BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 16, Number 1, March 1988

ESTIMATION OF MEANS OF MULTIVARIATE NORMAL MIXTURES

BY

JINE-PHONE CHOU (周金鳳)

Abstract. Assume $X=(X_1,\ldots,X_p)'$ has a p-variate density w. r. t. Lebesgue measure,

$$f(x)=\int\frac{1}{(2\pi)^{p/2}\sigma^p|\Sigma|^{1/2}}e^{-(1/2)(x-\theta)'(\Sigma^{-1}/\sigma^2)(x-\theta)}\,dF(\sigma)$$

where Σ is a known positive definite matrix, F is any known c. d. f. on $(0,\infty)$ and $p\geqslant 3$. For estimating θ under an arbitrary known quadratic loss function $\mathcal{L}_Q(\theta,\delta)=(\delta-\theta)'Q(\delta-\theta)$, Q a positive definite matrix, classes of minimax estimators based on order statistics are found.

1. **Introduction.** Since Stein (1955) first showed that the inadmissibility of the usual estimator for the mean of a multivariate normal distribution of dimension three or more, there has been considerable interest in the problem of multiparameter estimation for exponential families. Suppose X is a p-dimensional normal with identity covariance matrix, James and Stein (1961) showed that under sequared error loss, estimators of the form

$$\delta^{J-S}(X)=\left(1-\frac{a}{\|X\|^2}\right)X,$$

where $0<a<2(p-2)$ have lower risk than the usual estimator X, which is also the maximum likelihood estimator, the best unbiased estimator, and minimax. Furthermore, various improved normal means estimators have been shown to dominate the usual estimator uniformly under arbitrary quadratic losses. These results can be found in Berger (1974), Chou (1985).

Received by the editors July 17, 1987 and revised form November 3, 1987.
AMS 1980 subject classifications. 62C99, 62F10, 62H99.
Key words and phrases: Minimax estimator, multivariate normal mixture mean, quadratic loss.

Strawderman (1974) and Berger (1974, 1975) extended things in different directions. Strawderman considered the estimating problem for location parameters of certain spherically symmetric distribution, and Berger for the mean of a p-dimensional normal mixture X with density of the form

$$(1.1) \qquad f(x) = \int \frac{1}{(2\pi)^{p/2} \sigma^p |\Sigma|^{1/2}} e^{-(1/2)(x-\theta)'(\Sigma^{-1}/\sigma^2)(x-\theta)} dF(\sigma)$$

under the loss function, with Q a known positive $p \times p$ definite matrix,

$$(1.2) \qquad \mathcal{L}_Q(\theta, a) = (a-\theta)' Q(a-\theta).$$

With some regularities on the density (1.1), Berger (1975) found estimators given by

$$(1.3) \qquad \delta^*(X) = \left(I - \frac{r(X' \Sigma^{-1} Q^{-1} \Sigma^{-1} X) Q^{-1} \Sigma^{-1}}{X' \Sigma^{-1} Q^{-1} \Sigma^{-1} X}\right) X$$

where $r(\cdot)$ satisfies (i) $r(\cdot)$ is nondecreasing in $\cdot$ (ii) $r(\cdot)/\cdot$ is nonincreasing in $\cdot$ (iii) $0 \leq r \leq 2E_0(1/X' \Sigma^{-1} X)$, are minimax estimators, provided $p \geq 3$. Here E_0 denotes the expectation under the mean vector $\theta = 0$.

In this paper we do further extension of the problem considered by Berger (1975). All the densities and losses considered in this paper have the form (1.1), (1.2) respectively. Besides estimators given in (1.3), much more other minimax estimators, which are Stein-type estimators (Stein 1981), are found here. There are

$$(1.4) \qquad \delta(X) = X + B^{-1} g((B')^{-1} \Sigma^{-1} X),$$

where B is a matrix such that $B' B = Q$, $B \Sigma B' = A$ a diagonal matrix and the ith component of $g(z)$ is given by, with $(a \wedge b) = \min(a, b)$,

$$(1.5) \qquad g_i(z) = \begin{cases} \dfrac{-r\left(\sum_{j=1}^{p} z_j^2 \wedge w_{(k)}^2\right)}{\sum z_j^2 \wedge w_{(k)}^2} z_i, & \text{if } \quad |z_i| \leq w_{(k)} \\[2ex] \dfrac{-r\left(\sum z_j^2 \wedge w_{(k)}^2\right)}{\sum z_j^2 \wedge w_{(k)}^2} w_{(k)} \operatorname{sign}(z_i), & \text{if } \quad |z_i| > w_{(k)} \end{cases}$$

where $w_i = |z_i|$, $w_{(1)} < w_{(2)} < \cdots < w_{(p)}$, $k \geq 3$, and $r(\cdot)$ is nondecreasing in $\cdot$, $r(\cdot)/\cdot$ nonincreasing in $\cdot$ and $0 \leq r$

$\leq 2(k-2)/E\sigma^{-2}$. Note that minimax estimators proposed by Berger (1974) are exactly special cases of (1.4) for $k=p$.

It is easy to check that if $E\sigma^2$ and $E\sigma^{-2}$ are finite, then the usual estimator $\delta^0(X)=X$ is an equalizer and extended Bayes rule. Hence δ^0 is a minimax estimator, and an estimator δ is thus minimax if $\Delta_\delta(\theta)=R(\delta^0,\theta)-R(\delta,\theta)\geq 0$ for all θ. The technique, finding δ such that $\Delta_\delta(\theta)\geq 0$ for all θ, is used in this paper to construct minimax estimators. We know that minimax estimators been constructed in this way are at least as good as δ^0, and estimators δ for which the above inequality holds and with strict inequality for some θ improve upon δ^0.

2. **Minimax estimators of the mean vector of a multivariate normal mixture.** For notation, let $\|\cdot\|$ denote the Euclidean norm, and ∇_z $z=(z_1,\cdots,z_p)'$ be the vector differential operator of first partial derivative with ith component

$$\nabla_i=\frac{\partial}{\partial z_i}.$$

We define a real valued function $h(z)$, $z\in \boldsymbol{R}^p$, is almost differentiable if it is an indefinite integral of $\partial h/\partial z_i$ for all $i=1,\cdots,p$.

LEMMA 1. *Let $Z=(Z_1,\cdots,Z_p)'$ be a random vector in $\boldsymbol{R}^p$ with density, with respect to Lebesgue measure, $f(z)=e^{\mu\cdot z-M(\mu)-K(z)}I_E(z)$, E an open connected set in $\boldsymbol{R}^p$, $K(z)$ differentiable. If $f(z)$ approaches zero monotonically as z approaches the boundary of E alone the coordinate axes, then the identity*

$$E(\nabla_z K(Z)-\mu)\,g(Z)=E\nabla_z g(Z)$$

holds for any almost differential real valued function $g(\cdot)$ satisfying $E\|\nabla g(Z)\|<\infty$, and $E\|(\nabla_z K(Z)-\mu)\,g(Z)\|<\infty$ if $p>1$.

Proof. Chou (1985).

LEMMA 2. *Let X have a p-variate normal distribution with mean vector θ and known covariance matrix Σ, and let $g=(g_1,g_2,\cdots,g_p)'$ be such that g_i almost differentiable and $E|\partial g_j/\partial X_i|<\infty$, $E\|(X-\theta)g_j\|<\infty$ for all i and j. Then*

$$E[(X-\theta)'Q(X-\theta)-(X+g-\theta)'Q(X+g-\theta)] = E[-2\nabla_z\cdot Bg-(Bg)'Bg]$$

where $B'B=Q$, $Z=(B')^{-1}\Sigma^{-1}X$ *and* $\nabla_z\cdot Bg=\sum_{i=1}^{p}\partial(Bg)_i/\partial z_i$, $(Bg)_i$ *the ith component of Bg.*

Proof.

$$E[(X-\theta)'Q(X-\theta)-(X+g-\theta)'Q(X+g-\theta)] = E[-2(BX-B\theta)'Bg-(Bg)'(Bg)].$$

By using the identity in Lemma 1 with $Z=(B')^{-1}\Sigma^{-1}X$, the Lemma follows.

THEOREM 1. *Let X have a normal mixture with density of the form* (1.1), *where* $p\geq 3$. *Assume* $E\sigma^2$ *and* $E\sigma^{-2}$ *are finite. Let* δ *be of the form* (1.4), *then* δ *is a minimax estimator of the mean parameter* θ *under the quardratic loss function* (1.2).

Proof. It is sufficient to prove that $\Delta_\delta(\theta)=R(\delta^0,\theta)-R(\delta,\theta)\geq 0$, for all θ. Here $\delta^0(X)=X$ is a minimax estimator of θ. Since $r(\cdot)$ is monotonic, WLOG, we assume $r(\cdot)$ is differentiable and let $\dot{r}$ denote its first derivative. By applying Lemma 2, we have

$$\begin{aligned}\Delta_\delta(\theta) &= \iint[(x-\theta)'Q(x-\theta) - (x+B^{-1}g((B')^{-1}\Sigma^{-1}x)-\theta)' \cdot Q(x+B^{-1}g((B')^{-1}\Sigma^{-1}x)-\theta)]\cdot f_\sigma(x)\,dx\,dF(\sigma) \\ &= \iint(-2[\nabla_z\cdot g(z)]\sigma^2 - g(z)'g(z))f_\sigma(x)\,dx\,dF(\sigma)\end{aligned} \tag{2.1}$$

where $f_\sigma(x)$ is the density of a p-dimensional normal distribution with mean θ and covariance $\sigma^2\Sigma$, $Z=(B')^{-1}\Sigma^{-1}X$. Since

$$\nabla_z\cdot g(z) = -\left[\frac{(k-2)\,r(\sum z_j^2\wedge w_{(k)}^2)}{\sum_{j=1}^{p} z_j^2\wedge w_{(k)}^2} + 4\dot{r}\left(\sum_{j=1}^{p} z_j^2\wedge w_{(k)}^2\right)\right]$$

and $\dot{r}$ is nonnegative, let $s=\sum_{j=1}^{p} z_j^2\wedge w_{(k)}^2$, and b denote $\sup r(\cdot)$, then

$$\begin{aligned}
\Delta_\delta(\theta) &\geq \iint \left[2(k-2) - \frac{r(s)}{\sigma^2}\right] \frac{r(s)}{s/\sigma^2} f_\sigma(x)\,dx\,dF(\sigma) \\
&\geq \iint \left[2(k-2) - \frac{b}{\sigma^2}\right] \frac{r(s)}{s/\sigma^2} f_\sigma(x)\,dx\,dF(\sigma).
\end{aligned} \tag{2.2}$$

Now we are going to prove $h(\sigma) = \int (r(s)/(s/\sigma^2)) f_\sigma(x)\,dx$ is nondecreasing in σ. Let $Y = Z/\sigma = (1/\sigma)(B')^{-1} \Sigma^{-1} X$, $V_i = |Y_i|$ and $V_{(1)} < V_{(2)} < \cdots < V_{(p)}$, and let $d_\sigma^i(\cdot)$ denote the density of a normal distribution with mean μ_i/σ and variance $1/a_i$. Here μ_i is the ith element of $\mu = (\mu_1, \cdots, \mu_p)' = (B')^{-1} \Sigma^{-1}\theta$, a_i is the ith diagonal element of $A = B \Sigma B'$. Then

$$h(\sigma) = \int \frac{r(\sigma^2(\sum y_j^2 \wedge v_{(k)}^2))}{\sum y_j^2 \wedge V_{(k)}^2} \left[\prod_{i=1}^{p} d_\sigma^i(y_i)\right] dy.$$

Define

$$\tilde{h}(\sigma_1, \cdots, \sigma_p) = \int \frac{r((\sum \sigma_j/p)^2 (\sum y_j^2 \wedge V_{(k)}^2))}{\sum y_j^2 \wedge V_{(k)}^2} \left[\prod_{i=1}^{p} d_{\sigma_i}^i(y_i)\right] dy\,.$$

If for each i, with σ_j's $j \neq i$ fixed, $\tilde{h}(\sigma_1, \cdots, \sigma_p)$ is nondecreasing in σ_i, then $h(\sigma)$ is nondecreasing. Since other cases are similar it is sufficient te provo that for fixed $\sigma_2, \cdots, \sigma_p$, $\tilde{h}(\sigma_1) = \tilde{h}(\sigma_1, \sigma_2, \cdots, \sigma_p)$ is nondecreasing in σ_1. And $\tilde{h}(\sigma_1)$ is nondecreasing if, with $\sigma_2, \cdots, \sigma_p$ fixed.

$$\int_{-\infty}^{\infty} \frac{r\left(\left(\sigma_1/p + \sum_{j=2}^{p} \sigma_j/p\right)^2 (\sum y_j^2 \wedge V_{(k)}^2)\right)}{\sum y_j^2 \wedge V_{(k)}^2} d_{\sigma_1}^1(y_1)\,dy_1$$

is nondecreasing in σ_1 for any fixed $y_2, \cdots, y_p$. WLOG we assume $\mu_1 \geq 0$, since $\mu_1 \leq 0$, the argument is the same. For $\sigma_1 > \sigma_1'$, since $r(\cdot)$ is nondecreasing and $d_{\sigma_1}^1(y_1) = d_{\sigma_1'}^1(2m - y_1)$ where $m = \frac{1}{2}((\mu_1/\sigma_1) + (\mu_1/\sigma_1'))$, with $\bar{\sigma} = 1/p \sum_{i=1}^{p} \sigma_i$

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{r\left(\left((1/p)\sigma_1 + 1/p \sum_{j=2}^{p} \sigma_j\right)^2 (\sum y_j^2 \wedge V_{(k)}^2)\right)}{\sum y_j^2 \wedge V_{(k)}^2} d_{\sigma_1}^1(y_1)\,dy_1 \\
&\qquad - \int_{-\infty}^{\infty} \frac{r\left(\left((1/p)\sigma_1' + \sum_{j=2}^{p} \sigma_j/p\right)^2 (\sum y_j^2 \wedge V_{(k)}^2)\right)}{\sum y_j^2 \wedge V_{(k)}^2} d_{\sigma_1'}^1(y_1)\,dy_1 \\
&\geq \int_{-\infty}^{\infty} \frac{r((\bar{\sigma})^2 (\sum y_j^2 \wedge V_{(k)}^2))}{[\sum y_j^2 \wedge V_{(k)}^2]} (d_{\sigma_1}^1(y_1) - d_{\sigma_1'}^1(y_1))\,dy_1
\end{aligned}$$

$$= \bar{\sigma}^2 \int_{-\infty}^{m} \left(\frac{r((\bar{\sigma})^2(\sum y_j^2 \wedge V_{(k)}^2))}{\bar{\sigma}^2(\sum y_j^2 \wedge V_{(k)}^2)} - \frac{r(\bar{\sigma}^2(\sum \tilde{y}_j^2 \wedge \tilde{V}_{(k)}^2))}{\bar{\sigma}^2(\sum \tilde{y}_j^2 \wedge \tilde{V}_{(k)}^2)} \right) \cdot$$

$$(d_{\sigma_1}^1(y_1) - d_{\sigma_1'}^1(y_1))\, dy_1$$

$$\geq 0$$

where $\tilde{y}_1 = 2m - y_1$, $\tilde{y}_i = y_i$, $i \neq 1$, $\tilde{V}_i = |\tilde{y}_i|$ and $\tilde{V}_{(1)} < \tilde{V}_{(2)} < \cdots < \tilde{V}_{(p)}$. The last inequality follows from the facts that $d_{\sigma_1}^1(y_1) - d_{\sigma_1'}^1(y_1) \geq 0$, $\sum \tilde{y}_j^2 \wedge \tilde{V}_{(k)}^2 \geq \sum (y_j^2 \wedge V_{(k)}^2)$ for all $y_1 \leq m$, and $r(\cdot)/\cdot$ is nonincreasing in $\cdot$. Hence $h(\sigma)$ is nondecreasing in σ.

Using the nondecreasing properties of $h(\sigma)$ and $2(k-2) - b/\sigma^2$, it follows from (2.1) that

$$\Delta_\delta(\theta) \geq \int \left(2(k-2) - \frac{b}{\sigma^2}\right) h(\sigma)\, dF(\sigma)$$

$$\geq \int \left(2(k-2) - \frac{b}{\sigma^2}\right) dF(\sigma) \int h(\sigma)\, dF(\sigma).$$

Since $b = \sup r(\cdot)$, $0 \leq r \leq 2(k-2)/E\sigma^{-2}$, $\Delta_\delta(\theta) \geq 0$.

REMARK 1. In Theorem 1, the requirement of $p \geq 3$ is necessary for g with the form (1.5) to satisfy the conditions in Lemma 2.

REMARK 2. For X with density (1.1) and $p \geq 3$, $E_0 X' \Sigma^{-1} X$ and $E_0(1/X' \Sigma^{-1} X)$ are finite is equivalent to $E\sigma^2$ and $E\sigma^{-2}$ are finite. Precisely $E_0 X' \Sigma^{-1} X = pE\sigma^2$ and $E_0(1/X' \Sigma^{-1} X) = (1/(p-2))\, E\sigma^{-2}$. Hence in Theorem 1 the condition, $E\sigma^2$ and $E\sigma^{-2}$ are finite can be replaced by the others are finite.

We now give a couple of examples of the application of Theorem.

Example 1. Let X be p-dimensional normal with mean θ and known covariance Σ. Clearly X has a density of the form (1.1) with F being degenerate at 1. When $p \geq 3$, Theorem 1 applies. Note that the class of minimax estimators thus defined by Theorem 1 is essentially the class found in Chou (1985).

Example 2. Consider the following model for the $p \times 1$ observation vector y,

$$y = \theta + u$$

where $\theta = (\theta_1, \cdots, \theta_p)'$ is a unknown parameter and $u_{p\times 1} \sim N(0, \sigma^2 \Sigma)$ with Σ a known $p \times p$ positive definite matrix and $\sigma > 0$ a variable distributed according to a known c. d. f. F. For estimating the mean vector θ of Y, under the loss (1.2), classes of minimax estimators are found by applying Theorem 1.

REMARK 3. Some other multivariate normal mixture distributions, for example "double exponential" or the "Cauchy like" distributions, were given by Berger (1975). And for estimating the mean vector of these distributions under the loss (1.2), Theorem 1 applies.

REFERENCES

1. J. Berger, *Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss*, Technical Report 376, Purdue Uuiv. (1974).

2. J. Berger, *Minimax estimation of location vectors for a wide class of densities*, Ann. Statist., 3 (1975), 1318-1328.

3. J. P. Chou, *An identity for multidimensional Continuous exponential families and its applications*, To appear in Journal of Multivariate Analysis (1985)

4. W. James and C. Stein, *Estimation with quadratic loss*, Proceedings of the Fourh Berkeley Symposium on Mathematical Statistics and Probability, 1 (1961), 361-379, University of California Press.

5. C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proc. Third Berkeley Symp. Math. Statist. Prob., 1 (1955), 197-206, Univ. of California Press.

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

7. W. E. Strawderman, *Minimax estimation of location paramers for certain spherically symmetric distributions*, J. Multivariate Anal., 4 (1974), 225-264.

Institute of Statistics
Academia Sinica
Taipei, Taiwan 11529
Republic of China