MAXIMUM LIKELIHOOD ESTIMATORS AND LIKELIHOOD RATIO CRITERIA FOR MULTIVARIATE ELLIPTICALLY CONTOURED DISTRIBUTIONS

TECHNICAL REPORT NO. 1

T. W. ANDERSON and KAI-TAI FANG

SEPTEMBER 1982

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG29-82-K-0156

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

MAXIMUM LIKELIHOOD ESTIMATORS AND LIKELIHOOD RATIO CRITERIA FOR MULTIVARIATE ELLIPTICALLY CONTOURED DISTRIBUTIONS

TECHNICAL REPORT NO. 1

T. W. ANDERSON
STANFORD UNIVERSITY

and

KAI-TAI FANG
STANFORD UNIVERSITY AND INSTITUTE OF APPLIED MATHEMATICS,
ACADEMIA SINICA, BEIJING, CHINA

SEPTEMBER 1982

U. S. ARMY RESEARCH OFFICE

CONTRACT DAAG29-82-K-0156

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. Introduction.

If the characteristic function (c.f.) of an n-dimensional random vector $\underset{\sim}{x}$ has the form $\exp(i\underset{\sim}{t}'\underset{\sim}{\mu})\phi(\underset{\sim}{t}'\underset{\sim}{\Sigma}\underset{\sim}{t})$, where $\underset{\sim}{\mu}: n \times 1$, $\underset{\sim}{\Sigma}: n \times n$, $\mathrm{rk}\,\underset{\sim}{\Sigma} = k$, $\underset{\sim}{\Sigma} \geq 0$, and $\phi \in \Phi_k = \{\phi \mid \phi(t)$ is a c.f. such that $\phi(t_1^2 + \cdots + t_k^2)$ is a c.f. on $\mathbb{R}^k\}$, we say that $\underset{\sim}{x}$ is distributed according to an elliptically contoured distribution with parameters $\underset{\sim}{\mu}$, $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\phi}$, and write $\underset{\sim}{x} \sim EC_n(\underset{\sim}{\mu},\underset{\sim}{\Sigma},\phi)$.

Elliptically contoured distributions have been extended to the case of matrices by Dawid (1977, 1978), Chmielewski (1980), and Anderson and Fang (1982b).

Let $\underset{\sim}{X}$, $\underset{\sim}{M}$ and $\underset{\sim}{T}$ be $n \times p$ matrices. We express them in terms of elements, columns, and rows as

$$\underset{\sim}{X} = (x_{ij}) = (\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_p) = \begin{pmatrix} \underset{\sim}{x}'_{(1)} \\ \vdots \\ \underset{\sim}{x}'_{(n)} \end{pmatrix}, \qquad \underset{\sim}{x} = \mathrm{vec}\,\underset{\sim}{X}' ,$$

$$\underset{\sim}{M} = (\mu_{ij}) = (\underset{\sim}{\mu}_1,\ldots,\underset{\sim}{\mu}_p) = \begin{pmatrix} \underset{\sim}{\mu}'_{(1)} \\ \vdots \\ \underset{\sim}{\mu}'_{(n)} \end{pmatrix}, \qquad \underset{\sim}{\mu} = \mathrm{vec}\,\underset{\sim}{M}' ,$$

$$\underset{\sim}{T} = (t_{ij}) = (\underset{\sim}{t}_1,\ldots,\underset{\sim}{t}_p) = \begin{pmatrix} \underset{\sim}{t}'_{(1)} \\ \vdots \\ \underset{\sim}{t}'_{(n)} \end{pmatrix}, \qquad \underset{\sim}{t} = \mathrm{vec}\,\underset{\sim}{T}' .$$

Here $\underset{\sim}{x} = \mathrm{vec}\,\underset{\sim}{X}' = (\underset{\sim}{x}'_{(1)},\ldots,\underset{\sim}{x}'_{(n)})'$ with the corresponding meanings for $\underset{\sim}{\mu}$ and $\underset{\sim}{t}$.

If the c.f. of a random matrix $\underset{\sim}{X}$ has the form

$$(1.1) \qquad \exp(i \sum_{j=1}^{n} \underset{\sim}{t}'_{(j)} \underset{\sim}{\mu}_{(j)}) \phi(\underset{\sim}{t}'_{(1)} \underset{\sim}{\Sigma}_1 \underset{\sim}{t}_{(1)}, \ldots, \underset{\sim}{t}'_{(n)} \underset{\sim}{\Sigma}_n \underset{\sim}{t}_{(n)}) \quad ,$$

with $\underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_n \geq 0$, we say that $\underset{\sim}{X}$ is distributed according to a multivariate (rows) elliptically contoured distribution (MECD) and write $\underset{\sim}{X} \sim MEC_{n \times p}(\underset{\sim}{M}; \underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_n; \phi)$. In this paper we consider only the subclass of MECD in which the function ϕ satisfies

$$(1.2) \qquad \phi(t_1, \ldots, t_n) = \phi(t_1 + \cdots + t_n) \quad ;$$

we continue to denote MECD in this subclass by $MEC_{n \times p}(\underset{\sim}{M}; \underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_n; \phi)$.

Let $\underset{\sim}{u}^{(q)}$ denote a random vector which is uniformly distributed on the unit sphere in $\mathbb{R}^q$ and $\Omega_q(\|t\|^2)$ denote its c.f.. Schoenberg (1938) pointed out that a c.f. $\phi \in \Phi_m$ if and only if

$$(1.3) \qquad \phi(u) = \int_0^\infty \Omega_m(ur^2) dF(r)$$

for some distribution function $F(x)$ on $[0,\infty)$. Since $\Phi_m \supset \Phi_n$ for $m < n$ if the distribution function F of R is related to ϕ as in (1.3) with n substituted for m, then $\phi \in \Phi_m$, $m < n$, and there exists a distribution function $F^*(x)$ of R^* being related to ϕ as in (1.3) with F^* substituted for F. There exists a relationship between R and R^*, that is $R^* \overset{d}{=} Rb$, where $b \geq 0$, $b^2 \sim B(m/2, (n-m)/2)$, b is independent of R, and $x \overset{d}{=} y$ denotes $\underset{\sim}{x}$ and $\underset{\sim}{y}$ have the same distribution. (cf. Cambanis, Huang and Simons (1981).) For convenience we denote these relationship by $R \leftrightarrow \phi \in \Phi_n$ and $R^* \overset{d}{=} R\, b_{m/2,(n-m)/2} \leftrightarrow \phi \in \Phi_m$.

Throughout the paper we assume that $\underset{\sim}{X} \sim MEC_{n\times p}\,(\underset{\sim}{M};\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_n;\phi)$ with $n > p$, $\Sigma_1 > 0,\ldots,\underset{\sim}{\Sigma}_n > 0$, $\phi \in \Phi_{np}$ and $\underset{\sim}{X}$ has a density of the form

$$(1.4)\qquad \prod_{i=1}^{n} |\underset{\sim}{\Sigma}_i|^{-1/2}\, g\Big[\sum_{i=1}^{n} (\underset{\sim}{x}_{(i)}-\underset{\sim}{\mu}_{(i)})'\underset{\sim}{\Sigma}_i^{-1}(\underset{\sim}{x}_{(i)}-\underset{\sim}{\mu}_{(i)})\Big] \quad .$$

In this case $\underset{\sim}{X}$ can be expressed as following

$$(1.5)\qquad \underset{\sim}{X} = \underset{\sim}{M} + (R_1\underset{\sim}{A}'_{(1)}\underset{\sim}{u}_1^{(p)},\ldots,R_n\underset{\sim}{A}'_{(n)}\underset{\sim}{u}_n^{(p)})' \quad ,$$

where R_j, $j = 1,\ldots,n$, are independent of $\underset{\sim}{u}_j^{(p)}$, $j = 1,\ldots,n$, $\underset{\sim}{u}_1^{(p)},\ldots,\underset{\sim}{u}_n^{(p)}$ are independent, $\underset{\sim}{u}_j^{(p)} \overset{d}{=} \underset{\sim}{u}^{(p)}$, $\underset{\sim}{\Sigma}_j = \underset{\sim}{A}_j'\underset{\sim}{A}_j$ is a factorization of $\underset{\sim}{\Sigma}_j$, $j = 1,\ldots,n$, $R_1 \geq 0,\ldots,R_n \geq 0$ and

$$(R_1^2,\ldots,R_n^2) \overset{d}{=} R^2(d_1,\ldots,d_n) \quad ,$$

where $(d_1,\ldots,d_{n-1}) \sim D_n(p/2,\ldots,p/2;p/2)$, a Dirichlet distribution, $d_n = 1-d_1 - \cdots - d_{n-1}$, R is nonnegative and independent of $d_1,\ldots,d_n$, and

$$(1.6)\qquad R^2 \overset{d}{=} \sum_{\alpha=1}^{n} (\underset{\sim}{X}_{(\alpha)}-\underset{\sim}{\mu}_{(\alpha)})'\underset{\sim}{\Sigma}_{(\alpha)}^{-1}(\underset{\sim}{X}_{(\alpha)}-\underset{\sim}{\mu}_{(\alpha)}) \quad .$$

When $\underset{\sim}{\mu}_{(1)} = \cdots = \underset{\sim}{\mu}_{(n)} = \underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}_1 = \cdots = \underset{\sim}{\Sigma}_n = \underset{\sim}{\Sigma}$, we write $\underset{\sim}{X} \sim LEC_{n\times p}(\underset{\sim}{\mu},\underset{\sim}{\Sigma},\phi)$. In this case (1.4) reduces to

(1.7) $$|\Sigma|^{-n/2} g(\operatorname{tr} \Sigma^{-1} G) \ ,$$

where

(1.8) $$G = \sum_{\alpha=1}^{n} (x_{(\alpha)} - \mu)(x_{(\alpha)} - \mu)'$$

and (1.6) is

(1.9) $$R^2 \overset{d}{=} \operatorname{tr} \Sigma^{-1} G \quad .$$

The maximum likelihood estimators and the unbiased estimators of μ and Σ, and their distributions are studied in Section 2 and some discussion on the existence and uniqueness of maxima is given. In Section 3 we extend many basic likelihood ratio criteria from the normal case to the MECD case. We find that most of the criteria and their corresponding distributions are the same in the class of MECD. A similar discussion on testing the general linear hypothesis is given in Section 4.

Throughout the paper $N_n(\mu, \Sigma)$ denotes the n-dimensional normal distribution with mean μ and covariance matrix Σ; $B(\alpha_1, \alpha_2)$ denotes the Beta distribution with parameters α_1 and α_2; $D_m(\alpha_1, \ldots, \alpha_{m-1}; \alpha_m)$ denotes the Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_m$; $F(k,m)$ denotes F-distribution with k and m degrees of freedom. $W_p(\Sigma, n)$ denotes the Wishart distribution with $p \times p$ covariance matrix Σ and n degrees of freedom; $U_{p,m,n}$ denotes Wilks' distribution which is the distribution of the ratio $|G|/|G+H|$, where $G \sim W_p(\Sigma, n)$, $H \sim W_p(\Sigma, m)$, and H and G

are independent; $\underset{\sim}{I}_n$ denotes the $n \times n$ identity matrix; $\underset{\sim}{\varepsilon}_n$ denotes the $n \times 1$ vector with elements 1; $\text{rk}\,\underset{\sim}{A}$ denotes the rank of the matrix $\underset{\sim}{A}$ and $\text{tr}\,\underset{\sim}{A}$ denotes the trace of $\underset{\sim}{A}$.

2. Estimation of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ and their distributions.

Assume $\underset{\sim}{X} \sim LEC_{n \times p}(\underset{\sim}{\mu},\underset{\sim}{\Sigma},\phi)$ and $\underset{\sim}{X}$ has a density

$$(2.1) \qquad |\underset{\sim}{\Sigma}|^{-n/2}\, g(\text{tr}\,\underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}) \quad ,$$

where $\underset{\sim}{G}$ is defined by (1.8). Let

$$(2.2) \qquad \underset{\sim}{W} = \sum_{j=1}^{n} (\underset{\sim}{x}_{(j)}-\bar{\underset{\sim}{x}})(\underset{\sim}{x}_{(j)}-\bar{\underset{\sim}{x}})' \quad ,$$

where

$$(2.3) \qquad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} \underset{\sim}{x}_{(j)} \quad .$$

Then $\underset{\sim}{W} = \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{X}$, with $\underset{\sim}{D} = I_n - (1/n)\underset{\sim}{\varepsilon}_n\underset{\sim}{\varepsilon}_n'$.

<u>Lemma 1</u>. If $n > p$, then $\underset{\sim}{G}$ and $\underset{\sim}{W}$ are positive definite with probability one.

<u>Proof</u>. From the assumption and (1.5) we have

$$(2.4) \qquad \underset{\sim}{X} = \underset{\sim}{\varepsilon}_n\underset{\sim}{\mu}' + R\underset{\sim}{U}\underset{\sim}{A} \quad ,$$

where $\underset{\sim}{U}: n \times p$, $\text{vec}\,\underset{\sim}{U} = \underset{\sim}{u}^{(np)}$, $A: p \times p$ and $\underset{\sim}{A}'\underset{\sim}{A} = \underset{\sim}{\Sigma} > 0$. Hence

$$G = (X-\varepsilon_n\mu')'(X-\varepsilon_n\mu') \overset{d}{=} R^2 A'U'UA \overset{d}{=} R^2 A'Y'YA/\mathrm{tr}\, Y'Y \quad ,$$

where $Y{:}n \times p$ with i.i.d. elements distributed according to $N(0,1)$. Since $Y'Y > 0$ with probability one and $P(R^2 > 0) = 1$, then $P(G > 0) = 1$. The property for W is proved similarly. Q.E.D.

In order to obtain the maximum likelihood estimators of μ and Σ we need the following lemma.

Lemma 2. Assume that $g(\cdot)$ is a decreasing and differentiable function such that $cg(x_1^2 + \cdots + x_N^2)$ is a density in $\mathbb{R}^N$, where c is a constant. Then the function

$$h(x) \equiv x^{N/2} g(x) \qquad \text{for} \qquad x \geq 0 \tag{2.5}$$

has a maximum at some finite $x_0 > 0$, and x_0 is a solution of

$$g'(x) + (N/2x)g(x) = 0 \quad . \tag{2.6}$$

Proof. Since $cg(x_1^2 + \cdots + x_N^2)$ is a density, we have

$$\infty > \int \cdots \int_{\mathbb{R}^N} g(x_1^2 + \cdots + x_N^2)dx_1 \cdots dx_N = \frac{\pi^{N/2}}{\Gamma(N/2)} \int_0^\infty x^{N/2-1} g(x)dx$$

using the transformation to spherical coordinates (Anderson (1958), Chapter 7, Problem 4, for example), and

$$2^{-N/2}(2x)^{N/2}g(2x) = x^{N/2}g(2x) \le x^{N/2-1}\int_x^{2x} g(t)dt \le \int_x^{2x} t^{N/2-1}g(t)dt \to 0$$

as $x \to \infty$, i.e., $h(x) = x^{N/2}g(x) \to 0$ as $x \to \infty$. By this fact and $h(0) = 0$, $h(x) \ge 0$ for $\forall x \ge 0$, the first assertion of the theorem follows. Now assume that $h(x)$ has its maximum at $x_0 > 0$; then

$$0 = h'(x_0) = x_0^{N/2}[(N/2x_0)g(x_0) + g'(x_0)] \quad ,$$

which completes the proof. Q.E.D.

From Lemma 2, when $N = np$, the function

$$f(\lambda) \equiv \lambda^{-np/2}g(p/\lambda) \tag{2.7}$$

arrives at its maximum at some finite λ_0. Throughout the paper we denote this λ_0 by $\lambda_{max}(g)$.

It is well-known that if $h'(x_0) = 0$ and $h''(x_0) < 0$ (if it exists), then $h(x)$ has a local maximum at x_0. If there exists a unique such x_0, then $h(x)$ has a unique maximum at x_0. In the proof of Lemma 2 we see that $h'(x_0) = 0$ if and only if x_0 is a solution of (2.6). Further if this x_0 satisfies the following inequality

$$g''(x_0) < (N(N+2)/4x_0^2)g(x_0) \quad ,$$

then $h''(x_0) < 0$.

Lemma 2 shows us that the equation (2.6) has at least one solution if $g(\cdot)$ satisfies the conditions of Lemma 2. For example, if $g(x) = \exp(-ax)$,

$a > 0$, then the equation (2.6) has a unique solution $x_0 = N/2a$.
If $g(x) = e^{-x}\exp(-e^{-x})$, the equation (2.6) becomes

$$e^{-x} = 1 - N/2x \quad .$$

It is easy to see that this equation has a unique solution $x_0(> N/2)$.
The following are some computations:

N	x_0
8	4.069529635
10	5.032816088
12	6.014691496
20	10.00045381
30	15.00000459
40	20.00000004

We can combine these two examples to obtain the following interesting example in which the equation (2.6) has two solutions. Taking $N = 6$, let

$$\begin{aligned} g(x) &= e^{-x}, && 0 \le x \le 6 , \\ &= e^{-x}\exp(-e^{-x}), && 6.014691496 \le x < \infty . \end{aligned}$$

and the values of $g(x)$ on interval $(6, 6.014691496)$ depend on a polynomial

$$p(x) = a_0x^3 + a_1x^2 + a_2x + a_3$$

such that

$$\begin{aligned} p(6) &= g(6), & P(6.014691496) &= g(6.014691496) , \\ p'(6) &= g'(6), & p'(6.014691496) &= g'(6.014691496) \quad . \end{aligned}$$

According to this requirement, we find

$$a_0 = 4.2951974, \quad a_1 = -77.4064525, \quad a_2 = 464.9942434, \quad a_3 = -931.0921074 ,$$

by calculation. It is easy to check that the $g(x)$ is decreasing and differentiable, and the corresponding equation (2.6) has two solutions: $x_{01} = 6$ and $x_{02} = 6.014691496$ (cf. the above first example with $a = 1$ and $N = 12$ and another example with $N = 12$). In this case we need to compare the values of $h(x)$ at x_{01} and x_{02}. We have

$$h(6) = 6^6 e^{-6} = 115.4686616$$

and

$$\begin{aligned} h(6.014691496) &= (6.014691496)^6 e^{-6.014691496} \exp(-e^{-6.014691496}) \\ &= 115.364451 \quad . \end{aligned}$$

Thus $h(x)$ arrives at its maximum at 6.

Now we come back to the maximum likelihood estimators of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$.

<u>Theorem 1</u>. Assume that $\underset{\sim}{X} \sim LEC_{n \times p}(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}, \phi)$ with $n > p$ and $\underset{\sim}{X}$ has a density (2.1). Further assume that $g(\cdot)$ is a decreasing and differentiable function. Then the maximum likelihood estimators of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ are $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{x}}$ and $\hat{\underset{\sim}{\Sigma}} = \lambda_{max}(g)\underset{\sim}{W}$.

<u>Proof</u>. It is easy to see that

$$\log L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}) = -(n/2) \log |\underset{\sim}{\Sigma}| + \log g[\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1} \underset{\sim}{W} + n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' \underset{\sim}{\Sigma}^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})] \quad .$$

The assumption that $g(\cdot)$ is decreasing and $\underset{\sim}{\Sigma} > 0$ shows us $\log L(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ arrives at its maximum at $\underset{\sim}{\mu} = \bar{\underset{\sim}{x}}$ and the concentrated likelihood is

$$(2.8) \qquad \log L(\bar{\underset{\sim}{x}},\underset{\sim}{\Sigma}) = -(n/2) \log |\underset{\sim}{\Sigma}| + \log g(\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{W}) \quad .$$

By Lemma 1, $\underset{\sim}{W} > 0$ with probability one. So there exists a nonsingular matrix $\underset{\sim}{C}$ with probability one such that $\underset{\sim}{C}\underset{\sim}{C}' = \underset{\sim}{W}$. Let $\tilde{\underset{\sim}{\Sigma}} = \underset{\sim}{C}^{-1}\underset{\sim}{\Sigma}\underset{\sim}{C}'^{-1}$. We then have

$$\log L(\bar{\underset{\sim}{x}},\underset{\sim}{\Sigma}) = -(n/2) \log |\tilde{\underset{\sim}{\Sigma}}| - (n/2) \log |\underset{\sim}{W}| + \log g(\mathrm{tr}\ \tilde{\underset{\sim}{\Sigma}}^{-1}) \quad .$$

There exists an orthogonal matrix $\underset{\sim}{\Gamma}$ such that

$$\underset{\sim}{\Gamma}'\tilde{\underset{\sim}{\Sigma}}\underset{\sim}{\Gamma} = \Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_p)$$

with $\lambda_1,\ldots,\lambda_p > 0$. Therefore

$$\begin{aligned} \log L(\bar{\underset{\sim}{x}},\underset{\sim}{\Sigma}) &= -(n/2) \sum_{i=1}^{p} \log \lambda_i + \log g(\sum_{i=1}^{p} 1/\lambda_i) - (n/2) \log |\underset{\sim}{W}| \\ &\equiv f(\lambda_1,\ldots,\lambda_p) - (n/2) \log |\underset{\sim}{W}| \quad , \end{aligned}$$

say. As $f(\lambda_1,\ldots,\lambda_2)$ is a symmetric function of $\lambda_1,\ldots,\lambda_p$, we have $\lambda_1 = \lambda_2 = \cdots = \lambda_p = \lambda$, say. The function $f(\lambda_1,\ldots,\lambda_p)$ reduces to

$$(2.9) \qquad f(\lambda) = -(np/2) \log \lambda + \log g(p/\lambda) \quad .$$

The theorem follows by Lemma 2. Q.E.D.

It is clear from $\log L(\mathbf{\mu},\mathbf{\Sigma})$ that $\bar{\mathbf{x}}$ and $\mathbf{W}$ are a sufficient set of statistics for $\mathbf{\mu}$ and $\tilde{\mathbf{\Sigma}}$.

From the proof of Theorem 1 we have

$$(2.10) \qquad \max_{\mathbf{\mu},\mathbf{\Sigma}>0} L(\mathbf{\mu},\mathbf{\Sigma}) = \lambda_{\max(g)}^{-np/2} g(p/\lambda_{\max(g)})|\mathbf{W}|^{-n/2} .$$

In the normal case

$$g(x) = (2\pi)^{-np/2} e^{-1/2x} ,$$

which satisfies the conditions of Theorem 1. As is well-known there exists a unique solution $\lambda_{\max(g)} = 1/n$, and $\hat{\mathbf{\Sigma}} = (1/n)\mathbf{W}$.

Now we want to obtain the distributions of $\bar{\mathbf{x}}$ and $\lambda_{\max(g)} \mathbf{W}$. If $\mathbf{X} \sim LEC_{n\times p}(\mathbf{0},\mathbf{\Sigma},\phi)$ with $P(\mathbf{X}=\mathbf{0})=0$, and $\mathbf{X} = \begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}$ with $\mathbf{X}_1 : m \times p$, the distribution of $\mathbf{X}_1'\mathbf{X}_1$ is denoted by $M_{p,2}(\mathbf{\Sigma};m/2;(n-m)/2;\phi)$. Anderson and Fang (1982b) have obtained the density of $M_{p,2}(\mathbf{\Sigma};m/2;(n-m)/2;\phi)$.

<u>Theorem 2</u>. Assume that $X \sim LEC_{n\times p}(\mathbf{\mu},\mathbf{\Sigma},\phi)$ with $n > p$ and $\mathbf{X}$ has a density (2.1). Then

(1) the joint density of $\bar{\mathbf{x}}$ and $\mathbf{W}$ is

$$(2.11) \qquad \frac{n^{p/2}\pi^{(n-1)p/2-p(p-1)/4}}{\prod_{j=1}^{p}\Gamma((n-j)/2)} |\mathbf{W}|^{(n-p)/2-1}|\mathbf{\Sigma}|^{-n/2} g(\mathrm{tr}\ \mathbf{\Sigma}^{-1}\mathbf{W}+n(\bar{\mathbf{x}}-\mathbf{\mu})'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}-\mathbf{\mu})) ;$$

(2) $\underset{\sim}{\bar{x}} \sim EC_p(\underset{\sim}{\mu},(1/n)\underset{\sim}{\Sigma},\phi)$ with $\phi \in \Phi_p \leftrightarrow R^* = Rb_{p/2,(n-1)p/2}$;

(3) $\underset{\sim}{W} \sim MG_{p,2}(\Sigma;(n-1)/2;\frac{1}{2};\phi)$, where $\phi \in \Phi_{(n-1)p} \leftrightarrow R^{**} = Rb_{(n-1)p/2,p/2}$ and the density of $\underset{\sim}{W}$ is

$$(2.12) \qquad \frac{2\pi^{np/2-p(p-1)/4}}{\Gamma(p/2) \prod_{j=1}^{p} \Gamma((n-j)/2)} |\underset{\sim}{\Sigma}|^{-n/2} |\underset{\sim}{W}|^{(n-p)/2-1} \int_0^\infty r^{p-1} g(r^2+\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{W})dr \quad .$$

<u>Proof</u>. Let $\underset{\sim}{B}$ be an $n \times n$ orthogonal matrix with the last row $(1/\sqrt{n},\ldots,1/\sqrt{n})$ and $\underset{\sim}{Y} = (y_{(1)},\ldots,\underset{\sim}{y}_{(n)})' = \underset{\sim}{B}\underset{\sim}{X}$. Then

$$y_{(n)} = \sqrt{n}\ \underset{\sim}{\bar{x}} \quad ,$$

$$\underset{\sim}{W} = \sum_{j=1}^{n-1} \underset{\sim}{y}_{(j)}\underset{\sim}{y}'_{(j)} \quad .$$

From Corollary 2 of Lemma 2 of Anderson and Fang (1982b) we have

$$\mathrm{vec}(\underset{\sim}{Y}) = \mathrm{vec}(\underset{\sim}{B}\underset{\sim}{X}) \sim EC_{np}(\underset{\sim}{\mu} \times (\underset{\sim}{B}\underset{\sim}{\varepsilon}_n),\underset{\sim}{\Sigma} \times (\underset{\sim}{B}\underset{\sim}{B}'),\phi) \quad .$$

Thus we have

$$\underset{\sim}{\mu} \otimes (\underset{\sim}{B}\underset{\sim}{\varepsilon}_n) = \underset{\sim}{\mu} \otimes (0,\ldots,0,1/\sqrt{n})'$$

and

$$\underset{\sim}{\Sigma} \otimes (\underset{\sim}{B}\underset{\sim}{B}') = \underset{\sim}{\Sigma} \otimes \underset{\sim}{I}_n \quad .$$

Using Lemma 13.3.1 of Anderson (1958) and $\underset{\sim}{y}_{(n)} = \sqrt{n}\bar{x}$, (2.11) follows.

Since $\underset{\sim}{W} = \underset{\sim}{X}'\underset{\sim}{D}\underset{\sim}{X}$ with $\underset{\sim}{D} = \underset{\sim}{I}_n - (1/n)\underset{\sim}{\varepsilon}_n\underset{\sim}{\varepsilon}_n'$, rk $\underset{\sim}{D} = n - 1$ and $\underset{\sim}{D}^2 = \underset{\sim}{D}$, $\underset{\sim}{W} \sim MG_{p,2}(\Sigma; (n-1)/2; \frac{1}{2}; \phi)$ from Theorem 9 of Anderson and Fang (1982b) and the density (2.12) follows from Section 5.1 (2)(B) of the same paper. The assertion (2) follows by Corollary 2 of Lemma 2 with $\underset{\sim}{B} = (1/n)\underset{\sim}{\varepsilon}_n'$ of Anderson and Fang (1982b). Q.E.D.

The next natural question is what are the unbiased estimators of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$? We need the following lemma:

Lemma 3. Assume $\underset{\sim}{x} \sim EC_n(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}, \phi)$ with $\phi \leftrightarrow R$ and $ER^2 < \infty$. Then

$$(2.13) \qquad E\underset{\sim}{x} = \underset{\sim}{\mu}, \quad \text{and} \quad E(\underset{\sim}{x}-\underset{\sim}{\mu})(\underset{\sim}{x}-\underset{\sim}{\mu})' = (ER^2/n)\underset{\sim}{\Sigma} \quad .$$

(cf. Cambanis, Huang and Simons (1981).)

Theorem 3. Assume that $\underset{\sim}{X} \sim EC_{n\times p}(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}, \phi)$ with $\phi \in \Phi_{np} \leftrightarrow R$ and $0 < ER^2 < \infty$. Then the unbiased estimators of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ are $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{x}}$ and

$$(2.14) \qquad \hat{\underset{\sim}{\Sigma}} = \frac{np}{(n-1)ER^2} \underset{\sim}{W} \quad ,$$

respectively.

Proof. From Theorem 2 and Lemma 3 the first assertion is obvious. Using the notation of the proof of Theorem 2, we have $\underset{\sim}{W} = \sum_{j=1}^{n-1} \underset{\sim}{y}_{(j)}\underset{\sim}{y}_{(j)}'$. If we can prove $E\underset{\sim}{y}_{(j)}\underset{\sim}{y}_{(j)}' = (ER^2/np)\underset{\sim}{\Sigma}$, $j = 1,\ldots,n-1$, then (2.14) follows. It is easy to show that $\underset{\sim}{y}_{(j)} \sim EC_p(\underset{\sim}{0}, \underset{\sim}{\Sigma}, \phi)$ with $\phi \in \Phi_p \leftrightarrow R^* = Rb_{p/2,(n-1)p/2}$,

$j = 1,\ldots,n-1$ (cf. Lemma 2, Anderson and Fang (1982b)). By Lemma 3 and the moments of Beta distribution

$$E\underset{\sim}{y}_{(j)}\underset{\sim}{y}'_{(j)} = (ER^{*2}/p)\underset{\sim}{\Sigma} = (ER^2/np)\underset{\sim}{\Sigma} \quad ,$$

which completes the proof. Q.E.D.

<u>Example 1</u>. (Multivariate t-distribution). Let $\underset{\sim}{X}$ be an $n \times p$ random matrix with i.i.d. rows distributed according to $N_p(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ and $s \sim \chi_\nu$ be independent of $\underset{\sim}{X}$. Let $\underset{\sim}{Y} = \sqrt{\nu}\underset{\sim}{X}/s$, then the rows of $\underset{\sim}{Y}$ and vec $\underset{\sim}{Y}'$ have the multivariate t-distributions. It is easy to verify the following facts (cf. Johnson and Kotz (1972) and Anderson and Fang (1982a)):

(1) The density of $\underset{\sim}{Y}$ is

$$\frac{\Gamma[(np+\nu)/2]\,|\underset{\sim}{\Sigma}|^{-n/2}}{\Gamma(\nu/2)(\pi\nu)^{np/2}}(1 + \nu^{-1}\,\mathrm{tr}\,\underset{\sim}{G}\underset{\sim}{\Sigma}^{-1})^{\;1/2(np+\nu)} \quad , \tag{2.15}$$

where

$$G = \sum_{j=1}^{n}(\underset{\sim}{y}_{(j)}-\underset{\sim}{\mu})(\underset{\sim}{y}_{(j)}-\underset{\sim}{\mu})' \quad .$$

(2) $\underset{\sim}{Y} \sim LEC_{n\times p}(\underset{\sim}{\mu},\underset{\sim}{\Sigma},\phi)$ with

$$g(x) = c(1+\nu^{-1}x)^{-1/2(np+\nu)} \quad , \tag{2.16}$$

where

$$c = \Gamma((np+\nu)/2)/\Gamma(\nu/2)/(\pi\nu)^{np/2} \;.$$

Therefore, $g(x)$ satisfies the conditions of Theorem 1 and $\lambda_{max}(g) = 1/n$.

(3) The maximum likelihood estimator of $\underset{\sim}{\Sigma}$ is $\underset{\sim}{W}/n$, where $\underset{\sim}{W}$ is defined by (2.2) with $\underset{\sim}{x}$ substituted for $\underset{\sim}{y}$.

(4) The unbiased estimator of $\underset{\sim}{\Sigma}$ is $((\nu-2)/(n-1)\nu)\ \underset{\sim}{W}$.

3. Likelihood ratio criteria.

In this section we study problems of testing hypotheses about mean vectors and covariance matrices for the elliptically contoured distributions. The likelihood criteria are obtained by a unified technique. We find that most of these likelihood ratio criteria are independent of the specific form of the density in the class of MECD.

3.1 Testing lack of correlation between sets of variates.

Assume that $\underset{\sim}{X} \sim LEC_{n\times p}(\underset{\sim}{\mu},\underset{\sim}{\Sigma},\phi)$ with $n > p$ and $\underset{\sim}{X}$ has a density (2.1) where $g(x)$ is a decreasing and differentiable function. Partition $\underset{\sim}{\mu}$, $\underset{\sim}{\Sigma}$, and $\underset{\sim}{W}$ as follows:

$$(4.1)\quad \underset{\sim}{\mu} = \begin{pmatrix} \underset{\sim}{\mu}^{(1)} \\ \vdots \\ \underset{\sim}{\mu}^{(q)} \end{pmatrix}, \quad \underset{\sim}{\Sigma} = \begin{pmatrix} \underset{\sim}{\Sigma}_{11} & \cdots & \underset{\sim}{\Sigma}_{1q} \\ \vdots & & \vdots \\ \underset{\sim}{\Sigma}_{q1} & \cdots & \underset{\sim}{\Sigma}_{qq} \end{pmatrix}, \quad \underset{\sim}{W} = \begin{pmatrix} \underset{\sim}{W}_{11} & \cdots & \underset{\sim}{W}_{1q} \\ \vdots & & \vdots \\ \underset{\sim}{W}_{q1} & & \underset{\sim}{W}_{qq} \end{pmatrix},$$

where $\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)}$ have $p_1,\ldots,p_q$ components, respectively, and $\underset{\sim}{\Sigma}_{ij}$ and $\underset{\sim}{W}_{ij}$ are $p_i \times p_j$ matrices, $i,j = 1,\ldots,q$. Let

$$(3.2) \qquad \underset{\sim}{\Sigma}_0 = \begin{pmatrix} \underset{\sim}{\Sigma}_{11} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ 0 & \underset{\sim}{\Sigma}_{22} & \cdots & \underset{\sim}{0} \\ \vdots & & & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{\Sigma}_{qq} \end{pmatrix} ;$$

then $\underset{\sim}{\Sigma}_0 > 0$ if and only if $\underset{\sim}{\Sigma}_{11} > 0, \ldots, \underset{\sim}{\Sigma}_{qq} > 0$. Testing lack of correlation between sets of variates is equivalent to testing the hypothesis $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_0$.

<u>Lemma 4</u>. Under the assumptions of Theorem 1

$$(3.3) \qquad \max_{\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_0 > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_0) = \lambda_{max}(g)^{-np/2} \left(\prod_{j=1}^{q} |\underset{\sim}{W}_{jj}|^{-n/2} \right) g(p/\lambda_{max}(g)) \quad .$$

<u>Proof</u>. To show the pattern we consider only the case of $q = 2$. By a method similar to that used in the proof of Theorem 1, we have

$$\max_{\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_0 > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_0) = \max_{\underset{\sim}{\Sigma}_0 > 0} L(\underset{\sim}{\bar{x}}, \underset{\sim}{\Sigma}_0)$$

$$= \max_{\substack{\lambda_j^{(i)} > 0 \\ j = 1, \ldots, p_i \\ i = 1,2}} \{ -(n/2) \prod_{j=1}^{p_1} \log \lambda_j^{(1)} - (n/2) \prod_{j=1}^{p_2} \log \lambda_j^{(2)} + \log g(\prod_{j=1}^{p_1} \lambda_j^{(1)^{-1}}$$

$$+ \prod_{j=1}^{p_2} \lambda_j^{(2)^{-1}}) - (n/2) \times (\log |\underset{\sim}{W}_{11}| + \log |\underset{\sim}{W}_{22}|) \} \quad ,$$

where $\lambda_1^{(1)}, \ldots, \lambda_{p_1}^{(1)}$ and $\lambda_1^{(2)}, \ldots, \lambda_{p_2}^{(2)}$ are the eigenvalues of $\underset{\sim}{\tilde{\Sigma}}_{11}$ and

$\tilde{\Sigma}_{22}$, respectively, and $\tilde{\Sigma}_{11} = C_1^{-1}\Sigma_{11}C_1'^{-1}$, $\tilde{\Sigma}_{22} = C_2^{-1}\Sigma_{22}C_2'^{-1}$, $C_1C_1' = W_{11}$ and $C_2C_2' = W_{22}$. As the above function in the braces is a symmetric function of $\lambda_1^{(1)},\ldots,\lambda_{p_1}^{(1)}$, $\lambda_1^{(2)},\ldots,\lambda_{p_2}^{(2)}$, these $\{\lambda_j^{(i)}\}$ must be equal. We have

$$\max_{\mu,\Sigma_0>0} L(\mu,\Sigma_0) = \max_{\lambda>0} \{-(np/2)\log\lambda + \log g(p/\lambda) - (n/2)(\log|W_{11}| + \log|W_{22}|)\}$$

$$= \log\{\lambda_{max}(g)^{-np/2} g(p/\lambda_{max}(g))|W_{11}|^{-n/2}|W_{22}|^{-n/2}\} \quad ,$$

which completes the proof. Q.E.D.

<u>Theorem 4</u>. Under the assumptions of Theorem 1, the likelihood ratio criterion for testing

(3.4) $$H:\Sigma_{ij} = 0, \quad i \neq j, \quad i,j = 1,\ldots,q \quad ,$$

is

(3.5) $$\tau = \frac{\max\limits_{\mu,\Sigma_0>0} L(\mu,\Sigma_0)}{\max\limits_{\mu,\Sigma>0} L(\mu\ \Sigma)} = \left(\frac{|W|}{\prod_{j=1}^{q}|W_{jj}|}\right)^{n/2}$$

and

(3.6) $$\tau^{2/n} \overset{d}{=} \prod_{j=2}^{q} v_j \quad ,$$

where $v_2,\ldots,q_q$ are independent and $v_j \sim U_{p_j,\bar{p}_j,n-\bar{p}_j}$ with $\bar{p}_j = \sum_{i=1}^{j-1} p_i$.

Proof. The criterion (3.5) is from (2.10) and (3.3). By (2.4), $\underset{\sim}{X} = \underset{\sim}{\varepsilon}_n \underset{\sim}{\mu}' + \underset{\sim}{R}\underset{\sim}{U}\underset{\sim}{A}$. Partition $\underset{\sim}{A}$ into $\underset{\sim}{A} = (\underset{\sim}{A}_1,\ldots,\underset{\sim}{A}_q)$ with $\underset{\sim}{A}_j : p \times p_j$, $j = 1,\ldots,q$. Then

$$\underset{\sim}{W} \overset{d}{=} R^2 \underset{\sim}{A}'\underset{\sim}{U}'\underset{\sim}{D}\underset{\sim}{U}\underset{\sim}{A} = R^2 \underset{\sim}{A}'\underset{\sim}{Y}'\underset{\sim}{D}\underset{\sim}{Y}\underset{\sim}{A}/\mathrm{tr}\ \underset{\sim}{Y}'\underset{\sim}{Y}$$

$$\underset{\sim}{W}_{jj} = R^2 \underset{\sim}{A}_j'\underset{\sim}{U}'\underset{\sim}{D}\underset{\sim}{U}\underset{\sim}{A}_j = R^2 \underset{\sim}{A}_j'\underset{\sim}{Y}'\underset{\sim}{D}\underset{\sim}{Y}\underset{\sim}{A}_j/\mathrm{tr}\ \underset{\sim}{Y}'\underset{\sim}{Y}, \qquad j = 1,\ldots,q\ ,$$

where $\underset{\sim}{D},\underset{\sim}{U}$ and $\underset{\sim}{Y}$ have the same meaning as before. Thus

$$\tau^{2/n} \overset{d}{=} |A'Y'DYA| / \prod_{j=1}^{q} |A_j'Y'DYA_j| \quad , \tag{3.7}$$

and the distribution of $\tau^{2/n}$ is independent of any specific form of the density in the class of MECD. In particular we can consider $\underset{\sim}{x}_{(j)} \sim N_p(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ independently, $j = 1,\ldots,n$. Hence (3.6) follows by Anderson (1958), Chapter 8. Q.E.D.

When $q = 2$, $p_1 = 1$ and $p_2 = p - 1$, testing the null hypothesis $\underset{\sim}{\Sigma}_{12} = \underset{\sim}{0}$ is equivalent to testing $\bar{R}^2 = 0$, where $\bar{R}$ is the multiple correlation coefficient. (cf. Theorem 5 of Anderson and Fang (1982b).)

3.2 Testing the hypothesis that a mean vector is equal to a given vector.

Consider the hypothesis $H:\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$. The likelihood criterion is

$$\tau = \max_{\underset{\sim}{\Sigma} > 0} L(\underset{\sim}{\mu}_0,\underset{\sim}{\Sigma}) / \max_{\underset{\sim}{\mu},\underset{\sim}{\Sigma} > 0} L(\underset{\sim}{\mu},\underset{\sim}{\Sigma}) \quad .$$

By a method similar to that used before we find

$$\max_{\underset{\sim}{\Sigma} > 0} L(\underset{\sim}{\mu}_0, \underset{\sim}{\Sigma}) = \lambda_{\max}(g)^{-np/2} g(p/\lambda_{\max}(g)) |\underset{\sim}{W}_0|^{-n/2} ,$$

where

$$\underset{\sim}{W}_0 = \sum_{j=1}^{n} (\underset{\sim}{x}_{(j)} - \underset{\sim}{\mu}_0)(\underset{\sim}{x}_{(j)} - \underset{\sim}{\mu}_0)' \quad .$$

Thus

$$\tau^{2/n} = \frac{|\underset{\sim}{W}|}{|\underset{\sim}{W}_0|} = \frac{|\underset{\sim}{W}|}{|\underset{\sim}{W} + n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)'|} = \frac{1}{1 + T^2/(n-1)} \quad ,$$

where

$$T^2 = n(n-1)(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)' \underset{\sim}{W}^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0) \quad ,$$

which is the same as in the normal case. Also, if the null hypothesis is true, the distribution of T^2 is the same as in the normal case. (cf. Theorem 6 of Anderson and Fang (1982b).)

3.3 Testing the hypothesis of equality of covariance matrices.

Now we consider the case that the sample are from q populations with mean vectors $\underset{\sim}{\mu}^{(1)}, \ldots, \underset{\sim}{\mu}^{(q)}$ and covariance matrices $\underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_q$, respectively. Assume that the joint distribution of the samples is a multivariate elliptically contoured distribution, i.e., the data matrix $\underset{\sim}{X} \sim MLE_{n \times p}(\underset{\sim}{M}; \underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_1, \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_2, \ldots, \underset{\sim}{\Sigma}_q, \ldots, \underset{\sim}{\Sigma}_q; \phi)$ with n_1 $\underset{\sim}{\Sigma}_1'$s n_2 $\underset{\sim}{\Sigma}_2's, \ldots, n_q$ $\underset{\sim}{\Sigma}_q'$s, $n_1 > p, \ldots, n_q > p$, $n_1 + n_2 + \cdots + n_q = n$ and the rows of

$\underset{\sim}{M}$ are n_1 $\underset{\sim}{\mu}^{(1)}$'s, n_2 $\underset{\sim}{\mu}^{(2)}$'s,...,$n_q$ $\underset{\sim}{\mu}^{(q)}$'s, successively. From (1.4) the density of $\underset{\sim}{X}$ is

$$(3.8)\qquad L(\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)};\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_q) = \prod_{i=1}^{q} |\underset{\sim}{\Sigma}_i|^{-n_i/2} g\left(\sum_{i=1}^{q} \operatorname{tr} \underset{\sim}{\Sigma}_i^{-1}\underset{\sim}{G}_i\right) ,$$

where

$$\underset{\sim}{G}_i = \sum_{j=\bar{n}_{i-1}+1}^{\bar{n}_i} (\underset{\sim}{x}_{(j)}-\underset{\sim}{\mu}^{(i)})(\underset{\sim}{x}_{(j)}-\underset{\sim}{\mu}^{(i)})'$$

with $n_0 = 0$, $\bar{n}_i = n_1 + \cdots + n_i$.

We wish to test the hypothesis

$$(3.9)\qquad H_1:\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \cdots = \underset{\sim}{\Sigma}_q \ .$$

<u>Lemma 5</u>. Under the assumptions of Theorem 1

$$(3.10)\qquad \max_{\underset{\sim}{\mu}_1^{(1)},\ldots,\underset{\sim}{\mu}^{(q)},\underset{\sim}{\Sigma}_1>0,\ldots,\underset{\sim}{\Sigma}_q>0} L(\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)};\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_q)$$

$$= \lambda_{\max}(g)^{-np/2}\left[\prod_{i=1}^{q} \left(\frac{n_i}{n}\right)^{pn_i/2} |\underset{\sim}{W}_i|^{-n_i/2}\right] g(p/\lambda_{\max}(g)) ,$$

where

$$\underset{\sim}{W}_i = \sum_{j=\bar{n}_{i-1}+1}^{\bar{n}_i} (\underset{\sim}{x}_{(j)}-\underset{\sim}{\bar{x}}^{(i)})(\underset{\sim}{x}_{(j)}-\underset{\sim}{\bar{x}}^{(i)})' ,$$

$$i = 1,\ldots,q \ ,$$

$$\underset{\sim}{\bar{x}}^{(i)} = (1/n_i)\sum_{j=\bar{n}_{i-1}+1}^{\bar{n}_i} \underset{\sim}{x}_{(j)} \ .$$

Proof. By a method similar to that used in the proof of Theorem 1, we have

$$\max_{\substack{\underset{\sim}{\mu}^{(i)},\underset{\sim}{\Sigma}_i > 0\\ i=1,\ldots,q}} \log L(\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)};\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_q)$$

$$= \max_{\underset{\sim}{\Sigma}_1 > 0,\ldots,\underset{\sim}{\Sigma}_q > 0} \log L(\bar{\underset{\sim}{x}}^{(1)},\ldots,\bar{\underset{\sim}{x}}^{(q)};\underset{\sim}{\Sigma}_1,\ldots,\underset{\sim}{\Sigma}_q)$$

$$= \max_{\substack{\lambda_j^{(i)} > 0, j=1,\ldots,n_i\\ i=1,\ldots,q}} \{-(1/2)\sum_{i=1}^{q}\sum_{j=1}^{p} n_i \log \lambda_j^{(i)}$$

$$+ \log g(\sum_{i=1}^{q}\sum_{j=1}^{p}\lambda_j^{(i)-1}) - (1/2)\log\sum_{i=1}^{q} n_i \log|\underset{\sim}{W}_i|\}$$

where $\lambda_1^{(i)},\ldots,\lambda_p^{(i)}$ are the eigenvalues of $\tilde{\underset{\sim}{\Sigma}}_i = \underset{\sim}{C}_i^{-1}\underset{\sim}{\Sigma}_i\underset{\sim}{C}_i'^{-1}$ and $\underset{\sim}{C}_i\underset{\sim}{C}_i' = \underset{\sim}{W}_i$. Because $n_i > p$, $\underset{\sim}{W}_i > 0$ with probability one by Lemma 1 ($i = 1,\ldots,q$). The function in the above braces is a symmetric function of $\lambda_1^{(i)},\ldots,\lambda_p^{(i)}$ for $i = 1,\ldots,q$. Therefore $\lambda_1^{(i)} = \cdots = \lambda_p^{(i)} \equiv \lambda^{(i)}$ and the quantity in the braces becomes

$$(3.11)\quad -(p/2)\sum_{i=1}^{q} n_i \log\lambda^{(i)} + \log g(p\sum_{i=1}^{q}\lambda^{(i)-1}) - (1/2)\sum_{i=1}^{q} n_i \log|\underset{\sim}{W}_i| \equiv L^* \quad ,$$

say. Let $\partial L^*/\partial\lambda^{(i)} = 0$, $i = 1,\ldots,q$. We have

$$\lambda^{(i)} = -\frac{2}{n_i}\,\frac{g'(p\sum_1^q \lambda^{(i)-1})}{g(p\sum_1^q \lambda^{(i)-1})}\,, \qquad i = 1,\ldots,q \quad ,$$

and $\lambda^{(i)}/\lambda^{(1)} = n_1/n_i$, $i = 1,\ldots,q$. Let $\lambda = \lambda^{(1)}n_1/n$, then $\lambda^{(i)} = n\lambda/n_i$, $i = 1,\ldots,q$, and (3.11) becomes

$$-(np/2)\log\lambda + \log g(p/\lambda) - (pn/2)\log n - (p/2)\sum_{i=1}^{n} n_i \log n_i - (1/2)\sum_{i=1}^{q} n_i \log|\underset{\sim}{W}_i|$$

which completes the proof. Q.E.D.

Similarly, we find

$$(3.12)\qquad \max_{\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)},\underset{\sim}{\Sigma}>0} L(\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)};\underset{\sim}{\Sigma},\ldots,\underset{\sim}{\Sigma})$$

$$= \lambda_{max}(g)^{-np/2} g(p/\lambda_{max}(g))|\underset{\sim}{W}_1 + \cdots + \underset{\sim}{W}_q|^{-n/2} .$$

Thus we obtain the following theorem:

Theorem 5. Under the above assumptions the likelihood ratio criterion for testing (3.9) is

$$(3.13)\qquad \tau_1 = \prod_{i=1}^{q} \left[\left(\frac{|\underset{\sim}{W}_i|}{|\sum_{j=1}^{q}\underset{\sim}{W}_j|}\right)^{n_i/2}\left(\frac{n}{n_i}\right)^{pn_i/2}\right]$$

which has the same distribution as in the normal case.

The distribution of τ_1 was discussed by Anderson (1958), Chapter 10. A sufficient set of statistics is $\bar{x}^{(1)},\ldots,\bar{x}^{(q)}$, $\underset{\sim}{W}_1,\ldots,\underset{\sim}{W}_q$.

3.4 Testing equality of several means.

Assuming $\underset{\sim}{\Sigma}_1 = \cdots = \underset{\sim}{\Sigma}_q = \underset{\sim}{\Sigma}$, say, we want to test the hypothesis

$$H_2 : \underset{\sim}{\mu}^{(1)} = \cdots = \underset{\sim}{\mu}^{(q)} \quad . \tag{3.14}$$

The corresponding likelihood ratio criterion is

$$\tau_2 = \frac{\max\limits_{\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)},\underset{\sim}{\Sigma}>0} L(\underset{\sim}{\mu}^{(1)},\ldots,\underset{\sim}{\mu}^{(q)};\underset{\sim}{\Sigma},\ldots,\underset{\sim}{\Sigma})}{\max\limits_{\underset{\sim}{\mu},\underset{\sim}{\Sigma}>0} L(\underset{\sim}{\mu},\ldots,\underset{\sim}{\mu};\underset{\sim}{\Sigma},\ldots,\underset{\sim}{\Sigma})} = \frac{|\underset{\sim}{W}|^{n/2}}{|\underset{\sim}{W}_1 + \cdots + \underset{\sim}{W}_q|^{n/2}} \tag{4.15}$$

where $\underset{\sim}{W}$ is defined by (2.2).

It is well-known that the hypothesis H_2 can be expressed as a linear hypothesis which will be discussed in the next section.

3.5 Testing equality of several means and covariance matrices.

We want to test

$$H : \underset{\sim}{\mu}^{(1)} = \cdots = \underset{\sim}{\mu}^{(q)} \quad \text{and} \quad \underset{\sim}{\Sigma}_1 = \cdots = \underset{\sim}{\Sigma}_q \quad . \tag{3.16}$$

From Lemma 10.3.1 of Anderson (1958), the likelihood ratio criterion for the hypothesis H is the product of the likelihood ratio criteria for H_1 and H_2, i.e.

$$\tau = \tau_1 \tau_2 = \prod_{i=1}^{q} \left(\frac{|\underset{\sim}{W}_i|}{|\underset{\sim}{W}|}\right)^{n_i/2} \left(\frac{n}{n_i}\right)^{pn_i/2} , \tag{3.17}$$

which has the same distribution as in the normal case.

3.6 Testing the hypothesis that a covariance matrix is proportional to a given matrix.

Now we come back to the case of $\underset{\sim}{X} \sim LEX_{n \times p}(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}, \phi)$ and want to test

$$H: \underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{\Sigma}_0, \quad \underset{\sim}{\Sigma}_0 > 0 , \tag{3.18}$$

where $\underset{\sim}{\Sigma}_0$ is given. First, we assume $\underset{\sim}{\Sigma}_0 = \underset{\sim}{I}_p$ and decompose the hypothesis H into $H_1: \underset{\sim}{\Sigma}$ is diagonal and H_2: the diagonal elements of $\underset{\sim}{\Sigma}$ are equal, given $\underset{\sim}{\Sigma}$ is diagonal. Let $\underset{\sim}{\Sigma}_1 = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$ and $\underset{\sim}{\Sigma}_2 = \mathrm{diag}(\sigma^2, \ldots, \sigma^2)$. It follows from Lemma 4 that

$$\max_{\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_1 > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_1) = \lambda_{max}(g)^{-np/2} g(p/\lambda_{max}(g)) \left(\prod_{j=1}^{p} w_{jj}\right)^{-n/2} , \tag{3.19}$$

where w_{jj} is the j-th diagonal element of $\underset{\sim}{W}$. Similarly

$$\max_{\underset{\sim}{\mu}, \sigma > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_2) = \lambda_{max}(g)^{-np/2} g(p/\lambda_{max}(g)) (\mathrm{tr}\, \underset{\sim}{W}/p)^{-np/2} . \tag{3.20}$$

Then we have

$$\tau_1 = \frac{\max\limits_{\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_1 > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_1)}{\max\limits_{\underset{\sim}{\mu}, \underset{\sim}{\Sigma} > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})} = \frac{|\underset{\sim}{W}|^{n/2}}{\prod_{j=1}^{p} w_{jj}^{n/2}} , \tag{3.21}$$

(3.22) $$\tau_2 = \frac{\max_{\underset{\sim}{\mu},\sigma > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_2)}{\max_{\underset{\sim}{\mu},\underset{\sim}{\Sigma}_1 > 0} L(\underset{\sim}{\mu}, \underset{\sim}{\Sigma}_1)} = \frac{\prod_{j=1}^{p} w_{jj}^{n/2}}{(\mathrm{tr}\ \underset{\sim}{W}/p)^{np/2}} ,$$

and

(3.23) $$\tau = \frac{|\underset{\sim}{W}|^{n/2}}{(\mathrm{tr}\ W/p)^{np/2}} .$$

When $\underset{\sim}{\Sigma}_0 \neq \underset{\sim}{I}_p$, the corresponding criterion is

(3.24) $$\tau = \frac{|\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{W}|^{n/2}}{(\mathrm{tr}\ \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{W}/p)^{np/2}}$$

which has the same distribution as in the normal case. (cf. Anderson (1958), Chapter 10).

3.7 Testing the hypothesis that a covariance matrix is equal to a given matrix.

We wish to test $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_0 > 0$, where $\underset{\sim}{\Sigma}_0$ is given. The corresponding likelihood ratio criterion is

(3.25) $$\tau = \lambda_{max}(g)^{np/2} g^{-1}(p/\lambda_{max}(g)) g(\mathrm{tr}\ \underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{W}) |\underset{\sim}{W} \underset{\sim}{\Sigma}_0^{-1}|^{n/2} .$$

Note that the distributions of τ are not the same in the class of MECD and depend on ϕ. Similarly, we find that the likelihood ratio criterion for testing $H: \underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_0$ and $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ is

$$(3.26)\quad \tau = \lambda_{max}(g)^{np/2} g^{-1}(p/\lambda_{max}(g))\, g(\mathrm{tr}\ \underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{W} + n(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu}_0)'\underset{\sim}{\Sigma}_0^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu}_0))\,|\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{W}|^{n/2} .$$

4. Testing the general linear hypothesis.

Consider the following multiple regression model:

$$(4.1)\quad \begin{cases} \underset{\sim}{Y}_{n\times p} = \underset{\sim}{X}_{n\times q}\underset{\sim}{B}_{q\times p} + \underset{\sim}{E}_{n\times p},\ p+q < n,\quad \mathrm{rk}\ \underset{\sim}{X} = q , \\ \underset{\sim}{E} \sim LEC_{n\times p}(\underset{\sim}{0},\underset{\sim}{\Sigma},\phi) \ \text{ with a density } |\underset{\sim}{\Sigma}|^{-n/2} g(\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{G}) \ \text{ and } \ \underset{\sim}{G} = \underset{\sim}{E}'\underset{\sim}{E} , \\ \text{where } \ g(\cdot) \ \text{ is a decreasing and differentiable function .} \end{cases}$$

We want to find the maximum likelihood estimates of $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}$. As

$$L(\underset{\sim}{B},\underset{\sim}{\Sigma}) = |\underset{\sim}{\Sigma}|^{-n/2} g(\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{G})$$

$$= |\underset{\sim}{\Sigma}|^{-n/2} g(\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{Y}-\underset{\sim}{X}\hat{\underset{\sim}{B}})'(\underset{\sim}{Y}-\underset{\sim}{X}\hat{\underset{\sim}{B}}) + \mathrm{tr}(\hat{\underset{\sim}{B}}-\underset{\sim}{B})\underset{\sim}{X}\Sigma^{-1}\underset{\sim}{X}'(\hat{\underset{\sim}{B}}-\underset{\sim}{B})') ,$$

where

$$(4.2)\qquad \hat{\underset{\sim}{B}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y} ,$$

the MLE of $\underset{\sim}{B}$ is the least squares estimate $\hat{\underset{\sim}{B}}$ (cf. Sec. 7, Anderson and Fang (1982b)). Maximizing

$$L(\hat{\underset{\sim}{B}},\underset{\sim}{\Sigma}) = |\underset{\sim}{\Sigma}|^{-n/2} g(\mathrm{tr}\ \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{Y}-\underset{\sim}{X}\hat{\underset{\sim}{B}})'(\underset{\sim}{Y}-\underset{\sim}{X}\hat{\underset{\sim}{B}}))$$

with respect to $\underset{\sim}{\Sigma}$ gives the maximum likelihood estimator

(4.3) $$\hat{\Sigma} = \lambda_{max}(g)(Y-X\hat{B})'(Y-X\hat{B})$$

and

(4.4) $$L(\hat{B},\hat{\Sigma}) = \max_{B,\Sigma>0} L(B,\Sigma) = \lambda_{max}(g)^{-np/2}|(Y-XB)'(Y-XB)|^{-n/2} g(p/\lambda_{max}(g)) \quad ,$$

where $\lambda_{max}(g)$ has the same meaning before.

Similarly, the unbiased estimators of B and Σ are $\hat{B}$ and

$$\hat{\Sigma} = \frac{np}{(n-q)ER^2}(Y-X\hat{B})'(Y-X\hat{B})$$

respectively, if the second moments of components of E exist. A sufficient set of statistics is B and Σ.

We want to test the following linear hypothesis:

(4.5) $$H{:}HB = C, \quad H{:}t \times q, \quad C{:}t \times p \quad \text{and} \quad \mathrm{rk}\, H = t < p \quad .$$

It is easy to show that

(4.6) $$\max_{HB=C,\Sigma>0} L(B,\Sigma) = \lambda_{max}(g)^{-np/2} g(p/\lambda_{max}(g))|S + T|^{-n/2} \quad ,$$

where

$$S = (Y-X\hat{B})'(Y-X\hat{B})$$

and

$$\underset{\sim}{T} = (\underset{\sim}{H}\hat{\underset{\sim}{B}}-\underset{\sim}{C})'(\underset{\sim}{H}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{H}')^{-1}(\underset{\sim}{H}\hat{\underset{\sim}{B}}-\underset{\sim}{C}) \quad .$$

Thus the likelihood ratio criterion for H is

$$\tau = \frac{\max\limits_{\underset{\sim}{B},\underset{\sim}{\Sigma} > 0} L(\underset{\sim}{B},\underset{\sim}{\Sigma})}{\max\limits_{\underset{\sim}{H}\underset{\sim}{B} = \underset{\sim}{C},\underset{\sim}{\Sigma} > 0} L(\underset{\sim}{B},\underset{\sim}{\Sigma})} = \frac{|S|^{-n/2}}{|S+T|^{-n/2}} \tag{4.7}$$

which has the same distribution as in the normal case (cf. Theorem 12 of Anderson and Fang (1982b).)

We summarize the above results as follows:

Theorem 6. Under the model (4.1) the maximum likelihood estimators of $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}$ are (4.2) and (4.3), respectively. The likelihood ratio statistic for testing the hypothesis (4.5) is (4.7) and the distribution of $\tau^{-2/n}$ is $U_{p,t,n-q}$ when the null hypothesis is true.

Acknowledgment. The second author was supported by Academia Sinica.

REFERENCES

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York.

Anderson, T. W., and Fang, K. T. (1982a), Distributions of quadratic forms and Cochran's Theorem for elliptically contoured distributions and their applications, Technical Report No. 53, ONR Contract N00014-75-C-0442, Department of Statistics, Stanford University, Stanford, California.

Anderson, T. W., and Fang, K. T. (1982b), On the thoery of multivariate elliptically contoured distributions and their applications, Technical Report No. 54, ONR Contract N00014-75-C-0442, Department of Statistics, Stanford University, Stanford, California.

Cambanis, S., Huang, S., and Simons, G. (1981), On the theory of elliptically contoured distributions, Journal of Multivariate Analysis, 11, 368-385.

Chmielewski, M. A. (1980), Invariant scale matrix hypothesis tests under elliptical symmetry, Journal of Multivariate Analysis, 10, 343-350.

Dawid, A. P. (1977), Spherical matrix distributions and a multivariate model, Journal of the Royal Statistical Society, B, 39, 254-261.

Dawid, A. P. (1978), Extendibility of spherical matrix distributions, Journal of Multivariate Analysis, 8, 559-566.

Johnson, N. J., and Kotz, S. (1972), Distributions in Statistics: Continuous Multivariate Distributions, John Wiley and Sons, New York.

Schoenberg, I. J. (1938), Metric spaces and completely monotone functions, Annals of Mathematics, 39, 811-841.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. W. Anderson and Kai-Tai Fang		8. CONTRACT OR GRANT NUMBER(s) DAAG29-82-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, Sequoia Hall Stanford University Stanford, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-19065-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE September 1982 13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Likelihood ratio criteria, maximum likelihood estimators, multiple regression model, multivariate elliptically contoured distributions.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we generalize the theory of maximum likelihood estimation and likelihood ratio criteria from normal distributions to multivariate elliptically contoured distributions for which the vector observations are not necessarily either normal or independent. We find that many usual likelihood ratio criteria and their null distributions are the same in the class of multivariate elliptically contoured distributions.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)