int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda w)^{-A-1}H(\lambda)\mathrm{d}\lambda$. Using these equalities and the condition (C3), we can get the inequality (3.2). ■

Let

$$D(w) = \int_0^{\lambda_0} \lambda^{\alpha+1}(1+\lambda w)^{-A+1}H'(\lambda)\mathrm{d}\lambda \Big/ \int_0^{\lambda_0} \lambda^{\alpha}(1+\lambda w)^{-A+1}H(\lambda)\mathrm{d}\lambda. \tag{3.4}$$

Then we can get an inequality about the bound of $\psi_{\pi}(w)$.

Lemma 2. *Assume the conditions (C1)-(C4). Then, the following inequality holds:*

$$\psi_\pi(w) \le \frac{p/2-1+D(w)}{A-p/2-D(w)}. \tag{3.5}$$

Proof. It is observed that

(3.6)
$$\psi_\pi(w) = \frac{w\int_0^{\lambda_0}\lambda^{\alpha+1}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda / \int_0^{\lambda_0}\lambda^{\alpha}(1+\lambda w)^{-A+1}H(\lambda)\mathrm{d}\lambda}{1-w\int_0^{\lambda_0}\lambda^{\alpha+1}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda / \int_0^{\lambda_0}\lambda^{\alpha}(1+\lambda w)^{-A+1}H(\lambda)\mathrm{d}\lambda}.$$

Using the similar equality to (3.3) gives that

$$\begin{aligned} w\frac{\int_0^{\lambda_0}\lambda^{\alpha+1}(1+\lambda w)^{-A}H(\lambda)\mathrm{d}\lambda}{\int_0^{\lambda_0}\lambda^{\alpha}(1+\lambda w)^{-A+1}H(\lambda)\mathrm{d}\lambda} =& -\frac{1}{A-1}\frac{\lambda_0^{\alpha+1}(1+\lambda_0 w)^{-A+1}H(\lambda_0)}{\int_0^{\lambda_0}\lambda^{\alpha}(1+\lambda w)^{-A+1}H(\lambda)\mathrm{d}\lambda} \\ &+\frac{\alpha+1}{A-1}+\frac{1}{A-1}D(w) \\ \le& \frac{\alpha+1}{A-1}+\frac{1}{A-1}D(w). \end{aligned}$$

Substituting this into (3.6), we get

$$\psi_\pi(w) \le \frac{\alpha+1+D(w)}{A-1-(\alpha-1)-D(w)},$$

which gives the inequality (3.5). ■

Using Lemmas 1, 2 and Proposition 4, we can obtain the conditions on $h(\cdot)$ for the minimaxity of the generalized Bayes estimator $\widehat{\boldsymbol{\theta}}_\pi^{GB}$. Following the notations as in W&Z, let

$$g(\nu) = -(1+\nu)h'(\nu)/h(\nu), \quad \text{for} \quad \nu = (1-\lambda)/\lambda, \tag{3.7}$$

where $\nu > \nu_0$ for $\nu_0 = (1-\lambda_0)/\lambda_0$. Since $\lambda H'(\lambda)/H(\lambda) = -(1+\nu)h'(\nu)/h(\nu)$, the functions $C(w)$ and $D(w)$ are rewritted based on ν as

$$\begin{aligned} C(w) =& \frac{\int_{\nu_0}^\infty(1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A}g(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^\infty(1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A}h(\nu)\mathrm{d}\nu} \\ &-\frac{\int_{\nu_0}^\infty(1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A}g(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^\infty(1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A}h(\nu)\mathrm{d}\nu}, \\ D(w) =& \frac{\int_{\nu_0}^\infty(1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A+1}g(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^\infty(1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A+1}h(\nu)\mathrm{d}\nu}. \end{aligned}$$

Let

$$c_0 = \sup_w C(w) \quad \text{and} \quad d_0 = \sup_w D(w).$$

From Lemma 1, $w^{c_0}\psi_\pi(w)$ is nondecreasing in w for $c = c_0$. From Proposition 4, it is seen that the generalized Bayes estimator is minimax if $0 \leq \psi_\pi(w) \leq 2(p-2-2c_0)/(n+2+2c_0)$. Also from Lemma 2, it follows that

$$\psi_\pi(w) \leq \frac{p/2-1+d_0}{A-p/2-d_0} = \frac{p-2+2d_0}{n+3-2K-2d_0}.$$

Hence, the minimaxity is guaranteed if K, c_0 and d_0 satisfy the inequality

$$\frac{p-2+2d_0}{n+3-2K-2d_0} \leq 2\frac{p-2-2c_0}{n+2+2c_0}, \tag{3.8}$$

which can be rewritten by (3.9) in the following proposition.

Proposition 6. *Assume the conditions (C1)-(C4). Then, the generalized Bayes estimator* $\widehat{\boldsymbol{\theta}}^{GB}_\pi(\boldsymbol{X}, S)$ *is minimax if* K, c_0 *and* d_0 *satisfy the inequality*

$$(p-2)(n+4-4K) - 2(2n+p+4-4K)c_0 - 2(n+2p-2)d_0 + 4c_0d_0 \geq 0. \tag{3.9}$$

It is not easy to get the values c_0 and d_0, and we need to approximate them. We below consider the three cases of $g(\nu)$.

(Case 1). Consider the case that $|g(\nu)|$ is bounded. Then, from (3.1) and (3.4), it is seen that $C(w) \leq \sup_{\nu>\nu_0} g(\nu) - \inf_{\nu>\nu_0} g(\nu)$ and $D(w) \leq \sup_{\nu>\nu_0} g(\nu)$. In this case, let

$$c = c_1 = \sup_{\nu>\nu_0} g(\nu) - \inf_{\nu>\nu_0} g(\nu) \quad \text{and} \quad d_1 = \sup_{\nu>\nu_0} g(\nu),$$

and the minimaxity is guaranteed if K, c_1 and d_1 satisfy the inequality (3.9) where c_0 and d_0 are replaced with c_1 and d_1.

(Case 2). Consider the case that $g(\nu)$ can be decomposed as

$$g(\nu) = g_1(\nu) + g_2(\nu), \tag{3.10}$$

where $g_1(\nu)$ is nondecreasing in λ and $g_2(\nu)$ is nonnegative and bounded. This case was discussed in W&Z, and we can here derive a slightly better condition since Proposition 4 gives the slightly wider condition. From the monotonicity of $g_1(\nu)$, it can be shown that

$$\begin{aligned}&\frac{\int_{\nu_0}^\infty (1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A} g_1(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^\infty (1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A} h(\nu)\mathrm{d}\nu}\\ &\qquad \leq \frac{\int_{\nu_0}^\infty (1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A} g_1(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^\infty (1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A} h(\nu)\mathrm{d}\nu},\end{aligned}$$

so that from (3.1),

$$
\begin{aligned}
C(w) \leq & \frac{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A}g_2(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A}h(\nu)\mathrm{d}\nu} \\
& - \frac{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A}g_2(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-2}[1+w/(1+\nu)]^{-A}h(\nu)\mathrm{d}\nu}, \\
\leq & \sup_{\nu>\nu_0} g_2(\nu).
\end{aligned} \tag{3.11}
$$

In this case, let

$$
c = c_2 = \sup_{\nu>\nu_0} g_2(\nu) \quad \text{and} \quad d_2 = \sup_{\nu>\nu_0}\{g_1(\nu)+g_2(\nu)\}.
$$

Thus, the minimaxity is guaranteed if K, c_1 and d_1 satisfy the inequality (3.9) where c_0 and d_0 are replaced with c_2 and d_2.

Since the minimaxity condition in this case will be applied to examples given below, we provide a more useful condition. For this aim, let $G_1 = \lim_{\nu\to\infty} g_1(\nu)$, and note that $d_2 \leq G_1 + c_2$ for $c_2 = \sup_{\nu>\nu_0} g_2(\nu)$. Then from the inequality (3.8), we derive the following sufficient condition for the minimaxity:

$$
\frac{(p-2+2G_1)+2c_2}{(n+3-2K-2G_1)-2c_2} \leq 2\frac{p-2-2c_2}{n+2+2c_2},
$$

which is rewritten as

$$
\begin{aligned}
& 2(p-2)(n+3-2K-2G_1)-(n+2)(p-2+2G_1) \\
& \quad -(3n+3p+2-4K-2G_1)(2c_2)+(2c_2)^2 \geq 0.
\end{aligned} \tag{3.12}
$$

Since $(2c_2)^2 \geq 0$, we can obtain a simple condition given by $0 \leq 2c_2 \leq \{2(p-2)(n+3-2K-2G_1)-(n+2)(p-2+2G_1)\}/(3n+3p+2-4K-2G_1)$, namely,

$$
0 \leq \sup_{\nu>\nu_0} g_2(\nu) \leq \frac{2(p-2)(n+3-2K-2G_1)-(n+2)(p-2+2G_1)}{2(3n+3p+2-4K-2G_1)}. \tag{3.13}
$$

where $G_1 = \lim_{\nu\to\infty} g_1(\nu)$.

(Case 3). Consider the case that $g(\nu)$ can be decomposed as $g(\nu) = g_1(\nu)+g_2(\nu)$, where $g_1(\nu)$ is nondecreasing in λ and $g_2(\nu)$ is nonnegative and nonincreasing in λ. Then, it is noted that

$$
\frac{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A}g_2(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}[1+w/(1+\nu)]^{-A}h(\nu)\mathrm{d}\nu} \leq \frac{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}g_2(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}h(\nu)\mathrm{d}\nu},
$$

so that from (3.11),

$$
C(w) \leq \frac{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}g_2(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-3}h(\nu)\mathrm{d}\nu} \equiv c_3.
$$

Similarly,

$$D(w) \le \min\left\{ d_2,\ \lim_{\nu\to\infty} g_1(\nu) + \frac{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-2} g_2(\nu)h(\nu)\mathrm{d}\nu}{\int_{\nu_0}^{\infty}(1+\nu)^{-\alpha-2}h(\nu)\mathrm{d}\nu} \right\} \equiv d_3.$$

Hence, the minimaxity is guaranteed if K, c_2 and d_2 satisfy the inequality (3.9) where c_0 and d_0 are replaced with c_3 and d_3.

Example 3. (Maruyama and Strawderman type priors) Consider the priors with $h(\nu) = C\nu^b(1+\nu)^{-a-b-2}$ for $\nu > \nu_0$ and a constant C. These priors have been treated by Maruyama and Strawderman (2005) and Wells and Zhou (2008) for $\nu_0 = 0$, where $-K = e + 1/2$. W&Z showed that the conditions (C1)-(C3) are satisfied when $n/2+e > a > -p/2-1$, and the function $g(\nu)$ defined in (3.7) is written as

$$g(\nu) = a + 2 - b/\nu.$$

These authors handled the case that $b \ge 0$, which means that $g(\nu)$ is nondecreasing, so that it corresponds to the case that $g_2(\nu) = 0$ in (3.10) and $G_1 = a+2$ in (3.13). Thus, the conditions that $b \ge 0$ and $0 < (p/2 + a + 1)/(n/2 + e - a) \le 2(p-2)/(n+2)$ for $\nu_0 \ge 0$ are imposed for the minimaxity.

The condition (3.13) allows us to handle the case that $b < 0$ where ν_0 is restricted to a positive number, namely $\nu_0 > 0$. In this case, let $G_1 = a + 2$ and $g_2(\nu) = -b/\nu$, and $g_2(\nu)$ is decreasing and $\sup_{\nu>\nu_0} g_2(\nu) = -b/\nu_0$. From (3.13), we obtain the condition that

$$0 < -\frac{b}{\nu_0} \le \frac{2(p-2)(n+2e-2a) - (n+2)(p+2+2a)}{3n+3p+4e-2a},$$

where $(p/2 + a + 1)/(n/2 + e - a) < 2(p-2)/(n+2)$.

Combining the conditions of the cases $b \ge 0$ and $b < 0$ gives the following condition for the minimaxity: $n/2 + e > a > -p/2 - 1$ and

$$\max\left\{-\frac{b}{\nu_0}, 0\right\} \le \frac{2(p-2)(n+2e-2a) - (n+2)(p+2+2a)}{3n+3p+4e-2a},$$

which is wider than the conditions given in the literature. ■

Example 4. (Generalized Student priors) Consider the priors with $\nu_0 = 0$ and

$$h(\nu) = C(\nu+1)^{\beta-\alpha-\gamma-(p-2)/2}\nu^{\gamma-\beta}\exp\{\gamma/\nu\},$$

which was handled by W&Z who treated the two cases (1) $\alpha \le 0$, $\beta \le 0$ and $\gamma < 0$, and (2) $\alpha \le 0$, $\beta > 0$ and $\gamma < 0$. W&Z showed that the conditions (C1)-(C3) are satisfied when $n+p+1-2K > 0$ and $\alpha > 2-p$, and the function $g(\nu)$ defined in (3.7) is written as

$$g(\nu) = \frac{p-2}{2} + \alpha + \frac{\beta}{\nu} + \frac{\gamma}{\nu^2}.$$

The first case (1) implies that $g(\nu)$ is increasing, so that $g_2(\nu) = 0$ in (3.10) and $G_1 = (p-2)/2+\alpha$ in (3.13). Hence the minimaxity condition (3.13) is guaranteed when $0 < (p-2+2G_1)/(n+3-2K-2G_1) \le 2(p-2)/(n+2)$, or

$$0 < \frac{2(p-2+\alpha)}{n+5-2K-p-2\alpha} \le \frac{2(p-2)}{n+2}.$$

Since the second case (2) means that $g(\nu)$ is not monotone, W&Z decomposed $g(\nu)$ as

$$g(\nu) = \gamma\left(\frac{1}{\nu}+\frac{\beta}{2\gamma}\right)^2 - \frac{\beta^2}{4\gamma} + \frac{p-2}{2} + \alpha = g_1(\nu) + g_2(\nu),$$

where

$$g_1(\nu) = \begin{cases} g(\nu) + \dfrac{\beta^2}{4\gamma} & \text{if} \quad \nu \le -\dfrac{2\gamma}{\beta}, \\ \dfrac{p-2}{2} + \alpha & \text{if} \quad \nu > -\dfrac{2\gamma}{\beta}, \end{cases}$$

$$g_2(\nu) = \begin{cases} -\dfrac{\beta^2}{4\gamma} & \text{if} \quad \nu \le -\dfrac{2\gamma}{\beta}, \\ \dfrac{\beta}{\nu} + \dfrac{\gamma}{\nu^2} & \text{if} \quad \nu > -\dfrac{2\gamma}{\beta}. \end{cases}$$

It is noted that $g_1(\nu)$ is increasing, $G_1 = (p-2)/2+\alpha$, and $0 < g_2(\nu) \le -\beta^2/(4\gamma)$. Hence from (3.13), we obtain the minimaxity condition that

$$0 < -\frac{\beta^2}{4\gamma} \le \frac{(p-2)(n+5-2K-p-2\alpha)-(n+2)(p-2+\alpha)}{3n+2p+4-4K-2\alpha}. \tag{3.14}$$

As stated in W&Z, the spherical multivariate Student-t priors with m degrees of freedom and a scale parameter τ corresponds to the case that $\alpha = (m-p+4)/2$, $\beta = [m(1-\tau)+2]/2$ and $\gamma = -m\tau/2$. In fact, given η, the conditional density of $\boldsymbol{\theta}$ is written as

$$\begin{aligned}\pi(\boldsymbol{\theta}|\eta) =& C_0 \int_0^\infty (\eta^2/\nu)^{p/2} h(\nu) \exp\{-\eta^2\|\boldsymbol{\theta}\|^2/(2\nu)\} \mathrm{d}\nu \\ =& C_1 \left(\frac{\eta^2}{\tau m}\right)^{p/2} \left(1 + \frac{\eta^2}{\tau m}\|\boldsymbol{\theta}\|^2\right)^{-(m+p)/2},\end{aligned}$$

for constants C_0 and C_1. In this case, the condition (3.14) can be simplified as

$$0 < \frac{[m(1-\tau)+2]^2}{8m\tau} \le \frac{2(p-2)(n+1-2K-m)-(n+2)(p+m)}{2(3n+3p-4K-m)},$$

where $2(p-2)(n+1-2K-m) > (n+2)(p+m)$ and $n+1-2K-m > 0$. For some special cases, we can get the corresponding conditions. In the case of $\tau = 1$ and $-2K = 3$, which was treated by W&Z, the condition is expressed as

$$\frac{1}{m} \le \frac{2(p-2)(n+4-m)-(n+2)(p+m)}{3n+3p+6-m}.$$

This condition is equivalent to each of the following ones:

$$(2p+n-2)m^2-(pn+6p-4n-15)m+3(n+p+2)\le 0,$$

or

$$p\ge \frac{(n-2)m^2+(4n+15)m+3(n+2)}{(n+6)m-2m^2-3}, \tag{3.15}$$

where $(n+6)m-2m^2-3>0$. When n goes to infinity, these inequalities are, respectively, given by

$$\frac{p-4-\sqrt{(p-4)^2-12}}{2}\le m\le \frac{p-4+\sqrt{(p-4)^2-12}}{2},$$
$$p\ge (m+3)(m+1)/m,$$

which suggest that the minimaxity can be guaranteed for small m or large p.

The case $m=1$ corresponds to the spherical multivariate Cauchy prior, and from (3.15), we get the condition $p\ge(8n+19)/(n+1)$. On the other hand, the condition (5.3) of W&Z with $m=1$ and $-2K=3$, where their (m,n) corresponds to our (n,m), is equivalent to the condition that there exists a constant b such that

$$\frac{2(n+2)}{(p-2)(n+4)}\left(1+2\frac{p-2}{n+2}\right)\le b\le (p-6)/(p-2),$$

which is guaranteed when $p\ge(8n+20)/n$. It can be seen that the condition $p\ge(8n+19)/(n+1)$ is slightly wider than $p\ge(8n+20)/n$. ■

4. Improvement on the James-Stein estimator

The problems of improving the MLE $\widehat{\boldsymbol{\theta}}_0$ by the estimators $\widehat{\boldsymbol{\theta}}_\psi$ have been studied in the previous sections. As a further dominance property, we here construct a class of the estimators $\widehat{\boldsymbol{\theta}}_\psi$ improving on the James-Stein estimator $\widehat{\boldsymbol{\theta}}^{JS}$ given by (1.1). This result was given by Kubokawa (1994) without a proof, because it is complicated in comparison with the case of a known variance σ^2. However, using the integral expression given in Section 2., we can prove it more simply. Thus, we here provide the simple proof using the risk expression (2.2) or (4.2) based on the integral.

Theorem 2. *Assume that the function $\psi(w)$ is absolutely continuous and satisfies the condition $\int_0^1 z^{n/2-1}|\psi'(w/z)|\mathrm{d}z<\infty$ and the following conditions:*

(a) *$\psi(w)$ is nondecreasing and $\lim_{w\to\infty}\psi(w)=a_0=(p-2)/(n+2)$,*

(b) *$\psi(w)\ge\psi_0(w)$, where*

$$\psi_0(w)=\int_0^w \frac{y^{p/2-1}}{(1+y)^{(n+p)/2+1}}\mathrm{d}y\Big/\int_0^w \frac{y^{p/2-2}}{(1+y)^{(n+p)/2+1}}\mathrm{d}y. \tag{4.1}$$

Then the estimator $\widehat{\boldsymbol{\theta}}_\psi$ dominates the James-Stein estimator $\widehat{\boldsymbol{\theta}}^{JS}=(1-a_0/W)\boldsymbol{X}$.

Proof. We begin with providing another expression of the risk (2.2). Let $\chi^2_p(\lambda)$ and χ^2_{n+2} be mutually independent random variables where $\chi^2_p(\lambda)$ has a noncentral chi-square distribution with p degrees of freedom and the noncentrality $\lambda = \|\boldsymbol{\theta}\|^2/\sigma^2$. Let us define $\tilde{W}$ by $\tilde{W} = \chi^2_p(\lambda)/\chi^2_{n+2}$. By incorporating the term S/σ^2 into the density of S/σ^2 in the expression (2.2), the risk function of $\widehat{\boldsymbol{\theta}}_\psi$ can be rewritten as

$$R(\omega, \widehat{\boldsymbol{\theta}}_\psi) = p + nE\left[\frac{1}{\tilde{W}}\mathcal{I}(\tilde{W})\right] = p + n\int_0^\infty \frac{1}{w}\mathcal{I}(w)f(w;\lambda)\mathrm{d}w, \tag{4.2}$$

where the density of $\tilde{W}$ is written by

$$f(w;\lambda) = \sum_{j=0}^\infty P_j(\lambda)\frac{c_j w^{p/2-1+j}}{(1+w)^{(n+p)/2+1+j}}$$

for $c_j = \Gamma((n+p)/2+j+1)/\{\Gamma(n/2+1)\Gamma(p/2+j)\}$ and the Poisson probability $P_j(\lambda) = (\lambda/2)^j e^{-\lambda/2}/j!$. From the condition (a), the risk of the James-Stein estimator is given by $R(\omega, \widehat{\boldsymbol{\theta}}^{JS}) = p + nE[\tilde{W}^{-1}\lim_{t\to\infty}\mathcal{I}(t\tilde{W})]$. Then, the risk difference $\Delta = R(\omega, \widehat{\boldsymbol{\theta}}^{JS}) - R(\omega, \widehat{\boldsymbol{\theta}}_\psi)$ can be written as

$$\begin{aligned}\Delta =& nE\left[\frac{1}{\tilde{W}}\left\{\lim_{t\to\infty}\mathcal{I}(t\tilde{W}) - \mathcal{I}(\tilde{W})\right\}\right]\\ =& nE\left[\frac{1}{\tilde{W}}\int_1^\infty \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{I}(t\tilde{W})\mathrm{d}t\right] = nE\left[\int_1^\infty \mathcal{I}'(t\tilde{W})\mathrm{d}t\right],\end{aligned}$$

which is also expressed by

$$\Delta = n\int_0^\infty\int_1^\infty \mathcal{I}'(tw)f(w;\lambda)\mathrm{d}t\mathrm{d}w,$$

where $\mathcal{I}'(w) = (\mathrm{d}/\mathrm{d}w)\mathcal{I}(w)$ and $f(w;\lambda)$ is defined below (4.2). Making the transformations $x = tw$ and $y = x/t$ in turn with $\mathrm{d}x = t\mathrm{d}w$ and $\mathrm{d}t/t = \mathrm{d}y/y$, we observe that

$$\begin{aligned}E\left[\int_1^\infty \mathcal{I}'(t\tilde{W})\right] &= \int_0^\infty\int_1^\infty \mathcal{I}'(tw)f(w;\lambda)\mathrm{d}t\mathrm{d}w\\ &= \int_0^\infty\int_1^\infty \mathcal{I}'(x)t^{-1}f(x/t;\lambda)\mathrm{d}t\mathrm{d}x\\ &= \int_0^\infty \mathcal{I}'(x)F(x;\lambda)\mathrm{d}x, \qquad (4.3)\end{aligned}$$

where $F(x;\lambda) = \int_0^x y^{-1}f(y;\lambda)\mathrm{d}y$. To provide the derivative $\mathcal{I}'(x)$, we can use the expression given in (2.9) which can be derived under the condition $\int_0^1 z^{n/2-1}|\psi'(x/z)|\mathrm{d}z < \infty$, and note that

$$x\int_0^1 z^{n/2-1}\psi'(x/z)\mathrm{d}z = x^{n/2+1}\int_x^\infty s^{-n/2-1}\psi'(s)\mathrm{d}s.$$

Then, the derivative $\mathcal{I}'(x)$ has the form

$$\begin{aligned}\mathcal{I}'(x) =&2\psi(x)\psi'(x) - 2a_0\psi'(x)\\ &- (n+p)x^{n/2}\int_x^\infty s^{-n/2-1}\psi'(s)\mathrm{d}s + 2\frac{n+p}{n+2}\psi'(x)\\ =&2\psi(x)\psi'(x) + 2\psi'(x) - (n+p)\int_0^1 z^{n/2-1}\psi'(x/z)\mathrm{d}z.\end{aligned} \tag{4.4}$$

By making the transformation $w = x/z$ with $\mathrm{d}w = \mathrm{d}x/z$, it is noted that

$$\int_0^\infty\int_0^1 z^{n/2-1}\psi'(x/z)\mathrm{d}zF(x;\lambda)\mathrm{d}x = \int_0^\infty \psi'(w)\int_0^1 z^{n/2}F(zw;\lambda)\mathrm{d}z\mathrm{d}w. \tag{4.5}$$

Combining (4.3), (4.4) and (4.5), we obtain the expression

$$\Delta = 2n\int_0^\infty \psi'(w)\left\{(\psi(w)+1)F(w;\lambda) - \frac{n+p}{2}\int_0^1 z^{n/2}F(zw;\lambda)\mathrm{d}z\right\}\mathrm{d}w.$$

Since $\psi'(w) \geq 0$ from the condition (a), it is seen that $\Delta \geq 0$ if $\psi(w)$ satisfies the inequality

$$\psi(w) \geq -1 + \frac{n+p}{2}\frac{\int_0^1 z^{n/2}F(zw;\lambda)\mathrm{d}z}{F(w;\lambda)}. \tag{4.6}$$

Taking the integration by parts gives

$$\begin{aligned}\int_0^1 z^{n/2}F(zw;\lambda)\mathrm{d}z =& \int_0^1\left\{\frac{\mathrm{d}}{\mathrm{d}z}\frac{2}{n+2}z^{n/2+1}\right\}F(zw;\lambda)\mathrm{d}z\\ =&\frac{2}{n+2}F(w;\lambda) - \frac{2}{n+2}\int_0^1 z^{n/2}f(zw;\lambda)\mathrm{d}z\\ =&\frac{2}{n+2}F(w;\lambda) - \frac{2}{n+2}\frac{1}{w^{n/2+1}}\int_0^w y^{n/2}f(y;\lambda)\mathrm{d}y,\end{aligned}$$

which is used to rewrite (4.6) as

$$\psi(w) \geq a_0 - \frac{n+p}{n+2}\frac{1}{w^{n/2+1}}\frac{\int_0^w y^{n/2}f(y;\lambda)\mathrm{d}y}{\int_0^w y^{-1}f(y;\lambda)\mathrm{d}y}. \tag{4.7}$$

Let $f(y) = f(y;0) = c_0y^{p/2-1}/(1+y)^{(n+p)/2+1}$ and note that $f(y;\lambda)/f(y)$ is increasing in y. Then from Lemma 3.1 of Kubokawa (2007), we obtain the inequality

$$\frac{\int_0^w y^{n/2}f(y;\lambda)\mathrm{d}y}{\int_0^w y^{-1}f(y;\lambda)\mathrm{d}y} \geq \frac{\int_0^w y^{n/2}f(y)\mathrm{d}y}{\int_0^w y^{-1}f(y)\mathrm{d}y},$$

which means that the inequality (4.7) holds if $\psi(w)$ satisfies the inequality

$$\psi(w) \geq a_0 - \frac{n+p}{n+2}\frac{1}{w^{n/2+1}}\frac{\int_0^w y^{n/2}f(y)\mathrm{d}y}{\int_0^w y^{-1}f(y)\mathrm{d}y}. \tag{4.8}$$

Finally, we shall verify that the r.h.s. of the inequality (4.8) is equal to $\psi_0(w)$ given by (4.1). Note that

$$
\begin{aligned}
\int_0^w y^{n/2} f(y)\mathrm{d}y &= \int_0^w \frac{y^{(n+p)/2-1}}{(1+y)^{(n+p)/2+1}}\mathrm{d}y \\
&= \int_0^{w/(1+w)} u^{(n+p)/2-1}\mathrm{d}u = \frac{2}{n+p}\left(\frac{w}{1+w}\right)^{(n+p)/2}. \qquad (4.9)
\end{aligned}
$$

By the integration by parts, it is also shown that

$$
\begin{aligned}
\int_0^w f(y)\mathrm{d}y &= \int_0^w \frac{y^{p/2-1}}{(1+y)^{(n+p)/2+1}}\mathrm{d}y = \int_0^w \left(\frac{y}{1+y}\right)^{p/2-1} \frac{\mathrm{d}}{\mathrm{d}y}\left(-\frac{2/(n+2)}{(1+y)^{n/2+1}}\right)\mathrm{d}y \\
&= -\frac{2}{n+2}\frac{w^{p/2-1}}{(1+w)^{(n+p)/2}} + \frac{p-2}{n+2}\int_0^w y^{-1} f(y)\mathrm{d}y. \qquad (4.10)
\end{aligned}
$$

Using the equations (4.9) and (4.10), we can verify that the r.h.s. of the inequality (4.8) is equal to

$$
a_0 - \frac{n+p}{n+2}\frac{1}{w^{n/2+1}}\frac{\int_0^w y^{n/2} f(y)\mathrm{d}y}{\int_0^w y^{-1} f(y)\mathrm{d}y} = \frac{\int_0^w f(y)\mathrm{d}y}{\int_0^w y^{-1} f(y)\mathrm{d}y},
$$

which is identical to (4.1). Therefore, the proof of Theorem 2 is complete. ■

From this theorem, we can derive several interesting estimators improving on the James-Stein estimator, including the generalized Bayes estimator. For the details, see Kubokawa (1994, 98) and Maruyama (1999).

Acknowledgments. The author is grateful to two referees for their valuable comments and helpful suggestions. The research of the author was supported in part by Grant-in-Aid for Scientific Research Nos. 15200021, 15200022 and 16500172 and in part by a grant from the 21st Century COE Program at Faculty of Economics, University of Tokyo.

REFERENCES

Alam, K. (1973). A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* , **1**, 517-525.

Baranchik, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* , **41**, 642-645.

Brown, L.D. (1971). Admissible estimators, recurrent diffusions, and insolvable boundary value problems. *Ann. Math. Statist.*, **42**, 855-904.

Eaton, M. (2004). Evaluating improper priors and the recurrence of symmetric Markov chains: an overview. In *A festschrift for Herman Rubin*, 5-20, IMS Lecture Notes Monogr. Ser., 45, Inst. Math. Statist., Beachwood, OH.

Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.*, **4**, 11-21.

Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-542.

Hudson, H.M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.*, **6**, 478-484.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 361-379. University of California Press, Berkeley.

Johnstone, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann. Statist.*, **12**, 1173-1198.

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

Kubokawa, T. (1998). The Stein phenomenon in simultaneous estimation: A review. In *Applied Statistical Science* III (S.E. Ahmed, M. Ahsanullah and B.K. Sinha, eds.), 143-173, NOVA Science Publishers, New York.

Kubokawa, T. (2007). Characterization of priors in the Stein problem. *J. Japn. Statist. Soc.*, **37**, 207-237.

Maruyama, Y. (1999). Improving on the James-Stein estimator. *Statist. & Decisions*, **14**, 137-140.

Maruyama, Y. and Strawderman, W.E. (2005). A new class of generalized Bayes minimax ridge regression estimators. *Ann. Statist.*, **33**, 1753-1770.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Baerkeley Symp. Math. Statist. Probab.*, **1**, 197-206. Univ. California Press, Berkeley.

Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proc. Prague Symp. Asymptotic Statist.*, 345-381.

Stein, C. (1977). Lectures on multivariate estimation theory. (In Russian.) In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad.(In Russian)

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135-1151.

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