Admissible minimax estimators of a mean vector of scale mixtures of multivariate normal distributions

Yuzo Maruyama

Abstract
The problem of estimating a mean vector of scale mixtures of multivariate normal distributions with the quadratic loss function is considered. For a certain class of these distributions, which includes at least multivariate-$t$ distributions, admissible minimax estimators are given.

Abbreviated Title
Admissible Minimax Estimators

Affiliation and Addresses
Center for Spatial Information Science,
Faculty of Economics, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-0033, Japan
E-mail: maruyama@csis.u-tokyo.ac.jp

AMS 1991 subject classifications
Primary 62C15; Secondary 62C20.

Key words and phrases
admissible; minimax; scale mixtures of multivariate normal distributions; the FKG inequality
1 Introduction

Since Stein (1956) showed that the usual minimax best equivariant estimator of a $p$-dimensional normal mean is inadmissible for $p \geq 3$, there has been much research on improving the best invariant estimator. James and Stein (1961) gave an explicit form of an improved estimator. Brown (1966) substantially extended Stein (1956) to a wide class of distributions. Explicit minimax improvements for the mean vector of a spherically symmetric distribution has been an important topic in this field. As a special case of spherically symmetric distributions, Strawderman (1974) introduced scale mixtures of multivariate normal distributions as follows. Let $X$ have density $f(\|x - \theta\|^2)$ where

$$f(\|x - \theta\|^2) = \int_0^\infty (2\pi)^{-p/2}v^{p/2} \exp \left( -\frac{\|x - \theta\|^2}{2} \right) g(v)dv,$$  

(1.1)

where $g(v)$ is a known probability density function. Strawderman (1974) proposed a class of improved minimax estimators of the mean vector and found some minimax generalized Bayes estimators. Generally Berger (1975), Bock (1985) and Brandwein and Strawderman (1978, 1991) found several classes of minimax estimators for spherically symmetric distributions. To date there seems to be no general results on the admissibility of the minimax estimator produced by these authors, while, in the normal case, a lot of admissible minimax estimators have been derived by Strawderman (1971), Alam (1973), Berger (1976), Fourdrinier et al. (1998) and Maruyama (1998). Hence it seems desirable to characterize a class of estimators which are both minimax and admissible.

In this paper, we consider the problem of estimating the mean vector $\theta$ of (1.1) with the quadratic loss function $\|\delta - \theta\|^2$. A prior density function considered here is the same one which leads admissibility and minimaxity in the normal case. We derive admissible minimax estimators for a certain subclass of these distributions, which includes at least multivariate-$t$ distributions with four or more degrees of freedom.

2 The Main Result

2.1 The construction of the generalized Bayes estimators

First of all, we derive generalized Bayes estimators for the following prior distribution. Let the conditional distribution of $\theta$ given $t$, $0 < t < 1$, be normal with mean 0 and covariance matrix $t^{-1}(1-t)I_p$ and a density function of $t$ is proportional to $t^{-a}(1-t)^b I_{(0,1)}(t)$ with $b > -1$. Thus the (generalized) density function $h_{a,b}(\theta)$ is

$$h_{a,b}(\theta) = \int_0^1 \left( \frac{t}{1-t} \right)^{p/2} \exp \left( -\frac{\|\theta\|^2}{2(1-t)} \right) t^{-a}(1-t)^b dt.$$

It is noted that the above prior distribution is proper for $a < 1$ and is improper for $a \geq 1$. As we consider the quadratic loss function, the generalized Bayes estimator can be expressed as

$$\delta_{a,b}(x) = \frac{\int \theta f(\|x - \theta\|^2) h_{a,b}(\theta) d\theta}{\int f(\|x - \theta\|^2) h_{a,b}(\theta) d\theta}.$$  

(2.1)
Hence we have the generalized Bayes estimator calculated as follows:

\[
\int_0^1 \left( \frac{\lambda}{1-\lambda} \right)^{p/2} \exp \left( -\frac{v\lambda\|\theta\|^2}{2(1-\lambda)} \right) v^{p/2-a+1} (1-\lambda)^b (1-\lambda + v\lambda)^{a-b-2} d\lambda.
\]

The integrals with respect to \(\theta\) of both the denominator and numerator of (2.1) are calculated as follows:

numerator = \[\int_{R^n} \theta \exp \left( -\frac{v\|x - \theta\|^2}{2} - \frac{v\lambda\|\theta\|^2}{2(1-\lambda)} \right) d\theta\]

= \[\int_{R^n} \theta \exp \left( -\frac{\lambda v}{2} \|x\|^2 \right) \exp \left( -\frac{v\|\theta - (1-\lambda)x\|^2}{2(1-\lambda)} \right) d\theta\]

= \[(2\pi)^{p/2} \exp \left( -\frac{\lambda v}{2} \|x\|^2 \right) \left( \frac{1-\lambda}{v} \right)^{p/2} (1-\lambda)x\]

and

\[
\text{denominator} = \int_{R^n} \exp \left( -\frac{v\|x - \theta\|^2}{2} - \frac{v\lambda\|\theta\|^2}{2(1-\lambda)} \right) d\theta
\]

= \[(2\pi)^{p/2} \exp \left( -\frac{\lambda v}{2} \|x\|^2 \right) \left( \frac{1-\lambda}{v} \right)^{p/2}.\]

Hence we have the generalized Bayes estimator \(\delta_{a,b}(x) = (1 - \phi_{a,b}(\|x\|^2)/\|x\|^2)x\), where

\[
\phi_{a,b}(w) = \int_0^1 \int_0^\infty \lambda^{p/2-a+1} v^{p/2-a+1} (1-\lambda)^b (1-\lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2) g(v) dv d\lambda
\]

Here we introduce the assumption on \(g\):

(I) \(g(s_1 t_1) g(s_2 t_2) \leq g(s_1 t_2) g(s_2 t_1)\) for \(s_1 \leq s_2, t_1 \leq t_2\).

It is noted that (I) is equivalent to that \(\beta^{-1} g(\beta^{-1} y)\) has monotone likelihood ratio in \(y\) when considered as a scale parameter family of distributions and that \(vg'(v)/g(v)\) is nonincreasing if \(g(v)\) is differentiable. In the following, some properties of the behavior of \(\phi_{a,b}(w)\) are given.

**Theorem 2.1.**

1. If \(g(v)\) satisfies the assumption (I), \(\phi_{a,b}(w)\) is monotone nondecreasing in \(w\) for \(b - a + 2 > 0\) and \(b \geq 0\).

2. \(\phi_{a,b}(w)\) is monotone nondecreasing in \(w\) for \(b - a + 2 = 0\) and \(b \geq 0\).

**Proof.** For simplicity of later calculations, let

\[
\Phi_k(w) = \int_0^1 \int_0^\infty \lambda^{p/2-a+k} v^{p/2-a+1} (1-\lambda)^b (1-\lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2) g(v) dv d\lambda,
\]
for $k = 0, 1$. Then the derivative of $\phi_{a,b}^g(w)$ is written as

\[
\frac{d}{dw} \phi_{a,b}^g(w) = \Phi_1(w) \Phi_0^{-1}(w)
\]

\[
-2^{-1}w\Phi_0^{-1}(w) \int_0^1 \int_0^\infty \lambda^{p/2-a+2}v^{p/2-a+2}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2}
\]

\[
\times \exp \left(-\frac{wv\lambda}{2}\right) g(v)dvd\lambda
\]

\[
+2^{-1}w\Phi_1(w)\Phi_0^{-2}(w) \int_0^1 \int_0^\infty \lambda^{p/2-a+1}v^{p/2-a+2}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2}
\]

\[
\times \exp \left(-\frac{wv\lambda}{2}\right) g(v)dvd\lambda.
\]

For $b > 0$, applying an integration by parts gives

\[
\int_0^1 \lambda^{p/2-a+2}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2)d\lambda = \frac{2}{wv} \left( \int_0^1 (p/2 - a + 2)\lambda^{p/2-a+1}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2)d\lambda
\]

\[
- \int_0^1 b\lambda^{p/2-a+2}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2)d\lambda
\]

\[
+ \int_0^1 (a - b - 2)(v - 1)\lambda^{p/2-a+2}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-3} \exp(-wv\lambda/2)d\lambda
\]

\[
= \frac{p}{wv} \int_0^1 \lambda^{p/2-a+1}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2)d\lambda
\]

\[
- \frac{2}{wv} \int_0^1 ((a - 2)(1 - \lambda) + bv\lambda)\lambda^{p/2-a+1}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-3} \exp(-wv\lambda/2)d\lambda.
\]

Similarly, we get

\[
\int_0^1 \lambda^{p/2-a+1}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2)d\lambda
\]

\[
= \frac{p - 2}{wv} \int_0^1 \lambda^{p/2-a}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-2} \exp(-wv\lambda/2)d\lambda
\]

\[
- \frac{2}{wv} \int_0^1 ((a - 2)(1 - \lambda) + bv\lambda)\lambda^{p/2-a}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a-b-3} \exp(-wv\lambda/2)d\lambda.
\]

Therefore we have

\[
\frac{d}{dw} \phi_{a,b}^g(w) = \Phi_0^{-2}(w) \left( \Phi_0(w) \int_0^1 \int_0^\infty ((a - 2)(1 - \lambda) + bv\lambda)\lambda^{p/2-a+1}(1 - \lambda)^b
\]

\[
\times (1 - \lambda + v\lambda)^{a-b-3} \exp(-wv\lambda/2)v^{p/2-a+1}g(v)dvd\lambda
\]

\[
- \Phi_1(w) \int_0^1 \int_0^\infty ((a - 2)(1 - \lambda) + bv\lambda)\lambda^{p/2-a}(1 - \lambda)^b
\]

\[
\times (1 - \lambda + v\lambda)^{a-b-3} \exp(-wv\lambda/2)v^{p/2-a+1}g(v)dvd\lambda \right).
\]
Making a transformation gives
\[
\frac{d}{dw} \phi_{a,b}^g(w) = \Phi_0^{-2}(w) \left( \int_0^1 \int_0^\infty \frac{(a - 2) + bt}{1 + t} \frac{\lambda}{1 - \lambda} G_w(t, \lambda) dt d\lambda \int_0^1 \int_0^\infty G_w(t, \lambda) dt d\lambda \right) - \int_0^1 \int_0^\infty \frac{(a - 2) + bt}{1 + t} \frac{1}{1 - \lambda} G_w(t, \lambda) dt d\lambda \int_0^1 \int_0^\infty \lambda G_w(t, \lambda) dt d\lambda,
\]
where
\[
G_w(t, \lambda) = t^{p/2-a+1} (1 + t)^{a-b-2} \lambda^{-2} (1 - \lambda)^{p/2} \exp(-wt(1 - \lambda)/2) g(\lambda^{-1}(1 - \lambda)t).
\]
Here FKG inequality due to Fortuin et al. (1971) is useful.

**Lemma 2.1 (FKG inequality).** Let \( \xi \) denote a probability density function with respect to \( \nu \) for a \( q \)-variate random vector. For two points \( y = (y_1, \cdots, y_q) \) and \( z = (z_1, \cdots, z_q) \), define
\[
\begin{align*}
    y \wedge z &= (y_1 \wedge z_1, \cdots, y_q \wedge z_q), \\
    y \vee z &= (y_1 \vee z_1, \cdots, y_q \vee z_q)
\end{align*}
\]
where \( a \wedge b = \min(a, b) \), \( a \vee b = \max(a, b) \). Suppose that \( \xi \) satisfies
\[
\xi(y) \xi(z) \leq \xi(y \vee z) \xi(y \wedge z) \tag{2.3}
\]
and that \( \alpha(y), \beta(y) \) are nondecreasing in each argument and that \( \alpha, \beta, \alpha \beta \) are integrable with respect to \( \xi \). Then
\[
\int \alpha \beta \xi d\nu \geq \int \alpha \xi d\nu \int \beta \xi d\nu.
\]

By the assumption \( I \), we have
\[
G_w(t_1, \lambda_1) G_w(t_2, \lambda_2) \leq G_w(t_1, \lambda_1) G_w(t_2, \lambda_2)
\]
for \( t_1 \leq t_2 \) and \( \lambda_1 \leq \lambda_2 \). In the case \( b = a + 2 \geq 0 \), \( (a - 2 + bt)/(1 + t)^{-1}(1 - \lambda)^{-1} \) is nondecreasing in \( t \) and \( \lambda \), \( \phi_{a,b}^g(w) \) is nondecreasing by Lemma 2.1. We note that, if \( b = a = 0 \), \( \phi_{a,b}^g(w) \) is nondecreasing without the assumption \( I \).

In the case \( b = 0 \), a slightly different calculation also gives that \( \phi_{a,b}^g(w) \) is nondecreasing. This completes the proof. \( \square \)

**Theorem 2.2.** If \( g(v) \) satisfies the assumption \( I \), \( \phi_{a,b}^g(w)/w \) is monotone nonincreasing in \( w \) for \( a - b - 2 \geq 0 \).

**Proof.** The derivative of \( \phi_{a,b}^g(w)/w \) is
\[
\frac{d}{dw} \left( \phi_{a,b}^g(w)/w \right) = 2^{-1} \Phi_0^{-2}(w) \left( \Phi_1(w) \int_0^1 \int_0^\infty \lambda^{p/2-a+1} v^{p/2-a+2} (1 - \lambda)^b \right.
\]
\[
\times (1 - \lambda + v\lambda)^{a-b-2} \exp \left( -\frac{wv\lambda}{2} \right) g(v) dv d\lambda
\]
\[
- \Phi_0(w) \int_0^1 \int_0^\infty \lambda^{p/2-a+2} v^{p/2-a+2} (1 - \lambda)^b (1 - \lambda + v\lambda)^{a-b-2}
\]
\[
\times \exp \left( -\frac{wv\lambda}{2} \right) g(v) dv d\lambda.
\]
Proof. We will show $\Psi$ where

$$\int_0^1 \int_0^\infty \lambda M_w(t, \lambda) d\lambda dt \int_0^1 \int_0^\infty t M_w(t, \lambda) d\lambda dt$$

where

$$M_w(t, \lambda) = \lambda^{-2}(1 - \lambda)^b t^{\frac{p}{2} - a + 1}(1 - \lambda + t)^{a - b - 2} \exp(-tw/2) \exp(-tw/2)$$

The inequality $a - b - 2 \geq 0$ and the assumption (I) imply that

$$M_w(t_1, \lambda_2)M_w(t_2, \lambda_1) \leq M_w(t_1, \lambda_1)M_w(t_2, \lambda_2),$$

for every $\lambda_1 \leq \lambda_2$ and $t_1 \leq t_2$. By Lemma 2.1, $\phi_{a,b}(w)/w$ is nonincreasing. \[ \square \]

**Theorem 2.3.** If $a < p/2 + 1$ and $b \geq \max(a - 2, 0)$, $\lim_{w \to \infty} \phi_{a,b}(w) = 2(p/2 - a + 1) \int_0^\infty v^{-1} g(v) dv$.

**Proof.** $\phi_{a,b}(w)$ can be expressed as

$$\phi_{a,b}(w) = \int_0^\infty \Psi(v, w)v^{-1} g(v) dv$$

where

$$\Psi_k(v, w) = (wv)^{p/2 - a + 1 + k} \int_0^1 \lambda^{p/2 - a + k}(1 - \lambda)^b(1 - \lambda + v\lambda)^{a - b - 2} \exp(-wv\lambda/2) d\lambda$$

for $k = 0, 1$. Making a transformation gives

$$\Psi_k(v, w) = (2w)^{p/2 - a + 1 + k} \int_0^\infty g(s, v) \exp(-ws) ds$$

where $g(s, v) = s^{p/2 - a + k}(1 - 2s/v)^b(1 - 2s/v + 2s)^{a - b - 2} I_{(0,v/2)}(s)$. Since $g(s, v) \sim s^{p/2 - a + k}$ as $s \to 0+$, from a Tauberian theorem, the well-known property of Laplace transformation, we have

$$\int_0^\infty g(s, v) \exp(-ws) ds \sim \Gamma(p/2 - a + 1 + k)w^{-(p/2 - a + 1 + k)}$$

as $w \to \infty$, which implies that $\lim_{w \to \infty} \Psi_k(v, w) = 2^{p/2 - a + 1 + k}\Gamma(p/2 - a + 1 + k)$. Next we will show $\Psi_k(v, w)$ is bounded above. For $2 \leq a < p/2 + 1$ and $b \geq a - 2$, $\Psi_k(v, w)$ is evaluated as

$$\Psi_k(v, w) \leq (wv)^{p/2 - a + 1 + k} \int_0^1 \lambda^{p/2 - a + k} \exp(-wv\lambda/2) d\lambda$$

$$\leq 2^{p/2 - a + 1 + k}\Gamma(p/2 - a + 1 + k).$$
In the case $a < 2$ and $b \geq 0$, $\Psi_k(v, w)$ is evaluated as

$$
\Psi_k(v, w) \leq w^{p/2-a+1+k}v^{p/2+k-1}\int_0^1 \lambda^{p/2+k-2} \exp(-wv\lambda/2)d\lambda
\leq w^{2-a+2p/2+k-1-1}G(p/2+k-1).
$$

Therefore $\Psi_k(v, w)$ for $a < p/2+1$ and $b \geq \max(a-2, 0)$ is bounded by a constant value which is not depend upon $v$ and $w$. By Lebesgue’s dominated convergence theorem, we have

$$
\lim_{w \to \infty} \phi_{a,b}^g(w) = \lim_{w \to \infty} \int_0^\infty \Psi_1(v, w)v^{-1}g(v)dv / \lim_{w \to \infty} \int_0^\infty \Psi_0(v, w)g(v)dv
= \int_0^\infty \lim_{w \to \infty} \Psi_1(v, w)v^{-1}g(v)dv / \int_0^\infty \lim_{w \to \infty} \Psi_0(v, w)g(v)dv
= 2(p/2 - a + 1) \int_0^\infty v^{-1}g(v)dv.
$$

### 2.2 Admissibility and Minimaxity

First we consider admissibility. In this estimation setting, Brown (1979) showed that if a generalized prior density $\pi(\theta)$ is the form of $\pi(\theta) \sim \|\theta\|^{\alpha}$ with $\alpha \leq 2-p$ as $\|\theta\| \to \infty$, the generalized Bayes estimator with respect to $\pi$ is admissible. Because we have, by using a Tauberian theorem,

$$
h_{a,b}(\theta) = \int_0^\infty s^{p/2-a}(1+s)^{a-b-2} \exp(-s\|\theta\|^2/2)ds
\sim \Gamma(p/2-a+1)2^{p/2-a+1}\|\theta\|^{2a-p-2},
$$
as $\|\theta\| \to \infty$, we have the following result.

**Theorem 2.4.** The estimator $\delta_{a,b}^g(X)$ is admissible if $a \leq 2$.

Next we consider minimaxity. For the shrinkage estimator $\delta\phi(x) = (1-\phi(\|x\|^2)/\|x\|^2)x$ with nondecreasing $\phi$, the sufficient conditions for minimaxity in the case of scale mixtures of multivariate normal distributions were given by Strawderman (1974) and Berger (1975).

**Theorem 2.5.** Assume that

1. $\phi(w)/w$ is monotone nonincreasing in $w$ and $0 \leq \phi(w) \leq 2(p-2)/\int_0^\infty vg(v)dv$, or
2. $0 \leq \phi(w) \leq 2(p-2)\int_0^\infty v^{p/2-1}g(v)dv / \int_0^\infty v^{p/2}g(v)dv$.

Then $\delta\phi(X)$ is minimax.

The assumption (ii) is a special case of Berger (1975)’s Theorem 3. Combining Theorem 2.1, 2.2, 2.3 and 2.5, we have the sufficient conditions for minimaxity of $\delta_{a,b}^g(x)$. Let

$$
a_1 = p/2 + 1 - (p-2)\left( \int_0^\infty v^{-1}g(v)dv \int_0^\infty vg(v)dv \right)^{-1}
$$
and
\[ a_2 = \frac{p}{2} + 1 - (p - 2) \int_0^\infty v^\frac{p}{2} - 1 g(v) dv \left( \int_0^\infty v^{-1} g(v) dv \int_0^\infty v^\frac{p}{2} g(v) dv \right)^{-1}. \]

It is noted that \( a_1 \) is clearly less than \( a_2 \).

**Theorem 2.6.**

1. If \( g(v) \) satisfies the assumption (I) and \( b = a - 2 \), \( \delta_{a,b}^g(x) \) is minimax for \( \max(a_1, 2) \leq a < \frac{p}{2} + 1 \).

2. \( \delta_{a,b}^g(x) \) is minimax for \( b = a - 2 \) and \( \max(a_2, 2) \leq a < \frac{p}{2} + 1 \).

3. If \( g(v) \) satisfies the assumption (I), \( \delta_{a,b}^g(x) \) is minimax for \( a_2 \leq a < \frac{p}{2} + 1 \) and \( b \geq \max(a - 2, 0) \).

Part 1 and 2 of Theorem 2.6 are revised versions of Strawderman (1974)'s result about minimax generalized Bayes estimators. In fact, \( h_{a,b}(\theta) \) for \( b = a - 2 \) is proportional to \( \|\theta\|^{2a-2-p} \), which equals the generalized prior density function considered in Strawderman (1974).

Combining Theorem 2.4 and 2.6, we have admissible minimax estimators.

**Theorem 2.7.**

1. If \( a_1 \leq 2 \) and \( g(v) \) satisfies the assumption (I), then \( \delta_{a,b}^g(x) \) is admissible and minimax for \( a = 2 \) and \( b = 0 \).

2. If \( a_2 \leq 2 \), then \( \delta_{a,b}^g(x) \) is admissible and minimax for \( a = 2 \) and \( b = 0 \).

3. If \( g(v) \) satisfies the assumption (I) and \( a_2 \leq 2 \), \( \delta_{a,b}^g(x) \) is admissible and minimax for \( a_2 \leq a \leq 2 \) and \( b \geq 0 \).

Theorem 2.7 is a generalization of Berger (1976) in which minimax admissible estimators in the multivariate normal case were given.

**Figure 1**: Ranges of values \((a, b)\) for minimaxity and admissibility.
Finally we consider the class of $g(v)$ satisfying the assumption $(I)$. For example, we easily check that $g_{\alpha,\beta}(v) = \beta^{\alpha} \Gamma(\alpha)^{-1} v^{\alpha-1} \exp(-\beta v)$ for $\alpha > 0$ and $\beta > 0$ satisfies the assumption. In particular $g_{\alpha,\beta}(v)$ for $\alpha = \beta = m/2$ corresponds to the case of multivariate-$t$ distribution whose degrees of freedom is $m$. If $m \geq 4$, we have $a_1 \leq 2$ and hence $\delta_{a,b}(X)$ for $a = 2$ and $b = 0$ is admissible and minimax. Furthermore if $m \geq p + 2$, we have $a_2 \leq 2$ and hence we can propose many admissible minimax estimators.

Acknowledgements

I am grateful to referees and the Editor for their valuable comments and helpful suggestions.

References


