CIRJE-F-749

Non-minimaxity of Linear Combinations of Restricted Location Estimators and Related Problems

Tatsuya Kubokawa
University of Tokyo

William E. Strawderman
Rutgers University

July 2010

Non-minimaxity of Linear Combinations of Restricted Location Estimators and Related Problems

Tatsuya Kubokawa *and William E. Strawderman†
University of Tokyo and *Rutgers University*

July 7, 2010

The estimation of a linear combination of several restricted location parameters is addressed from a decision-theoretic point of view. The corresponding linear combination of the best location equivariant and the unrestricted unbiased estimators is minimax. Since the locations are restricted, it is reasonable to use the linear combination of the restricted estimators such as maximum likelihood estimators. In this paper, a necessary and sufficient condition for such restricted estimators to be minimax is derived, and it is shown that the restricted estimators are not minimax when the number of the location parameters is large. The condition for the minimaxity is examined for some specific distributions. Finally, similar problems of estimating the product and sum of the restricted scale parameters are studied, and it is shown that similar non-dominance properties appear when the number of the scale parameters is large.

Key words and phrases: Decision theory, linear combination, location parameter, maximum likelihood estimator, restricted parameter, restricted estimator, scale parameter, truncated estimator.

1 Introduction

Point estimation of restricted parameters has been studied from a decision-theoretic point of view since Katz (1961) and Farrell (1964). This classical problem has been revisited by Marchand and Strawderman (2004, 2005), Kubokawa and Saleh (1998), Kubokawa (2004), Tsukuma and Kubokawa (2008) and others. One of the most interesting issues is whether the generalized Bayes estimator against the uniform prior over the restricted space is minimax or not. Hartigan (2004) recently considered the simultaneous estimation of a mean vector restricted to a convex set in a k-variate normal distribution and used the Gauss divergence theorem to show that the generalized Bayes estimator against the

*Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan E-Mail: tatsuya@e.u-tokyo.ac.jp
†Department of Statistics, Rutgers University, Piscataway, New Jersey, 08854-8019, E-mail: straw@stat.rutgers.edu

uniform prior is minimax. However, Kubokawa (2010) established the non-minimaxity of the generalized Bayes estimator in the context of the estimation of the sum of k restricted normal means when $k \geq 2$. This shows that there is a deifference in minimaxity considerations between these estimation problems. This paper is concerned with the latter problem, and we shall investgate the minimaxity of the maximum likelihood estimator or the restricted best equivariant estimator.

To explain instructively the problem treated here, consider the following simple model: Let $X_1, \ldots, X_k$ be mutually independent random variables such that for $i = 1, \ldots, k$, X_i has a normal distribution $\mathcal{N}(\mu_i, 1)$ where the mean μ_i is restricted to the space $\mu_i > 0$. Then, we want to consider the problem of estimating the sum of the means $\theta = \sum_{i=1}^k \mu_i$ relative to the quadratic loss function $(\hat{\theta} - \theta)^2$. It is noted that the risk function is the mean squared error (MSE). A benchmark estimator is the unrestricted estimator $\hat{\theta}^M = \sum_{i=1}^k X_i$, which is minimax and also the best location equivariant estimator. Since the means μ_i are restricted, however, the unrestricted minimax estimator has a drawback of taking negative values. To address this issue, two methods are available. One is the generalized Bayes estimator of θ against the uniform prior over the restricted spaces, and the other is the maximum likelihood estimator (MLE). Kubokawa (2010) recently developed the results that the generalized Bayes estimator is not minimax for $k \geq 2$ while it is minimax for $k = 1$. This fact raises the question whether the MLE is minimax or not for $k \geq 2$. The MLE is given by $\hat{\theta}^{TR} = \sum_{i=1}^k \max\{X_i, 0\}$, and intuitively it would be plausible that the MLE may be minimax, namely, it dominates $\sum_{i=1}^k X_i$ under the restriction $\mu_i > 0$ for $i = 1, \ldots, k$.

In this paper, we treat more general location families with restricted location parameters, and consider the problem of estimating linear combinations of the restricted location parameters relative to quadratic loss. The best location equivariant and unrestricted estimator of the linear combination is minimax, but inadmissible because the parameter space is restricted. Thus, it is reasonable to consider the linear combination of the truncated estimators which limit the unrestricted minimax estimators over the restricted space. The MLE in the case of the above normal distributions is an example of the truncated estimators. In Section 2, we derive a necessary and sufficient condition for the truncated estimator to be minimax. This condition implies that the truncated estimator is minimax for small k, but not minimax for large k. In Section 3, the necessary and sufficient condition is examined for scale mixtures of normal distributions including normal, t-, double exponential and logistic distibutions, symmetric unimodal distributions including a uniform distribution, and an exponential distribution. For example, the MLE of $\sum_{i=1}^k \mu_i$ in the above normal distributions is minimax for $k \leq 4$, while it is not minimax for $k \geq 5$. The behavior of MSE's of the MLE is illustrated in Figure 1 for $k = 1, 4, 5, 8$. It is also shown that the conditoin $k \leq 4$ is sufficient for the minimaxity of the MLE in estimation of any linear combination $\sum_{i=1}^k a_i \mu_i$.

Section 4 investigates whether the non-minimaxity or non-dominance property of the truncated estimators can be extended to the estimation of the restricted scale parameters. There we consider the two problems of estimating the product and the sum of the restricted scale parameters. Since estimation of the product of scales is invariant under the scale transformations, the best scale equivariant estimator is minimax. A necessary and sufficient condition is derived for the product of the truncated estimators to be minimax, and it is shown that it is not minimax for large k. For the estimation of the sum of scales, the invariance structure does not hold and we could not provide a minimax estimator. Instead, we handle an unbiased estimator for the sum of scales, derive a condition for the restricted estimator to dominate the unbiased estimator, and investigate the condition for exponential and uniform distributions.

2 Minimaxity and Non-minimaxity of the Truncated Estimator

2.1 Case of $k = 1$

We begin by deriving the bias and the mean squared error of the truncated estimator of a single positive location parameter. Let X be a random variable whose density function is given by $f(x-\mu)$ where μ is a location parameter restricted on the space $\{\mu \in \boldsymbol{R} | \mu > 0\}$. The unbiased estimator of μ is $\widehat{\mu}^U = X - c$ for $c = E[X-\mu] = \int uf(u)\mathrm{d}u$. Since this is the best location equivariant estimator, it is minimax. Since μ is positive, $\widehat{\mu}^U$ is improved on by the truncated estimator

$$\widehat{\mu}^{TR} = \max\{X - c, 0\},$$

and the bias, mean squared error and variance are denoted by $B(\mu) = Bias(\mu, \widehat{\mu}^{TR}) = E[\widehat{\mu}^{TR} - \mu]$, $M(\mu) = MSE(\mu, \widehat{\mu}^{TR}) = E[\big(\widehat{\mu}^{TR} - \mu\big)^2]$ and $V(\mu) = Var(\mu, \widehat{\mu}^{TR}) = E[\big(\widehat{\mu}^{TR} - E[\widehat{\mu}^{TR}]\big)^2]$, respectively. Let $F(z)$ be the distribution function of X, namely $F(z) = \int_{-\infty}^{z} f(u)\mathrm{d}u$.

Lemma 2.1 *The bias and mean squared error of the truncated estimator $\widehat{\mu}^{TR}$ of μ are expressed as*

$$B(\mu) = -\int_{-\infty}^{c-\mu} (z - c + \mu) f(z)\mathrm{d}z = \int_{-\infty}^{c-\mu} F(z)\mathrm{d}z, \tag{2.1}$$

$$M(\mu) = M_0 - \int_{-\infty}^{c-\mu} (z-c)(z-c+\mu) f(z)\mathrm{d}z - \mu B(\mu) = M_0 - 2L(\mu), \tag{2.2}$$

where $M_0 = E[(X - c - \mu)^2]$ and

$$L(\mu) = -\int_{-\infty}^{c-\mu} (z-c)F(z)\mathrm{d}z = \frac{1}{2}\Big\{\int_{-\infty}^{c-\mu} (z-c)(z-c+\mu) f(z)\mathrm{d}z + \mu B(\mu)\Big\}.$$

Also, the variance of $\widehat{\mu}^{TR}$ is $V(\mu) = M_0 - 2L(\mu) - \{B(\mu)\}^2$.

Proof. Let $Z = X - \mu$, and note that $\int_{-\infty}^{\infty}(z-c)f(z)\mathrm{d}z = 0$. Then,

$$E[\max(X-c, 0) - \mu] = E[\max(Z - c, -\mu)] = \int_{c-\mu}^{\infty} (z-c) f(z)\mathrm{d}z - \mu \int_{-\infty}^{c-\mu} f(z)\mathrm{d}z.$$

Since $\int_{c-\mu}^{\infty}(z-c)f(z)\mathrm{d}z = -\int_{-\infty}^{c-\mu}(z-c)f(z)\mathrm{d}z$, it is observed that $B(\mu) = -\int_{-\infty}^{c-\mu}(z-c+\mu)f(z)\mathrm{d}z$. Using integration by parts, we can show that $\int_{-\infty}^{c-\mu}(z-c+\mu)f(z)\mathrm{d}z = -\int_{-\infty}^{c-\mu}F(z)\mathrm{d}z$, which gives expression (2.1). Similarly,

$$\begin{aligned}E[\big(\max\{X-c,0\}-\mu\big)^2] =& E[\big(\max\{Z-c,-\mu\}\big)^2] \\ =& \int_{c-\mu}^{\infty}(z-c)^2 f(z)\mathrm{d}z + \mu^2\int_{-\infty}^{c-\mu}f(z)\mathrm{d}z.\end{aligned}$$

Since $\int_{c-\mu}^{\infty}(z-c)^2 f(z)\mathrm{d}z = \int_{-\infty}^{\infty}(z-c)^2 f(z)\mathrm{d}z - \int_{-\infty}^{c-\mu}(z-c)^2 f(z)\mathrm{d}z$, $M(\mu)$ can be rewritten as

$$\begin{aligned}M(\mu) =& M_0 - \int_{-\infty}^{c-\mu}(z-c)^2 f(z)\mathrm{d}z - \mu\int_{-\infty}^{c-\mu}(z-c)f(z)\mathrm{d}z \\ &+ \mu\Big\{\int_{-\infty}^{c-\mu}(z-c)f(z)\mathrm{d}z + \mu\int_{-\infty}^{c-\mu}f(z)\mathrm{d}z\Big\},\end{aligned}$$

which yields that $M(\mu) = M_0 - \int_{-\infty}^{c-\mu}(z-c)(z-c+\mu)f(z)\mathrm{d}z - \mu B(\mu)$. Using integration by parts, we can show that

$$\int_{-\infty}^{c-\mu}(z-c)(z-c+\mu)f(z)\mathrm{d}z = -2\int_{-\infty}^{c-\mu}(z-c)F(z)\mathrm{d}z - \mu\int_{-\infty}^{c-\mu}F(z)\mathrm{d}z, \tag{2.3}$$

which gives expression (2.2). The variance of $\widehat{\mu}^{TR}$ can be easily derived from (2.1) and (2.2). ■

From (2.1) and (2.2), the derivative of the bias $B(\mu)$, the MSE $M(\mu)$ and the variance $V(\mu)$ are given by

$$\begin{aligned}B'(\mu) =& -F(c-\mu) < 0, \\ M'(\mu) =& 2\mu F(c-\mu) > 0, \\ V'(\mu) =& 2\{\mu + B(\mu)\}F(c-\mu) > 0\end{aligned}$$

for $\mu > 0$. These means that $B(\mu)$ is decreasing in μ and $M(\mu)$ and $V(\mu)$ are increasing in μ. Thus, the maximum bias, the minimum MSE and the minimum variance are attained at $\mu = 0$, and they are given by $B(0) = -\int_{-\infty}^{c}(z-c)f(z)\mathrm{d}z$, $M(0) = M_0 - \int_{-\infty}^{c}(z-c)^2 f(z)\mathrm{d}z$ and $V(0) = M(0) - \{B(0)\}^2$.

Lemma 2.2 *The bias $B(\mu)$ is decreasing in μ for $\mu > 0$ with $B(0) > 0$ and $\lim_{\mu\to\infty} B(\mu) = 0$. The MSE $M(\mu)$ and the variance $V(\mu)$ are increasing in μ with $0 < M(\mu) < M_0$, $0 < V(\mu) < M_0$ and $\lim_{\mu\to\infty} M(\mu) = \lim_{\mu\to\infty} V(\mu) = M_0$. In particular, the truncated estimator $\widehat{\mu}^{TR}$ is minimax, and has a smaller varaince than $\widehat{\mu}^{U}$.*

Concerning the MLE of μ, it is given by $\widehat{\mu}^{ML} = \max\{X, 0\}$ under the condition that the density $f(x-\mu)$ has the mode at $x = \mu$. To get the bias and MSE of the MLE, it is noted that $0 = \int_{-\infty}^{\infty}(z-c)f(z)\mathrm{d}z = \int_{-\infty}^{-\mu}zf(z)\mathrm{d}z + \int_{-\mu}^{\infty}zf(z)\mathrm{d}z - c$ and

$M_0 = \int_{-\infty}^{\infty}(z-c)^2 f(z)\mathrm{d}z = \int_{-\infty}^{-\mu} z^2 f(z)\mathrm{d}z + \int_{-\mu}^{\infty} z^2 f(z)\mathrm{d}z - c^2$. Using these equalities and similar arguments as in Lemma 2.1, we can obtain

$$\begin{aligned}
B_{ML}(\mu) =& E[\max\{X,0\}-\mu] \\
=& \int_{-\mu}^{\infty} zf(z)\mathrm{d}z - \mu \int_{-\infty}^{-\mu} f(z)\mathrm{d}z \\
=& -\int_{-\infty}^{-\mu} (z+\mu) f(z)\mathrm{d}z + c, \\
M_{ML}(\mu) =& E[\{\max\{X,0\}-\mu\}^2] \\
=& \int_{-\mu}^{\infty} z^2 f(z)\mathrm{d}z + \mu^2 \int_{-\infty}^{-\mu} f(z)\mathrm{d}z \\
=& M_0 + c^2 + \int_{-\infty}^{-\mu} (\mu^2 - z^2) f(z)\mathrm{d}z.
\end{aligned}$$

Since $M'_{ML}(\mu) = 2\mu\int_{-\infty}^{-\mu} f(z)\mathrm{d}z > 0$, it can be observed that

$$M_{ML}(\mu) \le \lim_{\mu\to\infty} M_{ML}(\mu) = M_0 + c^2.$$

This implies that $\widehat{\mu}^{ML}$ is not minimax if $c \neq 0$. When $c=0$, it follows from Lemma 2.2 that $\widehat{\mu}^{ML}$ is minimax. In the case that the density function $f(x-\mu)$ is symmetric about μ, we have $c=0$, and the MLE $\max\{X,0\}$ is minimax..

2.2 Case of $k \ge 2$

We now investigate minimax estimation for a linear combination of positive location parameters. Let $X_1,\ldots,X_k$ be mutually independent random variables where X_i has density $f_i(x_i-\mu_i)$ for $\mu_i > 0$, and distribution function $\int_{-\infty}^{x_i} f_i(u-\mu_i)\mathrm{d}u = F_i(x_i-\mu_i)$ for $F_i(x) = \int_{-\infty}^{x} f_i(z)\mathrm{d}z$. Let us consider the linear combination of the locations given by

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

where $a_1,\ldots,a_k$ are real constants. The unbiased estimator of θ is $\hat{\theta}^U = \sum_{i=1}^k a_i(X_i - c_i)$ for $c_i = E[X_i - \mu_i]$ and it is minimax as shown in Kubokawa (2010). However, if for example, all of the a_i's are positive, it has the drawback of taking negative values with a positive probability because θ is positive. An alternative is a linear combination of the truncated estimators given by

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

where $\widehat{\mu}_i^{TR} = \max\{X_i - c_i, 0\}$ is a truncated estimator of μ_i. From a decision-theoretic point of view, an interesting query is whether the truncated estimator is minimax or not.

Let Λ_+ and Λ_- be subsets of $\{1,\ldots,k\}$ such that $\Lambda_+ \cup \Lambda_- = \{1,\ldots,k\}$ and

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

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

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

The MSE of the truncated estimator $\hat{\theta}^{TR}$ is written as

$$MSE(\boldsymbol{\mu}, \hat{\theta}^{TR}) = E[(\hat{\theta}^{TR}_+ - \theta_+)^2] + E[(\hat{\theta}^{TR}_- - \theta_-)^2] - 2E[\hat{\theta}^{TR}_+ - \theta_+]E[\hat{\theta}^{TR}_- - \theta_-].$$

Also, note that

$$E[(\hat{\theta}^{TR}_+ - \theta_+)^2] = \sum_{i\in\Lambda_+} a_i^2 M_i(\mu_i) + \sum_{i\in\Lambda_+}\sum_{\ell\in\Lambda_+,\ell\neq i} a_i a_\ell B_i(\mu_i)B_\ell(\mu_\ell),$$

where $M_i(\mu_i) = E[(\widehat{\mu}^{TR}_i - \mu_i)^2]$ and $B_i(\mu_i) = E[\widehat{\mu}^{TR}_i - \mu_i]$. Let $L_i(\mu_i) = -\int_{-\infty}^{c_i-\mu_i}(z - c_i)F_i(z)\mathrm{d}z$. Note that $L_i(\mu_i) > 0$ and that $MSE(\boldsymbol{\mu}, \hat{\theta}^U) = \sum_{i\in\Lambda_+} a_i^2 E[(X_i - c_i - \mu_i)^2] + \sum_{j\in\Lambda_-} a_j^2 E[(X_j - c_j - \mu_j)^2]$. Then, the difference of the MSEs of $\hat{\theta}^{TR}$ and $\hat{\theta}^U$ can be expressed as

$$\begin{aligned}
\Delta(\boldsymbol{\mu}) =& MSE(\boldsymbol{\mu}, \hat{\theta}^{TR}) - MSE(\boldsymbol{\mu}, \hat{\theta}^U) \\
=& -2\sum_{i\in\Lambda_+} a_i^2 L_i(\mu_i) + \sum_{i\in\Lambda_+}\sum_{\ell\in\Lambda_+,\ell\neq i} a_i a_\ell B_i(\mu_i)B_\ell(\mu_\ell) \\
& -2\sum_{j\in\Lambda_-} a_j^2 L_j(\mu_j) + \sum_{j\in\Lambda_-}\sum_{\ell\in\Lambda_-,\ell\neq j} a_j a_\ell B_j(\mu_j)B_\ell(\mu_\ell) \\
& +2\sum_{i\in\Lambda_+}\sum_{j\in\Lambda_-} a_i a_j B_i(\mu_i)B_j(\mu_j).
\end{aligned} \tag{2.6}$$

We first derive a necessary condition for the minimaxity of $\hat{\theta}^{TR}$. Note that $\lim_{\mu_i\to\infty} L_i(\mu_i) = 0$ and $\lim_{\mu_i\to\infty} B_i(\mu_i) = 0$. Let C_+ be a subset of Λ_+. When $\mu_i \to 0$ for all $i \in C_+$, $\mu_\ell \to \infty$ for all $\ell \in \Lambda_+\backslash C_+$ and $\mu_j \to \infty$ for all $j \in \Lambda_-$, the MSE difference $\Delta(\boldsymbol{\mu})$ converges to

$$\Delta(\boldsymbol{\mu}) \to -2\sum_{i\in C_+} a_i^2 L_i(0) + \sum_{i\in C_+}\sum_{\ell\in C_+,\ell\neq i} a_i a_\ell B_i(0)B_\ell(0).$$

A similar property holds for a subset C_- of Λ_-, and these give us a necessary condition for the minimaxity of $\hat{\theta}^{TR}$. Thus, if $\hat{\theta}^{TR}$ is minimax, then the following inequalities hold for all subsets $C_+ \subset \Lambda_+$ and $C_- \subset \Lambda_-$:

$$\begin{aligned}
\sum_{i\in C_+}\sum_{\ell\in C_+,\ell\neq i} a_i a_\ell B_i(0)B_\ell(0) \leq& 2\sum_{i\in C_+} a_i^2 L_i(0), \\
\sum_{j\in C_-}\sum_{\ell\in C_-,\ell\neq j} a_j a_\ell B_j(0)B_\ell(0) \leq& 2\sum_{j\in C_-} a_j^2 L_j(0).
\end{aligned} \tag{2.7}$$

We next show that condition (2.7) is sufficient. Since $(\mathrm{d}/\mathrm{d}\mu_i)L_i(\mu_i) = -\mu_i F_i(c_i - \mu_i)$ and $(\mathrm{d}/\mathrm{d}\mu_i)B_i(\mu_i) = -F_i(c_i - \mu_i)$, the derivative of $\Delta(\boldsymbol{\mu})$ with respect to μ_i for $i \in \Lambda_+$ can be written as

$$\frac{\partial}{\partial \mu_i}\Delta(\boldsymbol{\mu}) = 2F_i(c_i - \mu_i)\Big\{a_i^2\mu_i - a_i \sum_{\ell\in\Lambda_+,\ell\neq i} a_\ell B_\ell(\mu_\ell) - a_i \sum_{j\in\Lambda_-} a_j B_j(\mu_j)\Big\}.$$

Since the content of the above bracket is inceasing in μ_i, we can consider the two cases: (1) $(\partial/\partial\mu_i)\Delta(\boldsymbol{\mu}) \geq 0$ for all $\mu_i > 0$, or (2) there is a positive point $\mu_{i,0}$ such that $(\partial/\partial\mu_i)\Delta(\boldsymbol{\mu}) < 0$ for $0 < \mu_i < \mu_{i,0}$, and $(\partial/\partial\mu_i)\Delta(\boldsymbol{\mu}) \geq 0$ for all $\mu_i \geq \mu_{i,0}$. This means that

$$\Delta(\boldsymbol{\mu}) \leq \max\Big\{\lim_{\mu_i\to 0}\Delta(\boldsymbol{\mu}),\ \lim_{\mu_i\to\infty}\Delta(\boldsymbol{\mu})\Big\}.$$

Applying this argument for all $i \in \Lambda_+$ and all $j \in \Lambda_-$ and noting that

$$\sum_{i\in\Lambda_+}\sum_{j\in\Lambda_-} a_i a_j B_i(\mu_i)B_j(\mu_j) \leq 0,$$

we see that

$$\begin{aligned}\Delta(\boldsymbol{\mu}) \leq & \max_{C_+\subset\Lambda_+}\Big\{-2\sum_{i\in C_+} a_i^2 L_i(0) + \sum_{i\in C_+}\sum_{\ell\in C_+,\ell\neq i} a_i a_\ell B_i(0)B_\ell(0)\Big\} \\ & + \max_{C_-\subset\Lambda_-}\Big\{-2\sum_{j\in C_-} a_j^2 L_j(0) + \sum_{j\in C_-}\sum_{\ell\in C_-,\ell\neq j} a_j a_\ell B_j(0)B_\ell(0)\Big\},\end{aligned}$$

which implies that condition (2.7) is sufficient for the minimaxity of $\hat{\theta}^{TR}$. Hence, we have the following result.

Proposition 2.1 *The condition given in* (2.7) *is a necessary and sufficient condition for the truncated estimator* $\hat{\theta}^{TR}$ *to be minimax.*

Consider a special case that $f_1(z) = \cdots = f_k(z) = f(z)$ and $a_1 = \cdots = a_k = 1$, namely, $\theta = \mu_1 + \cdots + \mu_k$. Then, condition (2.7) is expressed as

$$\sum_{i\in C}\sum_{\ell\in C,\ell\neq i}\{B(0)\}^2 \leq 2\sum_{i\in C} L(0),$$

for all subsets C of $\{1,\ldots,k\}$, where

$$\begin{aligned} B(0) &= -\int_{-\infty}^{c}(z-c)f(z)\mathrm{d}z = \int_{-\infty}^{c} F(z)\mathrm{d}z, \\ L(0) &= \int_{-\infty}^{c}(z-c)^2 f(z)\mathrm{d}z = -\int_{-\infty}^{c-\mu}(z-c)F(z)\mathrm{d}z. \end{aligned} \tag{2.8}$$

This condition can be simplified as

$$(k-1)\{B(0)\}^2 \leq 2L(0), \tag{2.9}$$

and we get the following proposition.

Proposition 2.2 *Consider the case that* $f_1(z) = \cdots = f_k(z) = f(z)$ *and* $a_1 = \cdots = a_k = 1$*. A necessary and sufficient condition for the minimaxity of the truncated estimator* $\hat{\theta}^{TR}$ *is that* $(k-1)\{B(0)\}^2 \le 2L(0)$*. That is,* $\hat{\theta}^{TR}$ *is minimax for* $k \le 2L(0)/\{B(0)\}^2 + 1$ *and not minimax for* $k > 2L(0)/\{B(0)\}^2 + 1$*.*

The next result shows that the condition of Poposition 2.2 is sufficient (but not necessary) for minimaxity of $\hat{\theta}^{TR}$ for all linear combinations.

Proposition 2.3 *Suppose that* $f_1(z) = \cdots = f_k(z) = f(z)$ *and that the condition* $k \le 2L(0)/\{B(0)\}^2 + 1$ *is satisfied. Then* $\hat{\theta}^{TR}$ *is minimax for all linear combinations.*

Proof. We show that the condition (2.7) is satisfied for all subsets C of Λ_+. The same proof holds for subsets of Λ_-. Let the cardinality of Λ_+ be k_+ $(\le k)$ and the cardinality of a particular subset C be k^* $(\le k_+ \le k)$. We use the fact that for any set of positive (or negative) constants $(\sum_{i=1}^{\ell} a_i)^2 \le \ell \sum_{i=1}^{\ell} a_i^2$, and hence that

$$\sum_{i\in C}\sum_{j\in C, j\neq i} a_i a_j = (\sum_{i\in C} a_i)^2 - \sum_{i\in C} a_i^2 \le (k^*-1)\sum_{i\in C} a_i^2 \le (k-1)\sum_{i\in C} a_i^2.$$

The LHS of the first expression in (2.7) (since all $B_i(0)$'s and $L_i(0)$'s are equal) satisfies

$$B^2(0)\sum_{i\in C}\sum_{j\in C, j\neq i} a_i a_j \le B^2(0)(k-1)\sum_{i=1}^{k^*} a_i^2.$$

Now, since the assumption of the proposition is equivalent to $B^2(0)(k-1) \le 2L(0)$, we have that

$$B^2(0)\sum_{i\in C}\sum_{j\in C, j\neq i} a_i a_j \le B^2(0)(k-1)\sum_{i=1}^{k^*} a_i^2 \le 2L(0)\sum_{i=1}^{k^*} a_i^2,$$

which is the desired condition. This completes the proof. ■

This result shows that, in a certain sense, the sum of the means is a least favorable combination as far as minimaxity is concerned when all densities are the same. The same reasoning implies the slightly stronger conclusion that $\hat{\theta}^{TR}$ is minimax provided $k^* \le 2L(0)/\{B(0)^2\} + 1$ where $k^* = \max\{card(\Lambda_+), card(\Lambda_-)\}$. The condition is also necessary if $|a_i| = a$ for all i.

Finally, the variance of $\hat{\theta}^{TR}$ is written as

$$Var(\boldsymbol{\mu}, \hat{\theta}^{TR}) = \sum_{i=1}^{k} a_i^2 Var(\mu_i, \widehat{\mu}_i^{TR}),$$

which means that the variance of $\hat{\theta}^{TR}$ is less than or equal to that of $\hat{\theta}^{U}$ if for $i = 1, \ldots, k$, the variance of $\widehat{\mu}_i^{TR}$ is less than or equal to that of $X_i - c_i$. Thus, from Lemma 2.2, $\hat{\theta}^{TR}$ has a smaller variance than $\hat{\theta}^{U}$.

3 Examples of Non-minimaxity of Truncated or Maximum Likelihood Estimators

3.1 MLE in normal distributions

Let $X_1, \ldots, X_k$ be mutually independent random variables such that X_i has a normal distribution with mean μ_i and unit variance, namely, $X_i \sim \mathcal{N}(\mu_i, 1)$ for $\mu_i > 0$. The MLE of the linear combination $\theta = \mu_1 + \cdots + \mu_k$ is

$$\hat{\theta}^{ML} = \widehat{\mu}_1^{ML} + \cdots + \widehat{\mu}_k^{ML},$$

where $\widehat{\mu}_i^{ML} = \max\{X_i, 0\}$ is the MLE of μ_i.

From Lemma 2.1, the bias and mean squared error of $\widehat{\mu}_i^{ML}$ are expressed as

$$\begin{aligned} B(\mu_i) =& -\int_{-\infty}^{-\mu_i} (z + \mu_i)\phi(z)\mathrm{d}z \\ =& \phi(\mu_i) - \mu_i \Phi(-\mu_i), \\ M(\mu_i) =& 1 - \int_{-\infty}^{-\mu_i} z(z+\mu_i)\phi(z)\mathrm{d}z - \mu_i B(\mu_i) \\ =& 1 - \Phi(-\mu_i) - \mu_i B(\mu_i), \end{aligned} \tag{3.1}$$

where $\phi(z)$ and $\Phi(z)$ are density and distribution functions of the standard normal distribution. Then, $B(0) = 1/\sqrt{2\pi}$, $\lim_{\mu_i \to \infty} B(\mu_i) = 0$. Since $L(\mu_i) = \{\Phi(-\mu_i) + \mu_i B(\mu_i)\}/2$, it is observed that $L(0) = 1/4$ and $\lim_{\mu_i \to \infty} L(\mu_i) = 0$. Thus, $2L(0)/\{B(0)\}^2 = \pi$, and the following proposition follows from Propositions 2.2 and 2.3.

Proposition 3.1 *In the estimation of $\theta = \sum_{i=1}^k \mu_i$, a necessary and sufficient condition for the minimaxity of the MLE $\hat{\theta}^{ML}$ is that $k \leq \pi + 1$. That is, $\hat{\theta}^{ML}$ is minimax for $k \leq 4$ and not minimax for $k \geq 5$. Also, the condition $k \leq \pi + 1$ is sufficient for the minimaxity in estimation of any linear combination.*

It is interesting to illustrate the behaviors of the risk functions of the MLE for several values of k. For simplicity, let $\mu_1 = \cdots = \mu_k = \mu$ and consider the case of $\theta = \mu_1 + \cdots + \mu_k$. Then, the MSE of the MLE is given as

$$MSE(\mu, \hat{\theta}^{ML}) = kM(\mu) + k(k-1)\{B(\mu)\}^2,$$

where $M(\mu)$ and $B(\mu)$ are given in (3.1). Since the MSE of the unbiased estimator of θ is k, the ratio of the MSE's of the MLE and the unbiased estimator should be less or equal to one if the MLE is minimax. The ratio of the MSE's is illustrated in Figure 1 for $k = 1, 4, 5, 8$ and $0 < \mu < 3$. As seen from this figure, the ratio of the MSE's for $k = 5, 8$ exceeds one at $\mu = 0$.

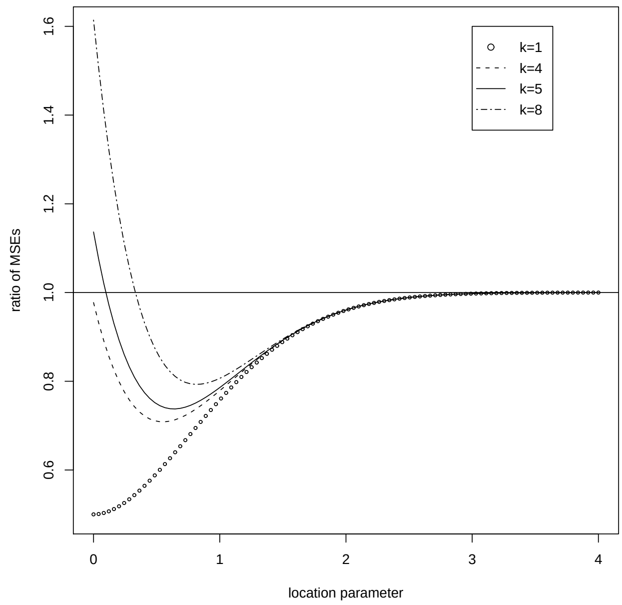

Figure 1: Ratio of MSE's of the MLE and the unbiased estimator, $MSE(\mu, \hat{\theta}^{ML})/MSE(\mu, \hat{\theta}^{U})$, for $k = 1, 4, 5, 8$ and $0 < \mu < 3$

3.2 Scale mixtures of normals distribution

Let $X_1, \ldots, X_k$ be mutually independent random variables such that X_i has a scale mixture of normals distribution, namely, the conditional distribution of X_i given V has the nornal distribution $\mathcal{N}(\mu_i, V)$ and V has a distribution G. Then, $B(0) = E[V^{1/2}]/(2\pi)^{1/2}$ and $2L(0) = E[V]/2$, so that

$$2L(0)/\{B(0)\}^2 = \pi E[V]/\{E[V^{1/2}]\}^2.$$

Hence from Propositions 2.2 and 2.3, we get the following proposition.

Proposition 3.2 *In the estimation of $\theta = \sum_{i=1}^{k} \mu_i$, a necessary and sufficient condition for the minimaxity of the MLE $\hat{\theta}^{ML}$ is that*

$$k \leq \pi E[V]/\{E[V^{1/2}]\}^2 + 1, \tag{3.2}$$

which is also is sufficient for the minimaxity in estimation of any linear combination.

The normal distribution given in Section 3.1 corresponds to the case where V is degenerate at one. Proposition 3.2 shows that the normal distribution is least favorable in the sense that, since $E[V]/\{E[V^{1/2}]\}^2 \geq 1$, the critical value of k (for minimaxity of $\hat{\theta}^{TR}$) is always at least as large for any scale mixtire of normals as it is for the normal distribution itself.

The scale mixtures of normal distribution includes t-, double exponential and logistic distributions, and Proposition 3.2 can be applied to these distributions.

[1] t-distribution. When $1/V$ has a chi-square with ν degrees of freedom divided by ν, the resulting distribution is a t-distribution with the density function

$$f_\nu(x-\mu_i) = C_\nu\big(1+(x-\mu_i)^2/\nu\big)^{-(\nu+1)/2},$$

for $C_\nu = (\nu\pi)^{-1/2}\Gamma((\nu+1)/2)/\Gamma(\nu/2)$ and $\mu_i > 0$. In this case, it is seen that

$$\begin{aligned}\pi\frac{E[V]}{\{E[V^{1/2}]\}^2} =&\pi\frac{E[(\nu/V)^{-1}]}{\{E[(\nu/V)^{-1/2}]\}^2} = \pi\frac{E[(\chi^2_\nu)^{-1}]}{\{E[(\chi^2_\nu)^{-1/2}]\}^2}\\ =&\pi\frac{2}{\nu-2}\Big(\frac{\Gamma(\nu/2)}{\Gamma((\nu-1)/2)}\Big)^2 \equiv K_\nu,\end{aligned}$$

so that the condition (3.2) becomes $k \le K_\nu + 1$. Using Stirling's formula, we can easily verify that $K_\nu \to \pi$ as $\nu \to \infty$, which corresponds to the case of the normal distribution. For $\nu = 3$, it is observed that $K_3 = \pi^2/2 = 4.929$ since $\Gamma(1/2) = \sqrt{\pi}$. Thus, in this case, $\hat\theta^{ML}$ is minimax for $k \le 5$ and not minimax for $k \ge 6$.

[2] Double exponential (or Laplace) distribution. When V has an exponential distibution mean 2 with the density $g(v) = 2^{-1}\exp\{-v/2\}$, Andrews and Mallows (1974) showed that the resulting distribution is a double exponential or Laplace distribution with the density

$$f(x-\mu_i) = 2^{-1}\exp\{-|x-\mu_i|\}.$$

See also West (1987), who extended the result to the exponential power family. In this case, it is seen that $E[V] = 2$ and $E[V^{1/2}] = \sqrt{\pi/2}$, which yields

$$\pi E[V]/\{E[V^{1/2}]\}^2 = 4.$$

Hence, the condition (3.2) becomes $k \le 5$.

[3] Logistic distribution. When V has the density function

$$g(v) = \sum_{j=1}^{\infty}(-1)^{j-1}j^2\exp\{-j^2v/2\},$$

Andrews and Mallows (1974) showed that the resulting distribution is the logistic distribution

$$f(x-\mu_i) = \exp\{-(x-\mu_i)\}[1+\exp\{-(x-\mu_i)\}]^{-2}.$$

In this case, it can be seen from Abramowitz and Stegun (1972, pp. 808) that

$$\begin{aligned}E[V] =&4\sum_{j=1}^{\infty}(-1)^{j-1}/j^2 = \pi^2/3,\\ E[V^{1/2}] =&\sqrt{2\pi}\sum_{j=1}^{\infty}(-1)^{j-1}/j = \sqrt{2\pi}\log(2),\end{aligned}$$

which yields

$$\pi E[V]/\{E[V^{1/2}]\}^2 = \frac{\pi^2}{6[\log(2)]^2} = 3.424.$$

Hence, the condition (3.2) becomes $k \le 4$.

3.3 Symmetric unimodal distributions

In this section, we study symmetric unimodal distributions and show that the uniform distribution on $[-a, a]$ is least favorable in the sense that the value of k giving minimaxity in Proposition 2.2 (and 3.1) is the minimum among all symmetric unimodal distributions, and that this value is given by 11/3. Hence for $k \leq 3$, $\hat{\theta}^{TR}$ is minimax for all linear combinations for all symmetric unimodal distributions. Here is the formal result.

Proposition 3.3 *Assume that the assumptions of Proposition 2.2* (*and 3.1*) *are satisfied, and in addition that all $f(z)$ is symmetric and unimodal. Then,*

(a) *If $k \leq 2L(0)/\{B(0)\}^2 + 1 = 2E[X^2]/\{E[|X|]\}^2 + 1$, $\hat{\theta}^{TR}$ is minimax for all linear combinations. The condition is also necessary for minimaxity for the sum of the means.*

(b) *If X has a uniform distribution on $[-a, a]$, $2E[X^2]/\{E|X|]\}^2 + 1 = 11/3$.*

(c) *For any symmetric unimodal distribution, $2E[X^2]/\{E|X|]\}^2 + 1 \geq 11/3$, so that the uniform distribution is least favorable in the sense indicated above.*

Proof. To prove part (a), note that for symmetric unimodal distributions $2L(0) = E[X^2]/2$, and $B(0) = E[|X|]/2$, so that the result follows. This calculation is also true for symmetric distributions for the estimator constructed from the truncated version of the X_i 's even though this estimator need not be the MLE.

Part (b) follows from part (a) by direct calculation since for a uniform distribution on $[-1, 1]$, $E[X^2] = 1/3$, and $E[|X|] = 1/2$. We note that the result is independent of the scale parameter.

Part (c) follows since all symmetric unimodal densities are scale mixtures of uniform distributions on $[-v, v]$, i.e., the distribution of $X|V$ has the uniform distribution on $[-V, V]$ and V has a density, $g(v)$ on $v > 0$. It follows that $E[X^2] = E[E[X^2|V] = E[V^2]/3$, and $E[|X|] = E[E[|X||V]] = E[V]/2$. Hence, $2L(0)/\{B(0)\}^2 = 2E[X^2]/\{E[|X|]\}^2 = (8/3)E[V^2]/\{E[V]\}^2 \geq 8/3$, since for positive random variables, $E[V^2]/\{E[V]\}^2 \geq 1$. Therefore, $2L(0)/\{B(0)\}^2 + 1 \geq 11/3$, which completes the proof. ■

It is interesting to compare the results of section 3.2 and this section. In general the smallest value of k that guarantees minimaxity for unimodal symmetric densities is $k = 3$, (attained for the uniform) while for scale mixtures of normals, it is $k = 4$ (attained for the normal itself). Note also since scale mixtures of normals are symmetric and unimodal, part (a), allows an alternative calculation of k.

3.4 Exponential distributions

Let $X_1, \ldots, X_k$ be mutually independent random variables such that X_i has an exponential distribution with location parameter μ_i, namely, the density function of X_i has the form $f_i(x_i - \mu_i) = \exp\{-(x_i - \mu_i)\}I(x_i > \mu_i)$ for $\mu_i > 0$, where $I(x_i > \mu_i)$ is the indicator function. It is noted that

$$\begin{aligned}\int_a^b z\exp\{-z\}\mathrm{d}z &= [-(z+1)\exp\{-z\}]_a^b,\\ \int_a^b z^2\exp\{-z\}\mathrm{d}z &= [-(z^2+2z+2)\exp\{-z\}]_a^b.\end{aligned} \tag{3.3}$$

Then $c_i = E[X_i - \mu_i] = 1$, and an unbiased estimator of μ_i is given by $\widehat{\mu}_i^U = X_i - 1$ with variance $E[(X_i - 1 - \mu_i)^2] = 1$. Since $\mu_i > 0$, it is reasonable to consider the truncated estimator $\widehat{\mu}_i^{TR} = \max(X_i - 1, 0)$. Noting that $X_i > \mu_i$, we can see that $\widehat{\mu}_i^{TR} = X_i - 1$ if $\mu_i > 1$, but $\widehat{\mu}_i^{TR} \geq X_i - 1$ if $0 \leq \mu_i \leq 1$. The truncated estimator of the linear combination $\theta = \mu_1 + \cdots + \mu_k$ is

$$\hat{\theta}^{TR} = \max(X_1 - 1, 0) + \cdots + \max(X_k - 1, 0).$$

From Lemma 2.1, the bias and mean squared error of $\widehat{\mu}_i^{ML}$ are expressed as

$$\begin{aligned}
B(\mu_i) =& - \int_0^{1-\mu_i} (z - 1 + \mu_i) \exp\{-z\} \mathrm{d}z I(0 < \mu_i < 1) \\
=& \{\exp\{\mu_i - 1\} - \mu_i\} I(0 < \mu_i < 1), \\
M(\mu_i) =& 1 - \int_0^{1-\mu_i} (z - 1)(z - 1 + \mu_i) \exp\{-z\} \mathrm{d}z I(0 < \mu_i < 1) - \mu_i B(\mu_i) \\
=& 1 - (1 - \mu_i)\{1 + \mu_i - 2\exp\{\mu_i - 1\}\} I(0 < \mu_i < 1).
\end{aligned}$$

Then, $B(0) = 1/e$, $\lim_{\mu_i \to \infty} B(\mu_i) = 0$. Since $L(\mu_i) = 2^{-1}(1 - \mu_i)\{1 + \mu_i - 2\exp\{\mu_i - 1\}\} I(0 < \mu_i < 1)$, it is observed that $L(0) = (1 - 2/e)/2$ and $\lim_{\mu_i \to \infty} L(\mu_i) = 0$. Thus, $2L(0)/\{B(0)\}^2 = e(e - 2) = 1.9524$, and from Propositions 2.2 and 2.3, it follows that in the estimation of $\theta = \sum_{i=1}^k \mu_i$, a necessary and sufficient condition for the minimaxity of the truncated estimator $\hat{\theta}^{TR}$ is that $k \leq e(e - 2) + 1 = (e - 1)^2$. That is, $\hat{\theta}^{TR}$ is minimax for $k \leq 2$ and not minimax for $k \geq 3$. If $k \leq 2$, $\hat{\theta}^{TR}$ is also minimax for any linear combination.

4 Extensions to estimation of restricted scale parameters

In this section, we treat the estimation of product and sum of the restricted scale parameters, and investigate whether a similar phenomenon as studied in the previous sections still holds.

4.1 Case of $k = 1$

Let X be a non-negative random variable having density function $\sigma^{-1} f(x/\sigma)$ with $\sigma > 1$. When the scale parameter σ is estimated relative to the quadratic loss function $(\hat{\sigma}/\sigma - 1)^2$, the scale equivariant estimator of σ is given by

$$\hat{\sigma}^C = cX.$$

The best scale equivariant and minimax estimator is $\hat{\sigma}^M = c^M X$, and the unbiased estimator is $\hat{\sigma}^U = c^U X$, where

$$c^M = E[Z]/E[Z^2], \quad c^U = 1/E[Z]$$

for $Z = X/\sigma$. Since $\sigma > 1$, it is reasonable to consider the truncation of cX at one, namely, the truncated estimator

$$\hat{\sigma}^{TR} = \max\{cX, 1\}.$$

Let $A(\sigma) = E[(\hat{\sigma}^{TR}/\sigma)^2] - E[(cX/\sigma)^2]$ and $B(\sigma) = E[\hat{\sigma}^{TR}/\sigma] - E[cX/\sigma]$. We begin with showing the following lemma which will be helpful for investigating properties of the risk functions.

Lemma 4.1 (1) *$A(\sigma)$ and $B(\sigma)$ are expressed as*

$$\begin{aligned} A(\sigma) =& \frac{1}{\sigma^2}\int_0^{1/c\sigma}(1-c^2\sigma^2z^2)f(z)\mathrm{d}z = 2c^2\int_0^{1/c\sigma} zF(z)\mathrm{d}z, \\ B(\sigma) =& \frac{1}{\sigma}\int_0^{1/c\sigma}(1-c\sigma z)f(z)\mathrm{d}z = c\int_0^{1/c\sigma} F(z)\mathrm{d}z, \end{aligned}$$

for $F(z) = \int_0^z f(x)\mathrm{d}x$.

(2) *$A(\sigma)$, $B(\sigma)$, $\sigma B(\sigma)$, $2B(\sigma) - A(\sigma)$ and $\sigma A(\sigma)/B(\sigma) - 2\sigma$ are decreasing in σ for $\sigma > 1$, and $\lim_{\sigma\to\infty} A(\sigma) = \lim_{\sigma\to\infty} B(\sigma) = 0$.*

(3) *$A(1) > B(1)$ for $c = c^M = E[Z]/E[Z^2]$.*

Proof. The part (1) can be verified by using integration by parts. For the part (2), the monotonicity of $A(\sigma)$ and $B(\sigma)$ can be seen since $A'(\sigma) = -2\sigma^{-3}F(1/c\sigma)$ and $B'(\sigma) = -\sigma^{-2}F(1/c\sigma)$. Also from the equality given in (1), $B(\sigma)\sigma = \int_0^{1/c\sigma}(1-c\sigma z)f(z)\mathrm{d}z$, so that $(B(\sigma)\sigma)' = -c\int_0^{1/c\sigma} zf(z)\mathrm{d}z < 0$, which shows that $\sigma B(\sigma)$ is decreasing. Since $2B'(\sigma) - A'(\sigma) = 2\sigma^{-3}F(1/c\sigma)(1-\sigma) < 0$ for $\sigma > 1$, it follows that $2B(\sigma) - A(\sigma)$ is decreasing for $\sigma > 1$. Leting $g(\sigma) = \sigma A(\sigma)/B(\sigma) - 2\sigma$, we can see that

$$g(\sigma) = 2c\sigma\frac{\int_0^{1/c\sigma} zF(z)\mathrm{d}z}{\int_0^{1/c\sigma} F(z)\mathrm{d}z} - 2\sigma,$$

and

$$\begin{aligned} g'(\sigma) =& 2c\frac{\int_0^{1/c\sigma} zF(z)\mathrm{d}z}{\int_0^{1/c\sigma} F(z)\mathrm{d}z} - 2 \\ & - \frac{2}{\sigma}\frac{(1/c\sigma)F(1/c\sigma)\int_0^{1/c\sigma} F(z)\mathrm{d}z - F(1/c\sigma)\int_0^{1/c\sigma} zF(z)\mathrm{d}z}{(\int_0^{1/c\sigma} F(z)\mathrm{d}z)^2} \\ =& 2\frac{\int_0^{1/c\sigma}(cz-1)F(z)\mathrm{d}z}{\int_0^{1/c\sigma} F(z)\mathrm{d}z} + \frac{2}{\sigma}\frac{F(1/c\sigma)\int_0^{1/c\sigma}(z-1/c\sigma)F(z)\mathrm{d}z}{(\int_0^{1/c\sigma} F(z)\mathrm{d}z)^2}, \end{aligned}$$

which is negative since $\sigma > 1$. Thus, $g(\sigma)$ is decreasing in σ.

Finally, for $c = E[Z]/E[Z^2]$, we show the inequality $A(1) > B(1)$, which is rewritten as $h(d) > 0$ where $h(d) = 2\int_0^d zF(z)\mathrm{d}z - d\int_0^d F(z)\mathrm{d}z$ for $d = 1/c$. Note that $h'(d) = dF(d) - \int_0^d F(z)\mathrm{d}z$ and $h''(d) = df(d)$. Since $h''(d) > 0$ and $h'(0) = 0$, it is observed that

$h'(d) \geq 0$, so that $h(d)$ is increasing in d. Since $h(0) = 0$, it is shown that $h(d) > 0$ for $d > 0$, proving the inequality given in the part (3). ■

The risk function of the truncated estimator $\hat{\sigma}^{TR}$ is expressed as

$$\begin{aligned} R(\sigma, \hat{\sigma}^{TR}) =& E[(\hat{\sigma}^{TR}/\sigma)^2] - 2E[\hat{\sigma}^{TR}/\sigma] + 1 \\ =& E[c^2 Z^2] + A(\sigma) - 2E[cZ] - 2B(\sigma) + 1 \\ =& R(\sigma, \hat{\sigma}^C) + A(\sigma) - 2B(\sigma). \end{aligned}$$

From Lemma 4.1, it follows that $A(\sigma) - 2B(\sigma)$ is increasing in $\sigma > 1$ and $\lim_{\sigma\to\infty}\{A(\sigma) - 2B(\sigma)\} = 0$, which means that $\hat{\sigma}^{TR}$ dominates $\hat{\sigma}^C$.

Proposition 4.1 *For any positive constant c, the estimator $\hat{\sigma}^C = cX$ is dominated by the truncated estimator $\hat{\sigma}^{TR} = \max\{cX, 1\}$.*

4.2 Case of $k \geq 2$

We now investigate whether the truncated estimators dominate the non-truncated estimators in the case of $k \geq 2$. Let $X_1, \ldots, X_k$ be mutually independent random variables where X_i has density $\sigma_i^{-1} f_i(x_i/\sigma_i)$ for $\sigma_i > 1$, and distribution function $\int_0^{x_i} \sigma_i^{-1} f_i(u/\sigma_i)\mathrm{d}u = F_i(x_i/\sigma_i)$ for $F_i(x) = \int_0^x f_i(z)\mathrm{d}z$. We consider two problems of estimating the product and the sum of the restricted scale parameters, given by

$$\begin{aligned} \eta =& \sigma_1 \times \sigma_2 \times \cdots \times \sigma_k, \\ \theta =& \sigma_1 + \sigma_2 + \cdots + \sigma_k, \end{aligned}$$

where estimators $\hat{\eta}$ and $\hat{\theta}$ are evaluated relative to the quadratic loss functions $(\hat{\eta}/\eta - 1)^2$ and $(\hat{\theta}/\theta - 1)^2$.

We first treat estimation of the product η. Along the line discussed in the case of $k = 1$, we consider the estimator $\hat{\sigma}_i^C = c_i X_i$ and the truncated estimator $\hat{\sigma}_i^{TR} = \max\{c_i X_i, 1\}$ for σ_i, which lead to the estimators

$$\begin{aligned} \hat{\eta}^C =& \hat{\sigma}_1^C \times \hat{\sigma}_2^C \times \cdots \times \hat{\sigma}_k^C, \\ \hat{\eta}^{TR} =& \hat{\sigma}_1^{TR} \times \hat{\sigma}_2^{TR} \times \cdots \times \hat{\sigma}_k^{TR}. \end{aligned}$$

Proposition 4.2 *The product of the truncated estimators $\hat{\eta}^{TR}$ dominates the non-truncated estimator $\hat{\eta}^C$ if and only if k and the c_i's satisfy the inequality*

$$\frac{\prod_{i=1}^k \{c_i^2 E[Z_i^2] + A_i(1)\} - \prod_{i=1}^k \{c_i^2 E[Z_i^2]\}}{\prod_{i=1}^k \{c_i E[Z_i] + B_i(1)\} - \prod_{i=1}^k \{c_i E[Z_i]\}} \geq 2, \tag{4.1}$$

where $A_i(\sigma_i) = E[(\hat{\sigma}_i^{TR}/\sigma_i)^2] - E[(c_i Z_i)^2]$ and $B_i(\sigma_i) = E[\hat{\sigma}_i^{TR}/\sigma_i] - E[c_i Z_i]$.

Proof. From the independence of $X_1, \ldots, X_k$, the risk function of $\hat{\eta}^{TR}$ is written as

$$R(\boldsymbol{\sigma}, \hat{\eta}^{TR}) = \prod_{i=1}^k \{c_i^2 E[Z_i^2] + A_i(\sigma_i)\} - 2\prod_{i=1}^k \{c_i E[Z_i] + B_i(\sigma_i)\} + 1, \tag{4.2}$$

for $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_k)$. Differentiating $R(\boldsymbol{\sigma}, \hat{\eta}^{TR})$ with respect to σ_j gives that

$$\frac{\partial}{\partial \sigma_j} R(\boldsymbol{\sigma}, \hat{\eta}^{TR}) = \frac{2}{\sigma_j^3} F_j(1/c_j\sigma_j) \Big\{ \sigma_j \prod_{i \neq j} \{c_i E[Z_i] + B_i(\sigma_i)\} - \prod_{i \neq j} \{c_i^2 E[Z_i^2] + A_i(\sigma_i)\} \Big\},$$

which is negative for $\sigma_j < \sigma_j^*$ and positive for $\sigma_j > \sigma_j^*$, where

$$\sigma_j^* = \prod_{i \neq j} \{c_i^2 E[Z_i^2] + A_i(\sigma_i)\} / \prod_{i \neq j} \{c_i E[Z_i] + B_i(\sigma_i)\}.$$

This implies that

$$R(\boldsymbol{\sigma}, \hat{\eta}^{TR}) \leq \max\{ \lim_{\sigma_j \to 1} R(\boldsymbol{\sigma}, \hat{\eta}^{TR}), \lim_{\sigma_j \to \infty} R(\boldsymbol{\sigma}, \hat{\eta}^{TR}) \}.$$

Repeating this argument shows the inequality

$$R(\boldsymbol{\sigma}, \hat{\eta}^{TR}) \leq \max\{ \lim_{\sigma_j \to 1, j=1,\ldots,k} R(\boldsymbol{\sigma}, \hat{\eta}^{TR}), \lim_{\sigma_j \to \infty, j=1,\ldots,k} R(\boldsymbol{\sigma}, \hat{\eta}^{TR}) \}.$$

Since $\lim_{\sigma_j \to \infty, j=1,\ldots,k} R(\boldsymbol{\sigma}, \hat{\eta}^{TR}) = R(\boldsymbol{\sigma}, \hat{\eta}^{C})$, we can see that the condition (4.1) is a necessary and sufficient condition for $\hat{\eta}^{TR}$ to dominate $\hat{\eta}^{C}$. ∎

In the case of $f_1(z) = \cdots = f_k(z) = f(z)$ and $c_1 = \cdots = c_k = c$, the condition (4.1) can be simplified as

$$\frac{[1 + A(1)/\{c^2 E[Z^2]\}]^k - 1}{[1 + B(1)/\{cE[Z]\}]^k - 1} \leq 2 \left(\frac{E[Z]}{cE[Z^2]} \right)^k, \tag{4.3}$$

where $A(\sigma)$ and $B(\sigma)$ are given in (1) of Lemma 4.1 and Z is a random variable having the density $f(z)$. When $\hat{\sigma}_i^C = c^M X_i$ is the best scale equivariant estimator, namely, $c^M = E[Z]/E[Z^2]$, the condition (4.3) can be described as

$$\frac{[1 + A(1)/\{c^M E[Z]\}]^k - 1}{[1 + B(1)/\{c^M E[Z]\}]^k - 1} \leq 2, \tag{4.4}$$

or

$$\left[\frac{1 + A(1)/\{c^M E[Z]\}}{1 + B(1)/\{c^M E[Z]\}} \right]^k \frac{1 - [1 + A(1)/\{c^M E[Z]\}]^{-k}}{1 - [1 + B(1)/\{c^M E[Z]\}]^{-k}} \leq 2. \tag{4.5}$$

From (3) of Lemma 4.1, it is noted that $A(1) > B(1)$, so that the l.h.s. of (4.5) exceeds 2 for large k. Thus, we get the following proposition.

Proposition 4.3 *Assume that $f_1(z) = \cdots = f_k(z) = f(z)$ and $c_1 = \cdots = c_k = c$. The best scale equivariant estimator of η is given by $\hat{\eta}^C = \prod_{i=1}^k (c^M X_i)$ for $c^M = E[Z]/E[Z^2]$, and the condition* (4.4) *or* (4.5) *is a necessary and sufficient condition for the truncated estimator $\hat{\eta}^{TR}$ to dominate $\hat{\eta}^C$. Further, $\hat{\eta}^{TR}$ does not dominate $\hat{\eta}^C$ for large k.*

We next treat the estimation of the sum of the restricted scales $\theta = \sum_{i=1}^k \sigma_i$. For $\hat{\sigma}_i^C = c_i X_i$ and $\hat{\sigma}_i^{TR} = \max\{c_i X_i, 1\}$, we consider the two estimators

$$\begin{aligned} \hat{\theta}^C =& \hat{\sigma}_1^C + \hat{\sigma}_2^C + \cdots + \hat{\sigma}_k^C, \\ \hat{\eta}^{TR} =& \hat{\sigma}_1^{TR} + \hat{\sigma}_2^{TR} + \cdots + \hat{\sigma}_k^{TR}. \end{aligned}$$

Proposition 4.4 *Assume that the constants c_i's satisfy that $c_i E[Z_i] \le 1$ for $1 = 1, \dots, k$. Then, the sum of the truncated estimators $\hat{\theta}^{TR}$ dominates the non-truncated estimator $\hat{\theta}^C$ if k and the c_i's satisfy the inequalities*

$$\sum_{j\neq i}^{k}\{B_j(1) + 2c_j E[Z_j] - 2\} \le 2 - \frac{A_i(1)}{B_i(1)}, \tag{4.6}$$

for $i = 1, \dots, k$.

Proof. The risk function of the truncated estimator $\hat{\theta}^{TR}$ is

$$\begin{aligned}
R(\boldsymbol{\sigma}, \hat{\theta}^{TR}) =& E[\{\sum_{i=1}^{k}(\hat{\sigma}_i^{TR} - \sigma_i)\}^2]/\theta^2 \\
=& \sum_{i=1}^{k}\sigma_i^2 E[(\hat{\sigma}_i^{TR}/\sigma_i - 1)^2]/\theta^2 \\
& + \sum_{i=1}^{k}\sum_{j\neq i}^{k}\sigma_i\sigma_j E[\hat{\sigma}_i^{TR}/\sigma_i - 1]E[\hat{\sigma}_j^{TR}/\sigma_j - 1]/\theta^2 \\
=& \sum_{i=1}^{k}\sigma_i^2\{c_i^2 E[Z_i^2] + A_i(\sigma_i) - 2c_i E[Z_i] - 2B_i(\sigma_i) + 1\}/\theta^2 \\
& + \sum_{i=1}^{k}\sum_{j\neq i}^{k}\sigma_i\sigma_j\{c_i E[Z_i] + B_i(\sigma_i) - 1\}\{c_j E[Z_j] + B_j(\sigma_j) - 1\}/\theta^2,
\end{aligned} \tag{4.7}$$

which is less than or equal to $R(\boldsymbol{\sigma}, \hat{\theta}^C)$ if and only if for any $\sigma_i > 1$, $i = 1, \dots, k$,

$$\sum_{i=1}^{k}\sigma_i B_i(\sigma_i)\Big\{\sigma_i\frac{A_i(\sigma_i)}{B_i(\sigma_i)} - 2\sigma_i + \sum_{j\neq i}^{k}\sigma_j\{B_j(\sigma_j) + 2c_j E[Z_j] - 2\}\Big\} \le 0. \tag{4.8}$$

Note that $c_j E[Z_j] - 1 \le 0$ from the condition of c_i's given in Proposition 4.4. From the monotonicity properties of $\sigma_i A_i(\sigma_i)/B_i(\sigma_i) - 2\sigma_i$ and $\sigma_i B_i(\sigma_i)$ given in (2) of Lemma 4.1, it follows that

$$\sigma_i\frac{A_i(\sigma_i)}{B_i(\sigma_i)} - 2\sigma_i \le \frac{A_i(1)}{B_i(1)} - 2,$$
$$\sigma_j\{B_j(\sigma_j) + 2c_j E[Z_j] - 2\} \le B_j(1) + 2c_j E[Z_j] - 2.$$

Hence, the condition (4.8) holds if for $i = 1, \dots, k$,

$$\frac{A_i(1)}{B_i(1)} - 2 + \sum_{j\neq i}^{k}\{B_j(1) + 2c_j E[Z_j] - 2\} \le 0,$$

which is given in (4.6). ■

It is noted that the inequality (4.8) is a necessary and sufficient condition for $\hat{\theta}^{TR}$ to dominate $\hat{\theta}^{C}$. We thus derive the necessary condition given by

$$\sum_{i=1}^{k} B_i(1)\Big\{\frac{A_i(1)}{B_i(1)} - 2 + \sum_{j\neq i}^{k}\{B_j(1) + 2c_j E[Z_j] - 2\}\Big\} \leq 0, \tag{4.9}$$

which is weaker than the sufficient condition given in Proposition 4.4.

As described in the case of $k = 1$, the best scale equivariant and minimax estimator of σ_i is $\hat{\sigma}_i^M = c_i^M X_i$ for $c_i^M = E[Z_i]/E[Z_i^2]$, and the unbiased estimator is $\hat{\sigma}^U = c_i^U X_i$ for $c_i^U = 1/E[Z_i]$. It is noted that both of c_i^M and c_i^U satisfy the assumption $c_i E[Z_i] \leq 1$ in Proposition 4.4.

In the case of $f_1(z) = \cdots = f_k(z) = f(z)$ and $c_1 = \cdots = c_k = c$, the necessary condition (4.9) can be simplified as

$$\frac{A(1)}{B(1)} - 2 + (k-1)\{B(1) + 2cE[Z] - 2\} \leq 0, \tag{4.10}$$

which is identical to the sufficient condition given in Proposition 4.4.

Proposition 4.5 *Assume that $f_1(z) = \cdots = f_k(z) = f(z)$ and $c_1 = \cdots = c_k = c$. For the constant c satisfying $cE[Z_1] \leq 1$, the sum of the truncated estimators $\hat{\theta}^{TR}$ dominates the non-truncated estimator $\hat{\theta}^{C}$ if and only if k satisfies the condition* (4.10).

4.3 Examples

We here provide two examples of exponential and uniform distributions. We first treat the case of exponential distributions. Let $X_1, \ldots, X_k$ be mutually independent random variables such that X_i has an exponential distribution with scale parameter σ_i, namely, the density function of X_i has the form $\sigma^{-1} f_i(x_i/\sigma) = \sigma^{-1}\exp\{-x_i/\sigma_i\} I(x_i > 0)$ for $\sigma_i > 1$. Moments and integrals of the exponential distribution can be computed by using (3.3). Then, $E[Z] = 1$, $E[Z^2] = 2$ and

$$\begin{aligned} A(\sigma) =& (1/\sigma + 1/2)e^{-2/\sigma} + 1/\sigma^2 - 1/2, \\ B(\sigma) =& e^{-2/\sigma}/2 + 1/\sigma - 1/2. \end{aligned}$$

Thus, $A(1) = (1 + 3e^{-2})/2$ and $B(1) = (1 + e^{-2})/2$ for $Z = X_1/\sigma_1$. Note that $c^M = E[Z]/E[Z^2] = 1/2$ and $c^U = E[Z] = 1$.

Concerning the estimation of the product of σ_i's, it is noted that

$$\frac{[1 + A(1)/\{c^M E[Z]\}]^k - 1}{[1 + B(1)/\{c^M E[Z]\}]^k - 1} = \frac{(2 + 3e^{-2})^k - 1}{(2 + e^{-2})^k - 1}.$$

Then from Proposition 4.2 and the condition (4.4), it follows that $\prod_{i=1}^k \max\{c^M X_i, 1\}$ dominates $\prod_{i=1}^k (c^M X_i)$ if and only if $(2 + 3e^{-2})^k - 1 \leq 2(2 + e^{-2})^k - 2$. Investigating this inequality numerically, we can see that this condition is satisfied for $k \leq 5$, while it does not hold for $k \geq 6$. From (4.2), the MSE function of the truncated estimator of the best scale equivariant estimator $\hat{\eta}^{TR}$ is written as

$$MSE(\sigma, \hat{\eta}^{TR}) = \Big[(1/2 + A(\sigma))^k - 2(1/2 + B(\sigma))^k + 1\Big]\sigma^{2k},$$

in the case of $\sigma_1 = \cdots = \sigma_k = \sigma$. The numerical behavior of the ratio of the MSE's, $MSE(\sigma, \hat{\eta}^{TR})/MSE(\sigma, \hat{\eta}^{C})$, is illustrated in Figure 2 for $k = 1, 5, 6, 10$ and $1 < \sigma < 5$. From this figure, we can observe that the truncated estimator does not dominate the non-truncated estimator for $k \geq 6$.

For the estimation of the sum of σ_i's, from (4.10), the necessary and sufficient condition is given by

$$\frac{k-1}{2}(1 + e^{-2} + 4c - 4) \leq \frac{1 - e^{-2}}{1 + e^{-2}}.$$

For $c = c^M = 1/2$, we have $1 + e^{-2} + 4c^M - 4 = e^{-2} - 1 < 0$, so that $\sum_{i=1}^k \max\{c^M X_i, 1\}$ always dominates $\sum_{i=1}^k (c^M X_i)$, which is the sum of the best scale equivariant estimators. For $c = c^U = 1$, $1 + e^{-2} + 4c^U - 4 = 1 + e^{-2} > 0$, so that $\sum_{i=1}^k \max\{c^U X_i, 1\}$ dominates the unbiased estimator $\sum_{i=1}^k (c^U X_i)$ if and only if $k - 1 \leq 2(1 - e^{-2})/(1 + e^{-2})^2$. Investigating the inequality numerically, we can see that $\sum_{i=1}^k \max\{c^U X_i, 1\}$ dominates $\sum_{i=1}^k (c^U X_i)$ for $k = 1, 2$, while this dominance does not hold for $k \geq 3$. From (4.7), in the case of $\sigma_1 = \cdots = \sigma_k = \sigma$, the MSE of the truncated estimator for the unbiased estimator is written as

$$R(\sigma, \hat{\theta}^{TR}) = \big[A(\sigma) - 2B(\sigma) + 1 + 2(k-1)\{B(\sigma)\}^2\big]k\sigma^2.$$

The ratio of the MSE's, $MSE(\sigma, \hat{\theta}^{TR})/MSE(\sigma, \hat{\theta}^{C})$, is illustrated in Figure 3 for $k = 1, 2, 3, 5$ and $1 < \sigma < 5$.

We next treat the case of uniform distributions, namely, let $X_1, \ldots, X_k$ be mutually independent random variables such that X_i has a uniform distribution with the density function $\sigma^{-1} f(x/\sigma)$ for $f(z) = I(0 < z < 1)$ where $\sigma_i > 1$. Since $E[Z] = 1/2$ and $E[Z^2] = 1/3$, we have $c^M = 3/2$ and $c^U = 2$. Also note that $A(1) = 2/(3c)$ and $B(1) = 1/(2c)$.

Concerning the estimation of the product of σ_i's, it is noted that

$$\frac{[1 + A(1)/\{c^M E[Z]\}]^k - 1}{[1 + B(1)/\{c^M E[Z]\}]^k - 1} = \frac{(1 + 16/27)^k - 1}{(1 + 4/9)^k - 1}.$$

Then from Proposition 4.2 and the condition (4.4), it follows that $\prod_{i=1}^k \max\{c^M X_i, 1\}$ dominates $\prod_{i=1}^k (c^M X_i)$ if and only if $(1 + 16/27)^k - 1 \leq 2(1 + 4/9)^k - 2$. Since this inequality is satisfied for $k \leq 6$, the dominance result holds for $k \leq 6$, while it does not hold for $k \geq 7$.

For the estimation of the sum of σ_i's, from (4.10), the necessary and sufficient condition is given by

$$(k-1)((2c)^{-1} + c - 2) \leq 2/3.$$

For $c = c^M = 3/2$, we have $(2c)^{-1} + c - 2 = -1/6$, thus $\sum_{i=1}^k \max\{c^M X_i, 1\}$ always dominates $\sum_{i=1}^k (c^M X_i)$. For $c = c^U = 2$, $(2c)^{-1} + c - 2 = 1/4$, so that $\sum_{i=1}^k \max\{c^U X_i, 1\}$ dominates $\sum_{i=1}^k (c^U X_i)$ if and only if $k - 1 \leq 8/3$ or $k \leq 3$, while this dominance property does not hold if $k - 1 > 8/3$ or $k \geq 4$.

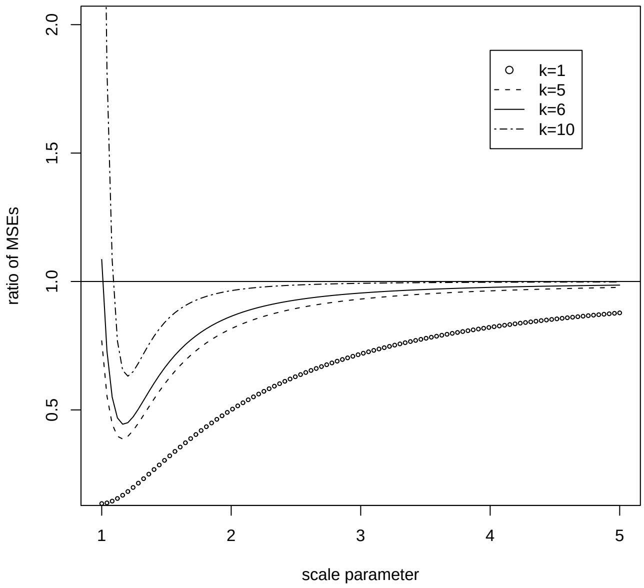

Figure 2: Ratio of MSE's of the truncated estimator $\hat{\eta}^{TR}$ and the best scale equivariant estimator for $k = 1, 5, 6, 10$ and $1 < \sigma < 5$

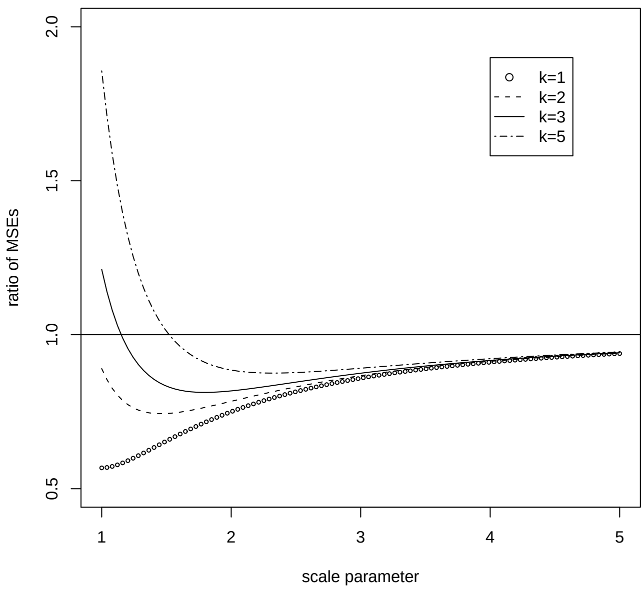

Figure 3: Ratio of MSE's of the truncated estimator $\hat{\theta}^{TR}$ and the unbiased estimator for $k = 1, 2, 3, 5$ and $1 < \sigma < 5$

5 Concluding Remarks

In this paper, we address the problem of estimation of a linear combination of restricted location parameters where estimators are evaluated through MSE. The linear combination of the best location equivariant estimators is a minimax estimator with a constant MSE, and it is used as a benchmark estimator. Since each location parameter is restricted to the positive real line, it is reasonable to consider the linear combination of the truncated estimators. Our interest is to investigate whether the restricted estimator remain minimax or not. In this paper, we have derived a necessary and sufficient condition for the restricted estimator to be minimax, and have examined this condition for some specific distributions. Especially, we have an interest in the minimaxity of the MLE in the estimation of the sum of the restricted means in k normal distributions, and we have established that the MLE is not minimax for $k \geq 5$, while it is still minimax for $k \leq 4$. This result corresponds to the result of Kubokawa (2010) who showed that the generalized Bayes estimator against the uniform prior over the restricted space is not minimax for $k \geq 2$. Thus, in the estimation of the sum of the restricted normal means, the MLE as well as the generalized Bayes estimator are not minimax for large k. It has been also shown that such a decision-theoretic phenomenon remains true in the estimation of the product and the sum of the restricted scale parameters.

In the context of the simultaneous estimation of k restricted normal means, Hartigan (2004) established that the generalzed Bayes estimator against the uniform prior over the restricted parameter space is still minimax, and it can be also shown that the MLE is minimax. However, the minimaxity of the generalized Bayes estimator and the MLE do not hold for large k in the context of estimation of the sum of the restricted means. This may be an interesting decision-theoretic phenomenon, and it is conjectured that such a property still hold for distributions other than the location and scale families treated in this paper.

Acknowledgments. This research was supported in part by Grant-in-Aid for Scientific Research Nos. 19200020 and 21540114 from Japan Society for the Promotion of Science.

References

[1] Abramowitz, M. and Stegun, C. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* Dover, New York.

[2] Andrews, D.F., and Mallows, C.L. (1974). Scale mixtures of normal distributions. *J. Royal Statist. Soc.*, **36**. 99-102.

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

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

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

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

[7] Kubokawa, T. (2010). Minimax estimation of linear combinations of restricted location parameters. Discussion Paper Series, CIRJE-F-723

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

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

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

[11] Tsukuma, H. and Kubokawa, T. (2008). Stein's phenomenon in estimation of means restricted to a polyhedral convex cone. *J. Multivariate Analysis*, **99**, 141-164.

[12] West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, **74**, 646-648.