Minimaxity in Estimation of Restricted Parameters

Tatsuya Kubokawa
The University of Tokyo

MINIMAXITY IN ESTIMATION OF RESTRICTED PARAMETERS

Tatsuya Kubokawa*

This paper is concerned with estimation of the restricted parameters in location and/or scale families from a decision-theoretic point of view. A simple method is provided to show the minimaxity of the best equivariant and unrestricted estimators. This is based on a modification of the known method of Girshick and Savage (1951) and can be applied to more complicated cases of restriction in the location-scale family. Classes of minimax estimators are also constructed by using the IERD method of Kubokawa (1994a,b): Especially, the paper succeeds in constructing such a class for estimating a restricted mean in a normal distribution with an unknown variance.

Key words and phrases: Decision theory, generalized Bayes estimator, location family, maximum likelihood estimator, minimaxity, restricted parameter, scale family.

1. Introduction

The point estimation of restricted parameters has been studied from a decision-theoretic point of view since Katz (1961), who showed that the generalized Bayes estimator of a restricted mean is minimax and admissible in a normal distribution with a known variance. Farrell (1964) established the minimaxity and admissibility in the general location family. This classical problem was recently revisited by Marchand and Strawderman (2004) who gave another proof for the minimaxity. However, the proof requires slightly complicated arguments, which inspired me to consider another simple method for the proof.

In this paper, we shall treat the following location and scale families with the parameters restricted to one-sided spaces: Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a set of random variables and $\boldsymbol{x} = (x_1, \dots, x_n)$ is an observation of $\boldsymbol{X}$.

[1] Location family. The density function of $\boldsymbol{X}$ is given by $f(\boldsymbol{x} - \mu)$ and the location parameter μ is restricted to

$$(1.1) \qquad A = \{\mu \mid \mu > a_0\} \quad \text{for known real } a_0,$$

where $\boldsymbol{x} - \mu$ means $(x_1 - \mu, \dots, x_n - \mu)$.

[2] Scale family. The density function of $\boldsymbol{X}$ is given by $\sigma^{-1} f(\boldsymbol{x}/\sigma)$ and the scale parameter σ is restricted to

$$(1.2) \qquad B = \{\sigma \mid \sigma > b_0\} \quad \text{for known real } b_0 > 0,$$

where $\boldsymbol{x}/\sigma$ means $(x_1/\sigma, \dots, x_n/\sigma)$.

[3] **Location-Scale family.** The density function of $\boldsymbol{X}$ is given by $\sigma^{-1}f((\boldsymbol{x}-\mu)/\sigma)$ and the location and scale parameters (μ, σ) are restricted to

$$(1.3) \qquad C = \{(\mu, \sigma) \,|\, \mu > c_0\sigma + a_0, \ \sigma > b_0\} \quad \text{for known real } c_0 > 0.$$

When $\boldsymbol{X}$ is a random sample from a density $p(x - \mu)$, the joint density of $\boldsymbol{X}$ is written by $f(\boldsymbol{x} - \mu) = p(x_1 - \mu) \times \cdots \times p(x_n - \mu)$. It is noted that the above setup includes the dependent cases, that is, $X_1, \ldots, X_n$ are not mutually independent.

In the unrestricted cases, it is well known that the best equivariant estimators of the location and/or scale parameters are minimax relative to invariant loss functions under the location and/or scale transformation groups. In the restricted case (1.1), Marchand and Strawderman (2004) demonstrated that the minimaxity property of the best location equivariant estimator still holds, but their proof requires slightly complicated arguments. In Section 2, we provide a simple proof for the minimaxity based on modification of the method of Girshick and Savage (1951). This new method can be also applied to the restricted scale problem (1.2). In Section 3, the method is used to establish the minimaxity in more general setups of the restricted location-scale problem (1.3), which may be a new result as long as I know.

We next address the problems of constructing classes of minimax estimators which include the maximum likelihood estimators and the generalized Bayes estimators against the uniform priors on the restricted parameter spaces. Marchand and Strawderman (2004) constructed such a class in the location family (1.1) by using the IERD method given by Kubokawa (1994a,b,98,99) and Kubokawa and Saleh (1998). While the method can be easily applied to the scale family (1.2), it is too hard to employ in the location-scale family (1.3). In Section 3.2, however, we can obtain a class of minimax estimators of a restricted mean in a random sample from a normal distribution whose canonical form is given by

$$X \sim \mathcal{N}(\theta, \sigma^2) \quad \text{and} \quad S \sim \sigma^2\chi_m^2$$

where χ_m^2 denotes the chi-square distribution with m degrees of freedom. As seen in the proof of Theorem 3..2, it is not easy to establish the result even in the distributional assumption of normality. The approach used there will be of benefit to us when other issues of estimating restricted means with the unknown variance are addressed in a future. Some dominance results in estimation of the restricted variance are given in Section 3.3.

Although the one-sided restrictions of the parameter spaces (1.1), (1.2) and (1.3) are handled in this paper, we can treat other types of one-sided restriction cases and provide the corresponding results for minimaxity and dominance. As shown in Casella and Strawderman (1981) and Marchand and Perron (2001), it is noted that the best equivariant and unrestricted estimators are not minimax in the case that the parameter spaces are restricted into bounded regions. This suggests that the unboundedness of restricted parameter spaces may be necessary for the minimaxity of the best equivariant and unrestricted estimators.

2. Estimation of Location and Scale Families

2.1. Minimaxity in the location family

We first deal with the estimation of the location parameter μ of the family $f(\boldsymbol{x} - \mu)$ where the location is restricted to the one-sided space

$$A = \{\mu \mid \mu > a_0\} \quad \text{for known real } a_0.$$

Estimator $\widehat{\mu}$ of μ is evaluated by the risk function $R(\mu, \widehat{\mu}) = E[L_\ell(\widehat{\mu}, \mu)]$ relative to the squared error loss

$$L_\ell(\widehat{\mu}, \mu) = (\widehat{\mu} - \mu)^2.$$

We begin with providing a simple proof for the minimaxity of the best location-equivariant estimator of μ, given by

$$\widehat{\mu}^U = \int_{-\infty}^{\infty} a f(\boldsymbol{X} - a) \mathrm{d}a \Big/ \int_{-\infty}^{\infty} f(\boldsymbol{X} - a) \mathrm{d}a.$$

This is called the Pitman estimator and is the generalized Bayes and unrestricted estimator against the Lebesgue measure $\mathrm{d}\mu$ on whole the real line. Although Marchand and Strawderman (2004) proved the minimaxity of $\widehat{\mu}^U$, their arguments seem slightly complicated. Girshick and Savage (1951) gave the nice proof of the best equivariant estimator for the unrestricted location parameter. We here provide a sophisticated proof by modifying the Girshick-Savage method to handle the restricted case.

Theorem 2..1. *The best equivariant and unrestricted estimator $\widehat{\mu}^U$ is minimax in the estimation issue on the restricted parameter space A relative to the L_ℓ-loss, and the minimax risk is given by $R_0 = R(\mu, \widehat{\mu}^U)$.*

Proof. Without any loss of generality, assume that $a_0 = 0$. Let $A_k = \{\mu \mid 0 < \mu < k\}$ for $k = 1, 2, \ldots$, and consider the sequence of prior distributions given by

$$\pi_k(\mu) = \begin{cases} k^{-1} & \text{if } \mu \in A_k \\ 0 & \text{otherwise,} \end{cases}$$

which yields the Bayes estimators

$$\widehat{\mu}_k^\pi = \widehat{\mu}_k^\pi(\boldsymbol{X}) = \int_{A_k} a f(\boldsymbol{X} - a) \mathrm{d}a \Big/ \int_{A_k} f(\boldsymbol{X} - a) \mathrm{d}a$$

with the Bayes risk function

$$(2.1) \qquad r_k(\pi_k, \widehat{\mu}_k^\pi) = \frac{1}{k} \int_{A_k} \int \{\widehat{\mu}_k^\pi(\boldsymbol{x}) - \mu\}^2 f(\boldsymbol{x} - \mu) \mathrm{d}\boldsymbol{x} \mathrm{d}\mu.$$

Since $r_k(\pi_k, \widehat{\mu}_k^\pi) \leq r_k(\pi_k, \widehat{\mu}^U) = R_0$, it is sufficient to show that $\liminf_{k\to\infty} r_k(\pi_k, \widehat{\mu}_k^\pi) \geq R_0$. Making the transformations $\boldsymbol{z} = \boldsymbol{x} - \mu$ and $t = a - \mu$ with $\mathrm{d}\boldsymbol{z} = \mathrm{d}\boldsymbol{x}$ and $\mathrm{d}t = \mathrm{d}a$ gives that

$$
\begin{aligned}
\widehat{\mu}_k^\pi(\boldsymbol{x}) - \mu &= \widehat{\mu}_k^\pi(\boldsymbol{z} + \mu) - \mu \\
&= \int_{A_k} (a - \mu) f(\boldsymbol{z} + \mu - a)\mathrm{d}a \Big/ \int_{A_k} f(\boldsymbol{z} + \mu - a)\mathrm{d}a \\
&= \int_{t+\mu \in A_k} t f(\boldsymbol{z} - t)\mathrm{d}t \Big/ \int_{t+\mu \in A_k} f(\boldsymbol{z} - t)\mathrm{d}t.
\end{aligned}
$$

(2.2)

Of importance in this proof is making the transformation $\xi = (2/k)(\mu - k/2)$ with $\mathrm{d}\xi = (2/k)\mathrm{d}\mu$, which rewrites the condition $0 < \mu < k$ as $|\xi| < 1$. Also the condition that $0 < t + \mu < k$ is expressed by the inequality $-(k/2)(\xi + 1) < t < (k/2)(1 - \xi)$. Let $A_k^* = \{t | -(k/2)(\xi + 1) < t < (k/2)(1 - \xi)\}$. Then the transformations are used in (2.2) and (2.1) to obtain that

$$
(2.3) \qquad \widehat{\mu}_k^\pi(\boldsymbol{x}) - \mu = \int_{A_k^*} t f(\boldsymbol{z} - t)\mathrm{d}t \Big/ \int_{A_k^*} f(\boldsymbol{z} - t)\mathrm{d}t = \widehat{\mu}_k^*(\boldsymbol{z}|\xi), \quad \text{(say)}
$$

and

$$
r_k(\pi_k, \widehat{\mu}_k^\pi) = \frac{1}{2} \int_{|\xi|<1} \int \{\widehat{\mu}_k^*(\boldsymbol{z}|\xi)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi.
$$

For a small $\varepsilon > 0$, the integral of a positive function $h(\xi)$ with respect to ξ is evaluated by

$$
\int_{|\xi|<1} h(\xi)\mathrm{d}\xi = \int_{|\xi|<1-\varepsilon} h(\xi)\mathrm{d}\xi + \int_{1-\varepsilon<|\xi|<1} h(\xi)\mathrm{d}\xi \geq \int_{|\xi|<1-\varepsilon} h(\xi)\mathrm{d}\xi,
$$

which is used to get that

$$
r_k(\pi_k, \widehat{\mu}_k^\pi) \geq \frac{1}{2} \int_{|\xi|<1-\varepsilon} \int \{\widehat{\mu}_k^*(\boldsymbol{z}|\xi)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi.
$$

The range of t in the integrals in $\widehat{\mu}_k^*(\boldsymbol{z}|\xi)$ given by (2.3) is $A_k^* = \{t | -(k/2)(\xi+1) < t < (k/2)(1 - \xi)\}$. Since $|\xi| < 1 - \varepsilon$, it is noted that $1 - \xi > 1 - (1 - \varepsilon) = \varepsilon > 0$ and $1 + \xi > 1 + (-1 + \varepsilon) = \varepsilon > 0$, which imply that the end points $(k/2)(1 - \xi)$ and $-(k/2)(1 + \xi)$ tend to infinity and minus infinity as $k \to \infty$ and then $\widehat{\mu}_k^*(\boldsymbol{z}|\xi)$ converges $\widehat{\mu}^U(\boldsymbol{z})$. Using the Fatou lemma, we obtain that

$$
\begin{aligned}
\liminf_{k\to\infty} r_k(\pi_k, \widehat{\mu}_k^\pi) &\geq \liminf_{k\to\infty} \frac{1}{2} \int_{|\xi|<1-\varepsilon} \int \{\widehat{\mu}_k^*(\boldsymbol{z}|\xi)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi \\
&\geq \frac{1}{2} \int_{|\xi|<1-\varepsilon} \int \left\{\liminf_{k\to\infty} \widehat{\mu}_k^*(\boldsymbol{z}|\xi)\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi \\
&= \frac{1}{2} \int_{|\xi|<1-\varepsilon} \mathrm{d}\xi \int \{\widehat{\mu}^U(\boldsymbol{z})\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \\
&= (1 - \varepsilon) R(\mu, \widehat{\mu}^U) = (1 - \varepsilon) R_0
\end{aligned}
$$

From the arbitrariness of $\varepsilon > 0$, it follows that $\liminf_{k\to\infty} r_k(\pi_k, \widehat{\mu}_k^\pi) \geq R_0$, completing the proof of Theorem 2..1. ∎

2.2.   A class of minimax estimators

The shortcoming of the crude minimax estimator $\widehat{\mu}^U$ is that it takes values outside the parameter space A with a positive probability. A simple modification is to truncate $\widehat{\mu}^U$ at the boundary of A as

$$\widehat{\mu}^{TR} = \max\{\widehat{\mu}^U, \ a_0\}.$$

It is easily seen that the truncated estimator $\widehat{\mu}^{TR}$ may dominate $\widehat{\mu}^U$. However, it leaves the undesirable property that it is not analytical or smooth. We shall construct classes of minimax estimators improving on $\widehat{\mu}^U$ which include the generalized Bayes and smooth estimators.

We first note that the minimax estimator $\widehat{\mu}^U$ is expressed as

$$\widehat{\mu}^U = X_1 - c(\boldsymbol{Y}), \quad \boldsymbol{Y} = (0, Y_2, \ldots, Y_n), \ Y_i = X_i - X_1,$$
$$c(\boldsymbol{Y}) = \int_{-\infty}^{\infty} uf(\boldsymbol{Y} + u)\mathrm{d}u \Big/ \int_{-\infty}^{\infty} f(\boldsymbol{Y} + u)\mathrm{d}u,$$

which suggests to consider the following form as estimators dominating $\widehat{\mu}^U$:

$$\widehat{\mu}_\phi = \widehat{\mu}_\phi(X_1, \boldsymbol{Y}) = X_1 - \phi(X_1 - a_0, \boldsymbol{Y}),$$

where $\phi(X_1, \boldsymbol{Y})$ is an absolutely continuous function. Marchand and Strawderman (2004) derived the conditions for the minimaxity of $\widehat{\mu}_\phi$, which is restated here with the proof, for the proof given here is done instructively and will be on the basis of the dominance results in the following sections.

Theorem 2..2.   *Assume that $\phi(x, \boldsymbol{y})$ satisfies the following conditions:*
(a) $\phi(x, \boldsymbol{y})$ is nondecreasing in x,
(b) $\lim_{x \to \infty} \phi(x, \boldsymbol{y}) = c(\boldsymbol{y})$,
(c) $\phi(x, \boldsymbol{y}) \geq \phi^m(x, \boldsymbol{y})$, where

$$(2.4) \qquad \phi^m(x, \boldsymbol{y}) = \int_{-\infty}^{x} uf(\boldsymbol{y} + u)\mathrm{d}u \Big/ \int_{-\infty}^{x} f(\boldsymbol{y} + u)\mathrm{d}u.$$

Then $\widehat{\mu}_\phi$ is a minimax estimator improving on $\widehat{\mu}^U$ relative to the L_ℓ-loss.

Proof. The IERD method provided by Kubokawa (1994a,b, 98, 99) is useful for the proof. The risk difference of the two estimators $\widehat{\mu}^U$ and $\widehat{\mu}_\phi$ is written by

$$\begin{aligned}
\Delta &= R(\mu, \widehat{\mu}^U) - R(\mu, \widehat{\mu}_\phi) \\
&= E\left[\{X_1 - c(\boldsymbol{Y}) - \mu\}^2 - \{X_1 - \phi(X_1 - a_0, \boldsymbol{Y}) - \mu\}^2\right]
\end{aligned}$$

which is, from the condition (b), expressed as

$$E\left[\{X_1 - \phi(X_1 - a_0 + t, \boldsymbol{Y}) - \mu\}^2 \Big|_{t=0}^{\infty}\right]$$
$$= E\left[\int_0^{\infty} \frac{\mathrm{d}}{\mathrm{d}t}\{X_1 - \phi(X_1 - a_0 + t, \boldsymbol{Y}) - \mu\}^2 \,\mathrm{d}t\right]$$
$$= \int\int_0^{\infty} \frac{\mathrm{d}}{\mathrm{d}t}\{x_1 - \phi(x_1 - a_0 + t, \boldsymbol{y}) - \mu\}^2 f(\boldsymbol{x} - \mu)\mathrm{d}t\mathrm{d}\boldsymbol{x}$$
$$= -2\int\int\int_{-a_0}^{\infty}\{x_1 - \phi(x_1 + t, \boldsymbol{y}) - \mu\}\,\phi'(x_1 + t, \boldsymbol{y})f(\boldsymbol{y} + x_1 - \mu)\mathrm{d}t\mathrm{d}x_1\mathrm{d}\boldsymbol{y},$$

where $\phi'(t, \boldsymbol{y}) = (\partial/\partial t)\phi(t, \boldsymbol{y})$. Making the transformations $z = x_1 - \mu + t$ and $u = z - t$ in turn with $\mathrm{d}z = \mathrm{d}x_1$ and $\mathrm{d}u = \mathrm{d}t$, we can rewrite Δ as

$$\Delta = -2\int\int\int_{-a_0}^{\infty}\{z - t - \phi(z + \mu, \boldsymbol{y})\}\,\phi'(z + \mu, \boldsymbol{y})f(\boldsymbol{y} + z - t)\mathrm{d}t\mathrm{d}z\mathrm{d}\boldsymbol{y}$$
$$= -2\int\int\int_{-\infty}^{z+a_0}\{u - \phi(z + \mu, \boldsymbol{y})\}\,\phi'(z + \mu, \boldsymbol{y})f(\boldsymbol{y} + u)\mathrm{d}u\mathrm{d}z\mathrm{d}\boldsymbol{y}.$$

Since $\phi'(z + \mu, \boldsymbol{y}) \geq 0$, it is sufficient to show that

(2.5) $$\phi(z + \mu, \boldsymbol{y}) \geq \int_{-\infty}^{z+a_0} uf(\boldsymbol{y} + u)\mathrm{d}u \Big/ \int_{-\infty}^{z+a_0} f(\boldsymbol{y} + u)\mathrm{d}u.$$

The condition (a) implies that

$$\phi(z + \mu, \boldsymbol{y}) \geq \phi(z + a_0, \boldsymbol{y}) \text{ for } \mu > a_0,$$

so that (2.5) is guaranteed by the condition (c). Therefore, the proof of Theorem 2..2 is complete. ∎

It is interesting to note that the class derived in Theorem 2..2 includes the generalized Bayes estimator against the uniform prior on A, given by

$$\widehat{\mu}^m = \int_A af(\boldsymbol{X} - a, \boldsymbol{Y})\mathrm{d}a \Big/ \int_A f(\boldsymbol{X} - a, \boldsymbol{Y})\mathrm{d}a$$
$$= X_1 - \phi^m(X_1 - a_0, \boldsymbol{Y})$$

for $\phi^m(x, \boldsymbol{y})$ is defined by (2.4). In fact, it is easy to check that $\phi^m(x, \boldsymbol{y})$ satisfies the conditions (a), (b) and (c) of Theorem 2..2. It is noted that these conditions can be satisfied without any other assumptions for the density $f(\boldsymbol{x} - \mu)$.

From the proof of Theorem 2..2, we notice that for the generalized Bayes estimator $\widehat{\mu}^m$,

(2.6) $$R(a_0, \widehat{\mu}^m) = R(a_0, \widehat{\mu}^U) = R_0 \quad \text{at} \quad \mu = a_0,$$

that is, $\widehat{\mu}^m$ and $\widehat{\mu}^U$ have the same risk at $\mu = a_0$. Since $\widehat{\mu}^m - \mu$ is written by

$$\widehat{\mu}^m - \mu = \int_{a_0}^{\infty} (a - \mu) f(\boldsymbol{x} - a)\mathrm{d}a \Big/ \int_{a_0}^{\infty} f(\boldsymbol{x} - a)\mathrm{d}a,$$

the risk function of $\widehat{\mu}^m$ is expressed by

$$R(\mu, \widehat{\mu}^m) = \int \left\{ \frac{\int_{a_0}^{\infty} (a - \mu) f(\boldsymbol{x} - a)\mathrm{d}a}{\int_{a_0}^{\infty} f(\boldsymbol{x} - a)\mathrm{d}a} \right\}^2 f(\boldsymbol{x} - \mu)\mathrm{d}\boldsymbol{x}$$

$$= \int \left\{ \frac{\int_{a_0 - \mu}^{\infty} t f(\boldsymbol{z} - t)\mathrm{d}t}{\int_{a_0 - \mu}^{\infty} f(\boldsymbol{z} - t)\mathrm{d}t} \right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z},$$

where the transformations $\boldsymbol{z} = \boldsymbol{x} - \mu$ and $t = a - \mu$ have been made. This expression provides the limiting value of the risk as

$$\lim_{\mu \to \infty} R(\mu, \widehat{\mu}^m) = \int \left\{ \widehat{\mu}^U(\boldsymbol{z}) \right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z} = R_0.$$

Together with (2.6), we can get the following property of the risk.

Proposition 2..1. *The risk function of the generallized Bayes estimator $\widehat{\mu}^m$ attains the minimax value R_0 at $\mu = a_0$ and when μ tending to infinity.*

Differentiating $R(\mu, \widehat{\mu}^m)$ with respect to μ gives that

$$\frac{\mathrm{d}}{\mathrm{d}\mu} R(\mu, \widehat{\mu}^m) = -2 \int \frac{\int_{a_0 - \mu}^{\infty} t f(\boldsymbol{z} - t)\mathrm{d}t f(\boldsymbol{z} + \mu)}{(\int_{a_0 - \mu}^{\infty} f(\boldsymbol{z} - t)\mathrm{d}t)^3} \int_{a_0 - \mu}^{\infty} (t + \mu) f(\boldsymbol{z} - t)\mathrm{d}t f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$$

$$= -2 \int \frac{\int_{a_0}^{\infty} (a - \mu) f(\boldsymbol{x} - a)\mathrm{d}a f(\boldsymbol{x})}{(\int_{a_0}^{\infty} f(\boldsymbol{x} - a)\mathrm{d}a)^3} \int_{a_0}^{\infty} a f(\boldsymbol{x} - a)\mathrm{d}a f(\boldsymbol{x} - \mu)\mathrm{d}\boldsymbol{x}$$

$$(2.7) \qquad = 2 \int \left\{ \mu - \frac{\int_{a_0}^{\infty} a f(\boldsymbol{x} - a)\mathrm{d}a}{\int_{a_0}^{\infty} f(\boldsymbol{x} - a)\mathrm{d}a} \right\} \frac{\int_{a_0}^{\infty} a f(\boldsymbol{x} - a)\mathrm{d}a}{(\int_{a_0}^{\infty} f(\boldsymbol{x} - a)\mathrm{d}a)^2} f(\boldsymbol{x}) f(\boldsymbol{x} - \mu)\mathrm{d}\boldsymbol{x},$$

which demonstrates that the derivative is negative at $\mu = a_0$ and positive for larger μ. This may suggest that there exists a point μ_0 such that the derivative $(\mathrm{d}/\mathrm{d}\mu)R(\mu, \widehat{\mu}^m)$ has one sign change at μ_0, in other words, the risk function of $\widehat{\mu}^m$ is decreasing in μ for $a_0 \le \mu < \mu_0$ and increasing for $\mu_0 < \mu < \infty$. This risk property can be verified for the normal distribution as given in the following example.

Example 2..1. Let $X_1, \ldots, X_n$ be a random sample from the normal distribution $\mathcal{N}(\mu, \sigma_0^2)$ for known variance σ_0^2. Let $X = \sum_{i=1}^{n} (X_i - a_0)/(\sigma_0 \sqrt{n})$, and X has $\mathcal{N}(\theta, 1)$ for $\theta = (\mu - a_0)/\sigma_0$, so that the problem is reduced to the estimation of θ based on X under the squared loss $(\hat{\theta} - \theta)^2$ where θ is restricted to the space

$$\theta > 0.$$

Then, the generalized Bayes estimator of θ against the Lebesgue measure $\mathrm{d}\theta I(\theta > 0)$ is expressed by

$$\hat{\theta}^m = \hat{\theta}^m(X) = \int_0^\infty \theta e^{-(X-\theta)^2/2}\mathrm{d}\theta \Big/ \int_0^\infty e^{-(X-\theta)^2/2}\mathrm{d}\theta.$$

Using the integration by parts, we note that

$$\int_0^\infty te^{-t^2/2+xt}\mathrm{d}t = -\int_0^\infty (-t+x)e^{-t^2/2+xt}\mathrm{d}t + x\int_0^\infty e^{-t^2/2+xt}\mathrm{d}t$$

$$= -\left[e^{-t^2/2+xt}\right]_{t=0}^\infty + x\int_0^\infty e^{-t^2/2+xt}\mathrm{d}t$$

(2.8)
$$= 1 + x\int_0^\infty e^{-t^2/2+xt}\mathrm{d}t,$$

which is used to rewrite $\hat{\theta}^m$ as

$$\hat{\theta}^m = X + g(X), \quad g(x) = \left[\int_0^\infty e^{-t^2/2+xt}\mathrm{d}t\right]^{-1}.$$

Then the derivative (2.7) is rewritten by

$$\frac{\mathrm{d}}{\mathrm{d}\theta}R(\theta,\hat{\theta}^m) = 2E\left[\{\theta - \{X+g(X)\}\}\{X+g(X)\}g(X)\right]$$

(2.9)
$$= -2E\left[(X-\theta)\{X+g(X)\}g(X)\right] - 2E\left[\{g(X)\}^2\{X+g(X)\}\right].$$

We shall show below that the derivative (2.9) has one sign change. From the equation (2.8), we first note that the derivative $g'(x) = \mathrm{d}g(x)/\mathrm{d}x$ is written as

$$g'(x) = -\frac{\int_0^\infty te^{-t^2/2+xt}\mathrm{d}t}{(\int_0^\infty e^{-t^2/2+xt}\mathrm{d}t)^2} = -\frac{1 + x\int_0^\infty e^{-t^2/2+xt}\mathrm{d}t}{(\int_0^\infty e^{-t^2/2+xt}\mathrm{d}t)^2}$$

(2.10)
$$= -\{g(x)\}^2 - xg(x) = -g(x)\{x+g(x)\}.$$

The Stein identity applies to the first term of the r.h.s. of (2.9) to rewrite it as

$$E\left[(X-\theta)\{X+g(X)\}g(X)\right] = E\left[\frac{\mathrm{d}}{\mathrm{d}X}\{(X+g(X))g(X)\}\right]$$

$$= E\left[g(X) + g(X)g'(X) + \{X+g(X)\}g'(X)\right]$$

$$= E\left[g(X) - \{g(X)\}^2\{X+g(X)\} - g(X)\{X+g(X)\}^2\right],$$

which is substituted into (2.9), and we get the expression

$$\frac{\mathrm{d}}{\mathrm{d}\theta}R(\theta,\hat{\theta}^m) = 2E\left[\{(X+g(X))^2 - 1\}g(X)\right]$$

(2.11)
$$= 2E\left[\left\{(\hat{\theta}^m(X))^2 - 1\right\}g(X)\right].$$

From Proposition 2..1, there exists a point θ_0 such that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}R(\theta,\hat{\theta}^m)|_{\theta=\theta_0} = 2\int\left\{(\hat{\theta}^m(x))^2 - 1\right\}g(x)f(x-\theta_0)\mathrm{d}x = 0,$$

for the normal density $f(x-\theta)$. For any $\theta > \theta_0$, we observe that

$$\int\left\{(\hat{\theta}^m(x))^2 - 1\right\}g(x)f(x-\theta)\mathrm{d}x$$
$$= \int\left\{(\hat{\theta}^m(x))^2 - 1\right\}g(x)f(x-\theta_0)\frac{f(x-\theta)}{f(x-\theta_0)}\mathrm{d}x.$$

It is here noted that $\hat{\theta}^m$ is increasing in X, since the derivative of $\hat{\theta}^m$, given by

$$\frac{\mathrm{d}}{\mathrm{d}X}\hat{\theta}^m = \frac{\int_0^\infty \theta^2 e^{-(X-\theta)^2/2}\mathrm{d}\theta \int_0^\infty e^{-(X-\theta)^2/2}\mathrm{d}\theta - (\int_0^\infty \theta e^{-(X-\theta)^2/2}\mathrm{d}\theta)^2}{(\int_0^\infty e^{-(X-\theta)^2/2}\mathrm{d}\theta)^2},$$

is nonnegative from Shwarz's inequality. Then from the monotonicity of $\hat{\theta}^m(x)$, it is noted that $\{(\hat{\theta}^m(x))^2 - 1\}g(x)$ has one sign change at some point x_0 from negative to positive. Since $f(x-\theta)/f(x-\theta_0)$ is increasing in x, we can show the inequality that

$$\int\left\{(\hat{\theta}^m(x))^2 - 1\right\}g(x)f(x-\theta_0)\frac{f(x-\theta)}{f(x-\theta_0)}\mathrm{d}x$$

(2.12)
$$> \int\left\{(\hat{\theta}^m(x))^2 - 1\right\}g(x)f(x-\theta_0)\mathrm{d}x\frac{f(x_0-\theta)}{f(x_0-\theta_0)},$$

which is zero (see Lemma 2.1 of Kubokawa (1994b)). This argument demonstrates that once the derivative $(\mathrm{d}/\mathrm{d}\theta)R(\theta,\hat{\theta}^m)$ becomes zero at θ_0, it holds positive for all $\theta > \theta_0$. In other words, the point θ_0 is uniquely determined. Hence, the risk function of $\hat{\theta}^m$ is decreasing in θ for $0 \le \theta < \theta_0$ and increasing for $\theta_0 < \theta < \infty$. ∎

2.3. Minimaxity in the scale family

The same arguments as in the previous subsections allow us to extend the results of the minimaxity to the scale family (1.2) of the density $\sigma^{-n}f(\boldsymbol{x}/\sigma)$ for scale parameter $\sigma > 0$. It is supposed that the scale σ is estimated by estimator $\hat{\sigma}$ relative to the entropy loss function

(2.13)
$$L_s(\hat{\sigma}/\sigma) = \hat{\sigma}/\sigma - \log\hat{\sigma}/\sigma - 1,$$

referred to as the Stein loss as well. The best scale-equivariant estimator $\hat{\sigma}^U$ is given by

$$\hat{\sigma}^U = \hat{\sigma}^U(\boldsymbol{X}) = \int_0^\infty \sigma^{-n-1}f(\boldsymbol{X}/\sigma)\mathrm{d}\sigma\Big/\int_0^\infty \sigma^{-n-2}f(\boldsymbol{X}/\sigma)\mathrm{d}\sigma,$$

where $\boldsymbol{X}/\sigma$ means $(X_1/\sigma, \ldots, X_n/\sigma)$. This is the unrestricted generalized Bayes estimator against the measure $\sigma^{-1}\mathrm{d}\sigma$ on whole the positive real line $\boldsymbol{R}_+$. Assume that the scale σ is restricted to the space

$$B = \{\sigma|\sigma > b_0\}.$$

Theorem 2..3. *The best equivariant and unrestricted estimator $\hat{\sigma}^U$ of σ is minimax under the entropy loss (2.13) in the estimation issue on the restricted parameter space B.*

Proof. Without any loss of generality, assume that $b_0 = 1$. Let $B_k = \{\sigma| 1 < \sigma < k\}$ for $k = 2, 3, \ldots$, and consider the sequence of prior distributions given by

$$\pi_k(\sigma)\mathrm{d}\sigma = \begin{cases} (\log k)^{-1}\sigma^{-1}\mathrm{d}\sigma & \text{if } \sigma \in B_k \\ 0 & \text{otherwise,} \end{cases}$$

which yields the Bayes estimators

$$\hat{\sigma}_k^{\pi} = \hat{\sigma}_k^{\pi}(\boldsymbol{X}) = \int_{B_k} b^{-n-1}f(\boldsymbol{X}/b)\mathrm{d}b \Big/ \int_{B_k} b^{-n-2}f(\boldsymbol{X}/b)\mathrm{d}b$$

with the Bayes risk function

$$r_k(\pi_k, \hat{\sigma}_k^{\pi}) = \frac{1}{\log k} \int_{B_k} \int L_s(\hat{\sigma}_k^{\pi}(\boldsymbol{x})/\sigma) \frac{1}{\sigma^{n+1}} f(\boldsymbol{x}/\sigma)\mathrm{d}\boldsymbol{x}\mathrm{d}\sigma$$

$$= \frac{1}{\log k} \int_{B_k} \int L_s(\hat{\sigma}_k^{\pi}(\sigma\boldsymbol{u})/\sigma) f(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \frac{1}{\sigma}\mathrm{d}\sigma,$$

where $\boldsymbol{u} = \boldsymbol{x}/\sigma$. Let $\eta = (2/\log k)\log\sigma - 1$ and $\mathrm{d}\eta = (2/\log k)\mathrm{d}\sigma/\sigma$. Then the condition that $\sigma \in B_k$ is expressed by $|\eta| < 1$, and $\hat{\sigma}_k^{\pi}(\sigma\boldsymbol{u})/\sigma$ is written as

$$\hat{\sigma}_k^{\pi}(\sigma\boldsymbol{u})/\sigma = \int_{1<\sigma s<k} s^{-n-1}f(\boldsymbol{u}/s)\mathrm{d}s \Big/ \int_{1<\sigma s<k} s^{-n-2}f(\boldsymbol{u}/s)\mathrm{d}s$$

$$= \int_{B_k^*} s^{-n-1}f(\boldsymbol{u}/s)\mathrm{d}s \Big/ \int_{B_k^*} s^{-n-2}f(\boldsymbol{u}/s)\mathrm{d}s = \hat{\sigma}_k^*(\boldsymbol{u}|\eta), \quad \text{(say)},$$

where B_k^* is the range of s in the integrals, given by

$$B_k^* = \left\{s\Big| -\frac{1+\eta}{2}\log k < \log s < \frac{1-\eta}{2}\log k\right\}.$$

Hence, the Bayes risk is rewritten by

$$r_k(\pi_k, \hat{\sigma}_k^{\pi}) = \frac{1}{2} \int_{|\eta|<1} \int L_s(\hat{\sigma}_k(\boldsymbol{u}|\eta))f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\mathrm{d}\eta$$

(2.14)
$$\geq \frac{1}{2} \int_{|\eta|<1-\varepsilon} \int L_s(\hat{\sigma}_k(\boldsymbol{u}|\eta))f(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\mathrm{d}\eta.$$

Since $|\eta| < 1 - \varepsilon$, it is noted that $1 - \eta > \varepsilon$ and $1 + \eta > \varepsilon$, which imply that the end points $(\log k)(1 - \eta)/2$ and $-(\log k)(1 + \eta)/2$ tend to infinity and minus infinity as $k \to \infty$. Therefore, the minimaxity of the estimator $\hat{\sigma}^U$ can be proved by using the same arguments as in the proof of Theorem 2..1. ∎

We next construct a class of minimax estimators of the form

$$(2.15) \qquad \hat{\sigma}_\phi = \hat{\sigma}_\phi(|X_1|, \boldsymbol{Z}) = |X_1|\phi(|X_1|/b_0, \boldsymbol{Z}), \quad \boldsymbol{Z} = \boldsymbol{X}/|X_1|$$

where $\phi(y, \boldsymbol{z})$ is an absolutely continuous function. This class includes the minimax estimator $\hat{\sigma}^U$ as the form

$$\hat{\sigma}^U = |X_1|c(\boldsymbol{Z}),$$
$$c(\boldsymbol{Z}) = \int_0^\infty v^{n-1} f(v\boldsymbol{Z})\mathrm{d}v \Big/ \int_0^\infty v^n f(v\boldsymbol{Z})\mathrm{d}v.$$

Theorem 2..4. *Assume that $\phi(y, \boldsymbol{z})$ satisfies the following conditions:*
(a) $\phi(y, \boldsymbol{z})$ is nonincreasing in y,
(b) $\lim_{y\to\infty} \phi(y, \boldsymbol{z}) = c(\boldsymbol{z})$,
(c) $\phi(y, \boldsymbol{z}) \leq \phi^m(y, \boldsymbol{z})$, where

$$(2.16) \qquad \phi^m(y, \boldsymbol{z}) = \int_0^y v^{n-1} f(v\boldsymbol{z})\mathrm{d}v \Big/ \int_0^y v^n f(v\boldsymbol{z})\mathrm{d}v.$$

Then $\hat{\sigma}_\phi$ is a minimax estimator improving on $\hat{\sigma}^U$ relative to the L_s-loss (2.13).

Proof. The same arguments as in the proof of Theorem 2..2 are used for the proof of this theorem, an outline of which is given here. The risk difference of the two estimators $\hat{\sigma}^U$ and $\hat{\sigma}_\phi$ is written by

$$\begin{aligned}
\Delta &= R(\sigma, \hat{\sigma}^U) - R(\sigma, \hat{\sigma}_\phi) \\
&= E\left[\int_1^\infty \frac{\mathrm{d}}{\mathrm{d}t} L_s(|X_1|\sigma^{-1}\phi(t|X_1|/b_0, \boldsymbol{Z}))\mathrm{d}t\right] \\
&= \int\int_{1/b_0}^\infty \{|x_1|/\sigma - 1/\phi(t|x_1|, \boldsymbol{z})\}\, |x_1|\phi'(t|x_1|, \boldsymbol{z})\sigma^{-n} f(\boldsymbol{x}/\sigma)\mathrm{d}t\mathrm{d}\boldsymbol{x} \\
&= \int\int\int_{1/b_0}^\infty \{|x_1|/\sigma - 1/\phi(t|x_1|, \boldsymbol{z})\}\, \phi'(t|x_1|, \boldsymbol{z})\sigma^{-n}|x_1|^n f(\sigma^{-1}|x_1|\boldsymbol{z})\mathrm{d}t\mathrm{d}x_1\mathrm{d}\boldsymbol{z}
\end{aligned}$$

where $\phi'(y, \boldsymbol{z}) = (\partial/\partial y)\phi(y, \boldsymbol{z})$. Making the transformations $\sigma y = |x_1|t$ and $v = y/t$, we can rewrite Δ as

$$\begin{aligned}
\Delta &= \int\int\int_{1/b_0}^\infty \{y/t - 1/\phi(\sigma y, \boldsymbol{z})\}\, \phi'(\sigma y, \boldsymbol{z})(y/t)^n f((y/t)\boldsymbol{z})(\sigma/t)\mathrm{d}t\mathrm{d}y\mathrm{d}\boldsymbol{z} \\
&= \int\int\int_0^{b_0 y} \{v - 1/\phi(\sigma y, \boldsymbol{z})\}\, \phi'(\sigma y, \boldsymbol{z})v^{n-1} f(v\boldsymbol{z})\sigma\mathrm{d}v\mathrm{d}y\mathrm{d}\boldsymbol{z},
\end{aligned}$$

which means that $\Delta \geq 0$ if $\phi'(\sigma y, \boldsymbol{z}) \leq 0$, and if

$$\phi(\sigma y, \boldsymbol{z}) \leq \int_0^{b_0 y} v^{n-1} f(v\boldsymbol{z}) \mathrm{d}v \bigg/ \int_0^{b_0 y} v^n f(v\boldsymbol{z}) \mathrm{d}v \quad \text{for} \quad \sigma > b_0.$$

These requirements are satisfied by the conditions in Theorem 2..4.   ∎

The class described in Theorem 2..4 includes the generalized Bayes estimator against the uniform prior $\mathrm{d}\sigma/\sigma$ on A, given by

$$\hat{\sigma}^m = \int_{b_0}^\infty b^{-n-1} f(\boldsymbol{X}/b) \mathrm{d}b \bigg/ \int_{b_0}^\infty b^{-n-2} f(\boldsymbol{X}/b) \mathrm{d}b$$
$$= |X_1| \phi^m(|X_1|/b_0, \boldsymbol{Z})$$

for $\phi^m(y, \boldsymbol{z})$ is defined by (2.16). In fact, it is easy to check that $\phi^m(y, \boldsymbol{z})$ satisfies the conditions (a), (b) and (c) of Theorem 2..4.

3. Estimation in the Location-Scale Family

3.1. Minimaxity of the best equivariant estimators

In this section, we treat the estimation of the restricted parameters in the location-scale family, which is more complicated than the location or scale families. Let random variable $\boldsymbol{X} = (X_1, \ldots, X_n)$ have the joint density

$$\boldsymbol{X} \sim \frac{1}{\sigma^n} f\left(\frac{\boldsymbol{x} - \mu}{\sigma}\right),$$

and assume that the parameters (μ, σ) is restricted to the space

$$(3.1) \qquad C = \{(\mu, \sigma) | \mu > c_0 \sigma + a_0, \ \sigma > b_0\},$$

where a_0, b_0 and c_0 are constants such that $c_0 \geq 0$, $b_0 \geq 0$ and $-\infty \leq a_0 < \infty$. The unrestricted case is described by $b_0 = c_0 = 0$ and $a_0 = -\infty$.

We treat the estimation of the location μ under the loss function

$$L_{\ell s}(\mu, \sigma; \widehat{\mu}) = (\widehat{\mu} - \mu)^2/\sigma^2.$$

The best location-scale equivariant estimator of μ is given by

$$\widehat{\mu}^U = \int_0^\infty \int_{-\infty}^\infty \mu \frac{1}{\sigma^{n+3}} f\left(\frac{\boldsymbol{X} - \mu}{\sigma}\right) \mathrm{d}\mu\mathrm{d}\sigma \bigg/ \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma^{n+3}} f\left(\frac{\boldsymbol{X} - \mu}{\sigma}\right) \mathrm{d}\mu\mathrm{d}\sigma,$$

which is the generalized Bayes estimator against the right-invariant Haar measure $\mathrm{d}\mu\mathrm{d}\sigma/\sigma$ with respect to the location-scale group (see Lehmann and Casella (1998)). .

Theorem 3..1. *The best location-scale equivariant and unrestricted estimator $\widehat{\mu}^U$ is minimax in the estimation issue of μ on the restricted parameter space C relative to the $L_{\ell s}$-loss.*

Proof. Without any loss of generality, let $a_0 = 0$ and $b_0 = 1$. Then C is given by $A = \{(\mu, \sigma) | \mu > c_0\sigma, \sigma > 1\}$. Let $C_k = \{(\mu, \sigma) | c_0\sigma < \mu < k, 1 < \sigma < k\}$ where k satisfies the condition that $\log k \geq c_0$. Consider the sequence of prior distributions given by

$$\pi_k(\mu, \sigma)\mathrm{d}(\mu, \sigma) = \begin{cases} \{k(\log k - c_0) + c_0\}^{-1}\sigma^{-1}\mathrm{d}\mu\mathrm{d}\sigma & \text{if } (\mu, \sigma) \in C_k \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathrm{d}(\mu, \sigma)$ means $\mathrm{d}\mu\mathrm{d}\sigma$. Then the Bayes estimators are given by

$$\widehat{\mu}_k^\pi = \widehat{\mu}_k^\pi(\boldsymbol{X}) = \int_{C_k} ab^{-n-3}f((\boldsymbol{X} - a)/b)\mathrm{d}(a, b) \Big/ \int_{C_k} b^{-n-3}f((\boldsymbol{X} - a)/b)\mathrm{d}(a, b)$$

with the Bayes risk function

$$r_k(\pi_k, \widehat{\mu}_k^\pi) = \frac{1}{k(\log k - c_0) + c_0} \int_{C_k} \int \left(\frac{\widehat{\mu}_k^\pi(\boldsymbol{x}) - \mu}{\sigma}\right)^2 \frac{1}{\sigma^{n+1}} f\left(\frac{\boldsymbol{x} - \mu}{\sigma}\right) \mathrm{d}\boldsymbol{x}\mathrm{d}(\mu, \sigma)$$

$$(3.2) \qquad = \frac{1}{k(\log k - c_0) + c_0} \int_{C_k} \int \left(\frac{\widehat{\mu}_k^\pi(\sigma\boldsymbol{z} + \mu) - \mu}{\sigma}\right)^2 f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}\frac{1}{\sigma}\mathrm{d}(\mu, \sigma)$$

where $\boldsymbol{z} = (\boldsymbol{x} - \mu)/\sigma$. Letting $t = (a - \mu)/\sigma$ and $s = b/\sigma$, we see that

$$\frac{\widehat{\mu}^\pi(\sigma\boldsymbol{z} + \mu) - \mu}{\sigma} = \frac{\int_{C_k}[(a - \mu)/\sigma](\sigma/b)^{n+3}f([\boldsymbol{z} - (a - \mu)/\sigma]\sigma/b)\mathrm{d}(a, b)}{\int_{C_k}(\sigma/b)^{n+3}f([\boldsymbol{z} - (a - \mu)/\sigma]\sigma/b)\mathrm{d}(a, b)}$$

$$(3.3) \qquad = \frac{\int_{C_k^*} ts^{-n-3}f((\boldsymbol{z} - t)s)\mathrm{d}(t, s)}{\int_{C_k^*} s^{-n-3}f((\boldsymbol{z} - t)s)\mathrm{d}(t, s)},$$

where $C_k^* = \left\{(t, s) \,\middle|\, c_0\sigma s < \sigma t + \mu < k, 1 < \sigma s < k\right\}$. Let $\xi = (2/k)\mu - 1$ and $\eta = (2/\log k)\log \sigma - 1$. Then, C_k^* is rewritten as

$$C_k^* = \left\{(t, s) | c_0s - \frac{1}{2}k^{(1-\eta)/2}(1 + \xi) < t < \frac{1}{2}k^{(1-\eta)/2}(1 - \xi),\right.$$

$$(3.4) \qquad \left. -\frac{1 + \eta}{2}\log k < \log s < \frac{1 - \eta}{2}\log k\right\}$$

and we denote the quantity (3.3) by $\widehat{\mu}^*(\boldsymbol{z}|\xi, \eta)$. Since the condition that $(\mu, \sigma) \in C_k$ is equivalently expressed by

$$(\xi, \eta) \in \left\{2c_0k^{-(1-\eta)/2} - 1 < \xi < 1, \; |\eta| < 1\right\},$$

the Bayes risk (3.2) is rewritten as

$$r_k(\pi_k, \widehat{\mu}_k^\pi) = \frac{(k/4)\log k}{k(\log k - c_0) + c_0} \int_{2c_0k^{(\eta-1)/2}-1<\xi<1, \; |\eta|<1} \int \{\widehat{\mu}^*(\boldsymbol{z}|\xi, \eta)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}(\xi, \eta)$$

$$\geq \frac{(k/4)\log k}{k(\log k - c_0) + c_0} \int_{2c_0k^{(\eta-1)/2}+\varepsilon-1<\xi<1-\varepsilon, \; |\eta|<1-\varepsilon} \int \{\widehat{\mu}^*(\boldsymbol{z}|\xi, \eta)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}(\xi, \eta).$$

Noting that $1 - \eta > \varepsilon$, $1 + \eta > \varepsilon$, $1 - \xi > \varepsilon$ and $1 + \xi > 2c_0 k^{(\eta-1)/2} + \varepsilon$, we see that the set C_k^* given in (3.4) contains the subset

$$(t, s) \in \left\{ c_0 s - a_0 - \frac{1}{2}\varepsilon k^{\varepsilon/2} < t < \frac{1}{2}\varepsilon k^{\varepsilon/2}, \ -\frac{\varepsilon}{2}\log k < \log s < \frac{\varepsilon}{2}\log k \right\},$$

which implies that all the end points of t and $\log s$ go to infinity or minus infinity as k tends to infinity, so that

$$\lim_{k \to \infty} \widehat{\mu}^*(\boldsymbol{z}|\xi, \eta) = \widehat{\mu}^U.$$

Hence, Fatou's lemma is used to evaluate the Bayes risk as

$$\liminf_{k \to \infty} r_k(\pi_k, \widehat{\mu}_k^\pi) \geq \frac{1}{4} \int_{|\xi|<1-\varepsilon, |\eta|<1-\varepsilon} \mathrm{d}(\xi, \eta) \int \{\widehat{\mu}^U(\boldsymbol{z})\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$$

$$= (1 - \varepsilon)^2 \int \{\widehat{\mu}^U(\boldsymbol{z})\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}.$$

Therefore, the minimaxity of the estimator $\widehat{\mu}^U$ can be proved by using the same arguments as in the proof of Theorem 2..1. ∎

3.2.   Improved minimax estimators

Since the estimator $\widehat{\mu}^U$ is outside the parameter space with a positive probability, it may be modified by the truncation at the boundary of C. For an estimator $\widehat{\mu} = \widehat{\mu}(\boldsymbol{X})$ of μ, consider the truncation rule

$$(3.5) \qquad\qquad [\widehat{\mu}(\boldsymbol{X})]^{TR} = \max\{\widehat{\mu}(\boldsymbol{X}), \ a_0 + b_0 c_0\},$$

since $\mu > c_0\sigma + a_0 > b_0 c_0 + a_0$ for $(\mu, \sigma) \in C$.

Proposition 3..1.   *The estimator $\widehat{\mu}(\boldsymbol{X})$ is improved on by the truncated one $[\widehat{\mu}(\boldsymbol{X})]^{TR}$.*

Proof.   The risk difference of the two estimators is written as

$$\begin{aligned}
\Delta =& R(\mu, \sigma; \widehat{\mu}) - R(\mu, \sigma; [\widehat{\mu}]^{TR}) \\
=& \frac{1}{\sigma^2} E\left[\left\{ (\widehat{\mu} - \mu)^2 - ((a_0 + b_0 c_0) - \mu)^2 \right\} I(\widehat{\mu} < a_0 + b_0 c_0) \right] \\
=& \frac{1}{\sigma^2} E\left[\{\widehat{\mu} - (a_0 + b_0 c_0)\} \{\widehat{\mu} + (a_0 + b_0 c_0) - 2\mu\} I(\widehat{\mu} < a_0 + b_0 c_0) \right] \\
\geq& \frac{1}{\sigma^2} E\left[\{\widehat{\mu} - (a_0 + b_0 c_0)\} \{\widehat{\mu} + (a_0 + b_0 c_0) - 2(a_0 + b_0 c_0)\} I(\widehat{\mu} < a_0 + b_0 c_0) \right],
\end{aligned}$$

which is nonnegative, and the proposition is verified. ∎

Using the truncation rule (3.5), we get the truncated estimator

$$\widehat{\mu}^{TR} = \max\{\widehat{\mu}^U, \ a_0 + b_0 c_0\},$$

improving on $\widehat{\mu}^U$. We want to make a class of minimax estimators including $\widehat{\mu}^{TR}$ and the generalized Bayes estimators, though it may be too difficult in the general setup. We thus consider the specific underlying distribution and restriction which allows us to construct such a class.

We here assume that $(X_1, \ldots, X_n)$ is a random sample from a normal distribution $\mathcal{N}(\mu, \sigma^2)$ where μ is restricted to the space $\mu > a_0$, which corresponds to the case that $b_0 = c_0 = 0$ in (3.1). Let $X = \sum_{i=1}^n (X_i - a_0)/\sqrt{n}$ and $S = \sum_{i=1}^n (X_i - \sum_{j=1}^n X_j/n)^2$, which are independently distributed as

$$(3.6) \qquad X \sim \mathcal{N}(\theta, \sigma^2) \quad \text{and} \quad S \sim \sigma^2 \chi_m^2$$

for $\theta = \sqrt{n}(\mu - a_0)$ and $m = n - 1$. Then the parameters are restricted to

$$(3.7) \qquad \theta > 0, \quad 0 < \sigma^2 < \infty,$$

and we consider the estimation of the mean θ under the restriction (3.7) relative to the loss $L_{\ell s}(\theta, \sigma^2; \hat{\theta}) = (\hat{\theta} - \theta)^2/\sigma^2$. The minimaxity of X follows from Theorem 3..1. To construct a class of minimax estimators improving on X, consider estimators of the form

$$(3.8) \qquad \hat{\theta}_\phi = X - \sqrt{S}\phi\left(X/\sqrt{S}\right),$$

for an absolutely continuous function ϕ.

Theorem 3..2. *Assume that $\phi(w)$ satisfies the following conditions:*
(a) $\phi(w)$ *is nondecreasing in w,*
(b) $\lim_{w \to \infty} \phi(w) = 0$,
(c) $\phi(w) \geq \phi^m(w)$, *where*

$$\phi^m(w) = \int_0^\infty \int_{-\infty}^w y e^{-v(1+y^2)/2} \mathrm{d}y \, v^{(m+1)/2} \mathrm{d}v \Big/ \int_0^\infty \int_{-\infty}^w e^{-v(1+y^2)/2} \mathrm{d}y \, v^{(m+1)/2} \mathrm{d}v$$

$$(3.9) \quad = -\frac{1}{m+1} \frac{(1+w^2)^{-(m+1)/2}}{\int_{-\infty}^w (1+x^2)^{-(m+1)/2-1} \mathrm{d}x}.$$

Then $\hat{\theta}_\phi$ is a minimax estimator improving on X relative to the $L_{\ell s}$-loss.

Proof. The same arguments as in the proof of Theorem 2..2 are used to rewrite the risk difference of the two estimators X and $\hat{\theta}_\phi$ as

$$\begin{aligned}
\Delta &= R(\theta, \sigma; X) - R(\theta, \sigma; \hat{\theta}_\phi) \\
&= E\left[\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\left\{X - \theta - \sqrt{S}\phi\left(\frac{X}{\sqrt{S}} + t\right)\right\}^2 / \sigma^2 \mathrm{d}t\right] \\
&= -2E\left[\int_0^\infty \left\{X - \theta - \sqrt{S}\phi\left(\frac{X}{\sqrt{S}} + t\right)\right\} \frac{\sqrt{S}}{\sigma^2}\phi'\left(\frac{X}{\sqrt{S}} + t\right) \mathrm{d}t\right].
\end{aligned}$$

Let $U = (X-\theta)/\sqrt{S}$ and $V = S/\sigma^2$. The joint density of (U, V) is $c_1 v^{(m+1)/2-1} e^{-v(1+u^2)/2}$ for the normalizing constant c_1. Then Δ is expressed by

$$\Delta = -2 \int \int \int_0^\infty \left\{ u - \phi(u + \lambda/\sqrt{v} + t) \right\} \phi'(u+\lambda/\sqrt{v}+t) c_1 v^{(m+1)/2} e^{-v(1+u^2)/2} \mathrm{d}t \mathrm{d}u \mathrm{d}v$$

where $\lambda = \theta/\sigma > 0$. Making the transformations $w = u+\lambda/\sqrt{v}+t$ and $y = -t+w$ with $\mathrm{d}w = \mathrm{d}u$ and $\mathrm{d}y = \mathrm{d}t$, we can rewrite Δ as

$$\Delta = -2 \int \int \int_0^\infty \left\{ w - \lambda/\sqrt{v} - t - \phi(w) \right\} \phi'(w) c_1 v^{(m+1)/2} e^{-v\{1+(w-\lambda/\sqrt{v}-t)^2\}/2} \mathrm{d}t \mathrm{d}w \mathrm{d}v$$

$$= -2 \int \int \int_{-\infty}^w \left\{ y - \lambda/\sqrt{v} - \phi(w) \right\} \phi'(w) c_1 v^{(m+1)/2} e^{-v\{1+(y-\lambda/\sqrt{v})^2\}/2} \mathrm{d}y \mathrm{d}v \mathrm{d}w.$$

Hence, it is seen that $\Delta \geq 0$ if $\phi'(w) \geq 0$ and if

$$\phi(w) \geq \frac{\int \int_{-\infty}^w (y - \lambda/\sqrt{v}) v^{(m+1)/2} e^{-v\{1+(y-\lambda/\sqrt{v})^2\}/2} \mathrm{d}y \mathrm{d}v}{\int \int_{-\infty}^w v^{(m+1)/2} e^{-v\{1+(y-\lambda/\sqrt{v})^2\}/2} \mathrm{d}y \mathrm{d}v} \equiv \phi_\lambda(w).$$

By using the integration by parts, $\phi_\lambda(w)$ is expressed by

$$\phi_\lambda(w) = -\frac{\int v^{(m-1)/2} e^{-v(1+w^2)/2+\sqrt{v}w\lambda} \mathrm{d}v}{\int \int_{-\infty}^w v^{(m+1)/2} e^{-v(1+x^2)/2+\sqrt{v}x\lambda} \mathrm{d}x \mathrm{d}v}$$

$$= -\frac{(1+w^2)^{-(m+1)/2} \int z^{(m-1)/2} e^{-z/2} e^{(w/\sqrt{1+w^2})\sqrt{z}\lambda} \mathrm{d}z}{\int_{-\infty}^w (1+x^2)^{-(m+1)/2-1} \int z^{(m+1)/2} e^{-z/2} e^{(x/\sqrt{1+x^2})\sqrt{z}\lambda} \mathrm{d}z \mathrm{d}x}$$

$$= -\frac{(1+w^2)^{-(m+1)/2} E[e^{(w/\sqrt{1+w^2})\sqrt{Z}\lambda}]}{\int_{-\infty}^w (1+x^2)^{-(m+1)/2-1} E[Z e^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda}] \mathrm{d}x},$$

where Z is a random variable having χ^2_{m+1}. To complete the proof, we need to show the inequality that

(3.10) $\phi_\lambda(w) \leq \lim_{\lambda \to 0} \phi_\lambda(w) \equiv \phi^m(w).$

We first show the inequality (3.10) in the case that $w < 0$. The inequality (3.10) is written by

$$\frac{(1+w^2)^{-(m+1)/2} E[e^{(w/\sqrt{1+w^2})\sqrt{Z}\lambda}]}{\int_{-\infty}^w (1+x^2)^{-(m+1)/2-1} E[Z e^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda}] \mathrm{d}x} \geq \frac{(1+w^2)^{-(m+1)/2}}{\int_{-\infty}^w (1+x^2)^{-(m+1)/2-1} E[Z] \mathrm{d}x},$$

equivalently,

$$\int_{-\infty}^w (1+x^2)^{-(m+1)/2-1} E[Z] \mathrm{d}x E[e^{(w/\sqrt{1+w^2})\sqrt{Z}\lambda}]$$

(3.11) $$\geq \int_{-\infty}^w (1+x^2)^{-(m+1)/2-1} E[Z e^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda}] \mathrm{d}x.$$

Since $x < w < 0$, the function $e^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda}$ is decreasing in Z, so that we get the inequality

$$E\left[Ze^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda}\right] \le E[Z]E\left[e^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda}\right].$$

Noting that for $x < w$,

$$x/\sqrt{1+x^2} \le w/\sqrt{1+w^2},$$

we can see that the inequality (3.11) holds for $w < 0$.

We next treat the case that $w > 0$. It is noted that the function $e^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda}$ can be expanded as

$$e^{(x/\sqrt{1+x^2})\sqrt{Z}\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \frac{w^j}{(1+w^2)^{j/2}} Z^{j/2},$$

and that $E[Z^{j/2}] = \Gamma((m+j+1)/2)2^{j/2}/\Gamma((m+1)/2)$. Taking these notes into account, we can see that the required inequality (3.11) is expressed by $I_{-\infty,w}(w) \ge 0$, where the notation $I_{a,b}(w)$ is defined by

$$I_{a,b}(w) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}\left(\frac{w^2}{1+w^2}\right)^{j/2}\Gamma\left(\frac{m+j+1}{2}\right)2^{j/2}\int_a^b \frac{1}{(1+x^2)^{(m+1)/2+1}}\mathrm{d}x$$

$$- \frac{2}{m+1}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}\Gamma\left(\frac{m+j+1}{2}+1\right)2^{j/2}\int_a^b \frac{x^j}{(1+x^2)^{(m+j+1)/2+1}}\mathrm{d}x.$$

The integrals in $I_{-\infty,w}(w)$ are decomposed as

$$\int_{-\infty}^{w} \frac{1}{(1+x^2)^{(m+1)/2+1}}\mathrm{d}x = \int_{-\infty}^{-w}\frac{1}{(1+x^2)^{(m+1)/2+1}}\mathrm{d}x + \int_{-w}^{w}\frac{1}{(1+x^2)^{(m+1)/2+1}}\mathrm{d}x,$$

$$\int_{-\infty}^{w}\frac{x^j}{(1+x^2)^{(m+j+1)/2+1}}\mathrm{d}x = \int_{-\infty}^{-w}\frac{x^j}{(1+x^2)^{(m+j+1)/2+1}}\mathrm{d}x + \int_{-w}^{w}\frac{x^j}{(1+x^2)^{(m+j+1)/2+1}}\mathrm{d}x,$$

and it is seen that

$$\int_{-w}^{w}\frac{x^j}{(1+x^2)^{(m+j+1)/2+1}}\mathrm{d}x = \begin{cases} 0 & \text{for odd } j, \\ 2\int_0^w \frac{x^j}{(1+x^2)^{(m+j+1)/2+1}}\mathrm{d}x & \text{for even } j, \end{cases}$$

Hence, $I_{-\infty,w}(w)$ is decomposed as $I_{-\infty,w}(w) = I_{-\infty,-w}(w) + I_{-w,w}(w)$, and $I_{-w,w}(w)$ is expressed by

$$I_{-w,w}(w) = 2\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}\left(\frac{w^2}{1+w^2}\right)^{j/2}\Gamma\left(\frac{m+j+1}{2}\right)2^{j/2}\int_0^w \frac{1}{(1+x^2)^{(m+1)/2+1}}\mathrm{d}x$$

$$(3.12) \qquad - \frac{4}{m+1}\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!}\Gamma\left(\frac{m+1}{2}+k+1\right)2^k\int_0^w \frac{x^{2k}}{(1+x^2)^{(m+1)/2+k+1}}\mathrm{d}x.$$

Since $w < 0$, from the arguments around (3.11), it follows that $I_{-\infty, -w}(w) \geq 0$, so that our final step is to show that $I_{-w,w}(w) \geq 0$. Making the transformations $t = x^2$ and $u = t/(1+t)$ in turn with $dx = dt/(2\sqrt{t})$ and $dt = du/(1-u)^2$, we demonstrate that

$$\int_0^w \frac{x^{2k}}{(1+x^2)^{(m+1)/2+k+1}} dx = \frac{1}{2} \int_0^{w^2} \frac{t^{k-1/2}}{(1+t)^{(m+1)/2+k+1}} dt$$

$$= \frac{1}{2} \int_0^A u^{1/2+k-1}(1-u)^{m/2} du,$$

where $A = w^2/(1+w^2)$, which is used to rewrite (3.12) as

$$I_{-w,w}(w) = \sum_{j=0}^\infty \frac{\lambda^j}{j!} A^{j/2} \Gamma\left(\frac{m+j+1}{2}\right) 2^{j/2} \int_0^A u^{1/2-1}(1-u)^{m/2} du$$

$$- \frac{2}{m+1} \sum_{k=0}^\infty \frac{\lambda^{2k}}{(2k)!} \Gamma\left(\frac{m+1}{2} + k + 1\right) 2^k \int_0^A u^{1/2+k-1}(1-u)^{m/2} du$$

$$= \sum_{k=0}^\infty \frac{\lambda^{2k+1}}{(2k+1)!} A^{1/2+k} \Gamma\left(\frac{m}{2} + k + 1\right) 2^{1/2+k} \int_0^A u^{1/2-1}(1-u)^{m/2} du$$

$$+ \sum_{k=0}^\infty \frac{\lambda^{2k}}{(2k)!} J_k(w),$$

where

$$J_k(w) = A^k \Gamma\left(\frac{m+1}{2} + k\right) 2^k \int_0^A u^{1/2-1}(1-u)^{m/2} du$$

$$- \frac{2}{m+1} \Gamma\left(\frac{m+1}{2} + k + 1\right) 2^k \int_0^A u^{1/2+k-1}(1-u)^{m/2} du.$$

We can thus complete the proof by showing the inequality that $J_k(w) \geq 0$. For the purpose, note that

$$\frac{\int_0^A u^{1/2+k-1}(1-u)^{m/2} du}{\int_0^A u^{1/2-1}(1-u)^{m/2} du} \leq \frac{\int_0^A u^{1/2+k-1} du}{\int_0^A u^{1/2-1} du} = \frac{1}{1+2k} A^k.$$

Then $J_k(w)$ can be evaluated as

$$J_k(w) \geq A^k \Gamma\left(\frac{m+1}{2} + k\right) 2^k \int_0^A u^{1/2-1}(1-u)^{m/2} du \left\{1 - \frac{m+1+2k}{(m+1)(1+2k)}\right\},$$

which is nonnegative. This shows that $I_{-w,w}(w) \geq 0$, and the proof of Theorem 3..2 is complete.   ∎

The generalized Bayes estimator of θ against the prior distribution $d\theta d\sigma^2/\sigma^2$ on $\theta > 0$ and $\sigma^2 > 0$ is given by

$$\hat{\theta}^m = \frac{\int_0^\infty \int_0^\infty \theta(\sigma^2)^{-(m+1)/2-2} e^{-\{(X-\theta)^2+S\}/2\sigma^2} d\theta d\sigma^2}{\int_0^\infty \int_0^\infty (\sigma^2)^{-(m+1)/2-2} e^{-\{(X-\theta)^2+S\}/2\sigma^2} d\theta d\sigma^2}$$

$$= X - \sqrt{S} \phi^m(X/\sqrt{S}),$$

where the function $\phi^m(w)$ is defined by (3.9). It can be seen that $\phi^m(w)$ satisfies all the conditions of Theorem 3..2, and the gneralized Bayes estimator $\hat{\theta}^m$ is minimax.

3.3.   Minimax estimation of the scale

We next look at the estimation of the scale σ under the restriction (3.1) relative to the loss $L_s(\hat{\sigma}/\sigma)$ given in (2.13). The best location-scale equivariant estimator of σ is given by

$$
\hat{\sigma}^U = \int_0^\infty \int_{-\infty}^\infty \sigma^{-n-1} f\left(\frac{\boldsymbol{X}-\mu}{\sigma}\right) \mathrm{d}\mu\mathrm{d}\sigma \Big/ \int_0^\infty \int_{-\infty}^\infty \sigma^{-n-2} f\left(\frac{\boldsymbol{X}-\mu}{\sigma}\right) \mathrm{d}\mu\mathrm{d}\sigma
$$

$$
= \int_0^\infty \sigma^{-n} \int_{-\infty}^\infty f\left(\frac{\boldsymbol{X}}{\sigma}-t\right) \mathrm{d}t\mathrm{d}\sigma \Big/ \int_0^\infty \sigma^{-n-1} \int_{-\infty}^\infty f\left(\frac{\boldsymbol{X}}{\sigma}-t\right) \mathrm{d}t\mathrm{d}\sigma.
$$

Using the same arguments as in the proofs of Theorems 2..3 and 3..1, we can verify the minimaxity of $\hat{\sigma}^U$.

Theorem 3..3.   *The best location-scale equivariant and unrestricted esti-mator $\hat{\sigma}^U$ is minimax in the estimation issue of σ on the restricted parameter space C relative to the L_s-loss.*

Since the estimator $\hat{\sigma}^U$ can take values outside the parameter space C, it should be truncated at the boundary of C. For an estimator $\hat{\sigma} = \hat{\sigma}(\boldsymbol{X})$ of σ, consider the truncation rule

(3.13) $$[\hat{\sigma}(\boldsymbol{X})]^{TR} = \max\{\hat{\sigma}(\boldsymbol{X}),\ b_0\}.$$

Proposition 3..2.   *The estimator $\hat{\sigma}(\boldsymbol{X})$ is improved on by the truncated one $[\hat{\sigma}(\boldsymbol{X})]^{TR}$ relative to the L_s-loss.*

Proof.   The risk difference of the two estimators is written as

$$
\begin{aligned}
\Delta =& R(\mu, \sigma; \hat{\sigma}) - R(\mu, \sigma; [\hat{\sigma}]^{TR}) \\
=& E\left[\left\{\left(\frac{\hat{\sigma}}{\sigma} - \log\hat{\sigma}\right) - \left(\frac{b_0}{\sigma} - \log b_0\right)\right\} I(\hat{\sigma} < b_0)\right] \\
=& E\left[\left\{\left(\frac{\hat{\sigma}}{b_0} - 1\right)\frac{b_0}{\sigma} - \log\frac{\hat{\sigma}}{b_0}\right\} I(\hat{\sigma} < b_0)\right] \\
\geq& E\left[\left(\frac{\hat{\sigma}}{b_0} - \log\frac{\hat{\sigma}}{b_0} - 1\right) I(\hat{\sigma} < b_0)\right],
\end{aligned}
$$

which is nonnegative, and the proposition is verified.   ∎

Using the truncation rule (3.13), we get the truncated estimator

$$
\hat{\sigma}^{TR} = \max\{\hat{\sigma}^U,\ b_0\},
$$

improving on $\hat{\sigma}^U$. To proceed further study on the minimax estimation, we shall specify the underlying distribution as

$$X \sim \mathcal{N}(\theta, \sigma^2) \quad \text{and} \quad S \sim \sigma^2 \chi_m^2,$$

given by (3.6).

When we consider the estimation of the variance σ^2 in the normal model, it is known that the unbiased estimator $\hat{\sigma}^{2U} = S/m$ is dominated by using the information contained in X, that is, Stein (1964) showed that

(3.14)
$$\hat{\sigma}^{2ST} = \min\left\{ \frac{S}{m}, \ \frac{S + X^2}{m+1} \right\}$$

dominates $\hat{\sigma}^{2U}$ relative to the L_s-loss. It is noted that $\hat{\sigma}^{2ST}$ is expressed by $\hat{\sigma}^{2ST} = S\psi^{ST}(X^2/S)$ for

(3.15)
$$\psi^{ST}(w) = \left\{ \frac{1}{m}, \ \frac{1+w}{m+1} \right\}.$$

Kubokawa (1994a, 99) extended the result to the class of the estimators

$$\hat{\sigma}_\psi^2 = S\psi(X^2/S)$$

and derived conditions on ψ for the dominance over $\hat{\sigma}^{2U}$.

Proposition 3..3. *Assume that $\psi(w)$ satisfies the following conditions:*
(a) $\psi(w)$ *is nondecreasing in w.*
(b) $\lim_{w \to \infty} \psi(w) = 1/m$.
(c) $\psi(w) \geq \psi^m(w)$, *where*

$$\psi^m(w) = \frac{1}{m+1} \int_0^w \frac{u^{1/2-1}}{(1+u)^{(m+1)/2}} du \Big/ \int_0^w \frac{u^{1/2-1}}{(1+u)^{(m+1)/2+1}} du.$$

Then $\hat{\sigma}_\psi^2$ is a minimax estimator improving on $\hat{\sigma}^{2U} = S/m$ relative to the L_s-loss.

Now we assume that the parameters are restricted to the space

(3.16)
$$\sigma^2 > 1, \quad -\infty < \mu < \infty.$$

and under this restriction, we want to consider to construct a class of estimators improving on the estimator $\hat{\sigma}_\psi^2$ given in Proposition 3..3. For the purpose, consider estimators of the form

$$\hat{\sigma}_{\phi,\psi}^2 = S\psi(X^2/S)\phi(S),$$

and we get the dominance result.

Theorem 3..4. *Assume that the functions $\psi(w)$ and $\phi(s)$ satisfy the following conditions:*

(a) $\psi(w) \leq 1/m$.

(b) $\phi(s)$ *is nonincreasing in s.*

(c) $\lim_{s\to\infty} \phi(s) = 1$.

(d) $\phi(s) \leq \phi_U^m(s)$, *where*

$$\phi_U^m(s) = m \int_0^s v^{m/2-1} e^{-v/2} dv \Big/ \int_0^s v^{m/2} e^{-v/2} dv.$$

Then $\hat{\sigma}_{\psi,\phi}^2 = S\psi(X^2/S)\phi(S)$ dominates $\hat{\sigma}_\psi^2 = S\psi(X^2/S)$ relative to the L_s-loss under the restriction (3.16).

Proof. The same arguments as in the proof of Theorem 2..2 are used for the proof of this theorem. The risk difference of the two estimators $\hat{\sigma}_\phi^2$ and $\hat{\sigma}_{\phi,\psi}^2$ is written by

$$
\begin{aligned}
\Delta =&R(\mu, \sigma; \hat{\sigma}_\psi^2) - R(\mu, \sigma; \hat{\sigma}_{\psi,\phi}^2)\\
=&E\left[\int_1^\infty \frac{d}{dt} L_s\left(\frac{S}{\sigma^2}\psi(X^2/S)\phi(tS)\right) dt\right]\\
=&E\left[\int\int_1^\infty \left\{\frac{S}{\sigma^2}\psi(X^2/S) - \frac{1}{\phi(tS)}\right\} S\phi'(tS)dt\right]\\
=&\int\int\int_1^\infty \left\{v\psi(u/v) - \frac{1}{\phi(\sigma^2 tv)}\right\} v\phi'(\sigma^2 tv)dt f_m(v)f_1(u;\lambda)dvdu,
\end{aligned}
$$

where $f_m(v)$ is the density of the chi-squared distribution with m degrees of freedom and $f_1(u;\lambda)$ denotes the density of the noncentral chi-square distribution with one degree of freedom and the noncentrality parameter $\lambda = \theta^2/\sigma^2$. Making the transformations $s = tv$ and $y = s/t$ in turn with $ds = tdv$ and $dy = (s/t^2)dt$, we can rewrite Δ as

$$
\begin{aligned}
\Delta =&\int\int\int_1^\infty \left\{(s/t)\psi(tu/s) - 1/\phi(\sigma^2 s)\right\} \phi'(\sigma^2 s)(s/t^2)f_m(s/t)f_1(u;\lambda)dtduds\\
=&\int\int\int_0^s \left\{y\psi(u/y) - 1/\phi(\sigma^2 s)\right\} \phi'(\sigma^2 s)f_m(y)f_1(u;\lambda)dyduds.
\end{aligned}
$$

Since $\phi'(s) \leq 0$, it is seen that $\Delta \geq 0$ if

$$
\phi(\sigma^2 s) \leq \frac{\int_0^s \int f_m(y)f_1(u;\lambda)dudy}{\int_0^s \int y\phi(u/y)f_m(y)f_1(u;\lambda)dudy}
$$

(3.17)
$$
= \frac{\int_0^s f_m(y)dy}{\int_0^s \int y\phi(u/y)f_m(y)f_1(u;\lambda)dudy}.
$$

From the condition (b) and the assumption that $\sigma^2 > 1$, we see that $\phi(\sigma^2 s) \leq \phi(s)$. Hence from the condition (a), it is sufficient to show that

$$
\phi(s) \leq \frac{\int_0^s f_m(y)dy}{\int_0^s \int y(1/m)f_m(y)f_1(u;\lambda)dudy} = m\frac{\int_0^s f_m(y)dy}{\int_0^s yf_m(y)dy},
$$

which is guaranteed by the condition (d), and the proof is complete.   ∎

Letting $\phi_U^{TR}(s) = \max(1, m/s)$ and considering $\phi_U^m(s)$, we see that $\phi_U^{TR}(s)$ and $\phi_U^m(s)$ satisfy the conditions (b), (c) and (d) of Theorem 3..4, so that we get improved procedures

$$S\psi(X^2/S) \times \phi_U^{TR}(S) \quad \text{and} \quad S\psi(X^2/S) \times \phi_U^m(S).$$

For instance, the Stein estimator $\hat{\sigma}^{2ST}$, given by (3.14), is dominated by the truncated estimator

$$\min\left\{\frac{S}{m}, \frac{S+X^2}{m+1}\right\} \times \max\left\{1, \frac{m}{S}\right\},$$

which is, from (3.13), further dominated by

$$\max\left[\min\left\{\frac{S}{m}, \frac{S+X^2}{m+1}\right\} \times \max\left\{1, \frac{m}{S}\right\}, \ 1\right].$$

When the parameters are restricted to the space

(3.18) $\sigma^2 < 1, \quad -\infty < \mu < \infty,$

consider the estimators of the form

$$\hat{\sigma}_\phi^{2TR} = S\min\left\{\phi(S, X^2/S), \ \psi^{ST}(X^2/S)\right\},$$

where $\psi^{ST}(w)$ is defined by (3.15). Then we derive a condition for $\hat{\sigma}_\phi^2$ to dominate the Stein estimator $\hat{\sigma}^{2ST}$ given by (3.14).

Theorem 3..5. *Assume that the functions $\phi(s, w)$ satisfy the following conditions:*
  (a) $\phi(s, w)$ *is nonincreasing in s.*
  (b) $\lim_{s\to 0} \phi(s, w) \geq \psi^{ST}(w)$.
  (c) $\phi(s, w) \geq \phi_L^m(s, w)$, *where*

(3.19) $\phi_L^m(s, w) = \int_s^\infty y^{(m+1)/2-1}e^{-y(1+w)/2}\,\mathrm{d}y \Big/ \int_s^\infty y^{(m+1)/2}e^{-y(1+w)/2}\,\mathrm{d}y.$

Then $\hat{\sigma}_\phi^{2TR} = S\min\{\phi(S, X^2/S), \psi^{ST}(X^2/S)\}$ dominates the Stein estimator $\hat{\sigma}^{2ST}$, given by (3.14), relative to the L_s-loss under the restriction (3.18).

Proof. The same arguments as in the proof of Theorem 3..4 are used for the proof of this theorem. Let $\Phi(s, w) = \min\left\{\phi(s, w), \ \psi^{ST}(w)\right\}$. Since

$\lim_{s\to 0}\Phi(s,w) = \psi^{ST}(w)$, the risk difference of the two estimators $\hat{\sigma}^{2ST}$ and $\hat{\sigma}_\phi^{2TR}$ is written by

$$\Delta = R(\mu,\sigma;\hat{\sigma}^{2ST}) - R(\mu,\sigma;\hat{\sigma}_\phi^{2TR})$$

$$= -E\left[\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}L_s\left(\frac{S}{\sigma^2}\Phi(tS,X^2/S)\right)\mathrm{d}t\right]$$

$$= -E\left[\int\int_0^1\left\{\frac{S}{\sigma^2} - \frac{1}{\Phi(tS,X^2/S)}\right\}S\Phi'(tS,X^2/S)\mathrm{d}t\right]$$

$$= -\int\int\int_0^1\left\{v - \frac{1}{\Phi(\sigma^2 tv,u/v)}\right\}v\Phi'(\sigma^2 tv,u/v)\mathrm{d}t\,f_m(v)f_1(u;\lambda)\mathrm{d}v\mathrm{d}u,$$

where $\Phi'(s,w) = (\mathrm{d}/\mathrm{d}s)\Phi(s,w)$. By the same arguments as in the proof of Theorem 3..4, Δ is rewritten as

$$\Delta = -\int\int\int_s^\infty\left\{y - 1/\Phi(\sigma^2 s,u/y)\right\}\Phi'(\sigma^2 s,u/y)f_m(y)f_1(u;\lambda)\mathrm{d}y\mathrm{d}u\mathrm{d}s$$

$$= -\int\int\int_s^\infty\left\{y - 1/\Phi(\sigma^2 s,w)\right\}y\Phi'(\sigma^2 s,w)f_m(y)f_1(wy;\lambda)\mathrm{d}y\mathrm{d}s\mathrm{d}w.$$

It is sufficient to show that $\Delta \geq 0$ in the case that $\Phi'(s,w) < 0$, that is, $\phi(s,w) < \psi^{ST}(w)$. Hereafter, we thus consider the case that $\Phi(s,w) = \phi(s,w)$. Since $\phi'(s,w) < 0$, it is seen that $\Delta \geq 0$ if

$$\phi(\sigma^2 s,w) \geq \int_s^\infty yf_m(y)f_1(wy;\lambda)\mathrm{d}y\bigg/\int_s^\infty y^2 f_m(y)f_1(wy;\lambda)\mathrm{d}y.$$

Let $f_1(u) = f_1(u;0)$ for $\lambda = 0$. Also let $E^*[\cdot]$ stand for the expectation with respect to the density $yf_m(y)f_1(wy)\mathrm{d}y/\int_s^\infty yf_m(y)f_1(wy)\mathrm{d}y$. From the fact that $f_1(u;\lambda)/f_1(u)$ is increasing in u, it follows that

$$E^*\left[\frac{f_1(wY;\lambda)}{f_1(wY)}\right]\times E^*[Y] \leq E^*\left[\frac{f_1(wY;\lambda)}{f_1(wY)}Y\right],$$

which rewritten as

$$\frac{\int_s^\infty yf_m(y)f_1(wy;\lambda)\mathrm{d}y}{\int_s^\infty y^2 f_m(y)f_1(wy;\lambda)\mathrm{d}y} \leq \frac{\int_s^\infty yf_m(y)f_1(wy)\mathrm{d}y}{\int_s^\infty y^2 f_m(y)f_1(wy)\mathrm{d}y}.$$

From the condition (a) and the assumption that $\sigma^2 < 1$, we see that $\phi(\sigma^2 s,w) \geq \phi(s,w)$. Hence, it is sufficient to show that

$$\phi(s) \geq \int_s^\infty yf_m(y)f_1(wy)\mathrm{d}y\bigg/\int_s^\infty y^2 f_m(y)f_1(wy)\mathrm{d}y,$$

for $s \in \{s\,|\,\phi(s,w) < \phi^{ST}(w)\}$, which is guaranteed by the condition (c), and the proof is complete. ∎

It can be seen that the function $\phi_L^m(s, w)$ given in (3.19) satisfies the conditions (a), (b) and (c) of Theorem 3..5. Hence we get the estimator

$$S \min \left\{ \phi_L^m(S, X^2/S), \ \psi^{ST}(X^2/S) \right\} = \min \left\{ S\phi_L^m(S, X^2/S), \ \hat{\sigma}^{2ST} \right\},$$

improving on the Stein estimator $\hat{\sigma}^{2ST}$. Since $\phi_L^m(S, X^2/S) \leq 1/S$, another simple improved estimator is given by $\min \left\{ 1, \ \hat{\sigma}^{2ST} \right\}$.

REFERENCES

Casella, G. and Strawderman, W.E. (1981). Estimating a bounded mean. *Ann. Statist.*, **9**, 870-879.

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.

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

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation.* 2nd ed., Springer-Verlag, New York,

Marchand, E. and Perron, F. (2001). Improving on the MLE of a bounded normal mean. *Ann. Statist.*, **29**, 1078-1093.

Marchand, E. and Strawderman, W.E. (2004). 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.*, to appear.

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.