Minimax Estimation of Means of Multivariate Normal Mixtures

JINE-PHONE CHOU

Academia Sinica, Taiwan

AND

WILLIAM E. STRAWDERMAN

Rutgers University*

Communicated by the Editors

Assume \( X = (X_1, \ldots, X_p)' \) is a normal mixture distribution with density w.r.t. Lebesgue measure,
\[
f(x) = \int \frac{1}{(2\pi)^{p/2} \sigma F} e^{-(x-\theta)' \Sigma^{-1}(x-\theta)/2\sigma^2} dF(\sigma),
\]
where \( \Sigma \) is a known positive definite matrix and \( F \) is any known c.d.f. on \((0, \infty)\). Estimation of the mean vector under an arbitrary known quadratic loss function \( L_{Q}(\theta, a) = (a - \theta)'Q(a - \theta), Q \) a positive definite matrix, is considered. An unbiased estimator of risk is obtained for an arbitrary estimator, and a sufficient condition for estimators to be minimax is then achieved. The result is applied to modifying all the Stein estimators for the means of independent normal random variables to be minimax estimators for the problem considered here. In particular the results apply to the Stein class of limited translation estimators.


1. INTRODUCTION

Let \( X = (X_1, \ldots, X_p)' \) be a random vector with density w.r.t. Lebesgue measure,
\[
f(x) = \int \frac{1}{(2\pi)^{p/2} \sigma F} e^{-(x-\theta)' \Sigma^{-1}(x-\theta)/2\sigma^2} dF(\sigma), \tag{1.1}
\]
Received March 7, 1988; revised May 9, 1990.


Key words and phrases: minimax estimator, multivariate normal mixture mean, quadratic loss.

* This research was done while the second author was visiting Academia Sinica. This research was supported by NSF Grant DMS-84-18416.
where $\Sigma$ is a known positive definite matrix and $F$ is any known c.d.f. on $(0, \infty)$. We consider the problem of estimating $\theta = (\theta_1, \ldots, \theta_p)'$, the mean vector, with loss function

$$L_0(\theta, a) = (a - \theta)'Q(a - \theta),$$  

(1.2)

where $Q$ is a positive definite matrix. We know that the usual estimator $\delta_0(X) = X$ is an equalizer, extended Bayes, and hence minimax estimator, provided $\text{E}\sigma^2$ and $\text{E}\sigma^{-2}$ are finite. Hence all estimators which improve upon the usual estimator $\delta_0(X)$ are minimax. We also know that a normal distribution with mean $\theta$ and covariance matrix $\Sigma$ is a special case of normal mixtures, for which the density is of the form (1.1) with $F$ being degenerate at 1. Estimation for normal mixture distributions was first studied by Strawderman [11] and also by Berger [3].

For the problem of estimating the mean of a $p$-dimensional normal distribution with squared error as loss, when the covariance matrix is the identity, Stein [9] first showed that the best invariant estimator is inadmissible, provided $p \geq 3$. Later James and Stein [8] presented a particularly simple estimator to improve upon the best invariant estimator. When the covariance matrix is known to be an arbitrary positive definite matrix and the loss is an arbitrary quadratic loss, classes of minimax estimators are found by Berger [2], Strawderman [12], Chou [5] and others. Strawderman [11] exploring in a different direction, found classes of minimax estimators for location parameters of certain spherically symmetric distributions. Berger [3] dealt with an extension combining Strawderman [11] and Berger [2]. He considered the estimation problem for the mean of a $p$-dimensional normal mixture distribution $X = (X_1, \ldots, X_p)'$ with density (1.1), under the loss (1.2). He proved that estimators given by

$$\delta^B(X) = \left( I - \frac{r(x'\Sigma^{-1}Q^{-1}x)Q^{-1}x}{x'\Sigma^{-1}Q^{-1}x} \right) X,$$

(1.3)

where $r(\cdot)$ satisfies (i) $r(\cdot)$ is nondecreasing in $\cdot$; (ii) $r(\cdot)/\cdot$ is nonincreasing in $\cdot$; (iii) $0 \leq r \leq 2/E(1/x'\Sigma^{-1}x)$ are minimax estimators, provided $p \geq 3$. Here $E_0$ denotes the expectation under the mean vector $\theta = 0$. In this paper, we consider Berger's [3] setup and find minimax estimators analogous to those of Stein [10] including his versions of the limited translation estimators.

For some $p$-dimensional random vector $Z$ with density $f(z) = e^{\mu z - M(\mu) - K(z)}I_F(z)$, Chou [5] established the identity $E_\mu(\nabla_z K(Z) - \mu) h(Z) = E_\mu \nabla_z h(Z)$ for some function $h$. In Section 2 through use of this identity, we derive an unbiased estimator of the risk function of a nearly arbitrary
estimator of the mean vector of a normal mixture, under the loss (1.2). Further, we find a sufficient condition for an estimator to be minimax. In Section 3 we apply the result of Section 2, and as a consequence all Stein’s [10] estimators or Hudson’s [7] estimators for the mean of a multivariate normal under squared loss can be modified to be minimax estimators for the problem considered here.

2. A SUFFICIENT CONDITION FOR ESTIMATORS TO BE MINIMAX

For notation, let \(|\cdot|\) denote the Euclidean norm, and \(\nabla_z z=(z_1, ..., z_p)'\) be the vector differential operator of first partial derivatives with \(i\)th component

\[
\nabla_i = \frac{\partial}{\partial z_i}.
\]

We define a function \(g=(g_1, ..., g_p)': \mathbb{R}^p \to \mathbb{R}^p\) to be almost differentiable if all its coordinate functions \(g_j\) are an indefinite integral of \(\partial g_j/\partial z_i\) for all \(i = 1, ..., p\).

**Lemma 1.** Let \(z\) be a random vector in \(\mathbb{R}^p\) with density, with respect to Lebesgue measure, \(f(z) = e^{(u'z - M(u) - K(z))'}I_E(z)\), \(E\) an open connected set in \(\mathbb{R}^p\), \(K(z)\) differentiable. If \(f(z)\) approaches zero monotonically as \(z\) approaches the boundary of \(E\) along the coordinate axes, then the identity

\[
E(\nabla_z K(Z) - \mu) h(Z) = E\nabla_z h(Z)
\]

holds for any almost differentiable real valued function \(h(\cdot)\) satisfying

\[
E|\nabla h(z)| < \infty, \quad \text{and} \quad E|\nabla_z K(Z) - \mu) h(Z)| < \infty \quad \text{if} \quad p > 1.
\]

**Proof.** Chou [5].

Lemma 1 will be used to obtain an unbiased estimator of the risk of a nearly arbitrary estimator of the mean vector of a multivariate normal mixture. The result is stated in the following lemma precisely.

**Lemma 2.** Let \(X=(X_1, ..., X_p)'\) be a normal mixture with density (1.1). Assume that \(g=(g_1, ..., g_p)': \mathbb{R}^p \to \mathbb{R}^p\) is an almost differentiable function with all its coordinate functions \(g_j\) satisfying \(E(|\partial g_j/\partial x_i| \mid \sigma^2) < \infty\) and
Then for estimating \( \theta \), the risk of the estimator \( X + g \) under the loss (1.2), is given by

\[
R(\theta, X + g) = \text{tr}(\Sigma Q) E \sigma^2 + E \{ 2 \sigma^2 \nabla_x \cdot Bg + \| Bg \|^2 \},
\]

where \( B'B = Q, \tilde{z} = (B')^{-1} \Sigma x, \) and \( \nabla_x \cdot Bg = \sum_{i=1}^p (\partial (Bg)/\partial z_i) \).

**Proof.** First note that if \( F \) is degenerate at some fixed \( \sigma \), then \( X \) is the normal distribution with mean \( \theta \) and covariance matrix \( \sigma^2 \Sigma \). By using the identity (2.1) with \( Z = (B')^{-1}(\sigma^2 \Sigma)^{-1} X \), we have

\[
E(BX - B\theta)'Bg = E\tilde{z}\cdot Bg = E\sigma^2 \tilde{z}\cdot Bg. \tag{2.2}
\]

For \( F \) being an arbitrary c.d.f. on \((0, \infty)\),

\[
R(\theta, X + g) = E(X + g - \theta)' Q(X + g - \theta)
= E(BX +Bg - B\theta)'(BX + Bg - B\theta)
= E[E(BX - B\theta + Bg)'(BX - B\theta + Bg) | \sigma]
= \text{tr}(\Sigma Q) E \sigma^2 + E \{ 2 \sigma^2 \nabla_x \cdot Bg + \| Bg \|^2 \}.
\]

The last equality follows from the facts that \( E \{(Bx - B\theta)'(Bx - B\theta) | \sigma\} = \text{tr}(\Sigma Q)\sigma^2 \) and (2.2). \( \Box \)

Lemma 3 is a slight extension of Anderson’s lemma [11], which will help us to achieve the main result in this section. Before stating the lemma, we define that a function \( f: \mathbb{R}^p \to \mathbb{R} \) is unimodal if the set \( K_u = \{ x : f(x) \geq u \} \) is convex for any \( u \geq 0 \).

**Lemma 3.** Let \( h \) and \( f \) be functions from \( \mathbb{R}^p \) to \( \mathbb{R}^+ \cup \{0\} \). Assume that \( h \) and \( f \) are symmetric about the origin, unimodal, and \( \int f(x) \, dx \) and \( \int h(x) \, dx \) are finite. Then

\[
\int h(x) \, f(x+ky) \, dx \geq \int h(x) \, f(x+y) \, dx
\]

for \( 0 \leq k \leq 1, \ y \in \mathbb{R}^p \).

**Proof.** Let \( v, \tilde{v} \) be measures with \( dv = f(x+ky) \, dx \), \( d\tilde{v} = f(x+y) \, dx \). From Anderson’s lemma,

\[
\int_{K_u} dv \geq \int_{K_u} d\tilde{v}
\]

for every \( u \geq 0 \), where \( K_u = \{ x : h(x) \geq u \} \). Thus

\[
v(K_u) = H(u) \geq \tilde{v}(K_u) = \tilde{H}(u).
\]
By the definition of the integral, we have
\[
\int h(x) f(x + ky) \, dx - \int h(x) f(x + y) \, dx
= \int h(x) \, dv - \int h(x) \, d\tilde{v}
= -\int_0^\infty u \, dH(u) + \int_0^\infty u \, d\tilde{H}(u) = \int_0^\infty u \, d(\tilde{H}(u) - H(u))
= \lim_{b \to \infty} h[\tilde{H}(b) - H(b)] - \lim_{a \to \infty} a[\tilde{H}(a) - H(a)] + \int_0^\infty [H(u) - \tilde{H}(u)] \, du
\geq 0.
\]
The last equality is due to integration by parts, and the last inequality follows from the facts that \( \int h(x) f(x) \, dx < \infty \), \( bH'(b) \to 0 \) as \( b \to \infty \), where \( H'(b) = \int_x f(x) \, dx \), and hence \( bH(b) \to 0 \) and \( b\tilde{H}(b) \to 0 \) as \( b \to \infty \).

Stein [10] proved for \( p \geq 3 \) and for \( g \) with \( 2V \cdot g + \|g\|^2 \leq 0 \), under squared error loss, that the estimator \( X + g(X) \) dominates the usual estimator \( X \) of the mean vector of a \( p \)-dimensional normal with covariance \( I \). Our main result is an analogous result for the mixture of normal case.

**Theorem 1.** Let \( X \) be a normal mixture with density (1.1), and let \( g \) be as given in Lemma 2. Assume that \( E\sigma^2 \) and \( E\sigma^{-2} \) are finite and that \( g \) satisfies the conditions that \( \|g\|^2 + 2V \cdot g \leq 0 \), \( g \) is homogeneous of degree \(-1\), and \( \|g\|^2 \) is unimodal. Then the estimator
\[
X + B^{-1}r\left(\frac{1}{\|g(Z)\|^2}\right) g(Z)
\]
is a minimax estimator under the loss (1.2), where \( B^*B = Q \), \( Z = (B^*)^{-1} \Sigma^{-1} X \), and \( r: \mathbb{R}^+ \to \mathbb{R} \) satisfies \( 0 \leq r \leq (E\sigma^{-2})^{-1} \), \( r(y)/y \) is nonincreasing and \( \|\nabla r(1/\|g(\cdot)\|^2)\| \cdot g(\cdot) \leq 0 \).

**Proof.** Since \( \delta^0(x) = X \) is a minimax estimator, an estimator \( \delta \) is thus minimax if \( \Delta_\delta(\theta) = R(\theta, \delta^0) - R(\theta, \delta) \leq 0 \) for all \( \theta \). From Lemma 2 and the properties of \( r \) and \( g \), it follows that
\[
\Delta_\delta(\theta) = E \left[ r^2 \left( \frac{1}{\|g(Z)\|^2} \right) \|g(Z)\|^2 + 2\sigma^2 \nabla_x \cdot r \left( \frac{1}{\|g(Z)\|^2} \right) g(Z) \right] \sigma^2
\leq E \left[ r^2 \left( \frac{1}{\|g(Z)\|^2} \right) \|g(Z)\|^2 + 2\sigma^2 r \left( \frac{1}{\|g(Z)\|^2} \right) \nabla_x g(Z) \right] \sigma^2
\leq E \left[ r^2 \left( \frac{1}{\|g(Z)\|^2} \right) \|g(Z)\|^2 \sigma^2 r \left( \frac{1}{\|g(Z)\|^2} \right) \|g(Z)\|^2 \right] \sigma^2.
\]
Let \( Y = Z / \sigma \) and \( a = \sup r(\cdot) \). Then

\[
A_\phi(\theta) \leq E \left[ E \left[ \frac{1}{\sigma^2} r^2 \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 - r \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \right] \right] \sigma^2
\]

\[
= E \left[ E \left[ r \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \left[ \frac{r(\sigma^2/\|g(Y)\|^2)}{\sigma^2} - 1 \right] \right] \sigma^2 \right]
\]

\[
\leq E \left[ E \left[ r \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \left[ \frac{a}{\sigma^2} - 1 \right] \right] \sigma^2 \right]. \tag{2.3}
\]

The first inequality in (2.3) follows from the fact that \( g \) is homogeneous of degree \(-1\).

Note that for given \( \sigma^2 \), \( Y \) has a normal distribution with mean \((B')^{-1} \Sigma^{-1} \theta / \sigma = \mu / \sigma\) and some covariance matrix which is not dependent on \( \sigma\). For fixed \( \sigma_o^2 \), \( r(\sigma_o^2 / \|g\|^2) \|g\|^2 \) is monotonic increasing in \( \|g\|^2 \) and hence is symmetric about 0 and unimodal, since \( \|g\|^2 \) is symmetric (since it is homogeneous of degree \(-1\)) and unimodal. By monotonicity of \( r(\cdot) \) it also follows that \( r(\sigma_o^2 / \|g\|^2) \|g\|^2 \leq r(\sigma^2 / \|g\|^2) \|g\|^2 \) for \( \sigma_o^2 < \sigma^2 \). Therefore, using Lemma 3,

\[
E_{\mu/\sigma_0} \left[ r \left( \frac{\sigma^2_o}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \right] \leq E_{\mu/\sigma_0} \left[ r \left( \frac{\sigma^2_1}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \right]
\]

\[
\leq E_{\mu/\sigma_1} \left[ r \left( \frac{\sigma^2_1}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \right];
\]

i.e.,

\[
E \left[ r \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \right] \sigma^2 \]

is increasing in \( \sigma^2 \). Hence, since \([a/\sigma^2 - 1]\) is decreasing in \( \sigma^2 \),

\[
E \left[ E \left\{ r \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \left[ \frac{a}{\sigma^2} - 1 \right] \right\} \sigma^2 \right] \]

\[
\leq E \left[ E \left\{ r \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \left[ \frac{a}{\sigma^2} - 1 \right] \right\} \right] E \left( \frac{a}{\sigma^2} - 1 \right)
\]

\[
\leq 0.
\]

The final inequality follows from the defining property of \( a \) and \( 0 \leq r \leq (E\sigma^{-2})^{-1} \).

Remark 1: Since \( E_0 X' \Sigma^{-1} X = p\sigma^2 \) and \( E_0 (1/X' \Sigma^{-1} X) = (1/(p-2)) E\sigma^{-2} \), the conditions of finiteness of \( E\sigma^2 \) and \( E\sigma^{-2} \) can be
replaced by finiteness of $E_0 X' \Sigma^{-1} X$ and $E_0 (1/X' \Sigma^{-1} X)$. Also, the condition $[\nabla r(1/\|g\|^2)] \cdot g \leq 0$ can be replaced by $\nabla \|g\|^2 \cdot g \geq 0$ if $r$ is a nondecreasing function.

Additionally, the condition that $\|g\|^2$ is unimodal may be weakened. What is required is that

$$E \left[ r \left( \frac{\sigma^2}{\|g(Y)\|^2} \right) \|g(Y)\|^2 \sigma^2 \right]$$

be nondecreasing in $\sigma^2$. By arguing coordinatewise, it suffices to assume that $\|g(Y)\|^2$ is unimodal in each coordinate separately and that $r(\cdot)$ is nondecreasing. Some of the examples in Section 3 require this modification of the theorem.

In Section 3, we apply our results to obtain a large collection of minimax estimators of $\theta$ which contains modifications of Stein's [10] estimators, modified Hudson's estimator [7] and modified versions of many estimators proposed before for the normal case.

APPLICATIONS

Let $g: \mathbb{R}^p \to \mathbb{R}^p$ be given by

$$g(X) = -\frac{1}{X'C} AX,$$

where

$$A \text{ is symmetric matrix with } 2A < \text{tr}(A) I \quad (3.2)$$

$$C = \left\{ \text{tr}(A) I - 2A \right\}^{-1} A^2. \quad (3.3)$$

Stein [10] showed that $g$ has the property $\|g\|^2 + 2\nabla \cdot g \leq 0$. Further, $g$ satisfies the conditions of Remark 1 if $\text{tr}(A) - \max_i(a_i) + 2 \min_i(a_i) \geq 0$, where $a_i$ are the eigenvalues of $A$.

**COROLLARY 1.** Assume that $X$ has the density (1.1) with $p \geq 3$. Let $A$ and $C$ be as given in (3.2) and (3.3) $B'B = Q$. Then for estimating $\theta$ under the loss (1.2), the estimator

$$\hat{\theta}(X) = X - r \left( \frac{X'\Sigma^{-1} B^{-1} C(B')^{-1} \Sigma^{-1} X)^2}{X'\Sigma^{-1} B^{-1} A^2(B')^{-1} \Sigma^{-1} X} \right)$$

$$\times \frac{1}{X'\Sigma^{-1} B^{-1} C(B')^{-1} \Sigma^{-1} X} B^{-1} A(B')^{-1} \Sigma^{-1} X$$
is minimax provided \( r(\cdot) \) is increasing \( \text{tr}(A) - 4 \max_i + 2 \min_i a_i > 0 \), \( r(Y)/Y \) is decreasing, and \( 0 < r(\cdot) < \left[ E(\sigma^{-2}) \right]^{-1} \).

**Proof.** The result follows directly from Theorem 1.

Note that the requirement of \( p \geq 3 \) is necessary for \( g \) of the form (3.1) to satisfy the conditions in Lemma 2.

**Remark 2.** The simplest minimax estimator of the form (3.4) is obtained by setting \( A = I \), and \( r(\cdot) \) a constant in Corollary 1. Note that minimax estimators proposed by Berger [3] have the form (1.3) which are exactly the special case of (3.4) with \( A = I \). If \( X \) is a normal distribution, that is, \( F \) is degenerate at \( 1 \), this class of minimax estimators is the same as in Berger [4] or Chou [5].

**Remark 3.** By an appropriate choice of \( A \) (see [10]), the minimax estimators (3.4) are based on a three-term symmetric moving average.

Here is the development for the mixture case of another solution of \( \| g \|^2 + 2V \cdot g \leq 0 \) given by Stein [10]. The resulting class of estimators contain versions of limited translation rules.

Let \( W_i = | Y_i |, i = 1, \ldots, p \), with \( W_{(1)} < W_{(2)} < \cdots < W_{(p)} \), and let \( k \) be a positive integer \( k > 2 \). Define \( g(Y) = (g_i(Y)) \) by

\[
g_i(Y) = \begin{cases} 
-\frac{Y_i}{\sum_{j=1}^{p} Y_j^2 \wedge W_{(k)}^2} & \text{if } |Y_i| \leq W_{(k)} \\
\frac{W_{(k)} \text{sign } Y_i}{\sum_{j=1}^{p} Y_j^2 \wedge W_{(k)}^2} & \text{if } |Y_i| > W_{(k)}.
\end{cases}
\]

Here \( a \wedge b = \min(a, b) \).

**Corollary 2.** Assume that \( X \) has the density (1.1) with \( p \geq 3 \). For estimating \( \theta \) under the loss (1.2), the estimator

\[
X + B^{-1} r \left( \frac{1}{\| g(B^T X) \|^2} \right) g((B')^{-1} \Sigma^{-1} X),
\]

where \( B'B = Q \), \( g \) is as given in (3.5), and \( r(\cdot) \) satisfies the conditions in Theorem 1, is minimax.

**Proof.** The result follows directly from Remark 1, and the proof is omitted.

**Remark 4.** By appropriate choice of \( r \), the minimax estimator proposed by Chou [5], which improve upon the naive estimator \( \delta^0(X) = X \) based on the order of data, are special cases of (3.6).
By a modification of the solution to $\|g\|^2 + 2\nabla \cdot g \leq 0$ given by Hudson [7], we can construct minimax estimators to improve upon the usual estimator $\delta_0(X) = X$ by shrinking the usual estimator towards a point determined by the average of data. Define $g(Y) = (g_i(Y))$ with

$$g_i(Y) = -\frac{Y_i - \bar{Y}}{\sum_{j=1}^{p} (Y_j - \bar{Y})^2}, \quad i = 1, \ldots, p, \quad (3.7)$$

where $\bar{Y} = \sum_{i=1}^{p} (1/p) Y_i$.

**Corollary 3.** Assume that $X$ has the density (1.1) with $p \geq 4$. For estimating $\theta$ under the loss (1.2), the estimator

$$X + B^{-1} r\left(\frac{1}{\|g((B')^{-1} \Sigma^{-1} X)\|} \right) g((B')^{-1} \Sigma^{-1} X),$$

where $B'B = Q$, $g$ is as given in (3.7) and $r(\cdot)$ satisfies the conditions in Theorem 1, is minimax.

**Proof.** The result follows directly from Theorem 1.

Note that in Corollary 3, we need $p \geq 4$ to ensure finiteness of $E(\|\nabla g_i\| | \sigma^2)$.

**Remark 5.** It is obvious that minimax estimators can be obtained by shrinking the usual estimator $\delta_0(X) = X$ toward a fixed point, say $a = (a_1, \ldots, a_p)'$. In other words, $g(Y) = (g_i(Y))$ is defined with $g_i(Y) = -(Y_i - a_i) / \sum_{j=1}^{p} (Y_j - a_j)^2$. And note that in this case, we only need $p \geq 3$.

**References**


