ARO 26394.7-MA

(2)

NONNORMAL MULTIVARIATE DISTRIBUTIONS: INFERENCE BASED ON ELLIPTICALLY CONTOURED DISTRIBUTIONS

BY

T. W. Anderson

TECHNICAL REPORT NO. 28

July 1992

U. S. Army Research Office

Contract DAAL03-89-K-0033

Theodore W. Anderson, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Nonnormal Multivariate Distributions: Inference based on Elliptically Contoured Distributions

By

T. W. Anderson

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

Technical Report No. 28

July 1992

U. S. Army Research Office

Contract DAAL03-89-K-0033

Theodore W. Anderson, Project Director

Department of Statistics

Stanford University

Stanford, California

Nonnormal Multivariate Distributions: Inference based on Elliptically Contoured Distributions *

T. W. Anderson†

Stanford University

1. Introduction.

The classical or conventional multivariate analysis is based largely on the multivariate normal distribution. This probability model fits many, though not all, sets of continuous multivariate data. The theory and methodology of inference for this model is highly developed and has been exposited extensively. The nature of the normal distribution permits considerable analysis in terms of conventional matrix algebra. The fact that the parameter set consists of a vector and a matrix that can be interpreted as the mean of the observation vector and its covariance matrix makes inference relatively easy to interpret and simplifies the analysis. Of course, these advantages of simplicity are also disadvantages of inflexibility that restrict the applicability. It is useful therefore, to extend multivariate probability distributions beyond the normal class.

In this lecture we shall describe a larger class of distributions, thus augmenting the scope of analysis. The set of nonnormal distributions, of course, is very wide; we cannot hope to cover more than a portion of this field. In the International Symposium on Multivariate Analysis and Its Applications held in Hong Kong in March 1992 Ingram Olkin gave a paper entitled "Multivariate Nonnormal Distributions." The topics included bivariate binomial distributions, bivariate Poisson distributions, bivariate exponential distributions, and multivariate distributions with given marginal distributions; none of these subjects will be included in this present paper. The title of

*The first C.G. Khatri Memorial Lecture given at Pennsylvania State University, May 8, 1992.

†Reaearch supported by the U. S. Army Research Office Contract No. DAAL03-89-K-0033 at Stanford University

William Cleveland's presentation to the Hong Kong conference was "Computer Intensive Methods and Graphical Methods for Analyzing Multivariate Data;" his approach was that of data analysis - another subject I am not including here.

My paper is devoted to the exposition of elliptically contoured distributions and statistical inference appropriate to such distributions. This class of distributions, provides more flexibility, specifically, it permits nontrivial kurtosis; the marginal distributions can have long tails. At the same time much of the structure of the normal distribution is retained.

As we shall see, many of the statistical methods appropriate to normal parent distributions are also suitable for a more general class of elliptically contoured distributions, but since the kurtosis in an elliptically contoured distribution may be quite different from the null kurtosis of the normal other methods are often needed. Such methods may be more robust than normal methods, which are usually based on linear and quadratic functions of the observations. Not only does this larger class of distributions call for new methods, the class forms an excellent framework in which to study and evaluate robust procedures.

It can be expected that in the future much more attention will be paid to the elliptically contoured distribution. This paper will point to some important aspects.

Chmielewski (1981) has given a review of the papers on spherically contoured and elliptically contoured distributions that appeared before 1980. He mentions Maxwell (1860), Bartlett (1934), and Hartman and Wintner(1940) as three of the earliest papers. Kelker (1970) developed some of the properties of spherically and elliptically contoured distributions. A recent summary is given in Fang and Zhang (1990). See also Fang and Anderson (1990).

2. The Normal Distribution.

2.1. General

The normal distribution of a (nondegenerate) p-component random vector $\boldsymbol{X} = (X_1, \cdots, X_p)'$ has a density which can be written

$$\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Lambda}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})}, \tag{2.1}$$

where $\boldsymbol{\nu}$ is a p-component vector and $\boldsymbol{\Lambda}$ is a positive definite matrix. Integration shows that the mean vector and covariance matrix of $\boldsymbol{X}$ are

$$\mathcal{E}\boldsymbol{X} = \boldsymbol{\nu}, \qquad \mathcal{E}(\boldsymbol{X} - \mathcal{E}\boldsymbol{X})(\boldsymbol{X} - \mathcal{E}\boldsymbol{X})' = \boldsymbol{\Lambda}, \tag{2.2}$$

respectively, There is mnemonic advantage in re-naming this vector $\boldsymbol{\nu}$ and this matrix $\boldsymbol{\Lambda}$ as $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$, respectively. Hence the density of $\boldsymbol{X}$ is

$$\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}; \tag{2.3}$$

we write $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The characteristic function of $\boldsymbol{X}$ is

$$\mathcal{E}e^{i\boldsymbol{t}'\boldsymbol{X}} = e^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}+i\boldsymbol{t}'\boldsymbol{\mu}}. \tag{2.4}$$

The moments of $\boldsymbol{X}$ up to order 4 are

$$\mathcal{E}\boldsymbol{X} = \boldsymbol{\mu}, \;\; \mathcal{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = \boldsymbol{\Sigma}, \tag{2.5}$$

$$\mathcal{E}(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)' = 0, \tag{2.6}$$

$$\mathcal{E}(X_i - \mu_i)(X_j - \mu_j,)(X_k - \mu_k)(X_l - \mu_l) = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} \tag{2.7}$$

Every moment of odd order is 0. The contours of constant density are ellipses

$$(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = \text{const.} \tag{2.8}$$

2.2. The spherical normal distribution

Let $\boldsymbol{A}$ be any nonsingular matrix satisfying

$$\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Sigma}. \tag{2.9}$$

Define

$$\boldsymbol{Y} = \boldsymbol{A}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}). \tag{2.10}$$

Then the density of $\boldsymbol{Y}$ is

$$\frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}}. \tag{2.11}$$

The characteristic function of $\boldsymbol{Y}$ is

$$\mathcal{E}e^{i\boldsymbol{t}'\boldsymbol{Y}} = e^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}}. \tag{2.12}$$

The moments of $\boldsymbol{Y}$ of order up to 4 are

$$\mathcal{E}\boldsymbol{Y} = \boldsymbol{0}, \qquad \mathcal{E}\boldsymbol{Y}\boldsymbol{Y}' = \boldsymbol{I}_p. \tag{2.13}$$

$$\mathcal{E}Y_iY_jY_k = 0, \tag{2.14}$$

$$\mathcal{E}Y_iY_jY_kY_l = \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}, \tag{2.15}$$

where $\delta_{ii} = 1$ and $\delta_{ij} = 0$, $i \neq j$. Every moment of odd order is 0. The contours of constant density are spheres centered at the origin.

Define

$$R^2 = ||\boldsymbol{Y}||^2 = \boldsymbol{Y}'\boldsymbol{Y}, \tag{2.16}$$

$$\boldsymbol{U} = \frac{1}{||\boldsymbol{Y}||}\boldsymbol{Y} = \frac{1}{R}\boldsymbol{Y}. \tag{2.17}$$

Then

$$R^2 \stackrel{\mathrm{d}}{=} \chi^2_p, \tag{2.18}$$

where (2.18) means R^2 is distributed as χ^2_p, the chi-squared random variable with p degrees of freedom. The density of $W = R^2$ is

$$\frac{1}{2^{p/2}\Gamma(p/2)} w^{\frac{1}{2}p-1} e^{-\frac{1}{2}w}. \tag{2.19}$$

The vector $\boldsymbol{U}$ has the uniform distribution on the unit sphere

$$\boldsymbol{u}\boldsymbol{u}' = 1; \tag{2.20}$$

that is, the distribution of $\boldsymbol{PU}$ is that of $\boldsymbol{U}$ for any (fixed) orthogonal matrix $\boldsymbol{P}$. We write this as

$$\boldsymbol{PU} \stackrel{d}{=} \boldsymbol{U}. \tag{2.21}$$

The scalar R and the vector $\boldsymbol{U}$ are independent. We can represent $\boldsymbol{Y}$ as

$$\boldsymbol{Y} \stackrel{d}{=} R\boldsymbol{U}, \tag{2.22}$$

and we can represent $\boldsymbol{X}$ as

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + R\boldsymbol{A}\boldsymbol{U}. \tag{2.23}$$

If $\boldsymbol{\Sigma}$ is of rank r, then we can write $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$ with $\boldsymbol{A}\; p \times r$. If $\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{I}_r)$, we can represent $\boldsymbol{X}$ as $\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{\nu}$, where $R^2 \sim \chi^2_r$ and $\boldsymbol{U}$ has the uniform distribution on $\boldsymbol{u}\boldsymbol{u}' = 1$ in r dimensions. The characteristic function of $\boldsymbol{X}$ is (2.4).

The moments of $\boldsymbol{Y}$ are the products of the corresponding moments of R and those of $\boldsymbol{U}$. Since the first two moments of R^2 are

$$\mathcal{E}R^2 = p, \qquad \mathcal{E}R^4 = p(p+2), \tag{2.24}$$

the odd-ordered moments of $\boldsymbol{U}$ are 0,

$$\mathcal{E}\boldsymbol{U}\boldsymbol{U}' = \frac{1}{\mathcal{E}R^2}\mathcal{E}\boldsymbol{Y}\boldsymbol{Y}' = \frac{1}{p}\boldsymbol{I}_p, \tag{2.25}$$

$$\mathcal{E}U_iU_jU_kU_l = \frac{1}{p(p+2)}(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \tag{2.26}$$

3. Elliptically Contoured Distributions.

3.1. Spherical distributions.

Analogous to the normal distributions a spherical distribution in general can be characterized in several ways as follows:

1. If $\boldsymbol{Y}$ $(p \times 1)$ has a density, it is of the form $g(\boldsymbol{y}'\boldsymbol{y})$, where $g(\boldsymbol{y}'\boldsymbol{y}) \geq 0$ and

$$\int_{-\infty}^{\infty} \cdots \int_{\infty}^{\infty} g(\boldsymbol{y}'\boldsymbol{y}) d\boldsymbol{y} = 1. \tag{3.1}$$

Contours of constant densities are spheres: $\boldsymbol{y}'\boldsymbol{y} = \text{const}$. However, $\boldsymbol{Y}$ may have a spherical distribution even though a density does not exist. For example, the vector $\boldsymbol{U}$ defined in Section 2 has a spherical distribution.

2. For every orthogonal matrix $\boldsymbol{P}$

$$\boldsymbol{Y} \stackrel{\mathrm{d}}{=} \boldsymbol{PY}. \tag{3.2}$$

If Y has a density, (3.2) follows from the form $g(\boldsymbol{y}'\boldsymbol{y})$.

3. Property 2 implies that the characteristic function of $\boldsymbol{Y}$ has the form

$$\mathcal{E}e^{i\boldsymbol{t}'\boldsymbol{Y}} = \phi(\boldsymbol{t}'\boldsymbol{t}). \tag{3.3}$$

4. The random vector $\boldsymbol{Y}$ has the representation

$$\boldsymbol{Y} \stackrel{\mathrm{d}}{=} R\boldsymbol{U}, \tag{3.4}$$

where $R \geq 0$, $\boldsymbol{U}$ has the uniform distribution on $\boldsymbol{u}'\boldsymbol{u} = 1$, and R and $\boldsymbol{U}$ are independent. The density of R is found from $g(\boldsymbol{y}'\boldsymbol{y})$ by transforming to polar coordinates and integrating out the $p-1$ angles. [See Anderson, (1984), Problems 1 to 4, Chapter 7, for example]. The resulting density is

$$f(r) = \frac{2\pi^{\frac{1}{2}p-1}}{\Gamma(p/2)} r^{p-1} g(r^2). \tag{3.5}$$

We note that $\mathcal{E}R^h < \infty$ if and only if

$$\int_0^{\infty} r^{h+p-1} g(r^2) dr < \infty. \tag{3.6}$$

We can write the characteristic function of $\boldsymbol{Y}$ as

$$\phi(\boldsymbol{t}'\boldsymbol{t}) = \mathcal{E}e^{iR\boldsymbol{t}'\boldsymbol{U}} = \int_0^{\infty} \omega(r^2 \boldsymbol{t}'\boldsymbol{t}) f(r) dr, \tag{3.7}$$

where

$$\omega(\boldsymbol{s}'\boldsymbol{s}) = \mathcal{E}e^{i\boldsymbol{s}'\boldsymbol{U}} \tag{3.8}$$

is the characteristic function of $\boldsymbol{U}$.

5. If $||\boldsymbol{a}|| = ||\boldsymbol{b}||$, then

$$\boldsymbol{a}'\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{b}'\boldsymbol{Y}. \tag{3.9}$$

We shall denote the distribution of $\boldsymbol{Y}$ with characteristic function (3.3) as $S_p(\phi)$.

3.2. Elliptically contoured distributions in general.

Define

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Y}, \tag{3.10}$$

where $\boldsymbol{A}$ is a nonsingular matrix such that

$$\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Lambda}. \tag{3.11}$$

1. The density of $\boldsymbol{X}$ is

$$|\boldsymbol{\Lambda}|^{-1} g[(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})]. \tag{3.12}$$

3. The characteristic function of $\boldsymbol{X}$ is

$$Ee^{i\boldsymbol{s}'\boldsymbol{X}} = e^{i\boldsymbol{s}'\boldsymbol{\mu}}\phi(\boldsymbol{s}'\boldsymbol{\Lambda}\boldsymbol{s}). \tag{3.13}$$

4.

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{R}\boldsymbol{A}\boldsymbol{U}, \tag{3.14}$$

where $\boldsymbol{R}$ and $\boldsymbol{U}$ were defined above.

Contours of constant density are

$$(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = \text{const.} \tag{3.15}$$

We shall denote the distribution of $\boldsymbol{X}$ with characteristic function (3.13) as $EC_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}; \phi)$.

3.3. Moments.

The moments of $\boldsymbol{X}$ can be found from the moments of R and $\boldsymbol{U}$, which are independent. The moments of $\boldsymbol{U}$ were given in Section 2. We find

$$\mathcal{E}\boldsymbol{X} = \boldsymbol{\mu}, \tag{3.16}$$

$$\mathcal{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = \frac{\mathcal{E}R^2}{p}\boldsymbol{\Lambda} = \boldsymbol{\Sigma}, \tag{3.17}$$

say,

$$\mathcal{E}(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k) = 0, \tag{3.18}$$

In fact, all moments of $\boldsymbol{X} - \boldsymbol{\mu}$ of odd order are 0. The fourth-order moments are obtained from (2.26) and (3.10) as

$$\begin{aligned} &\mathcal{E}(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l) \\ &= \frac{\mathcal{E}R^4}{p(p+2)}(\lambda_{ij}\lambda_{kl} + \lambda_{ik}\lambda_{jl} + \lambda_{il}\lambda_{jk}) \\ &= \frac{\mathcal{E}R^4}{(\mathcal{E}R^2)^2}\frac{p}{p+2}(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}). \end{aligned} \tag{3.19}$$

The first moments of R are related to the characteristic function $\phi(\cdot)$ by

$$\mathcal{E}R^2 = -2p\phi'(0), \tag{3.20}$$

$$\mathcal{E}R^4 = 4p(p+2)\phi''(0). \tag{3.21}$$

The fourth cumulant of the i-th component of $\boldsymbol{X}$ standardized by its standard deviation is

$$\begin{aligned} \frac{\mathcal{E}(X_i - \mu_i)^4 - 3[\mathcal{E}(X_i - \mu_i)^2]^2}{[\mathcal{E}(X_i - \mu_i)^2]^2} &= \frac{\frac{3\mathcal{E}R^4}{p(p+2)} - 3\left(\frac{\mathcal{E}R^2}{p}\right)^2}{\left(\frac{\mathcal{E}R^2}{p}\right)^2} \\ &= 3\left[\frac{\mathcal{E}R^4}{(\mathcal{E}R^2)^2}\frac{p}{p+2} - 1\right] \\ &= 3\left[\frac{\phi''(0)}{[\phi'(0)]^2} - 1\right] \\ &= 3\kappa, \end{aligned} \tag{3.22}$$

say. Note that the fourth cumulant is 3κ for every component of $\boldsymbol{X}$. The fourth cumulant of X_i, X_j, X_k, and X_l is

$$\begin{aligned} \kappa_{ijkl} &= \mathcal{E}(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l) - (\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}) \\ &= \kappa(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}). \end{aligned} \tag{3.23}$$

3.4. Marginal and conditional distributions.

The characteristic function of a linear function of $\boldsymbol{X}$, say $\boldsymbol{Z} = \boldsymbol{B}\boldsymbol{X}$, is

$$\begin{aligned} \mathcal{E}e^{i\boldsymbol{t}'\boldsymbol{Z}} &= \mathcal{E}e^{i\boldsymbol{t}'\boldsymbol{B}\boldsymbol{X}} \\ &= e^{i\boldsymbol{t}'\boldsymbol{B}\boldsymbol{\mu}}\phi(\boldsymbol{t}'\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}'\boldsymbol{t}) \end{aligned} \tag{3.24}$$

by use of (3.13). This shows that $\boldsymbol{Z}$ has the distribution $EC_p(\boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}'; \phi)$. In particular, if

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{3.25}$$

where $\boldsymbol{X}_1$ has p_1 components, then $\boldsymbol{X}_1$ has the distribution $EC_p(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}; \phi)$.

We can also characterize marginal distributions in terms of the representation (3.14). Consider

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \end{pmatrix} \stackrel{\mathrm{d}}{=} R\boldsymbol{U} = R\begin{pmatrix} \boldsymbol{U}^{(1)} \\ \boldsymbol{U}^{(2)} \end{pmatrix}, \tag{3.26}$$

where $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{U}^{(1)}$ have p_1 components and $\boldsymbol{Y}^{(2)}$ and $\boldsymbol{U}^{(2)}$ have p_2 components ($p_1 + p_2 = p$). Then $R_1^2 = \boldsymbol{Y}^{(1)\prime}\boldsymbol{Y}^{(1)}$ has the distribution of $R^2\boldsymbol{U}^{(1)\prime}\boldsymbol{U}^{(1)}$, and

$$\boldsymbol{U}^{(1)\prime}\boldsymbol{U}^{(1)} = \frac{\boldsymbol{U}^{(1)\prime}\boldsymbol{U}^{(1)}}{\boldsymbol{U}'\boldsymbol{U}} \stackrel{\mathrm{d}}{=} \frac{\boldsymbol{Y}^{(1)\prime}\boldsymbol{Y}^{(1)}}{\boldsymbol{Y}'\boldsymbol{Y}}. \tag{3.27}$$

In the case $\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{I}_p)$ (3.27) has the beta distribution, say $B(p_1, p_2)$, with density

$$\frac{\Gamma(p/2)}{\Gamma(p_1/2)\Gamma(p_2/2)} z^{\frac{1}{2}p_1-1}(1-z)^{\frac{1}{2}p_2-1}, \ 0 \le z \le 1. \tag{3.28}$$

Hence, for arbitrary $S_p(\phi)$

$$\boldsymbol{Y}^{(1)} \stackrel{\mathrm{d}}{=} R_1\boldsymbol{V}, \tag{3.29}$$

where $R_1^2 \stackrel{d}{=} R^2 b$, $b \sim B(p_1, p_2)$, $\boldsymbol{V}$ has the uniform distribution on $\boldsymbol{v}'\boldsymbol{v} = 1$ in p_1 dimensions, and R^2, b, and $\boldsymbol{V}$ are independent. All marginal distributions are elliptically contoured.

Now suppose that $\boldsymbol{A}$ satisfying (3.11) is lower triangular. It can be partitioned as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{21} \end{pmatrix}. \tag{3.30}$$

Then the first p_1 rows of (3.14) yield

$$\begin{aligned} \boldsymbol{X}^{(1)} &\stackrel{d}{=} \boldsymbol{\mu}^{(1)} + \boldsymbol{A}_{11}\boldsymbol{Y}^{(1)} \\ &\stackrel{d}{=} \boldsymbol{\mu}^{(1)} + \boldsymbol{A}_{11}R_1\boldsymbol{V}. \end{aligned} \tag{3.31}$$

Suppose $\boldsymbol{X}$ has the density

$$\begin{aligned} &g[(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})] \qquad (3.32) \\ &= g\left\{\left[\boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)} - \boldsymbol{B}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)})\right]' \boldsymbol{\Lambda}_{11.2}\left[\boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)} - \boldsymbol{B}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)})\right] + Q_2\right\}, \end{aligned}$$

where $\boldsymbol{B} = \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}, \boldsymbol{\Lambda}_{11.2} = \boldsymbol{\Lambda}_{11} - \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21}$, and

$$Q_2 = (\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)})'\boldsymbol{\Lambda}_{22}^{-1}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)}). \tag{3.33}$$

[See Anderson (1984), Chapter 2, Problem 58, and Theorem A. 3.1.] The conditional density of $\boldsymbol{X}^{(1)}$ given $\boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}$ is (3.32) divided by $g_2(Q_2)$, where $g_2(\cdot)$ is the marginal density of $\boldsymbol{X}^{(2)}$ at $\boldsymbol{x}^{(2)}$. Note that $g_{1.2}(\cdot)$ is the density of an elliptically contoured distribution. From (3.32) it follows that

$$\mathcal{E}(\boldsymbol{X}^{(1)}|\boldsymbol{x}^{(2)}) = \boldsymbol{\mu}^{(1)} + \boldsymbol{B}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)}), \tag{3.34}$$

$$\operatorname{Var}(\boldsymbol{X}^{(1)}|\boldsymbol{x}^{(2)}) = h(\boldsymbol{x}^{(2)})\boldsymbol{\Lambda}_{11.2}, \tag{3.35}$$

where $h(\boldsymbol{x}^{(2)})$ is a nonnegative function of $\boldsymbol{x}^{(2)}$. Note that the conditional expectation of $\boldsymbol{X}^{(1)}$ given $\boldsymbol{x}^{(2)}$ is the same as for the normal distribution and the conditional covariance matrix is proportional to that to that for the normal. In this sense the structure of the normal distribution is maintained.

3.5. Examples.

1. *The multivariate t-distribution.* Suppose $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_p)$, $ms^2 \overset{d}{=} \chi^2_m$, and $\boldsymbol{Z}$ and s^2 are independent. Define $\boldsymbol{Y} = (1/s)\boldsymbol{Z}$. Then the density of $\boldsymbol{Y}$ is

$$\frac{\Gamma(\frac{m+p}{2})}{\Gamma(\frac{m}{2})m^{p/2}\pi^{p/2}}\left(1+\frac{\boldsymbol{y}'\boldsymbol{y}}{m}\right)^{-\frac{m+p}{2}} \tag{3.36}$$

and

$$\frac{R^2}{p} = \frac{||\boldsymbol{Y}||^2}{p} \sim F_{p,n} = \frac{m}{p}\frac{\chi^2_p}{\chi^2_m}. \tag{3.37}$$

If $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Y}$, the density of $\boldsymbol{X}$ is

$$\frac{\Gamma(\frac{m+p}{2})}{\Gamma(\frac{m}{2})m^{p/2}\pi^{p/2}}|\boldsymbol{A}|^{\frac{1}{2}}\left[1+\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{m}\right] \tag{3.38}$$

2. *Contaminated normal.* The contaminated normal distribution is a mixture of two normal distributions with proportional covariance matrices and the same mean vector. The density can be written

$$(1-\varepsilon)\frac{1}{(2\pi)^{p/2}|\boldsymbol{A}|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} + \varepsilon\frac{1}{(2\pi)^{p/2}|\boldsymbol{A}|^{\frac{1}{2}}}e^{-\frac{1}{2c}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \tag{3.39}$$

where $c > 0$ and $0 \le \varepsilon \le 1$. Usually ε is rather small and c rather large.

4. Sampling.

4.1. The density and characteristic function

A random sample from $EC_p(\boldsymbol{\mu}, \Lambda; \phi)$ consists of N vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N$. The density of the sample is

$$|\Lambda|^{-\frac{N}{2}}\prod_{\alpha=1}^{N} g[(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})]. \tag{4.1}$$

The characteristic function of the sample is

$$\mathcal{E}e^{i\sum_{\alpha=1}^{N}\boldsymbol{t}'_\alpha\boldsymbol{X}_\alpha} = \prod_{\alpha=1}^{N}\left[e^{i\boldsymbol{t}'_\alpha\boldsymbol{\mu}}\phi(\boldsymbol{t}'_\alpha\Lambda\boldsymbol{t}_\alpha)\right] = e^{i\sum_{\alpha=1}^{n}\boldsymbol{t}'_\alpha\boldsymbol{\mu}}\prod_{\alpha=1}^{N}\phi(\boldsymbol{t}'_\alpha\Lambda\boldsymbol{t}_\alpha). \tag{4.2}$$

In the case of the normal distribution the density and characteristic function are based on

$$g(w) = \frac{1}{(2\pi)^{p/2}} e^{-w/2}, \tag{4.3}$$

$$\phi(v) = e^{-v/2}. \tag{4.4}$$

The density of $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N$ is

$$\begin{aligned} & \prod_{\alpha=1}^{N} \frac{1}{(2\pi)^{p/2} |\Lambda|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})' \Lambda^{-1} (\boldsymbol{x}_\alpha - \boldsymbol{\mu})\right] & (4.5) \\ = \; & (2\pi)^{-pN/2} |\Lambda|^{-N/2} \exp\left[-\frac{1}{2} \operatorname{tr} \Lambda^{-1} \sum_{\alpha=1}^{N} (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\right] \\ = \; & (2\pi)^{-pN/2} |\Lambda|^{-N/2} \exp\left[-\frac{1}{2} \left\{\operatorname{tr} \Lambda^{-1} \boldsymbol{A} + N(\bar{\boldsymbol{x}} - \boldsymbol{\mu})' \Lambda^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu})\right]\right\}, \end{aligned}$$

where

$$\boldsymbol{A} = \sum_{\alpha}^{N} (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})', \tag{4.6}$$

$$\bar{\boldsymbol{x}} = \frac{1}{N} \sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha. \tag{4.7}$$

Display (4.5) shows that $\boldsymbol{A}$ and $\bar{\boldsymbol{x}}$ are sufficient for Λ and $\boldsymbol{\mu}$ and $\boldsymbol{A}$ and $\bar{\boldsymbol{x}}$ are independent. In fact $\bar{\boldsymbol{x}} \sim N[\boldsymbol{\mu}, (1/N)\Lambda]$ and $A \sim W(\Lambda, n)$, where $W(\Lambda, n)$ denotes the Wishart distribution with covariance matrix Λ and $n = N - 1$ degrees of freedom. That $\boldsymbol{A}$ and $\bar{\boldsymbol{x}}$ are sufficient statistics and are independent is due to the fact that $g(w)$ is exponential. These properties do not hold for other elliptically contoured distributions.

4.2. The asymptotic distribution of the sample mean and covariance matrix

We define the sample covariance matrix as

$$\boldsymbol{S} = \frac{1}{n} \boldsymbol{A}, \tag{4.8}$$

where $n = N - 1$ is the number of degrees of freedom. Then the sample mean and covariance matrix are unbiased estimators of the model mean and covariance matrix:

$$\mathcal{E}\bar{\boldsymbol{x}} = \boldsymbol{\mu}, \qquad \mathcal{E}\boldsymbol{S} = \boldsymbol{\Sigma}. \tag{4.9}$$

By the law of large numbers they are consistent estimators as $N \to \infty$:

$$\bar{\boldsymbol{x}} \xrightarrow{\text{P}} \boldsymbol{\mu}, \qquad \boldsymbol{S} \xrightarrow{\text{P}} \boldsymbol{\Sigma}. \tag{4.10}$$

The covariances of $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$ are

$$\operatorname{Cov}(\bar{\boldsymbol{x}}) = \frac{1}{N}\boldsymbol{\Sigma}, \qquad \mathcal{E}(s_{ij} - \sigma_{ij})(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) = \mathbf{0} \tag{4.11}$$

$$\operatorname{Cov}(s_{ij}, s_{kl}) = \frac{\kappa}{N}(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) + \frac{1}{n}(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}). \tag{4.12}$$

Then as $N \to \infty$

$$n\operatorname{Cov}(s_{ij}, s_{kl}) \to (1+\kappa)(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) + \kappa\sigma_{ij}\sigma_{kl}, \tag{4.13}$$

It will be convenient to use more matrix algebra. Define $\operatorname{vec}\boldsymbol{B}$, $\boldsymbol{B} \otimes \boldsymbol{C}$ (the Kronecker product) and $\boldsymbol{K}_{mn}$ (the commutator matrix) by

$$\operatorname{vec}\boldsymbol{B} = \operatorname{vec}(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n) = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_n \end{bmatrix}, \tag{4.14}$$

$$\boldsymbol{B} \otimes \boldsymbol{C} = \begin{bmatrix} b_{11}\boldsymbol{C} & \cdots & b_{1n}\boldsymbol{C} \\ \vdots & & \vdots \\ b_{m1}\boldsymbol{C} & \cdots & b_{mn}\boldsymbol{C} \end{bmatrix}, \tag{4.15}$$

$$\boldsymbol{K}_{mn}\operatorname{vec}\boldsymbol{B} = \operatorname{vec}\boldsymbol{B}'. \tag{4.16}$$

See, for example Magnus and Neudecker (1979). We can rewrite (4.13) as

$$\begin{aligned} n\operatorname{Cov}(\operatorname{vec}\boldsymbol{S}) &= \mathcal{E}(\operatorname{vec}\boldsymbol{S} - \operatorname{vec}\boldsymbol{\Sigma})(\operatorname{vec}\boldsymbol{S} - \operatorname{vec}\boldsymbol{\Sigma})' \\ &\to (\kappa+1)(\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + \kappa\operatorname{vec}\boldsymbol{\Sigma}(\operatorname{vec}\boldsymbol{\Sigma})'. \end{aligned} \tag{4.17}$$

Then

$$\sqrt{n}\begin{bmatrix} (\bar{\boldsymbol{x}}-\boldsymbol{\mu})' \\ \operatorname{vec}\boldsymbol{S} - \operatorname{vec}\boldsymbol{\Sigma} \end{bmatrix} \tag{4.18}$$
$$\xrightarrow{\text{d}} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & (\kappa+1)(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \kappa \operatorname{vec}\boldsymbol{\Sigma}\,(\operatorname{vec}\boldsymbol{\Sigma})' \end{pmatrix}\right]$$

by the central limit theorem for independent identically distributed random vectors (with finite fourth moments). This statement forms the basis for large-sample inference.

4.3. Functions of sample covariances

Define

$$\boldsymbol{s} = \operatorname{vec}\boldsymbol{S}, \qquad \boldsymbol{\sigma} = \operatorname{vec}\boldsymbol{\Sigma}. \tag{4.19}$$

Consider $\boldsymbol{f}(\boldsymbol{s})$, a vector-valued function. Under the usual regularity conditions

$$\begin{aligned} \sqrt{n}\,[\boldsymbol{f}(\boldsymbol{s}) - \boldsymbol{f}(\boldsymbol{\sigma})] &= \frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\sqrt{n}(\boldsymbol{s}-\boldsymbol{\sigma}) + o_{\text{p}}(1) \\ &\xrightarrow{\text{d}} N\left\{0, \frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}[2(1+\kappa)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \kappa\boldsymbol{\sigma}\boldsymbol{\sigma}']\left(\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\right)'\right\}. \end{aligned} \tag{4.20}$$

Functions of the sample covariance matrix are also asymptotically normally distributed.

Note that if $[\partial \boldsymbol{f}(\boldsymbol{\sigma})/\partial\boldsymbol{\sigma}']\boldsymbol{\sigma} = 0$ the covariance matrix in (4.20) is simply a multiple of the covariance matrix when sampling from a normal distribution. Suppose $\boldsymbol{f}(\cdot)$ is scale invariant (homogeneous of degree zero); that is,

$$\boldsymbol{f}(c\boldsymbol{s}) = \boldsymbol{f}(\boldsymbol{s}), \qquad \forall c > 0, \qquad \forall\, \boldsymbol{S}\ \text{p.d.}. \tag{4.21}$$

Then

$$0 = \frac{\partial \boldsymbol{f}(c\boldsymbol{s})}{\partial c} = \frac{\partial \boldsymbol{f}(c\boldsymbol{s})}{\partial \boldsymbol{s}'}\frac{\partial (c\boldsymbol{s})}{\partial c} = \frac{\partial \boldsymbol{f}(c\boldsymbol{s})}{\partial \boldsymbol{s}'}\boldsymbol{s}; \tag{4.22}$$

that is, (for $c = 1$)

$$\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\boldsymbol{\sigma} = 0. \tag{4.23}$$

Then

$$\sqrt{n}[\boldsymbol{f}(\boldsymbol{s}) - \boldsymbol{f}(\boldsymbol{\sigma})] \xrightarrow{\text{d}} N\left\{0,\ 2(1+\kappa)\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\left[\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\right]'\right\}; \tag{4.24}$$

that is,

$$\frac{\sqrt{n}}{\sqrt{1+\kappa}}[\boldsymbol{f}(\boldsymbol{s}) - \boldsymbol{f}(\boldsymbol{\sigma})] \xrightarrow{\text{d}} N\left\{0,\ 2\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\left[\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\right]'\right\}. \tag{4.25}$$

Note that the normal distribution in (4.25) does not depend on κ [that is, $g(\cdot)$].

This result applies to any sequence of random positive definite matrices $\boldsymbol{W}_n$, such that $\boldsymbol{W}_n \xrightarrow{\text{p}} \boldsymbol{\Omega}$ and

$$\sqrt{n}(\operatorname{vec}\boldsymbol{w}_n - \operatorname{vec}\boldsymbol{\omega}) \xrightarrow{\text{d}} N[0, \tau_1(\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp})(\boldsymbol{\Omega} \otimes \boldsymbol{\Omega}) + \tau_2 \boldsymbol{w}\boldsymbol{w}'], \tag{4.26}$$

where $\boldsymbol{w}_n = \operatorname{vec}\boldsymbol{W}_n$ and $\boldsymbol{\omega} = \operatorname{vec}\boldsymbol{\Omega}$. Then

$$\sqrt{n}[\boldsymbol{f}(\boldsymbol{w}_n) - \boldsymbol{f}(\boldsymbol{\omega})] \xrightarrow{\text{d}} N\left\{0, 2\tau_1\frac{\partial \boldsymbol{f}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}'}(\boldsymbol{\Omega} \otimes \boldsymbol{\Omega})\left[\frac{\partial \boldsymbol{f}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}'}\right]'\right\}. \tag{4.27}$$

Tyler (1983) gave the above result in Theorem 1.

Example. Correlation coefficients. Let

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}, \quad \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\ \sigma_{jj}}} \tag{4.28}$$

be the sample and model coefficients. The limit distribution of

$$\frac{\sqrt{n}}{\sqrt{1+\kappa}}(r_{ij} - s_{ij}), \qquad i,j = 1,\cdots,p, \tag{4.29}$$

is the same as for $\boldsymbol{S}$ having a Wishart distribution.

Example. Eigenvectors and ratios of eigenvalues. The eigenvalues of the sample covariance matrix satisfy

$$|\boldsymbol{S} - \lambda\boldsymbol{I}| = 0 \tag{4.30}$$

The eigenvectors satisfy

$$\boldsymbol{S}\boldsymbol{x} = \lambda_i\boldsymbol{x}, \qquad i = 1,\cdots,p. \tag{4.31}$$

For $p = 2$ there is an angle θ such that

$$S\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}. \tag{4.32}$$

The normalized eigenvectors are (cos θ, sin θ) and (− sin θ, cos θ). The angle θ and the ratio of eigenvalues λ_1/λ_2 are scalar invariant. Hence, they have the same asymptotic normal distribution after correcting for the kurtosis as when sampling from the normal distribution.

4.4. Likelihood ratio criteria.

For normal distributions usually

$$-2\log LRC \xrightarrow{\mathrm{d}} \chi_f^2 \tag{4.33}$$

under the null hypothesis H. Consider a scalar function $h(\boldsymbol{s})$ such that

$$h(\boldsymbol{\sigma}) = 0, \quad \frac{\partial h(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'} = \mathbf{0}, \qquad \boldsymbol{\sigma} \in H. \tag{4.34}$$

Then

$$\begin{aligned} nh(\boldsymbol{s}) &= \frac{1}{2}\sqrt{n}(\boldsymbol{s}-\boldsymbol{\sigma})'\frac{\partial^2 h(\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\sqrt{n}(\boldsymbol{s}-\boldsymbol{\sigma}) + o_p(1) \\ &\xrightarrow{\mathrm{d}} \sum_i \nu_i \chi_i^2, \end{aligned} \tag{4.35}$$

where ν_i are the characteristic roots of

$$\frac{1}{2}\frac{\partial^2 h(\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}[2(1+\kappa)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \kappa\boldsymbol{\sigma}\boldsymbol{\sigma}'] \tag{4.36}$$

and χ_i^2 denotes χ^2 with 1 d.f.

Suppose h is scale invariant; that is,

$$h(c\boldsymbol{s}) = h(\boldsymbol{s}), \qquad \forall c > 0, \ \forall \boldsymbol{s} \text{ p.d.} \tag{4.37}$$

Then

$$0 = \frac{\partial^2 h(c\boldsymbol{s})}{\partial c^2} = c^2\boldsymbol{s}'\frac{\partial h(\boldsymbol{s})}{\partial\boldsymbol{s}\partial\boldsymbol{s}'}\boldsymbol{s}. \tag{4.38}$$

For $c = 1$ we obtain

$$\sigma' \frac{\partial^2 h(\sigma)}{\partial \sigma \partial \sigma'} \sigma = 0. \tag{4.39}$$

Here ν_i are the characteristic roots of

$$(1+\kappa) \frac{\partial^2 h(\sigma)}{\partial \sigma \partial \sigma'} (\Sigma \otimes \Sigma). \tag{4.40}$$

If

$$-2 \log LRC = nh(s) \xrightarrow{\mathrm{d}} \sum_i \nu_i \chi_i^2 \tag{4.41}$$

under normality, then for an elliptically contoured distribution with kurtosis κ

$$\frac{-2 \log LRC}{1+\kappa} \xrightarrow{\mathrm{d}} \sum_i \nu_i \chi_i^2. \tag{4.42}$$

Example. Sphericity. Consider the null hypothesis

$$H : \boldsymbol{\Lambda} = \text{const}\, \boldsymbol{I}_p. \tag{4.43}$$

Under normality the likelihood ratio criterion is

$$LRC = \left[\frac{|\boldsymbol{S}|}{\left(\frac{\operatorname{tr} \boldsymbol{S}}{p}\right)^p} \right]^{n/2}, \tag{4.44}$$

which is clearly scale invariant, and

$$-2 \log LRC = n[p \log(\operatorname{tr} \boldsymbol{S}) - \log |\boldsymbol{S}| - p \log p]. \tag{4.45}$$

Then

$$\frac{-2 \log LRC}{1+\kappa} \xrightarrow{\mathrm{d}} \chi_f^2, \tag{4.46}$$

where the number of degrees of freedom is $f = \frac{1}{2}p(p+1) - 1$.

Many hypotheses in multivariate analysis are invariant with respect to some group of linear transformations. For example, the hypothesis (4.43) is invariant with respect to transformations $\boldsymbol{X} \to c\boldsymbol{Q}\boldsymbol{X}$, where $\boldsymbol{Q}$ is orthogonal. If the group of transformations includes multiplication by a constant, the likelihood ratio criterion will satisfy (4.37).

Tyler (1983) has an alternative approach to testing hypotheses. Suppose a null hypothesis is defined by $\boldsymbol{k}(\boldsymbol{\sigma}) = \mathbf{0}$, where $\boldsymbol{k}(\cdot)$ has q components and satisfies the usual regularity condtions and (4.21). A Wald test can be based on

$$\frac{n}{1+\kappa}\boldsymbol{k}(\boldsymbol{s})'\left\{\frac{\partial \boldsymbol{k}(\boldsymbol{s})}{\partial \boldsymbol{s}'}(\boldsymbol{S}\otimes\boldsymbol{S})\left[\frac{\partial \boldsymbol{k}(\boldsymbol{s})}{\partial \boldsymbol{s}'}\right]'\right\}^{-1}\boldsymbol{k}(\boldsymbol{s}). \tag{4.47}$$

Tyler showed that this statistic has a limiting χ^2_q-distribution under the null hypothesis. A function that is asymptotically equivalent to (4.47) is

$$h(\boldsymbol{s}) = \frac{n}{1+k}\boldsymbol{k}(\boldsymbol{s})'\left\{\frac{\partial \boldsymbol{k}(\boldsymbol{s})}{\partial \boldsymbol{s}'}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\left[\frac{\partial \boldsymbol{k}(\boldsymbol{s})}{\partial \boldsymbol{s}'}\right]'\right\}^{-1}\boldsymbol{k}(\boldsymbol{s}), \tag{4.48}$$

which satisfies (4.34).

4.5. Estimation of the kurtosis parameter.

To apply the large-sample distribution theory derived for normal distributions to problems of inference for elliptically contoured distributions it is necessary to know or estimate the kurtosis parameter κ. Note that

$$\begin{aligned}\mathcal{E}[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^2 &= \mathcal{E}(\boldsymbol{Y}'\boldsymbol{Y})^2 \\ &= p\mathcal{E}Y_i^4 + p(p-1)(\mathcal{E}Y_i^2)^2 \\ &= p(3\kappa+p+2). \end{aligned} \tag{4.49}$$

We see that

$$\begin{aligned} M &= \frac{1}{N}\sum_{\alpha=1}^{N}[(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})]^2 \\ &\xrightarrow{\mathrm{p}} p(3\kappa+p+2) \end{aligned} \tag{4.50}$$

and

$$\frac{M-p(p+2)}{3p}\xrightarrow{\mathrm{p}}\kappa. \tag{4.51}$$

Mardia (1970) proposed the left-hand side of (4.51), say $\hat{\kappa}$, as a consistent estimator of κ. The convergence in (4.25) and (4.42) is valid when κ is replaced by the estimator $\hat{\kappa}$.

5. Estimation of Covariance Parameters.

5.1. Maximum likelihood estimation

We have considered using $\boldsymbol{S}$ as an estimator of $\boldsymbol{\Sigma} = (\mathcal{E}R^2/p)\boldsymbol{\Lambda}$. When the parent distribution is normal, $\boldsymbol{S}$ is the sufficient statistic invariant with respect to translations and hence is the efficient unbiased estimator. Now we study other estimators.

We consider first the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ when the form of the density $g(\cdot)$ is known. The logarithm of the likelihood function is

$$\log L = -\frac{N}{2}\log|\boldsymbol{\Lambda}| + \sum_{\alpha=1}^{N}\log g[(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})]. \tag{5.1}$$

The derivatives of $\log L$ with respect to the components of $\boldsymbol{\mu}$ are

$$\frac{\partial \log L}{\partial \boldsymbol{\mu}} = -2\sum_{\alpha=1}^{N}\frac{g'[(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})]}{g[(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})]}\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu}). \tag{5.2}$$

Setting this vector of derivatives to $\mathbf{0}$ leads to the equation

$$\sum_{\alpha=1}^{N}\frac{g'[(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})]}{g[(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})]}\boldsymbol{x}_\alpha = \hat{\boldsymbol{\mu}}\sum_{\alpha=1}^{N}\frac{g'(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})]}{g[(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})]}. \tag{5.3}$$

Setting to 0 the derivatives of $\log L$ with respect to the elements of Λ^{-1} gives

$$\hat{\boldsymbol{\Lambda}} = -\frac{2}{N}\sum_{\alpha=1}^{N}\frac{g'[(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})]}{g[(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})]}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'. \tag{5.4}$$

The estimator $\hat{\boldsymbol{\Lambda}}$ is a kind of weighted average of the rank 1 matrices $(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}})'$. In the normal case [$g(\boldsymbol{y})$ given by (4.3)] the weights are $1/N$. In most cases (5.3) and (5.4) cannot be solved explicitly, but the solution may be approximated by iterative methods.

The covariance matrix of the limiting normal distribution of $\sqrt{N}(\operatorname{vec}\hat{\boldsymbol{\Lambda}} - \operatorname{vec}\boldsymbol{\Lambda})$ is

$$\operatorname{Cov}(\operatorname{vec}\hat{\boldsymbol{\Lambda}}) = \sigma_{1g}(\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp})(\boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda}) + \sigma_{2g}\operatorname{vec}\boldsymbol{\Lambda}(\operatorname{vec}\boldsymbol{\Lambda})', \tag{5.5}$$

where

$$\sigma_{1g} = \frac{p(p+2)}{4\mathcal{E}\left[\frac{g'(R^2)}{g(R^2)}R^2\right]^2}, \tag{5.6}$$

$$\sigma_{2g} = -\frac{2\sigma_{1g}(1-\sigma_{1g})}{2+p(1-\sigma_{1g})}. \tag{5.7}$$

See Tyler (1982).

Example. Multivariate t. If the density $g(\boldsymbol{y}'\boldsymbol{y})$ is given by (3.36), then

$$\sigma_{1g} = \frac{p+m+2}{p+m}, \qquad \sigma_{2g} = \frac{2}{m}\sigma_{1g}. \tag{5.8}$$

Note that $1+\kappa = (m-2)/(m-4)$; that is, $\kappa = 2/(m-4)$. As $m \to \infty$, $\kappa \to 0$, $\sigma_{1g} \to (p+2)/p$, and $\sigma_{2g} \to 0$; these are the values for $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

5.2. Robust estimators.

Maronna (1976) has studied robust estimators or M-estimators. Set

$$d_\alpha^2 = (\boldsymbol{x}_\alpha - \tilde{\boldsymbol{\mu}})'\boldsymbol{V}^{-1}(\boldsymbol{x}_\alpha - \tilde{\boldsymbol{\mu}})' \tag{5.9}$$

for a vector $\tilde{\boldsymbol{\mu}}$ and a positive definite matrix $\boldsymbol{V}$. Suppose that $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{V}$ also satisfy

$$\frac{1}{N}\sum_{\alpha=1}^{N} u_1(d_\alpha)(\boldsymbol{x}_\alpha - \tilde{\boldsymbol{\mu}}) = \boldsymbol{0}, \tag{5.10}$$

$$\frac{1}{N}\sum_{\alpha=1}^{N} u_2(d_\alpha^2)(\boldsymbol{x}_\alpha - \tilde{\boldsymbol{\mu}})(\boldsymbol{x}_\alpha - \tilde{\boldsymbol{\mu}})' = \boldsymbol{V} \tag{5.11}$$

for $u_1(d)$ and $u_2(d^2)$ nonnegative, nondecreasing, and continuous for $d \geq 0$ such that $du_1(d)$ and $d^2u_2(d^2)$ are bounded. [Maronna (1976) gives two other conditions on $u_1(\cdot)$ and $u_2(\cdot)$.] Then $\tilde{\boldsymbol{\mu}}$ estimates $\boldsymbol{\mu} = \mathcal{E}\boldsymbol{X}$ and $\boldsymbol{V}$ estimates $\frac{1}{\gamma}\boldsymbol{\Lambda} = \boldsymbol{\Omega}$, say, where γ satisfies

$$\mathcal{E}\gamma R^2 u_2(\gamma R^2) = p. \tag{5.12}$$

These estimators have an asymptotic normal distribution. The covariance matrix of the limiting distribution of $\sqrt{N}[\text{vec}\,\boldsymbol{V} - \text{vec}\,\boldsymbol{\Omega}]$ has the same form as (4.17) and (5.5); it is

$$\sigma_{1u}(\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp})(\boldsymbol{\Omega} \otimes \boldsymbol{\Omega}) + \sigma_{2u}\,\text{vec}\,\boldsymbol{\Omega}(\text{vec}\,\boldsymbol{\Omega})', \tag{5.13}$$

where

$$\sigma_{1u} = \frac{(p+2)^2\psi_1}{(2\psi_1+p)^2}, \quad \sigma_{2u} = \frac{\psi_1 - 1}{\psi_2^2} - 2\frac{\psi_1(\psi_2-1)[(p+4)\psi_2+p]}{\psi_2^2(2\psi_2+p)^2}, \tag{5.14}$$

$$\psi_1 = \frac{\mathcal{E}[\gamma R^2 u_2(tR^2)]^2}{p(p+2)}, \tag{5.15}$$

$$\psi_2 = \frac{\gamma \mathcal{E} R^2 [u_2(\gamma R^2) + \gamma R^2 u_2'(\gamma R^2)]}{p}. \tag{5.16}$$

See Tyler (1982). Note that if in (5.13) we replace $\boldsymbol{V}$ by $\gamma \boldsymbol{V}$ and $\boldsymbol{\Omega}$ by $\gamma \boldsymbol{\Omega} = \boldsymbol{\Lambda}$, the coefficients σ_{1u} and σ_{2u} are unchanged.

Tyler (1983) has given a table of values of $\sigma_1\,(= \sigma_{1g}$ or $\sigma_{1u})$, $\sigma_2\,(= \sigma_{2g}$ or $\sigma_{2u})$, and γ for several estimators and several elliptically contoured distributions; part of his table is reproduced as Table 1 below. The estimators include maximum likelihood for the multivariate t-distribution with 1 and 5 degrees of freedom [ML:T(1) and ML:T(5)] and Huber-type estimates with $u_1(\cdot)$ and $u_2(\cdot)$ defined by

$$u_1(d) = \begin{cases} 1, & d \leq r, \\ \frac{r}{d}, & d > r, \end{cases} \tag{5.17}$$

$$\beta u_2(d^2) = \begin{cases} 1, & d^2 \leq r^2, \\ \frac{r^2}{d^2}, & d^2 > r^2. \end{cases} \tag{5.18}$$

A Huber-type estimator is denoted by HUB(q), where $q = \Pr\{\chi_p^2 > r\}$ and β is determined by $\mathcal{E}\chi_p^2 u_2(\chi_p^2) = p$. The distributions are the multivariate t-distributions with 1 and 5 degrees of freedom, the contaminated normal (CN) with $\varepsilon = .1$ and $c = 9$, and the normal.

Table 1

$p = 2$

	$T(1)$			$T(5)$			CN			Normal		
	σ_1	σ_2	γ	σ_1	σ_2	γ	σ_1	σ_2	γ	σ_1	σ_2	γ
ML:T(1)	1.67	3.33	1.00	1.48	0.89	1.75	1.45	0.69	1.73	1.43	0.46	2.02
ML:T(5)	2.28	5.84	0.28	1.29	0.51	1.00	1.28	0.51	1.03	1.11	0.05	1.31
HUB(.5)	1.70	3.73	0.65	1.50	1.41	0.92	1.46	1.10	0.89	1.44	0.98	1.00
HUB(.1)	2.15	5.10	0.29	1.32	0.57	0.78	1.23	0.36	0.83	1.09	0.09	1.00
$\boldsymbol{S}$	∞	∞	–	3.00	2.00	0.60	2.77	1.77	0.56	1.00	0.00	1.09

$p = 10$

	σ_1	σ_2	γ	σ_1	σ_2	γ	σ_1	σ_2	γ	σ_1	σ_2	γ
ML:T(1)	1.18	2.37	1.10	1.17	0.50	1.17	1.16	0.18	1.10	1.16	0.05	1.22
ML:T(5)	1.28	2.89	0.60	1.13	0.45	1.00	1.11	0.22	1.00	1.09	0.02	1.15
HUB (.5)	1.23	2.68	1.09	1.15	0.63	1.07	1.09	0.16	0.95	1.08	0.12	1.00
HUB (.1)	1.48	3.55	0.50	1.21	0.50	0.85	1.07	0.11	0.91	1.01	0.01	1.08
S	∞	∞	-	3.00	2.00	0.60	2.77	1.77	0.56	1.00	0.00	1.00

The two values, σ_1 and σ_2, for the maximum likelihood estimator ML:T(1) are the smallest for the distribution T(1) although the values for HUB(.5) are only slightly larger. Similarly, the values for ML:T(5) are the smallest for T(5), but the values for ML:T(1) and HUB(.1) are close. The values for HUB(.1) are smallest for CN. Of course, $\boldsymbol{S}$ is best for the normal and HUB(.1) is close. $\boldsymbol{S}$ is not a valid estimator for T(1) because the second moment of $\boldsymbol{X}$ does not exist, and $\boldsymbol{S}$ is not accurate for T(5) and CN. We see that $\boldsymbol{S}$ is not a very robust estimator.

6. Spherical Matrix Distributions

The observations $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ constitute an $N \times p$ matrix

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_N' \end{bmatrix}. \tag{6.1}$$

Consider an $N \times p$ random matrix $\boldsymbol{Y}$. We define the following classes of matrices:

Left-spherical

$$\boldsymbol{Q}_N \boldsymbol{Y} \stackrel{\mathrm{d}}{=} \boldsymbol{Y} \qquad \forall\, \boldsymbol{Q}_N, \tag{6.2}$$

Right-spherical

$$\boldsymbol{Y} \boldsymbol{Q}_p \stackrel{\mathrm{d}}{=} \boldsymbol{Y} \qquad \forall\, \boldsymbol{Q}_p, \tag{6.3}$$

Vector-spherical

$$\boldsymbol{Q}_{Np} \operatorname{vec} \boldsymbol{Y} \stackrel{\mathrm{d}}{=} \operatorname{vec} \boldsymbol{Y} \qquad \forall\, \boldsymbol{Q}_{Np}, \tag{6.4}$$

where $\boldsymbol{Q}_m$ denotes an orthogonal matrix of order m.

If $\boldsymbol{Y}$ is vector-spherical and has a density, it is also left-spherical and right-spherical and $\boldsymbol{Y}'$ is also vector-spherical because the density has the form

$$\begin{aligned} g[(\operatorname{vec}\boldsymbol{Y})'\operatorname{vec}\boldsymbol{Y}] &= g\left(\sum_{\alpha=1}^{N}\sum_{i=1}^{p} y_{i\alpha}^2\right) = g(\operatorname{tr}\boldsymbol{Y}\boldsymbol{Y}') \\ &= g(\operatorname{tr}\boldsymbol{Y}'\boldsymbol{Y}) = g[(\operatorname{vec}\boldsymbol{Y}')'\operatorname{vec}\boldsymbol{Y}'], \end{aligned} \tag{6.5}$$

An example is the case of all of the elements of $\boldsymbol{Y}$ being independent $N(0,1)$ variables; in that case $g(w) = (2\pi)^{-pN/2}e^{-w/2}$.

Define

$$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{A}' + \boldsymbol{\epsilon}_N\boldsymbol{\mu}', \tag{6.6}$$

where $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Lambda}$ and $\boldsymbol{\epsilon}_N' = (1,\cdots,1)$. Since (6.6) is equivalent to $\boldsymbol{Y} = (\boldsymbol{X} - \boldsymbol{\epsilon}_N\boldsymbol{\mu}')(\boldsymbol{A}')^{-1}$, and $(\boldsymbol{A}')^{-1}\boldsymbol{A}^{-1} = \boldsymbol{\Lambda}^{-1}, \boldsymbol{X}$ has the density

$$|\boldsymbol{\Lambda}|^{-N/2}g[\operatorname{tr}(\boldsymbol{X} - \boldsymbol{\epsilon}_N\boldsymbol{\mu}')'\boldsymbol{\Lambda}^{-1}(\boldsymbol{X} - \boldsymbol{\epsilon}_N\boldsymbol{\mu}')] = |\boldsymbol{\Lambda}|^{-N/2}g\left[\sum_{\alpha-1}^{N}(\boldsymbol{X}_\alpha - \boldsymbol{\mu})\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}_\alpha - \boldsymbol{\mu})\right]. \tag{6.7}$$

From (6.5) we deduce that $\operatorname{vec}\boldsymbol{Y}$ has the representation

$$\operatorname{vec}\boldsymbol{Y} \stackrel{\mathrm{d}}{=} R\operatorname{vec}\boldsymbol{U}, \tag{6.8}$$

where $w = R^2$ has the density

$$\frac{\pi^{\frac{1}{2}Np}}{\Gamma(Np/2)}\omega^{\frac{1}{2}NP-1}g(\omega), \tag{6.9}$$

$\operatorname{vec}\boldsymbol{U}$ has the uniform distribution on $\sum_{\alpha-1}^{N}\sum_{i=1}^{p}u_{i\alpha}^2 = 1$, and R and $\operatorname{vec}\boldsymbol{U}$ are independent. The covariance matrix of $\operatorname{vec}\boldsymbol{Y}$ is

$$\mathcal{E}\operatorname{vec}\boldsymbol{Y}(\operatorname{vec}\boldsymbol{Y})' = \frac{\mathcal{E}R^2}{Np}\boldsymbol{I}_{Np} = \frac{\mathcal{E}R^2}{Np}(\boldsymbol{I}_p\otimes\boldsymbol{I}_N), \tag{6.10}$$

Since $\operatorname{vec}\boldsymbol{F}\boldsymbol{G}\boldsymbol{H} = (\boldsymbol{H}'\otimes\boldsymbol{F})\operatorname{vec}\boldsymbol{G}$ for any conformable matrices $\boldsymbol{F},\boldsymbol{G}$, and $\boldsymbol{H}$, we can write (6.6) as

$$\operatorname{vec}\boldsymbol{X} = (\boldsymbol{A}\otimes\boldsymbol{I}_N)\operatorname{vec}\boldsymbol{Y} + \boldsymbol{\mu}\otimes\boldsymbol{\epsilon}_N. \tag{6.11}$$

Thus

$$\begin{aligned}
\mathcal{E}\,\mathrm{vec}\,\boldsymbol{X} &= \boldsymbol{\mu}\otimes\boldsymbol{\epsilon}_N, && (6.12)\\
\mathrm{Cov}\,(\mathrm{vec}\,\boldsymbol{X}) &= (\boldsymbol{\Lambda}\otimes\boldsymbol{I}_N)\,\mathrm{cov}\,(\mathrm{vec}\,\boldsymbol{Y})(\boldsymbol{\Lambda}'\otimes\boldsymbol{I}_N)\\
&= \boldsymbol{\Lambda}\otimes\boldsymbol{I}_N, && (6.13)\\
\mathcal{E}(\text{ row of }\boldsymbol{X}) &= \boldsymbol{\mu}', && (6.14)\\
\mathrm{Cov}(\text{row of }\boldsymbol{X}) &= \frac{\mathcal{E}R^2}{Np}\boldsymbol{\Lambda}. && (6.15)
\end{aligned}$$

The rows of $\boldsymbol{X}$ are uncorrelated (though not necessarily independent). From (6.11) we obtain

$$\begin{aligned}
\mathrm{vec}\,\boldsymbol{X} &\overset{\mathrm{d}}{=} R(\boldsymbol{\Lambda}\otimes\boldsymbol{I}_N)\mathrm{vec}\,\boldsymbol{U}+\boldsymbol{\mu}\otimes\boldsymbol{\epsilon}_N, && (6.16)\\
\boldsymbol{X} &\overset{\mathrm{d}}{=} R\boldsymbol{U}\boldsymbol{\Lambda}'+\boldsymbol{\epsilon}_N\boldsymbol{\mu}'. && (6.17)
\end{aligned}$$

Since $\boldsymbol{X}-\boldsymbol{\epsilon}_N\boldsymbol{\mu}' = (\boldsymbol{X}-\boldsymbol{\epsilon}_N\bar{\boldsymbol{x}}')+\boldsymbol{\epsilon}_N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'$, we can write the density of $\boldsymbol{X}$ as

$$|\boldsymbol{\Lambda}|^{-N/2}g[\mathrm{tr}\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\epsilon}_N\bar{\boldsymbol{x}}')'(\boldsymbol{X}-\boldsymbol{\epsilon}_N\bar{\boldsymbol{x}}')+N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})], \qquad (6.18)$$

where $\bar{\boldsymbol{x}} = (1/N)\boldsymbol{X}'\boldsymbol{\epsilon}_N$. This shows that a sufficient set of statistics for $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ is $\bar{\boldsymbol{x}}$ and $n\boldsymbol{S} = (\boldsymbol{X}-\boldsymbol{\epsilon}_N\bar{\boldsymbol{x}}')'(\boldsymbol{X}-\boldsymbol{\epsilon}_N\bar{\boldsymbol{x}}')$ as for the normal distribution.

The maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ are

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}, \qquad (6.19)$$

$$\hat{\boldsymbol{\Lambda}} = \frac{p}{w_g}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})', \qquad (6.20)$$

where w_g maximizes $w^{Np/2}g(w)$ [Anderson, Fang, and Hsu (1986)]. Note that $\hat{\boldsymbol{\mu}}$ is the same estimator as for the normal and $\hat{\boldsymbol{\Lambda}}$ is a multiple of the estimator of $\boldsymbol{\Lambda}$ in the normal case.

Theorem. Let $\boldsymbol{f}(\boldsymbol{X})$ be a vector-valued function of $\boldsymbol{X}$ such that

$$\boldsymbol{f}(\boldsymbol{X}+\boldsymbol{\epsilon}_N\boldsymbol{\nu}') = \boldsymbol{f}(\boldsymbol{X}) \qquad \forall\,\boldsymbol{\nu}, \qquad (6.21)$$

and

$$\boldsymbol{f}(c\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X}) \qquad \forall\,c\neq 0. \qquad (6.22)$$

Then the distribution of $\boldsymbol{f}(\boldsymbol{X})$, where $\boldsymbol{X}$ has the arbitrary density (6.7), is the same as the distribution of $\boldsymbol{f}(\boldsymbol{X})$, where $\boldsymbol{X}$ has the normal density (6.7).

Proof. Substitution of the representation (6.11) into $\boldsymbol{f}(\boldsymbol{X})$ gives

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{X}) &= \boldsymbol{f}[R(\boldsymbol{A} \otimes \boldsymbol{I}_N)\mathrm{vec}\,\boldsymbol{U} + \boldsymbol{\mu} \otimes \boldsymbol{\epsilon}_N] & (6.23) \\ &= \boldsymbol{f}[R(\boldsymbol{A} \otimes \boldsymbol{I}_N)\mathrm{vec}\,\boldsymbol{U}] & (6.24) \end{aligned}$$

by (6.20) and

$$f(\boldsymbol{X}) = f[(\boldsymbol{A} \otimes \boldsymbol{I}_n)\mathrm{vec}\,\boldsymbol{U}] \tag{6.25}$$

by (6.21). □

Any statistic satisfying (6.20) and (6.21) has the same distribution for all $g(\cdot)$. Hence, if its distribution is known for the normal case, the distribution is valid for all elliptically contoured distributions.

Anderson and Fang (1990b) gave the examples of the correlation coefficients and the multiple correlation coefficient. They also showed that when $\boldsymbol{\mu} = \boldsymbol{0}$ the distribution of Hotelling's $T^2 = N\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\boldsymbol{x}$ does not depend on $g(\cdot)$. Any likelihood ratio criterion under normality that is scale invariant and location-invariant in the sense of (6.20) has the same distribution for $g(\cdot)$. The sphericity criterion is an example.

Any function of the sufficient set of statistics that is translation invariant, that is, that satisfies (6.20), is a function of $\boldsymbol{S}$. Thus inference concerning $\boldsymbol{\Sigma}$ can be based on $\boldsymbol{S}$.

Anderson, Fang, and Hsu (1986) considered the likelihood ratio criterion for testing the null hypothesis $(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \in \omega$ in the model $(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \in \Omega$. Suppose $(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \in \omega$ implies $(\boldsymbol{\mu}, c\boldsymbol{\Lambda}) \in \omega$ and $(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \in \Omega$ implies $(\boldsymbol{\mu}, c\boldsymbol{\Lambda}) \in \Omega\ \forall\, c > 0$. Then the likelihood ratio criterion for arbitrary $g(\cdot)$ is the same as the LRC for normal $g(\cdot)$.

Suppose further that $\omega = \omega_m \times \omega_l$, $\Omega = \Omega_m \times \Omega_l$, $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \omega_m$ implies $\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 \in \omega_m$, $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \Omega_m$ implies $\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 \in \Omega$, $\boldsymbol{\mu} \in \omega_m$ implies $c\boldsymbol{\mu} \in \omega_m$, and $\boldsymbol{\mu} \in \Omega_m$ implies $c\boldsymbol{\mu} \in \Omega_m\ \forall\, c$. Then if the distribution of the LRC does not depend on $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ under normality, it does not depend on $g(\cdot)$ or on $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. Anderson, Fang, and Hsu gave several examples including the test for lack of correlation between sets of variates. They pointed out that the result also applies to tests of equality of several covariance matrices.

This class of vector elliptically contoured distributions shows that the sampling theory for the normal distribution is valid for a much wider class of distributions. Several papers in Fang and Anderson (1990) show that many properties of the normal can be extended to this class. The disadvantage of these models is that except for the normal the observations are dependent, though uncorrelated. The advantage is that the similarity to the normal is exact rather than asymptotic.

7. References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.

Anderson, T. W., and Kai-Tai Fang (1990a), On the theory of multivariate elliptically contoured distributions and their applications, in *Statistical Inference in Elliptically Contoured and Related Distributions* (Kai-Tai Fang and T. W. Anderson, eds.), Allerton Press, Inc., New York, 1990, 1–23.

Anderson, T. W., and Kai-Tai Fang (1990b), Inference in multivariate elliptically contoured distributions based on maximum likelihood, in *Statistical Inference in Elliptically Contoured and Related Distributions* (Kai-Tai Fang and T. W. Anderson, eds.), Allerton Press, Inc., New York, 1990, 201–216.

Anderson, T. W., Kai-Tai Fang, and Huang Hsu (1986), Maximum-likelihood estimates and likelihood ratio criteria for multivariate elliptically contoured distributions, *The Canadian Journal of Statistics,* **14**, 55–59.

Bartlett, M. S. (1934). The vector representation of a sample. *Proceedings of the Cambridge Philosophical Society,* **30**, 327-340.

Chmielewski, M. A. (1981). Elliptically symmetric distributions: A review and bibliography. *International Statistical Review,* **49**, 67-74.

Fang, Kai-Tai, and T. W. Anderson, eds. (1990). *Statistical Inference in Elliptically Contoured and Related Distributions,* Allerton Press, Inc., New York.

Fang, Kai-Tai, and Yao-Ting Zhang (1990) *Generalized Mutivariate Analysis*, Springer-Verlag, Berlin.

Hartman, P., and A. Wintner (1940). On the spherical approach to the normal distribution law. *American Journal of Mathematics*, **62**, 759-779.

Kelker, Douglas (1970), Distribution theory of spherical distributions and a location scale parameter generalization, *Sankhyā*, Series A, **32**, 419–430.

Magnus, J. R., and H. Neudecker (1979). The commutation matrix: some properties and applications *Annals of Statistics* **7**, 381–394.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika,* **57**, 519–530.

Maronna, Ricardo Antonio (1976). Robust M-estimators of multivariate location and scatter, *Annals of Statistics,* **4**, 51–67.

Maxwell, J. C. (1860). Illustration of the dynamical theory of gases – Part I. On the motions and collisions of perfectly elastic bodies. *Tayler's philosophical Magazine.* **19**, 19-32.

Muirhead, Robb J., and C. M. Waternaux (1980). Asympototic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika,* **37**, 31–43.

Tyler, David E. (1982). Radial estimates and the test for sphericity, *Biometrika,* **69**, 429–436.

Tyler, David E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika,* **70**, 411–420.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
July 1992

3. REPORT TYPE AND DATES COVERED
Technical Report

4. TITLE AND SUBTITLE
Nonnormal Multivariate Distributions: Inference Based on Elliptically Contoured Distributions

5. FUNDING NUMBERS
ARO
DAAL03-89-K-0033

6. AUTHOR(S)
T. W. Anderson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Statistics
Sequoia Hall
Stanford University
Stanford, CA 94305-4065

8. PERFORMING ORGANIZATION REPORT NUMBER
28

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
28

11. 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.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The class of elliptically contoured distributions, which includes multivariate t-distributions and contaminated normal distributions, serves as a useful generalization of the class of normal multivariate distributions. The density, marginal and conditional densities, and moments of an elliptically contoured distribution are related in a simple fashion to those of a normal distribution. The asymptotic normal distributions of the sample mean and covariance matrix are developed and are compared with the asymptotic distributions of the maximum likelihood estimators of the parameters of an elliptically contoured distribution. The class of elliptically contoured distributions serves as a model for evaluating other robust estimators. Many test procedures for normal distributions are easily modified for the elliptically contoured distributions. Further generalizations are discussed.

14. SUBJECT TERMS
Nonnormal multivariate distributions, elliptically contoured distributions, multivariate t-distributions, robust estimators.

15. NUMBER OF PAGES
27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL