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

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
MAXIMUM LIKELIHOOD ESTIMATORS AND LIKELIHOOD RATIO CRITERIA
FOR MULTIVARIATE ELLIPTICALLY CONTOUNED 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

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
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.
1. Introduction.

If the characteristic function (c.f.) of an $n$-dimensional random vector $\mathbf{x}$ has the form $\exp(\mathbf{it'}\mu)\phi(t'\Sigma t)$, where $\mu: n \times 1$, $\Sigma: n \times n$, $\text{rk} \Sigma = k$, $\Sigma \succeq 0$, and $\phi \in \Phi_k = \{\phi | \phi(t)\}$ 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 $\mathbf{x}$ is distributed according to an elliptically contoured distribution with parameters $\mu$, $\Sigma$ and $\phi$, and write $\mathbf{x} \sim EC_n(\mu, \Sigma, \phi)$.

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

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

$$
\mathbf{X} = (x_{ij}) = (x_1, \ldots, x_p) = \begin{bmatrix}
\begin{array}{c}
x'_1 \\
\vdots \\
x'_n
\end{array}
\end{bmatrix}, \quad \mathbf{x} = \text{vec } \mathbf{X}',
$$

$$
\mathbf{M} = (\mu_{ij}) = (\mu_1, \ldots, \mu_p) = \begin{bmatrix}
\begin{array}{c}
\mu'_1 \\
\vdots \\
\mu'_n
\end{array}
\end{bmatrix}, \quad \mathbf{\mu} = \text{vec } \mathbf{M}',
$$

$$
\mathbf{T} = (t_{ij}) = (t_1, \ldots, t_p) = \begin{bmatrix}
\begin{array}{c}
t'_1 \\
\vdots \\
t'_n
\end{array}
\end{bmatrix}, \quad \mathbf{t} = \text{vec } \mathbf{T}'.
$$

Here $\mathbf{x} = \text{vec } \mathbf{X}' = (x'_1, \ldots, x'_n)'$ with the corresponding meanings for $\mathbf{\mu}$ and $\mathbf{t}$. 

If the c.f. of a random matrix \( \mathbf{X} \) has the form

\[
\exp(i \sum_{j=1}^{n} t_j^j \sum_{j=1}^{n} \phi(t_j^j) \epsilon_{j/1} \epsilon_{j/2} \epsilon_{j/3} \epsilon_{j/4} \epsilon_{j/5} \epsilon_{j/6}),
\]

with \( \epsilon_{1}, \ldots, \epsilon_{n} \geq 0 \), we say that \( \mathbf{X} \) is distributed according to a multivariate (rows) elliptically contoured distribution (MECD) and write

\( \mathbf{X} \sim \text{MEC}_{n \times p}(\mathbf{M}; \epsilon_{1}, \ldots, \epsilon_{n}; \phi). \) In this paper we consider only the subclass of MECD in which the function \( \phi \) satisfies

\[
\phi(t_1, \ldots, t_n) = \phi(t_1 + \cdots + t_n); 
\]

we continue to denote MECD in this subclass by \( \text{MEC}_{n \times p}(\mathbf{M}; \epsilon_{1}, \ldots, \epsilon_{n}; \phi). \)

Let \( \mathbf{u} \sim \text{q} \) 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

\[
\phi(u) = \int_0^\infty \Omega_m(ur^2)dF(r)
\]

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 \( \phi \) 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 \( \phi \) 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 \( x \) and \( y \) have the same
distribution. (cf. Cambanis, Huang and Simons (1981).) For convenience we denote these relationship by \( R \leftrightarrow \phi \in \phi_n \) and \( R^* \overset{d}{=} R \overset{b_{m/2}}{\rightarrow} R \overset{h_{m/2},(n-m)/2}{\rightarrow} \phi \in \phi_m \).

Throughout the paper we assume that \( X \overset{\sim}{\sim} \text{MEC}_{n \times p} (M; \Sigma_{\sim 1}, \ldots, \Sigma_{\sim n}; \phi) \) with \( n > p, \Sigma_{\sim 1} > 0, \ldots, \Sigma_{\sim n} > 0, \phi \in \phi_{np} \) and \( X \) has a density of the form

\[
(1.4) \quad \prod_{i=1}^{n} |\Sigma_{\sim i}|^{-1/2} \exp \left\{ \frac{1}{2} \sum_{i=1}^{n} (x_{\sim i} - \mu_{\sim i})^T \Sigma_{\sim i}^{-1} (x_{\sim i} - \mu_{\sim i}) \right\}.
\]

In this case \( X \) can be expressed as following

\[
(1.5) \quad X \overset{\sim}{\sim} M + (R_{\sim j} \overset{\sim}{\sim} (1))_{\sim 1} \overset{\sim}{\sim} u_{(p)}^{(p)} + \cdots + (R_{\sim n} \overset{\sim}{\sim} (n))_{\sim n} \overset{\sim}{\sim} u_{(p)}^{(p)},
\]

where \( R_{\sim j}, j = 1, \ldots, n, \) are independent of \( u_{\sim j}^{(p)}, j = 1, \ldots, n, \)

\( u_{\sim 1}^{(p)}, \ldots, u_{\sim n}^{(p)} \) are independent, \( u_{\sim j}^{(p)} \overset{d}{=} u_{\sim j}^{(p)}, \Sigma_{\sim j} = A_{\sim j}^T A_{\sim j} \) is a factorization of \( \Sigma_{\sim j}, j = 1, \ldots, n, R_{\sim 1} \geq 0, \ldots, R_{\sim n} \geq 0 \) and

\[
(R_{\sim 1}^2, \ldots, R_{\sim n}^2) \overset{d}{=} R^2(d_1, \ldots, d_n),
\]

where \((d_1, \ldots, d_{n-1}) \overset{d}{=} D_{n}(p/2, \ldots, p/2;p/2),\) a Dirichlet distribution,

\( d_n = 1-d_1 - \cdots - d_{n-1}, \quad R \) is nonnegative and independent of \( d_1, \ldots, d_n, \)

and

\[
(1.6) \quad R^2 \overset{d}{=} \sum_{\alpha=1}^{n} (X_{\sim \alpha} - \mu_{\sim \alpha})^T \Sigma_{\sim \alpha}^{-1} (X_{\sim \alpha} - \mu_{\sim \alpha}).
\]

When \( \mu_{\sim 1} = \cdots = \mu_{\sim n} = \mu \) and \( \Sigma_{\sim 1} = \cdots = \Sigma_{\sim n} = \Sigma, \) we write \( X \overset{\sim}{\sim} \text{LEC}_{n \times p}(\mu, \Sigma, \phi). \) In this case (1.4) reduces to
(1.7) \[ |\Sigma|^{-n/2} \exp\left(\text{tr} \Sigma^{-1} \mathbf{g}\right), \]

where

(1.8) \[ \mathbf{g} = \sum_{\alpha=1}^{n} (\mathbf{x}(\alpha) - \bar{\mu})(\mathbf{x}(\alpha) - \bar{\mu})', \]

and (1.6) is

(1.9) \[ \mathbf{K}^2 \overset{\text{d}}{=} \text{tr} \Sigma^{-1} \mathbf{g}. \]

The maximum likelihood estimators and the unbiased estimators of \( \bar{\mu} \) and \( \bar{\Sigma} \), 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 MECO case. We find that most of the criteria and their corresponding distributions are the same in the class of MECO. A similar discussion on testing the general linear hypothesis is given in Section 4.

Throughout the paper \( N_n(\bar{\mu}, \bar{\Sigma}) \) denotes the \( n \)-dimensional normal distribution with mean \( \bar{\mu} \) and covariance matrix \( \bar{\Sigma} \); \( B(\alpha_1, \alpha_2) \) denotes the Beta distribution with parameters \( \alpha_1 \) and \( \alpha_2 \); \( D_m(\alpha_1, \ldots, \alpha_m; \alpha_m) \) denotes the Dirichlet distribution with parameters \( \alpha_1, \ldots, \alpha_m \); \( F(k, m) \) denotes the \( F \)-distribution with \( k \) and \( m \) degrees of freedom. \( W_p(\Sigma, n) \) denotes the Wishart distribution with \( p \times p \) covariance matrix \( \Sigma \) and \( n \) degrees of freedom; \( U_{p,m,n} \) denotes Wilks' distribution which is the distribution of the ratio \( |\bar{G}|/|\bar{G}+\bar{H}| \), where \( \bar{G} \sim W_p(\Sigma, n) \), \( \bar{H} \sim W_p(\Sigma, m) \), and \( \bar{G} \) and \( \bar{H} \)
are independent; \( I_n \) denotes the \( n \times n \) identity matrix; \( \varepsilon_n \) denotes the \( n \times 1 \) vector with elements 1; \( \text{rk} A \) denotes the rank of the matrix \( A \) and \( \text{tr} A \) denotes the trace of \( A \).

2. Estimation of \( \Psi \) and \( \Sigma \) and their distributions.

Assume \( X \sim \text{LEC} \mathbb{R}^{n \times p} (\mu, \Sigma, \phi) \) and \( X \) has a density

\[
\text{det}(\Sigma)^{-n/2} g(\text{tr} \Sigma^{-1} g),
\]

where \( g \) is defined by (1.8). Let

\[
W = \sum_{j=1}^{n} (x(j) - \bar{x})(x(j) - \bar{x})',
\]

where

\[
\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x(j).
\]

Then \( W = X'DX \), with \( D = I_n - (1/n) \varepsilon_n \varepsilon_n' \).

**Lemma 1.** If \( n > p \), then \( G \) and \( W \) are positive definite with probability one.

**Proof.** From the assumption and (1.5) we have

\[
X = \varepsilon_n u' + \text{RUA},
\]

where \( u \sim \mathbb{R}^{n \times p}, \text{vec} u = u \sim (np), A : p \times p \) and \( A'A = \Sigma > 0 \). Hence
\[ G = (X - \xi_n u')' (X - \xi_n u') \overset{\sim}{\sim} R^2 A' U' U A \overset{\sim}{\sim} R^2 A' Y' Y A / \text{tr} \ Y' Y, \]

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 \( \mu \) and \( \Sigma \)
we need the following lemma.

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

\[ (2.5) \quad h(x) \equiv x^{N/2} g(x) \quad \text{for} \quad x \geq 0 \]

has a maximum at some finite \( x_0 > 0 \), and \( x_0 \) is a solution of

\[ (2.6) \quad g'(x) + (N/2x)g(x) = 0. \]

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

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

using the transformation to spherical coordinates (Anderson (1958), Chapter 7, Problem 4, for example), and
\[ 2^{-N/2} (2x)^{N/2} g(2x) = x^{N/2} g(2x) \leq x^{N/2-1} \int_{x}^{2x} g(t) dt \leq \int_{x}^{2x} t^{N/2-1} g(t) dt \to 0 \]

as \( x \to \infty \), i.e., \( h(x) = x^{N/2} g(x) \to 0 \) as \( x \to \infty \). By this fact and \( h(0) = 0 \), \( h(x) \geq 0 \) for \( \forall x \geq 0 \), the first assertion of the theorem follows.

Now assume that \( h(x) \) has its maximum at \( x_0 > 0 \); then

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

which completes the proof. Q.E.D.

From Lemma 2, when \( N = np \), the function

\[ f(\lambda) \equiv \lambda^{-np/2} g(p/\lambda) \]

arrives at its maximum at some finite \( \lambda_0 \). Throughout the paper we denote this \( \lambda_0 \) by \( \lambda_{\text{max}}(g) \).

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

\[ g''(x_0) < (N(N+2)/4x_0^2) g(x_0) , \]

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

\[
e^{-x} = 1 - N/2x.
\]

It is easy to see that this equation has a unique solution \( x_0 (> N/2) \).

The following are some computations:

<table>
<thead>
<tr>
<th>N</th>
<th>( x_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4.069529635</td>
</tr>
<tr>
<td>10</td>
<td>5.032816088</td>
</tr>
<tr>
<td>12</td>
<td>6.014691496</td>
</tr>
<tr>
<td>20</td>
<td>10.00045381</td>
</tr>
<tr>
<td>30</td>
<td>15.000000459</td>
</tr>
<tr>
<td>40</td>
<td>20.000000004</td>
</tr>
</tbody>
</table>

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

\[
g(x) = e^{-x}, \quad 0 \leq x \leq 6,
\]

\[
e^{-x} \exp(-e^{-x}), \quad 6.014691496 \leq x < \infty.
\]

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

\[
p(6) = g(6), \quad p(6.014691496) = g(6.014691496),
\]

\[
p'(6) = g'(6), \quad p'(6.014691496) = g'(6.014691496).
\]
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

\[
\begin{align*}
  h(6.014691496) &= (6.014691496)^6 e^{-6.014691496} \exp(-e^{-6.014691496}) \\
                      &= 115.364451.
\end{align*}
\]

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

Now we come back to the maximum likelihood estimators of \( \mu \) and \( \Sigma \).

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

**Proof.** It is easy to see that

\[
\log L(\mu, \Sigma) = -(n/2) \log |\Sigma| + \log g[\text{tr} \Sigma^{-1} \bar{W} + n(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)] \]

9
The assumption that $g(\cdot)$ is decreasing and $\tilde{\Sigma} > 0$ shows us $\log L(\tilde{\mu}, \tilde{\Sigma})$ arrives at its maximum at $\tilde{\mu} = \tilde{x}$ and the concentrated likelihood is

\begin{equation}
(2.8) \quad \log L(\tilde{x}, \tilde{\Sigma}) = -(n/2) \log |\tilde{\Sigma}| + \log g(\text{tr } \tilde{\Sigma}^{-1} \tilde{W}) .
\end{equation}

By Lemma 1, $\tilde{W} > 0$ with probability one. So there exists a nonsingular matrix $C$ with probability one such that $CC' = \tilde{W}$. Let $C^{-1} = \tilde{\Sigma}^{-1/2} \tilde{C}^{-1}$. We then have

\begin{equation}
\log L(\tilde{x}, \tilde{\Sigma}) = -(n/2) \log |C^{1/2} \tilde{\Sigma}^{1/2}| - (n/2) \log |\tilde{W}| + \log g(\text{tr } \tilde{\Sigma}^{-1}) .
\end{equation}

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

\begin{equation}
\tilde{\Gamma}' \tilde{\Sigma} \tilde{\Gamma} = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)
\end{equation}

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

\begin{equation}
\log L(\tilde{x}, \tilde{\Sigma}) = -(n/2) \sum_{i=1}^{p} \log \lambda_i + \log g(\sum_{i=1}^{p} 1/\lambda_i) - (n/2) \log |\tilde{W}|
\end{equation}

\begin{equation}
= f(\lambda_1, \ldots, \lambda_p) - (n/2) \log |\tilde{W}| ,
\end{equation}

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

\begin{equation}
(2.9) \quad f(\lambda) = -(np/2) \log \lambda + \log g(p/\lambda) .
\end{equation}

The theorem follows by Lemma 2. Q.E.D.
It is clear from \( \log L(\mu, \Sigma) \) that \( \bar{x} \) and \( \bar{w} \) are a sufficient set of statistics for \( \mu \) and \( \Sigma \).

From the proof of Theorem 1 we have

\[
\max_{\mu, \Sigma > 0} L(\mu, \Sigma) = \lambda_{\text{max}}(g) \left( \frac{\lambda_{\text{max}}(g)}{p} \right)^{-np/2} g(p/\lambda_{\text{max}}(g)) |\bar{w}|^{-n/2}.
\]

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 \( \lambda_{\text{max}}(g) = 1/n \), and \( \Sigma = (1/n) \bar{w} \).

Now we want to obtain the distributions of \( \bar{x} \) and \( \lambda_{\text{max}}(g) \bar{w} \).

If \( X \sim \text{LEC}_{n \times p}(0, \Sigma, \phi) \) with \( P(X = 0) = 0 \), and \( X = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \) with \( Y_1 : m \times p \), the distribution of \( X' X \) is denoted by \( M_{p,2}(\Sigma; m/2; (n-m)/2; \phi) \). Anderson and Fang (1982b) have obtained the density of \( M_{p,2}(\Sigma; m/2; (n-m)/2; \phi) \).

**Theorem 2.** Assume that \( X \sim \text{LEC}_{n \times p}(\mu, \Sigma, \phi) \) with \( n > p \) and \( X \) has a density \( (2.1) \). Then

\[
(1) \text{ the joint density of } \bar{x} \text{ and } \bar{w} \text{ is}
\]

\[
\frac{n^{p/2}(n-1)p/2-p(p-1)/4}{\prod_{j=1}^{p} \Gamma((n-j)/2)} |\bar{w}|^{(n-p)/2-1} |\bar{\Sigma}|^{-n/2} g(\text{tr} \Sigma^{-1} \bar{W} \phi (\bar{x}-\mu) \Sigma^{-1} (\bar{x}-\mu));
\]
(2) \( \bar{x} \sim EC_p(\mu, (1/n)\Sigma, \phi) \) with \( \phi \in \Phi_p \leftrightarrow R^* = R_{p/2, (n-1)p/2}; \)

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

\[
(2.12) \quad \frac{2^{np/2-p(p-1)/4}}{\Gamma(p/2) \prod \Gamma((n-j)/2)} \left| \Sigma \right|^{-n/2} |W|^{(n-p)/2-1} \int_0^r r^{p-1} g(r^2 + tr \Sigma^{-1}w) dr .
\]

Proof. Let \( B \) be an \( n \times n \) orthogonal matrix with the last row \( (1/\sqrt{n}, \ldots, 1/\sqrt{n}) \) and \( \bar{Y} = (y_1, \ldots, y_n)' = BX. \) Then

\[
y(n) = \sqrt{n} \bar{x} ,
\]

\[
\bar{W} = \sum_{j=1}^{n-1} y(j)y(j)' .
\]

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

\[
vec(\bar{Y}) = vec(\bar{X}) \sim EC_{np}(\mu \times (B\Sigma_n), \Sigma \times (BB^'), \phi) .
\]

Thus we have

\[
\mu \otimes (B\Sigma_n) = \mu \otimes (0, \ldots, 0, 1/\sqrt{n})',
\]

and

\[
\Sigma \otimes (BB') = \Sigma \otimes I_n .
\]

Using Lemma 13.3.1 of Anderson (1958) and \( \bar{y}(n) = \sqrt{n} \bar{x}, \) (2.11) follows.
Since $\bar{W} = X'DX$ with $D = I_n - (1/n)\bar{\varepsilon}_n \bar{\varepsilon}_n'$, $\text{rk } D = n - 1$ and $D^2 = D$, $\bar{W} \sim \text{MG}_{p,2}(\Sigma; (n-1)/2, 1/2; \phi)$ 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 $B = (1/n)\bar{\varepsilon}_n'$ of Anderson and Fang (1982b). Q.E.D.

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

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

$$E_{\Sigma} = \mu, \text{ and } E_{\Sigma} (x-\mu) (x-\mu)' = (\text{ER}^2/n) \Sigma.$$ (cf. Cambanis, Huang and Simons (1981).)

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

$$E_{\Sigma} = \frac{\text{np}}{(n-1)\text{ER}^2} \bar{W},$$

respectively.

**Proof.** From Theorem 2 and Lemma 3 the first assertion is obvious. Using the notation of the proof of Theorem 2, we have $\bar{W} = \sum_{j=1}^{n-1} \bar{y}_{(j)} \bar{y}_{(j)}'$. If we can prove $E_{\bar{\Sigma}} \bar{y}_{(j)} \bar{y}_{(j)}' = (\text{ER}^2/\text{np}) \Sigma$, $j = 1, \ldots, n-1$, then (2.14) follows.

It is easy to show that $y_{(j)} \sim \text{EC}_p(0, \Sigma, \phi)$ with $\phi \leftrightarrow R^* = R_{p/2, (n-1)p/2}$. 

13
\[ j = 1, \ldots, n-1 \quad (\text{cf. Lemma 2, Anderson and Fang (1982b)}). \] By Lemma 3 and the moments of Beta distribution

\[ E y_{(j)}^r = (ER^2/p) \Sigma = (ER^2/np) \Sigma, \]

which completes the proof. Q.E.D.

**Example 1.** (Multivariate \( t \)-distribution). Let \( X \) be an \( n \times p \) random matrix with i.i.d. rows distributed according to \( N_p(\mu, \Sigma) \) and \( s \sim X' \Sigma X \) be independent of \( X \). Let \( \tilde{Y} = \sqrt{\tilde{X}/s} \), then the rows of \( \tilde{Y} \) and \( \text{vec} \, \tilde{Y}' \) 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 \( \tilde{Y} \) is

\[
E Y_{(j)} \Sigma^{-1/2} = \frac{\Gamma[(np+n)/2]|\Sigma|^{-n/2}}{\Gamma(n/2)(\pi np/2)^{np/2}} (1 + \Sigma^{-1} \text{tr} \, G^{-1} \Sigma^{-1} )^{1/2(np+n)},
\]

where

\[ G = \sum_{j=1}^n (y_{(j)} - \mu)(y_{(j)} - \mu)' \]

2. \( \tilde{Y} \sim \text{LEC}_{n \times p}(\mu, \Sigma, \phi) \) with

\[
g(x) = c(1 + \Sigma^{-1} x)^{-1/2(np+n)},
\]

where

\[ c = \frac{\Gamma((np+n)/2)/\Gamma(n/2)}{(\pi np/2)^{np/2}}.\]
Therefore, \( g(x) \) satisfies the conditions of Theorem 1 and \( \lambda_{\text{max}}(g) = 1/n \).

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

(4) The unbiased estimator of \( \Sigma \) is \( (\nu-2)/(n-1)\nu \) \( \hat{\Sigma} \).

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 \( X \sim \text{LEC}_{n \times p}(\mu, \Sigma, \phi) \) with \( n > p \) and \( X \) has a density (2.1) where \( g(x) \) is a decreasing and differentiable function. Partition \( \mu, \Sigma, \) and \( \hat{\Sigma} \) as follows:

\[
\begin{align*}
\mu &= \begin{pmatrix} \mu^{(1)} \\ \vdots \\ \mu^{(q)} \end{pmatrix}, \\
\Sigma &= \begin{pmatrix} \Sigma_{11} & \ldots & \Sigma_{1q} \\ \vdots & \ddots & \vdots \\ \Sigma_{q1} & \ldots & \Sigma_{qq} \end{pmatrix}, \\
\hat{\Sigma} &= \begin{pmatrix} \hat{\Sigma}_{11} & \ldots & \hat{\Sigma}_{1q} \\ \vdots & \ddots & \vdots \\ \hat{\Sigma}_{q1} & \ldots & \hat{\Sigma}_{qq} \end{pmatrix},
\end{align*}
\]

where \( \mu^{(1)}, \ldots, \mu^{(q)} \) have \( p_1, \ldots, p_q \) components, respectively, and \( \Sigma_{ij} \) and \( \hat{\Sigma}_{ij} \) are \( p_i \times p_j \) matrices, \( i,j = 1, \ldots, q \). Let
\[
\Sigma_0 = \begin{pmatrix}
\Sigma_{11} & 0 & \cdots & 0 \\
0 & \Sigma_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Sigma_{qq}
\end{pmatrix};
\]

then \( \Sigma_0 > 0 \) if and only if \( \Sigma_{11} > 0, \ldots, \Sigma_{qq} > 0 \). Testing lack of correlation between sets of variates is equivalent to testing the hypothesis \( \Sigma = \Sigma_0 \).

**Lemma 4.** Under the assumptions of Theorem 1

\[
\max_{\mu, \Sigma_0 > 0} L(\mu, \Sigma_0) = \lambda_{\max}(g)^{-np/2} \left( \prod_{j=1}^{q} \left| w_{jj} \right|^{n/2} \right) g(p/\lambda_{\max}(g)).
\]

**Proof.** 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

\[
\max_{\mu, \Sigma_0 > 0} L(\mu, \Sigma_0) = \max_{\Sigma_0 > 0} L(\bar{x}, \Sigma_0)
\]

\[
= \max_{\lambda_j > 0} \left\{ -(n/2) \prod_{j=1}^{p_1} \log \lambda_j^{(1)} - (n/2) \prod_{j=1}^{p_2} \log \lambda_j^{(2)} + \log g(p_1) \lambda_1^{(1)-1}
\right. \\
+ \left. \prod_{j=1}^{p_2} \lambda_j^{(2)-1} - (n/2) \times (\log |w_{11}| + \log |w_{22}|) \right\},
\]

where \( \lambda_1^{(1)}, \ldots, \lambda_{p_1}^{(1)} \) and \( \lambda_1^{(2)}, \ldots, \lambda_{p_2}^{(2)} \) are the eigenvalues of \( \Sigma_{11} \) and
\[ \tilde{\Sigma}_{22}, \text{ respectively, and } \tilde{\Sigma}_{11} = C_1^{-1} \tilde{\Sigma}_{11} C_1^{-1}, \tilde{\Sigma}_{22} = C_2^{-1} \tilde{\Sigma}_{22} C_2^{-1}, \]

\[ C_1 C_1' = \tilde{W}_{11} \text{ and } C_2 C_2' = \tilde{W}_{22}. \text{ As the above function in the braces is a symmetric function of } \lambda^{(1)}, \ldots, \lambda^{(1)}_{p_1}, \lambda^{(2)}, \ldots, \lambda^{(2)}_{p_2}, \text{ these } \lambda^{(i)}_{j}\text{ must be equal. We have} \]

\[
\max_{\mu, \Sigma > 0} L(\mu, \Sigma) = \max_{\lambda > 0} \{-(np/2) \log \lambda + \log g(p/\lambda) - (n/2)(\log |W_{11}| + \log |W_{22}|)\} \\
= \log \left\{ \frac{\lambda_{\max}(g)^{np/2} g(p/\lambda_{\max}(g)) |W_{11}|^{-n/2} |W_{22}|^{-n/2}}{\prod_{j=1}^q |W_{jj}|} \right\},
\]

which completes the proof. Q.E.D.

Theorem 4: Under the assumptions of Theorem 1, the likelihood ratio criterion for testing

\[(3.4) \quad H: \tilde{\Sigma}_{ij} = 0, \quad i \neq j, \quad i,j = 1, \ldots, q, \]

is

\[(3.5) \quad \tau = \frac{\max_{\mu, \Sigma \gg 0} L(\mu, \Sigma)}{\max_{\mu, \Sigma > 0} L(\mu, \Sigma)} = \left( \frac{|W|}{\prod_{j=1}^q |W_{jj}|} \right)^{n/2}, \]

and

\[(3.6) \quad \tau^2/n \sim \prod_{j=2}^q v_j, \]

where \(v_2, \ldots, q\) are independent and \(v_j \sim U_{p_j, p_j, n-p_j}\) with \(p_j = j-1 \sum_{i=1}^{j-1} p_i\).
Proof. The criterion (3.5) is from (2.10) and (3.3). By (2.4),
\[ X = \varepsilon_n \mu' + R U A. \]
Partition \( \mathcal{A} \) into \( \mathcal{A} = (A_1, \ldots, A_q) \) with \( A_j : p \times p_j, j = 1, \ldots, q. \) Then

\[ W \overset{d}{=} R^2 A' U' D U A = R^2 A' Y' D Y A / \text{tr} \ Y' Y \]
\[ \overset{\sim}{W} \sim j R^2 A' U' D U A \overset{\sim}{=} R^2 A' Y' D Y A / \text{tr} \ Y' Y, j = 1, \ldots, q, \]

where \( D, U \) and \( Y \) have the same meaning as before. Thus

\[ (3.7) \quad \tau^{2/n} \overset{d}{=} \left| A' Y' D Y A \right| / \prod_{j=1}^{q} \left| A_j' Y' D Y A_j \right|, \]

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

\[ x(j) \overset{\sim}{\sim} N(\mu, \Sigma) \] 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 \( \Sigma_{12} = 0 \) 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 \( H: \mu = \mu_0. \) The likelihood criterion is

\[ \tau = \max_{\Sigma > 0} L(\mu_0, \Sigma) / \max_{\mu, \Sigma > 0} L(\mu, \Sigma). \]

By a method similar to that used before we find
\[
\max_{\Sigma > 0} L(\mu_0, \Sigma) = \lambda_{\max}(g)^{-np/2} g(p/\lambda_{\max}(g)) |\bar{W}_0|^{-n/2},
\]

where
\[
\bar{W}_0 = \sum_{j=1}^{n} (\bar{x}_j - \mu_0)(\bar{x}_j - \mu_0)',
\]

Thus
\[
\tau^2/n = \frac{|\bar{W}|}{|\bar{W}_0|} = \frac{|W|}{|W + n(\bar{x} - \mu_0)(\bar{x} - \mu_0)'|} = \frac{1}{1 + \tau^2/(n-1)},
\]

where
\[
\tau^2 = n(n-1)(\bar{x} - \mu_0)'W^{-1}(\bar{x} - \mu_0),
\]

which is the same as in the normal case. Also, if the null hypothesis is true, the distribution of \( \tau^2 \) 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 \( \mu^{(1)}, \ldots, \mu^{(q)} \) and covariance matrices \( \Sigma_1, \ldots, \Sigma_q \), respectively. Assume that the joint distribution of the samples is a multivariate elliptically contoured distribution, i.e., the data matrix
\[
X \sim \text{MLE}_{n \times p} (M; \Sigma_1, \ldots, \Sigma_q, \Sigma_1, \ldots, \Sigma_q, \ldots, \Sigma_q ; \phi)
\]
with \( n_1 \Sigma_1', \ldots, n_q \Sigma_q' \), \( n_1 > p, \ldots, n_q > p, n_1 + n_2 + \cdots + n_q = n \) and the rows of
\( M \) are \( n_1 \mu_{(1)}', s, n_2 \mu_{(2)}', s, \ldots, n_q \mu_{(q)}', s \), successively. From (1.4) the density of \( X \) is

\[
L(\mu_{(1)}, \ldots, \mu_{(q)} ; \Sigma_1, \ldots, \Sigma_q) = \prod_{i=1}^q |\Sigma_i|^{-n_i/2} g\left( \frac{\text{tr} \Sigma_i^{-1} G_i}{\Sigma_i} \right),
\]

where

\[
G_i = \sum_{j=\bar{n}_{i-1}+1}^{\bar{n}_i} (x(j) - \bar{x}(i))(x(j) - \bar{x}(i)),
\]

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

We wish to test the hypothesis

\[
H_1 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_q.
\]

Lemma 5. Under the assumptions of Theorem 1

\[
\max_{\mu_{(1)}', \ldots, \mu_{(q)}', \Sigma_1 > 0, \ldots, \Sigma_q > 0} L(\mu_{(1)}, \ldots, \mu_{(q)} ; \Sigma_1, \ldots, \Sigma_q)
\]

\[
= \lambda_{\max}(g)^{-np/2} \prod_{i=1}^q \frac{n_i}{n}^{-n_i/2} g(p/\lambda_{\max}(g)),
\]

where

\[
W_i = \sum_{j=\bar{n}_{i-1}+1}^{\bar{n}_i} (x(j) - \bar{x}(i))(x(j) - \bar{x}(i))', \quad i = 1, \ldots, q,
\]

\[
\bar{x}(i) = \frac{1}{n_i} \sum_{j=\bar{n}_{i-1}+1}^{\bar{n}_i} x(j).
\]
Proof. By a method similar to that used in the proof of Theorem 1, we have

$$\max_{\mu^{(1)}, \ldots, \mu^{(q)}; \Sigma_1, \ldots, \Sigma_q} \log L(\mu^{(1)}, \ldots, \mu^{(q)}; \Sigma_1, \ldots, \Sigma_q)$$

$$= \max_{\Sigma_1 > 0, \ldots, \Sigma_q > 0} \log L(\tilde{x}^{(1)}, \ldots, \tilde{x}^{(q)}; \tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_q)$$

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

$$+ \log g(p \sum_{i=1}^{q} \lambda_j^{(i)}) - \frac{1}{2} \sum_{i=1}^{q} n_i \log |W_i| \right\}$$

where $\lambda_1^{(i)}, \ldots, \lambda_p^{(i)}$ are the eigenvalues of $\tilde{\Sigma}_i = \Sigma_i^{-1} \Sigma_i^{-1}$ and $\Sigma_i^{-1} \Sigma_i^{-1} = W_i$. Because $n_i > p$, $W_i > 0$ 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 $\lambda_1^{(i)} = \cdots = \lambda_p^{(i)} = \lambda^{(i)}$ and the quantity in the braces becomes

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

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

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

21
and $\lambda^{(i)} / \lambda^{(1)} = n_i / n_1$, $i = 1, \ldots, q$. Let $\lambda = \lambda^{(1)} n_1 / n$, then

$$\lambda^{(i)} = n \lambda / n_i, \quad i = 1, \ldots, q,$$

and (3.11) becomes

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

which completes the proof. Q.E.D.

Similarly, we find

$$\begin{align*}
\max_{\mu^{(1)}, \ldots, \mu^{(q)}, \Sigma, \tilde{\Sigma}} & L(\bar{y}^{(1)}, \ldots, \bar{y}^{(q)}; \Sigma, \tilde{\Sigma}) \\
= & \lambda^{\max} (g)^{-np/2} g(p/\lambda^{\max}(g)) |W_1|^{1/2} \cdots |W_q|^{1/2}. \\
= & \lambda^{\max} (g)^{-np/2} g(p/\lambda^{\max}(g)) |W_1 + \cdots + W_q|^{-n/2}.
\end{align*}$$

Thus we obtain the following theorem:

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

$$\tau_1 = \prod_{i=1}^q \left( \frac{|W_i|}{n_i} \right)^{n_i/2} \left( \frac{p}{n_i} \right)^{pn_i/2}$$

which has the same distribution as in the normal case.

The distribution of $\tau_1$ was discussed by Anderson (1958), Chapter 10. A sufficient set of statistics is $\bar{x}^{(1)}, \ldots, \bar{x}^{(q)}, \tilde{W}_1, \ldots, \tilde{W}_q$. 

22
3.4 Testing equality of several means.

Assuming \( \Sigma_1 = \cdots = \Sigma_q = \Sigma \), say, we want to test the hypothesis

\[
(3.14) \quad H_2 : \mu_1^{(1)} = \cdots = \mu_q^{(q)} .
\]

The corresponding likelihood ratio criterion is

\[
(4.15) \quad \tau_2 = \frac{\max_{\mu_1^{(1)}, \cdots, \mu_q^{(q)}, \Sigma > 0} L(\mu_1^{(1)}, \cdots, \mu_q^{(q)}; \Sigma, \cdots, \Sigma)}{\max_{\mu_1, \Sigma > 0} L(\mu_1, \cdots, \mu_q; \Sigma, \cdots, \Sigma)} = \frac{|W|^{n/2}}{|W_1 + \cdots + W_q|^{n/2}}
\]

where \( W \) 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

\[
(3.16) \quad H : \mu_1^{(1)} = \cdots = \mu_q^{(q)} \quad \text{and} \quad \Sigma_1 = \cdots = \Sigma_q .
\]

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.
\[(3.17) \quad \tau = \tau_1 \tau_2 = \prod_{i=1}^{q} \left( \frac{|W_i|}{|\tilde{W}|} \right)^{n_i/2} \left( \frac{n}{n_i} \right)^{p n_i/2}, \]

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 \( X \sim \text{LEX}_{n \times p} (\mu, \Sigma, \phi) \) and want to test

\[(3.18) \quad H: \Sigma \sim \sigma^2 \Sigma_0, \quad \Sigma_0 > 0, \]

where \( \Sigma_0 \) is given. First, we assume \( \Sigma_0 \sim \mathbb{I}_p \) and decompose the hypothesis \( H \) into \( H_1: \Sigma \) is diagonal and \( H_2: \) the diagonal elements of \( \Sigma \) are equal, given \( \Sigma \) is diagonal. Let \( \Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_p) \) and \( \Sigma_2 = \text{diag}(\sigma^2, \ldots, \sigma^2) \). It follows from Lemma 4 that

\[(3.19) \quad \max_{\mu, \Sigma_1 > 0} L(\mu, \Sigma_1) = \lambda_{\max}(g)^{-np/2} g(p/\lambda_{\max}(g)) \left( \prod_{j=1}^{p} w_{jj} \right)^{-n/2}, \]

where \( w_{jj} \) is the \( j \)-th diagonal element of \( \tilde{W} \). Similarly

\[(3.20) \quad \max_{\mu, \sigma > 0} L(\mu, \Sigma_2) = \lambda_{\max}(g)^{-np/2} g(p/\lambda_{\max}(g)) (\text{tr} \tilde{W}^p)^{-np/2}. \]

Then we have

\[(3.21) \quad \tau_1 = \frac{\max_{\mu, \Sigma_1 > 0} L(\mu, \Sigma_1)}{\max_{\mu, \Sigma > 0} L(\mu, \Sigma)} = \left( \frac{|\tilde{W}|}{\prod_{j=1}^{p} w_{jj}} \right)^{n/2}, \]
\[(3.22) \quad \tau_2 = \max_{\mu, \sigma > 0} \frac{L(\mu, \Sigma_2)}{\max_{\mu, \Sigma_1 > 0} L(\mu, \Sigma_1)} = \left( \prod_{j=1}^{p} w_{jj} \right)^{n/2} \left( \text{tr} \frac{W}{\Sigma} \right)^{np/2}, \]

and

\[(3.23) \quad \tau = \frac{|w|^{n/2}}{(\text{tr} \frac{W}{\Sigma})^{np/2}}. \]

When $\Sigma_0 \neq I_p$, the corresponding criterion is

\[(3.24) \quad \tau = \frac{|\Sigma_0^{-1} W|^{n/2}}{(\text{tr} \frac{\Sigma_0^{-1} W}{\Sigma_0})^{np/2}} \]

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 $\Sigma = \Sigma_0 > 0$, where $\Sigma_0$ is given. The corresponding likelihood ratio criterion is

\[(3.25) \quad \tau = \lambda_{\max}(g)^{np/2} g^{-1}(p/\lambda_{\max}(g)) g(\text{tr} \frac{\Sigma_0^{-1} W}{\Sigma_0}) |w|^{-1/2}. \]

Note that the distributions of $\tau$ are not the same in the class of MECD and depend on $\phi$. Similarly, we find that the likelihood ratio criterion for testing $H_0: \Sigma = \Sigma_0$ and $\mu = \mu_0$ is

25
\( (3.26) \quad \tau = \lambda_{\text{max}}(g) \sqrt{2np/2} g^{-1}(p/\lambda_{\text{max}}(g)) g(\text{tr} \Sigma_0^{-1} W + n(\bar{x} - \mu_0)' \Sigma_0^{-1}(\bar{x} - \mu_0)) |\Sigma_0^{-1} W|^{n/2}. \)

4. **Testing the general linear hypothesis.**

Consider the following multiple regression model:

\[
Y \sim \mathcal{N}(X, \Sigma, \phi) \quad \text{with density} \quad |\Sigma|^{-n/2} g(\text{tr} \Sigma^{-1} G) \quad \text{and} \quad G = E'E,
\]

where \( g(\cdot) \) is a decreasing and differentiable function.

We want to find the maximum likelihood estimates of \( \Sigma \) and \( \Sigma \). As

\[
L(B, \Sigma) = |\Sigma|^{-n/2} g(\text{tr} \Sigma^{-1} G)
\]

\[
= |\Sigma|^{-n/2} g(\text{tr} \Sigma^{-1}(Y-XB)'(Y-XB) + \text{tr}(\hat{B}-B)X\Sigma^{-1}X'(\hat{B}-B)'),
\]

where

\[
(4.2) \quad \hat{B} = (X'X)^{-1}X'Y,
\]

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

\[
L(\hat{B}, \Sigma) = |\Sigma|^{-n/2} g(\text{tr} \Sigma^{-1}(Y-X\hat{B})'(Y-X\hat{B}))
\]

with respect to \( \Sigma \) gives the maximum likelihood estimator
\[(4.3) \quad \hat{\Sigma} = \lambda_{\text{max}}(g) (Y-X\hat{B})' (Y-X\hat{B}) \]

and

\[(4.4) \quad L(\hat{B}, \hat{\Sigma}) = \max_{B, \Sigma > 0} L(B, \Sigma) = \lambda_{\text{max}}(g)^{-np/2} \left| (Y-X\hat{B})' (Y-X\hat{B}) \right|^{-n/2} g(p/\lambda_{\text{max}}(g)) , \]

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

Similarly, the unbiased estimators of \( \tilde{B} \) and \( \tilde{\Sigma} \) are \( \hat{B} \) and

\[ \hat{\Sigma} = \frac{np}{(n-q)\text{ER}} (Y-X\hat{B})' (Y-X\hat{B}) \]

respectively, if the second moments of components of \( E \) exist. A sufficient set of statistics is \( \tilde{B} \) and \( \tilde{\Sigma} \).

We want to test the following linear hypothesis:

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

It is easy to show that

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

where

\[ S = (Y-X\hat{B})' (Y-X\hat{B}) \]
and

\[ T = (\hat{H}_B - C)'(\tilde{H}(X'X)^{-1}\hat{H}')^{-1}(\tilde{H}_B - C) \].

Thus the likelihood ratio criterion for \( H \) is

\[
(4.7) \quad \tau = \frac{\max_{B, \Sigma > 0} L(B, \Sigma)}{\max_{\tilde{H}_B = C, \Sigma > 0} \tilde{L}(B, \Sigma)} = \frac{|S|^{-n/2}}{|S + T|^{-n/2}}
\]

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 \( B \) and \( \Sigma \) 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


### Title
Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions

#### Author(s)
T. W. Anderson and Kai-Tai Fang

#### Performing Organization Name and Address
Department of Statistics, Sequoia Hall
Stanford University
Stanford, California

#### Controlling Office Name and Address
U.S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

#### Distribution Statement (of this report)
Approved for public release; distribution unlimited.

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

#### Abstract
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.