1984

Contributions To Multivariate Analysis Based On Elliptic T Model

Brajendra Chandra Sutradhar

CONTRIBUTIONS TO MULTIVARIATE ANALYSIS BASED ON ELLIPTIC T MODEL

by

Brajendra Chandra Sutradhar

Department of Statistical and Actuarial Sciences

Submitted in partial fulfillment
of the requirements for the degrée of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
July, 1984

ABSTRACT

Theoretical developments in multiviarate analysis are primarily based on the assumption of multivariate normality and very little is known for other cases. The aim of the present work is to generalize results of multivariate analysis based on a class of elliptic distributions, more specifically the subclass of the multivariate t-distributions with suitable parameters, rather than the usual normality assumption. The multivariate normal distribution belongs to the class of elliptic as well as the subclass of t-distribution.

The major contributions of the thesis are:

(a) An elliptic set-up for uncorrelated samples is proposed.

(b) The distributions of sample mean and covariance matrix are derived.

(c) Classification problem is studied for the elliptic set-up.

The elliptic class is further specialized to a subclass of multivariate t-distributions, whose characteristic function, conditional distributions etc. are derived. Also the above problems (b) and (c) are studied. In addition the following problems have been solved:

(i) null and non-null distributions of quadratic forms (analogue of non-cnetral chi-square).

(ii) estimation of location, scale and degrees of freedom parameters of the t-distribution and the sampling properties of the estimators.

(iii) orthogonal factor analysis when both observed error and unobserved factors follow multivariate t-distributions

(iv) estimation of parameters and testing of hypothesis for a regression model with error variable having a multivariate t-distribution.

ACKNOWLEDGEMENTS

I am greatly indebted to Professor Mir Maswood Ali, my research supervisor, for suggesting the problem, his constant encouragement and his keen interest in my work. His valuable comments and helpful suggestions have considerably improved the present thesis.

I would like to express my thanks to Professors V.M. Joshi and M.S. Haq for making valuable suggestions and comments on the preliminary draft of the thesis.

I wish to thank the University of Western Ontario for financial assistance during my graduate studies. I would also like to extend my thanks to Joyce Collins for her excellent typing of a most difficult manuscript.

Finally, I acknowledge the patient support of my family throughout.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Historical Background

Theoretical developments in the area of classical multivariate analysis are primarily based on the assumption that the underlying distributions are multivariate normal. The unfortunate feature of this branch of statistics is that in other cases very little is known about the exact distribution theory. One of the important properties of the multivariate normal distribution is that the distribution has constant probability density function (p.d.f.) or what are called equiprobable surfaces on ellipsoids. This property is exploited extensively in theoretical work. Another important property is that a suitable linear transformation transforms the multivariate normal with elliptic contours of equiprobable surfaces to the special case of multivariate normal distribution with spherical equiprobable surfaces or simply to the spherical normal distributions whose components are independent. In many theoretical works this transformation is also extensively used to study distribution of statistics etc., since spherical normal distributions are rotationally invariant.

In the recent literature, as a generalization of the multivariate normal model, considerable amount of attention has been focused on the class of distributions sharing the property of having equiprobable surfaces on homothetic ellipsoids. These distributions are usually referred to as elliptically symmetric or simply elliptic distributions. This class of distributions has been used in many theoretical contexts, especially in robustness studies where it is of interest to know the behaviour of, say, some tests developed under normality assumptions in non-normal situations restricting the non-normality to members of the class of elliptic distributions of the continuous type.

Many properties of spherical distribution have been studied by several authors in connection with robustness studies. For example, Efron (1969) shows that Student's t-distribution remains unchanged if we assume an underlying spherically symmetric distribution rather than a normal distribution. Box (1952) notes that the usual F-statistics have the same null distribution for all spherically symmetric distributions. This robust property of F-statistics has been proved by Thomas (1970), Kariya and Eaton (1977) more specifically in connection with the study of linear models. Among others Kelker (1970) studies several important properties of elliptically symmetric distribution such as conditions for the existence of a probability density function, the marginal distributions, the conditional means, and several characterizations of the normal distribution among this class.

Elliptic distributions have found applications in practical as well as theoretical contexts in several areas. In the early nineteenth century Gauss encountered the so-called Gaussian distribution which belongs to the class of spherically symmetric distributions, in connection with the study of the theory of errors of observations in astronomy. Maxwell (1860) encountered the same model in connection with the study of velocities of molecules [see for example Section 2.1-2.2, Mathai and Pedrzoli (1977)].

Many authors, for example, Maxwell (1860), Bartlett (1934), and Hartman and Wintner (1940), Kelker (1970), Thomas (1970), and Nash and Klamkin (1976) have discussed the following important characterization of the normal law: If $Z = (z_1 \ldots, z_n)$ is a sample of n independently and identically distributed random variables, then Z has a spherically symmetric distribution if and only if Z has a Gaussian distribution.

Lord (1954) discusses the moments, cumulants, and characteristic function of spherical distributions. He also discusses the multivariate normal; Cauchy and exponential distributions. Furthermore he reported an application of the theory of spherical distributions in some problems of astronomy. He notes that the projected spherical distribution on a subspace of lower dimensions is also a spherical distribution. Hence the distribution of a globular star-cluster is often spherical and can be inferred from its projection on the 'plane of the sky'.

Box (1953) as well as Mandelbrot (1963) use the spherical distribution model in connection with price-change data. Mehta (1967) uses the model in quantum theory; Gilliland (1968) shows the use of the same distribution model for impact distributions in bombing.

Fraser (1951) finds the probability of at least one hit when an automatic gun is used against a moving target. On the assumption that the target is a Gaussian diffuse target (k-dimensional) and the position of the trajectory is distributed according to a Gaussian distribution about the point of aim, he shows that the probability of a 'hit' as a function of the point of aim also has the form of a Gaussian diffuse distribution. That is, it is a constant times a Gaussian probability density function of the point of aim. The point of aim is a random variable in k-dimensions and is called a prediction according to him.

Moreover, values under any two different co-ordinates for the k-dimensional predictions are assumed to be independent; and the target has been assumed to have a circular outline. Thus he uses the spherically symmetric distribution model to find the probability of at least one hit when an automatic gun is used against a moving target.

Dawid (1978) gives characterization of left spherical, right spherical and spherical distributions of random matrices which are submatrices of arbitrary large arrays sharing the same orthogonal invariance.

Zellner (1976) considers a single variable regression model with multivariate t-error distribution with known degrees of freedom (ν_o). Thus the response variable in Zellner's model has multivariate t-distribution, which is a special case of the elliptically symmetric distribution. Dunnett and Sobel (1954) also encountered a multivariate t-model and they encountered the model in the context of certain multiple decision problems.

Chmielewski (1980a,b) considers tests involving the scale matrix Σ. Some cases considered are equality of k scale matrices, sphericity, block diagonal structure, and equicorrelatedness. In all cases the null distributions of certain invariant test statistics are exactly those found under normal theory. For tests of equality and sphericity the non-null distributions are also those derived under normal theory.

Box (1953) states that those tests which are uniformly most powerful under normality assumptions are also uniformly most powerful for any spherical distribution. Recently Kariya (1980, Vol. 1) shows that Hotelling's T^2-test for testing location parameter $\theta = 0$ in the one-sample problem is robust against departures from normality. It is still UMPI in an elliptical class of distributions and the null distribution under any member of the elliptic distributions is the same as that under normality. In a separate study Kariya (1981, No. 6) gives necessary and sufficient conditions for the null distribution of a test statistic to remain the same in the class of elliptically symmetric distributions. He also shows that in certain special cases, the usual MANOVA tests are still uniformly most powerful invariant in the class of elliptically symmetric distributions.

Much of what has been discussed so far relates to distribution theory, statistical inference, physical science problems, econometrics & to the study of linear models. Elliptically symmetric distributions, however, have been considered in several other areas such as minimax estimation, stochastic processes, pattern recognition, fiducial inference, and probability inequalities. An excellent bibliography on elliptically symmetric distribution with its applications in these areas is to be found in Chmielewski (1981). In view of these reviews, it has not been felt necessary to survey these areas any further in the present thesis.

1.2 Elliptically Symmetric Distributions

A p-component random vector X is said to have a p-dimensional elliptic symmetric distribution if the probability density of X is constant on every homothetic ellipsoidal surface centred around a fixed point say $\Theta = (\Theta_1 \ldots, \Theta_p)$ [see for example Cramer, H. (1946), p. 288]. More specifically X is said to be elliptically distributed with location parameter Θ and a pxp positive definite scale matrix Λ (not necessarily covariance matrix), if the probability density function of X can be written in the form

$$f(x,\Theta,\Lambda) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\Theta)^T\Lambda^{-1}(x-\Theta)\} \tag{1.1}$$

[see for example Kelker (1970)], where K_p is a normalising constant, y^T stands for transpose of y.

If $\Lambda = I$ in (1.1), then X is said to be spherically distributed around the centre $\Theta = (\Theta_1, \ldots, \Theta_p)$. Thus the transition from elliptically symmetric to spherically symmetric and vice versa are made through appropriate affine transformations. The transformation $\Lambda^{-\frac{1}{2}}(x-\Theta) = Z$ in (1.1) yields the p.d.f. of the spherically distributed random variable Z of the form

$$f(z) = K_p \, g^{(Z^T Z)} \qquad (1.2)$$

Conversely, the affine transformation $x = \Theta + \Lambda^{\frac{1}{2}} Z$ in (1.2) produces the elliptically symmetric distribution of x having the probability density function (p.d.f.) of the form given in (1.1).

The elliptic random variable X with p.d.f. (1.1) and the spherical random variable Z with p.d.f. (1.2) are absolutely continuous variable. In what follows we will confine our discussion to the case of distribution of the absolutely continuous type.

The following are some simple examples of spherically and elliptically symmetric distributions:

(i) The standardized multinormal distribution with p.d.f.

$$f(z) = (2\pi\sigma^2)^{-\frac{p}{2}} \ \mathrm{Exp}\{-\frac{1}{2\sigma^2} z^T z\}$$

(ii) The standardized multivariate t-distribution with ν degrees of freedom and p.d.f.

$$f(z) = \frac{\nu^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}} \Gamma\left(\frac{\nu}{2}\right)} \{\nu + z^T z\}^{-\frac{\nu+p}{2}}$$

(iii) The standardized multivariate Cauchy distribution with parameter 'a' and p.d.f.

$$f(z) = C_p(a^2 + z^T z)^{-\frac{p+1}{2}},$$

where C_p is a constant

(iv) The standardized multivariate exponential distribution with parameter 'b' and p.d.f.

$$f(z) = \pi^{\frac{1-p}{2}} (2b)^{-p} \operatorname{Exp}\{-(z^T z)^{\frac{1}{2}}/b\}/\Gamma\left(\frac{p+1}{2}\right)$$

(v) The multivariate normal distribution with mean vector θ and covariance matrix Λ

$$f(x,\theta,\Lambda) = (2\pi)^{-\frac{p}{2}}|\Lambda|^{-\frac{1}{2}}\operatorname{Exp}\{-\tfrac{1}{2}(x-\theta)^T\Lambda^{-1}(x-\theta)\}$$

(vi) The multivariate t-distribution with mean vector θ, scale parameter Λ (not covariance matrix) and degrees of freedom ν

$$f(x,\theta,\Lambda,\nu) = \frac{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\Lambda|^{-\frac{1}{2}}\{\nu+(x-\theta)^T\Lambda^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}}$$

When ν=1, this is a multivariate Cauchy distribution.

Note that the distribution of Z given by (1.2) is invariant under any orthogonal transformation, i.e., if P is any orthogonal matrix of order pxp, then the distribution of Y=PZ is identical to the distribution of Z. The characteristic function of X whose p.d.f. is given by (1.1) has the form

$$\phi(t_1, \ldots, t_p) = \operatorname{Exp}(it^T\theta)\Psi(t^T\Lambda t)$$

[see Kelker (1970)], it follows that the characteristic function of Z with p.d.f. (1.2) has the form

$$\phi(t_1,\ldots,t_p) = \Psi(|t|)$$

[see also Lord (1954)], where Ψ is a function and $|t| = (t_1^2+\ldots+t_p^2)^{\frac{1}{2}}$.

The aim of the present thesis is to study the generalization of problems of multivariate analysis when normality assumption is replaced by the assumption of the class of elliptically symmetric distributions of the continuous type. For inferences, hypothesis testing and for computations of probabilities etc., we consider a specific form of the distributions belonging to this class. In particular we study the multivariate t-distribution of the form

$$f(x) = K_p|\Lambda|^{-\frac{1}{2}}\{\nu+(x-\theta)^T\Lambda^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}} \tag{1.3}$$

which accommodates the multivariate normal distribution by letting $\nu\to\infty$, as well as the multivariate Cauchy distribution by letting $\nu=1$. Hence, in the present work, primarily, we employ this subclass, namely, the multivariate t-distribution as a generalization of the multivariate normal distribution rather than the general class of elliptic distributions in the hope of obtaining sharper (specific) results.

1.3 Contributions of the Present Thesis

The following are the contributions in the present thesis:

1. An elliptic distribution set-up, namely,

$$f(x_1,\ldots\ldots,x_n) = K|\Lambda|^{-\frac{n}{2}} g\{\sum_{j=1}^{n} (\underset{\sim}{x}_j-\underset{\sim}{\theta})^T\Lambda^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\theta})\} \qquad (1.4)$$

is considered, with $\underset{\sim}{X}_j$ as the jth (j=1,...,n) p-dimensional random variable. While in the usual multivariate analysis $\underset{\sim}{X}_j$'s are assumed to be independent observations and termed as a sample of size n, in the present work, it is assumed that they are not necessarily independent but always pairwise uncorrelated. Without any ambiguity, we will assume $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ in (1.4) to be pairwise uncorrelated and for simplicity we will still call $X_1,\ldots,X_n$ a sample of size n. For this model (1.4) we derive

(i) the distribution of the sample mean vector

$$\underset{\sim}{\bar{x}} = (\bar{x}_1,\ldots,\bar{x}_i,\ldots,\bar{x}_p)^T, \text{ where } \bar{x}_i = \sum_{j=1}^{n} x_{ij}/n$$

(ii) the distribution of the sample sum of product matrix namely $V = \sum_{j=1}^{n} (\underset{\sim}{x}_j-\bar{x})(\underset{\sim}{x}_j-\bar{x})^T$

(a generalization of the usual Wishart distribution).

2. We deal with the classification problem for the general elliptic class. In particular,

(i) we deal with the classification of a sample observation into one of two elliptically symmetric populations with known parameters, and we

(ii) derive the distribution of the discriminant function and its moments

(iii) calculate probability of misclassifications

(iv) apply these results to the particular case of multivariate t-distribution.

3. A joint multivariate t-distribution of the elliptic type

$$f(x_1,\ldots,x_n) = K_{np}|\Sigma|^{-\frac{n}{2}}\{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}}, \tag{1.5}$$

$\nu>2$, with X_j having mean θ and covariance Σ, is proposed. This model accommodates the usual case when the samples $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ are assumed to be independently and identically distributed according to $N_p(\theta,\Sigma)$ by letting $\nu\to\infty$. For this model (1.5) we derive

(i) the distribution of the sample mean

(ii) the distribution of the sample sum of product matrix

(iii) the first and second moments of the distribution of sample sum of product matrix.

(iv) both non-null and null distributions of a quadratic form

(v) the joint distribution of the sample mean vector and sample sum of product matrix.

In addition we

(vi) estimate the parameters of the multivariate t-distribution

(vii) discuss the statistical properties of the estimators.

4. For a particular sample, the model (1.5) has the form

$$f(x) = K_p|\Sigma|^{-\frac{1}{2}}\{(\nu-2) + (x-\theta)^T\Sigma^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}} \tag{1.6}$$

which is the usual multivariate t-distribution with location parameter θ, covariance matrix Σ for $\nu > 2$, and ν degrees of freedom. For this distribution (1.6), we derive

(i) the conditional distribution

(ii) the characteristic function; and

(iii) show some applications of the characteristic function.

5. We consider the orthogonal factor model under the assumption that both observed errors and unobserved factors follow multivariate t-distribution and we estimate all parameters of the model including factor loading matrix.

6. Finally we consider a regression model with several response variables under the assumption that the error has a multivariate t-distribution of the form (1.5). The following problems have been dealt with for this regression model:

(i) the parameters of the model, namely, the regression parameters as well as the scale parameters and the degrees of freedom of the error variable are estimated.

(ii) the estimation procedure is illustrated by an actual stock market data taken from the New York Stock Exchange.

(iii) the properties of the estimators are shown, and

(iv) a test for the regression parameters has been discussed.

CHAPTER 2

DISTRIBUTION THEORY FOR ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

A p-component random vector X is elliptically distributed with location parameter θ and a pxp positive definite scale matrix Λ (not necessarily a covariance matrix), if the probability density function of X can be written in the form

$$f(x,\theta,\Lambda) = K_p|\Lambda|^{-1}g\{(x-\theta)^T\Lambda^{-1}(x-\theta)\} \qquad (2.1)$$

where x^T means transpose of x, K_p is a normalising constant. Several authors have studied this distribution. Among them we mention Kelker (1970), Muirhead (1982) who deal with among other things, the marginal distribution as well as distribution of linear combinations of component variables. In the following section we state some well-known results on the above-mentioned distributions.

2.1 Marginal Distribution, Distribution of Linear Combinations

The proof of the following theorem on marginals of elliptically symmetric distributions is straightforward [cf. Kelker (1970), Muirhead (1982)].

Theorem 2.1

If X a p-dimensional random variable has the probability density function (p.d.f.)

$$f(x,\theta,\Lambda) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\theta)^T\Lambda^{-1}(x-\theta)\},$$

where θ and Λ are location and scale parameters respectively, and if x, θ and Λ are partitioned as

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \ \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \text{ and } \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix},$$

where X_1 and X_2 are of order (mxl) and (p-m)xl and Λ_{11}, Λ_{22} are of order (mxm) and (p-m)x(p-m) respectively, then the marginal p.d.f. of X_1 is given by

$$f(x_1) = K|\Lambda_{11}|^{-\frac{1}{2}}g_m\{(x_1-\theta_1)^T\Lambda_{11}{}^{-1}(x_1-\theta_1)\}, \quad (2.2)$$

where the function g_m is determined only by the form of g and by the number of components in X and is independent of θ, Λ.

The distribution of linear combination of component variables of an elliptic distribution is once again elliptic (Kelker, 1970). However, we give a precise statement and derivation of the distribution of such linear combinations.

Theorem 2.2

If X has the p.d.f.

$$f(x,\theta,\Lambda) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\theta)^T\Lambda^{-1}(x-\theta)\}, \quad (2.3)$$

and P is any mxp matrix of rank m, m≤p, then the p.d.f. of Y=PX is given by

$$f(y) = K|P\Lambda P^T|^{-\frac{1}{2}}g_m\{(y-P\theta)^T(P\Lambda P^T)^{-1}(y-P\theta)\} \quad (2.4)$$

Proof:

Let $U = \begin{pmatrix} Y \\ W \end{pmatrix} = \begin{pmatrix} P \\ Q \end{pmatrix} X = AX$, so that $A = \begin{pmatrix} P \\ Q \end{pmatrix}$ is a pxp non-singular matrix. Then (analogous to the normal case) using (2.3) we have the p.d.f. of U as

$$f(u) = K_p|A\Lambda A^T|^{-\frac{1}{2}} g\{(u-A\theta)^T(A\Lambda A^T)^{-1}(u-A\theta)\} \tag{2.5}$$

Now it follows from theorem 2.1 that Y=PX has the p.d.f. given by

$$f(y) = K|P\Lambda P^T|^{-\frac{1}{2}} g_m\{(y-P\theta)^T(P\Lambda P^T)^{-1}(y-P\theta)\} \tag{2.6}$$

2.2 A proposed elliptic distribution set-up for uncorrelated samples (not necessarily independent)

Consider a vector random variable where each component of the random variable is a particular characteristic. Let there be n observations (not necessarily independent) such that each observation is a realization of the p-component vector random variable. Also let x_{ij} denote the jth $(j=1,\ldots\ldots,n)$ observation of the ith $(i=1,\ldots\ldots,p)$ characteristic. Then the jth observation containing the p-characteristics may be written as $\underset{\sim}{x}_j=(x_{1j},\ldots,x_{pj})^T$. Thus we consider the vector variables $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ as n samples, where $\underset{\sim}{X}_j=(X_{1j},\ldots,X_{pj})^T$, $j=1,\ldots,n$.

We now assume that $\underset{\sim}{X}_j$ are identically distributed for all $j=1,\ldots,n$. However, instead of making the usual assumption of the joint independence of $X_1,X_2,\ldots,X_n$ we assume that they are uncorrelated but not necessarily independent. In fact we make the stronger assumption that the joint p.d.f. of $X_1,X_2,\ldots,X_n$ is given by

$$f(x_1,\ldots,x_n) = K_{np}|\Lambda|^{-\frac{n}{2}} g\{\sum_{j=1}^{n} (\underset{\sim}{x}_j-\underset{\sim}{\theta})^T\Lambda^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\theta})\} \tag{2.7}$$

where K_{np} is an appropriate normalising constant. In the above model, we have introduced the location parameter θ and the scale parameter Λ for $\underset{\sim}{X}_j$ for all $j=1,2,\ldots,n$.

Let X^* be a stacked variable defined as $X^*=(\underset{\sim}{X}_1^T,\ldots,\underset{\sim}{X}_j^T,\ldots,\underset{\sim}{X}_n^T)^T$ with parameters $\theta^*=(1_n\otimes\theta)$, $1_n=(1,1,\ldots,1)^T_{nx1}$; and $\Lambda^*=I_n\otimes\Lambda$, where the Kronecker product $A\otimes B$ of $A = \left[a_{ij}\right]_{mxn}$ and B_{sxt} is defined as the (ms,nt) matrix $A\otimes B = \left[a_{ij}B\right]$. (See for example Neudecker (1969)). Then the model (2.7) can be written in the form

$$f(x^*) = K_{np}|\Lambda^*|^{-\frac{1}{2}} g\{(x^*-\theta^*)^T\Lambda^{*^{-1}}(x^*-\theta^*)\}\ . \tag{2.8}$$

Now it follows from (2.8) and theorem 2.1 that each X_j is identically distributed with p.d.f. of the form

$$f(x_j) = K|\Lambda|^{-\frac{1}{2}} g_p\{(x_j-\theta)^T\Lambda^{-1}(x_j-\theta)\}\ ,$$

which is (2.1).

Note that if $\mathrm{cov}(x_j) = \Sigma$ (say) exists then $\Sigma = \alpha\Lambda$, with α a scaler independent of θ and Λ [cf. Kelker, 1970, p. 419].

We remark that it is well-known that the component samples $X_1,X_2,\ldots,X_n$, having joint p.d.f. given by (2.7), are independent if and only if X_j $(j=1,\ldots,n)$ is distributed according to the p-variate normal distribution with mean θ and covariance matrix $\Sigma = \Lambda$. The familiar case of independent samples from the normal distribution is thus a member of the class of distributions having the p.d.f. of the form (2.7).

Independent samples having the multivariate normal p.d.f. is the usual case considered in the literature. However, several authors, for example, Newcomb (1886), Jeffreys (1957, pp. 64-5), Mandelbrot (1963), Box and Tiao (1962), and Praetz (1972) have indicated the importance of non-normal distribution in regression analysis. Zellner (1976) considered a linear regression model with the uncorrelated (not independent) error variables having multivariate t-distribution which is a special case of the model (2.7). Zellner's (1976) model arose among other situations in connection with the study of stock market problems. Dunnett and Sobel (1954) encountered a multivariate t-model in the context of certain multiple decision problems. That model is also a particular case of the elliptic model (2.7).

2.3 Distribution of the sample mean vector

Theorem 2.3

If p-dimensional random variables $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ has the joint p.d.f.

$$f(x_1^T,\ldots,x_j^T,\ldots,x_n^T) = K_{np}|\Lambda|^{-\frac{n}{2}}g\{\sum_{j=1}^{n}(x_j-\theta)^T\Lambda^{-1}(x_j-\theta)\} \quad , \quad (2.9)$$

then the probability density function of $\bar{X}$ is given by

$$f(\bar{x}) = K|\Lambda/n|^{-\frac{1}{2}}g_p\{(\bar{x}-\theta)^T\left(\frac{\Lambda}{n}\right)^{-1}(\bar{x}-\theta)\} \quad (2.10)$$

Proof:

Let $X^* = (x_{11},\ldots,x_{p1},x_{12},\ldots,x_{p2},\ldots,x_{1n},\ldots,x_{pn})$, and $\theta^* = (\underset{\sim}{1}_n \otimes \underset{\sim}{\theta})$, where $\underset{\sim}{1}_{nx1} = (1,1,\ldots,1)^T$. Then, the joint p.d.f. of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ can be written as

$$f(x^*) = K_{np}|I_n \otimes \Lambda|^{-\frac{1}{2}} g\{(x^*-\theta^*)^T(I_n \otimes \Lambda)^{-1}(x^*-\theta^*)\}$$

Next we consider the linear combinations

$$Y^* = PX^* = \frac{1}{n}(I_p,\ldots,I_p)X^* \quad , \tag{2.11}$$

where in (2.11) P is a pxnp matrix. The result is immediate from the theorem 2.2.

2.4 Distribution of sample sum of products matrix

It has been noted earlier that the distribution of linear combination of component variables has been studied by many authors, but there has been no work on the distribution of the sample covariance matrix. The following theorems may be viewed as a generalization of the usual Wishart distribution.

Theorem 2.4

If p-dimensional random variables $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_j,\ldots,\underset{\sim}{Z}_n$ has the p.d.f.

$$f(z_1^T,\ldots,z_j^T,\ldots,z_n^T) = K_{np}\; g(\sum_{n=1}^{n} z_j^T z_j) \quad , \tag{2.12}$$

then the distribution of the elements of the matrix

$$M_1 = \sum_{j=1}^{n} z_j z_j^T$$

$$= ((\sum_{j=1}^{n} z_{ij} z_{kj}))$$

$$= ((m_{ik})) \quad \text{(say)} \quad ,$$

is given by

$$f\{(m_{ik} \,/\, i \le k,\ i=1,\ldots,p;\ k=1,\ldots,p)\}$$

$$= C_{np}\, g\,(\text{trace } M_1)\ |M_1|^{\frac{n-p-1}{2}} \quad , \qquad (2.13)$$

where $z_j = (z_{1j},\ldots,z_{pj})^T$, and C_{np} is a normalising constant such that

$$\int_{m_1} f(m_1)\,dm_1 = 1 \quad .$$

Proof:

Consider the given matrix

$$M_1 = \begin{bmatrix} \sum_{j=1}^{n} z_{1j}^2 & \cdot & \cdot & \cdot & \sum_{j=1}^{n} z_{1j}z_{pj} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \sum_{j=1}^{n} z_{1j}z_{pj} & \cdot & \cdot & \cdot & \sum_{j=1}^{n} z_{pj}^2 \end{bmatrix} = \begin{bmatrix} m_{11} & \cdot & \cdot & \cdot & m_{1p} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ m_{p1} & \cdot & \cdot & \cdot & m_{pp} \end{bmatrix} \quad \text{(say)}$$

and let us find the distribution of the elements of the matrix M_1 when $Z_1,\ldots,Z_j,\ldots,Z_n$ has the joint p.d.f. given by (2.12), namely,

$$f(z_1^T,\ldots,z_j^T,\ldots,z_n^T) = K_{np}\, g\{\sum_{j=1}^{n} z_j^T z_j\}$$

This joint p.d.f. can be rewritten as

$$f(z_1,\ldots,z_i,\ldots,z_p) = K_{np}\, g\{\sum_{i=1}^{p} z_i z_i^T\}, \qquad (2.14)$$

where $\underset{\sim}{z}_i = (z_{i1},\ldots,z_{ij},\ldots,z_{in})$. Since M_1 is a symmetric matrix, the total number of distinct elements in M_1 is $p(p+1)/2$. We now wish to obtain the joint distribution of these distinct elements.

Le us make an orthogonal transformation from z's to y's in the following manner:

$$\left.\begin{aligned}
y_1^T &= z_1^T \\
y_2^T &= z_2^T - b_{21}y_1^T \\
&\cdots\cdots\cdots\cdots \\
y_i^T &= z_i^T - b_{i(i-1)}y_{i-1}^T - \cdots - b_{i1}y_1^T \\
&\cdots\cdots\cdots\cdots \\
y_p^T &= z_p^T - b_{p(p-1)}y_{p-1}^T - \cdots - b_{p1}y_1^T,
\end{aligned}\right\} \qquad (2.15)$$

where the scalars b's are so determined that $y_i y_\ell^T = 0$ for $i \neq \ell$ and for $i, \ell = 1,2,\ldots,p$. Now for $r < i$, $y_r y_i^T = y_r z_i^T - b_{ir} y_r y_r^T = 0$, therefore $b_{ir} = (y_r z_i^T) / (y_r y_r^T)$.

$$\left.\begin{aligned}
&\text{Let } b^*_{ir} = (y_r z_i^T) / \sqrt{y_r y_r^T} = b_{ir}\sqrt{y_r y_r^T}, \\
&\text{so that } b^{*2}_{ir} = b^2_{ir} y_r y_r^T
\end{aligned}\right\} \qquad (2.16)$$

Now by (2.15) and (2.16), we get

$$z_i z_i^T = (y_i + \sum_{r=1}^{i-1} b_{ir} y_r)(y_i^T + \sum_{r=1}^{i-1} b_{ir} y_r^T)$$

$$= y_i y_i^T + \sum_{r=1}^{i-1} b_{ir}^{*2} , \qquad (2.17)$$

for all i and $r < i$. Again for $i < \ell$, by (2.15) and (2.16), we can write

$$z_i z_\ell^T = (y_i + \sum_{r=1}^{i-1} b_{ir} y_r)(y_\ell^T + \sum_{r=1}^{\ell-1} b_{\ell r} y_r^T)$$

$$= (b_{\ell i} y_i y_i^T + \sum_{r=1}^{i-1} b_{ir} b_{\ell r} y_r y_r^T$$

$$= b_{\ell i}^* \sqrt{y_i y_i^T} + \sum_{r=1}^{i-1} b_{ir}^* b_{\ell r}^* \qquad (2.18)$$

Next using (2.17) and (2.18) we decompose M_1, the sample sum of product matrix as in the following:

$$M_1 = \begin{bmatrix} z_1 z_1^T & z_1 z_2^T & \cdots & z_1 z_p^T \\ \vdots & & & \vdots \\ z_p z_1^T & z_p z_2^T & \cdots & z_p z_p^T \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{y_1 y_1^T} & 0 & 0 & \cdots & 0 \\ b_{21}^* & \sqrt{y_2 y_2^T} & 0 & \cdots & 0 \\ b_{31}^* & b_{32}^* & \sqrt{y_3 y_3^T} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ b_{p1}^* & b_{p2}^* & b_{p3}^* & \cdots & \sqrt{y_p y_p^T} \end{bmatrix} \begin{bmatrix} \sqrt{y_1 y_1^T} & b_{21}^* & b_{31}^* & \cdots & b_{p1}^* \\ 0 & \sqrt{y_2 y_2^T} & b_{32}^* & \cdots & b_{p2}^* \\ 0 & 0 & \sqrt{y_3 y_3^T} & \cdots & b_{p3}^* \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{y_p y_p^T} \end{bmatrix}$$

(2.19)

Now by (2.16) we can write

$$b^* = (b^*_{i1},\ldots,b^*_{ir},\ldots b^*_{i,i-1})$$

$$= \left(\frac{y_1}{\sqrt{y_1y_1^T}},\ldots,\frac{y_r}{\sqrt{y_ry_r^T}},\ldots,\frac{y_{i-1}}{\sqrt{y_{i-1}y_{i-1}^T}}\right)^T z_i^T \tag{2.20}$$

$$= C_{(i-1)xn}z_i^T \quad \text{(say)}$$

Analogous to multivariate normal distribution, consider $z_1^T,\ldots,z_{i-1}^T$ to be fixed, so that $y_1^T,\ldots,y_{i-1}^T$ are also fixed. Then, each b^*_{ir} $(r = 1,\ldots,i-1)$ can be regarded as a linear function of the elements of z_i^T, i.e., $b^* = Cz_i^T$ as in (2.20). Rows of the matrix C in (2.20) are mutually orthogonal, because $y_ry_i^T = 0$ for $r < i$, by the transformation (2.15). Therefore, we may find a $(n-i+1)xn$ matrix P (say) such that $\begin{pmatrix} C \\ P \end{pmatrix} = D$, where D is an nxn orthogonal matrix. Then (2.20) can be extended as

$$(b^*, \xi)^T = (b^*_{i1},\ldots,b^*_{i,i-1},\xi_{i,i},\xi_{i,i+1},\ldots,\xi_{in})^T$$

$$= Dz_i^T \tag{2.21}$$

Since by (2.14) the joint probability differential for np elements of $Z = ((z_{ij}))$ is

$$K_{np}\, g\{\sum_{r=1}^{i-1} z_rz_r^T + \sum_{r=i}^{p} z_rz_r^T\} \prod_{i=1}^{p}\prod_{j=1}^{n} dz_{ij} ,$$

therefore, by (2.21), this probability differential is equivalent to

$$k_{np}\, g\{\sum_{i=1}^{p}\sum_{r=1}^{i-1} b^{*2}_{ir} + \sum_{i=1}^{p}\sum_{r=i}^{n} \xi_{ir}^2\} \prod_{i=1}^{p}\prod_{r=1}^{i-1} db^*_{ir} \prod_{i=1}^{p}\prod_{r=i}^{n} d\xi_{ir} \tag{2.22}$$

By the assumption that the samples are uncorrelated in the model (2.7), it follows that $\underset{\sim}{z}_1,\ldots,\underset{\sim}{z}_j,\ldots,\underset{\sim}{z}_n$ in (2.12) are pairwise uncorrelated. Now a comparison of the form (2.22) with the form (2.12) implies that b^*_{ir} $(r = 1,\ldots,i-1)$ and ξ_{ir} $(r = i,i+1,\ldots,n)$ are uncorrelated.

Next let us make the following polar transformation for all i

$$b^*_{ir} = b^*_{ir} \qquad r = 1,2,\ldots\ldots,i-1$$

$$\xi_{ii} = R_i^{1/2} \cos\theta_{i1}\cos\theta_{i2},\ldots,\cos\theta_{i,n-1}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\xi_{in} = R_i^{1/2} \sin\theta_{i1} \;,$$

so that $\prod_{r=1}^{i-1} db^*_{ir} \prod_{r=i}^{n} d\xi_{ir} = K_i' R_i^{\frac{n-(i-1)}{2}-1} dR_i \prod_{r=1}^{i-1} db^*_{ir}$,

with $\sum_{r=i}^{n} \xi_{ir}^2 = R_i$, and K_i' is a constant independent of R_i but depends on θ's. Then (2.22) reduces to

$$C_{np}\, g\, \{\sum_{i=1}^{p} \sum_{r=1}^{i-1} b^{*2}_{ir} + \sum_{i=1}^{p} R_i\} \prod_{i=1}^{p} R_i^{\frac{n-(i-1)}{2}-1} dR_i \prod_{i-1}^{p} \prod_{r=1}^{i-1} db^*_{ir} \;, \tag{2.23}$$

where $C_{np} = K_{np} \prod_{i=1}^{p} K_i'$, a constant.

From (2.17) and (2.21) it follows that $\sum_{r=i}^{n} \xi_{ir}^2 = y_i y_i^T$. Since again $\sum_{r=i}^{n} \xi_{ir}^2 = R_i$, by the above polar transformation, therefore, from (2.19) it follows that

$$\sum_{i=1}^{p} \sum_{r=1}^{i-1} b^{*2}_{ir} + \sum_{i=1}^{p} R_i = \sum_{i=1}^{p} \sum_{r=1}^{i-1} b^{*2}_{ir} + \sum_{i=1}^{p} y_i y_i^T = \text{trace } M_1 \tag{2.24}$$

Because

$$\prod_{i=1}^{p} (R_i)^{\frac{n-p-1}{2}} = \prod_{i=1}^{p} (y_i y_i^T)^{\frac{n-p-1}{2}} = |M_1|^{\frac{n-p-1}{2}} ,$$

therefore,

$$\prod_{i=1}^{p} (R_i)^{\frac{n-i+1}{2}-1} = \prod_{i=1}^{p} (R_i)^{\frac{n-p-1}{2}} \prod_{i=1}^{p} (R_i)^{\frac{p-i}{2}}$$

$$= |M_1|^{\frac{n-p-1}{2}} \prod_{i=1}^{p} (R_i)^{\frac{p-i}{2}} \qquad (2.25)$$

Now using (2.24) and (2.25) in (2.23), we can write the probability differential as

$$C_{np} \, g\{\text{trace } M_1\} |M_1|^{\frac{n-p-1}{2}} \prod_{i=1}^{p} (R_i)^{\frac{p-i}{2}} \prod_{i=1}^{p} dR_i \prod_{i=1}^{p} \cdot \prod_{r=1}^{i-1} db_{ir}^{*}$$

Next using the identity $\prod_{i=1}^{p} dR_i \prod_{i=1}^{p} \prod_{r=1}^{i-1} db_{ir}^{*} = |J| \prod_{i \le j} m_{ij}$ (as in the usual Wishart distribution), where $|J| = \prod_{i=1}^{p} (R_i)^{-\frac{p-i}{2}}$, we can write the last probability differential as

$$C_{np} \, g(\text{trace } M_1) |M_1|^{\frac{n-p-1}{2}} \prod_{i \le j} dm_{ij} \qquad (2.26)$$

which is the probability distribution of the elements of Matrix $M_1 = \sum_{j=1}^{n} z_j z_j^T$. In (2.26) C_{np} is a normalising constant such that

$$\int C_{np} \, g(\text{trace } M_1) |M_1|^{\frac{n-p-1}{2}} \prod_{i \le j} dm_{ij} = 1 .$$ Hence the theorem.

We remark that Anderson (1958, p. 319) gives a p.d.f. for the elements of the sample raw sum of product matrix when the underlying population is spherically symmetric as in the above theorem. But his derivation appears to be a sketchy one.

However, we now derive the distribution of the sample corrected sum of product matrix as in the following theorem.

<u>Theorem 2.4.1</u>

If p-dimensional random variables $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_j,\ldots,\underset{\sim}{Z}_n$ has the p.d.f.

$$f(z_1^T,\ldots,z_j^T,\ldots,z_n^T) = K_{np}g(\sum_{j=1}^{n} z_j^T z_j) ,$$

then the p.d.f. of the elements of the matrix $M = \sum_{j=1}^{n} (z_j-\bar{z})(z_j-\bar{z})^T$ is given by

$$C_{n'p}g_{n'p}(\text{trace } M)|M|^{\frac{n'-p-1}{2}} ,$$

where $n' = n-1$, and $g_{n'p}(\cdot)\alpha \int g(\sum_{j=1}^{n} z_j^T z_j) \prod_{i=1}^{p} dz_{in}$.

Proof:

Let $V_i^T = CZ_i^T$, where $V_i = (V_{i1},\ldots,V_{in})^T$ and C is an orthogonal matrix such that the last row of C is $(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{n}},\ldots,\frac{1}{\sqrt{n}})$. Then it follows that

$$f(v_i^T,\ldots,v_j^T,\ldots,v_n^T) = K_{np}g(\sum_{j=1}^{n} v_j^T v_j) , \tag{2.27}$$

and $M = \sum_{j=1}^{n} (z_j-\bar{z})(z_j-\bar{z})^T$

$$= \sum_{j=1}^{n-1} v_j v_j^T$$

$$= \sum_{j=1}^{n'} v_j v_j^T$$

Next it follows from (2.27) and the definition of $g_{n'p}(\cdot)$ that

$$f(v_1^T,\ldots,v_j^T,\ldots,v_{n'}^T) = K_{n'p}\; g_{n'p}(\sum_{j=1}^{n} v_j^T v_j) .$$

Now the theorem 2.4 yields the p.d.f. given by the present theorem, for the elements of the matrix $M = \sum_{j=1}^{n} (z_j-\bar{z})(z_j-\bar{z})$.

Theorem 2.5

If $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ has the joint probability density

$$f(x_1^T,\ldots,x_j^T,\ldots x_n^T) = k_{np}\; g\{\sum_{j=1}^{n} (\underset{\sim}{x}_j-\underset{\sim}{\theta})^T \Lambda^{-1} (\underset{\sim}{x}_j-\underset{\sim}{\theta})\} \quad , \qquad (2.28)$$

then the distribution of the elements of

$$V = \sum_{j=1}^{n} (\underset{\sim}{x}_j-\bar{x})(\underset{\sim}{x}_j-\bar{x})^T$$

$$= ((\sum_{j=1}^{n} (\underset{\sim}{x}_{ij}-\bar{x}_i)(\underset{\sim}{x}_{kj}-\bar{x}_k)))$$

$$= ((V_{ik})) \;\; \text{(say)} \; ,$$

is given by

$$f\{v_{ik} \;/\; i \le k, i,k=1,\dots,p\} = C_{n'p}\, g_{n'p}(\text{trace } \Lambda^{-1}V)|\Lambda|^{-\frac{n'}{2}}|V|^{\frac{n'-p-1}{2}} \qquad (2.29)$$

where $\underset{\sim}{x}_j = (x_{1j},\dots,x_{pj})^T$, $n' = n-1$, and $C_{n'p}$ is a normalising constant such that

$$\int_V f(v)dv = 1$$

Proof:

This theorem directly follows from theorem 2.4.1. We sketch the additional steps below.

Let $Z_j = \Lambda^{-\frac{1}{2}}(X_j-\theta)$. Then the p.d.f. given by (2.28) may be expressed as

$$f(z_1^T,\dots,z_n^T) = K_{np}\, g\, \{\sum_{j=1}^{n} z_j^T z_j\} ,$$

which is the same as the p.d.f. given by (2.12). The above transformation also implies that

$$z_j-\bar{z} = \Lambda^{-\frac{1}{2}}(x_j-\bar{x}) .$$

Therefore $M = \sum_{j=1}^{n'} (z_j-\bar{z})(z_j-\bar{z})^T$ in the theorem 2.4.1 can be written as

$$M = \sum_{j=1}^{n} (z_j-\bar{z})(z_j-\bar{z})^T$$

$$= \Lambda^{-\frac{1}{2}} \sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T \Lambda^{-\frac{1}{2}}$$

$$= \Lambda^{-\frac{1}{2}} V \Lambda^{-\frac{1}{2}} ,$$

because by definition $V = \sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T$.

It then follows that

$$\left.\begin{aligned} &\text{trace } M = \text{trace } \Lambda^{-1}V \,, \\ \text{and} \quad &|M| = |\Lambda^{-1}||V| \end{aligned}\right| \qquad (2.30)$$

Now using (2.30) in theorem 2.4.1, we get the p.d.f. of the elements of

$V = \sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T$ as

$$C_{n'p}\, g_{n'p}\, (\text{trace } \Lambda^{-1}V)\, |V|^{-\frac{n'-p-1}{2}} |\Lambda|^{\frac{n'-p-1}{2}} |J| \,,$$

where $|J|$ stands for Jacobian of the transformation $Z = \Lambda^{-\frac{1}{2}}(X-\theta)$, which is given by $|\Lambda|^{-\frac{p+1}{2}}$. Hence we obtain the probability density function of the elements of the matrix $V = \sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T$ as

$$C_{n'p}\, g_{n'p}\, (\text{trace } \Lambda^{-1}V)\, |\Lambda|^{-\frac{n'}{2}} |V|^{\frac{n'-p-1}{2}} \,,$$

which is (2.29).

CHAPTER 3

DISCRIMINANT ANALYSIS

The problem of classification or discrimination arises when an investigator makes a number of measurements on an individual and wishes to classify the individual into one of several categories on the basis of these measurements. In brief, one may state the problem as: given an individual with certain measurements; if several populations exist from which this individual may have come, the question is, from which population did it arise?

There is a vast literature on classification and discrimination. In order to classify an observation into one of the populations, Fisher (1936) suggested as a basis for classification decisions the use of a discriminant function linear in the components of the observation. Other bases for classification have included likelihood ratio tests (see for instance, Anderson, 1958), information theory (Kullback, 1959) and Bayesian techniques (see, for example Geisser, 1964). In all cases, sampling theories have been considered under the assumption that the populations involved are multivariate normal.

The purpose of this section is to examine the classification problem for a wider class of distributions, namely, for elliptical class of distributions. We assume that the location and the common scale parameter of two elliptically symmetric populations are known.

3.1 Classification of an observation into one of two elliptically symmetric populations with known parameters

3.1.1 Classification criterion

Consider the p.d.f. (2.1) of the elliptically distributed random variable $\underset{\sim}{X}$ given by

$$f(x) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\theta)^T\Lambda^{-1}(x-\theta)\} , \quad (3.1)$$

where $\Lambda > 0$ is a positive definite matrix, $0 < (x-\theta)^T\Lambda^{-1}(x-\theta) < \infty$, and $g\{(x-\theta)^T\Lambda^{-1}(x-\theta)\}$ is a monotonic decreasing function of $(x-\theta)^T\Lambda^{-1}(x-\theta)$.

We think of two specific (known) elliptical populations π_1 and π_2 (say) with parameters (θ_1,Λ) and (θ_2,Λ) respectively, from any of which our individual $\underset{\sim}{x} = (x_1,x_2,\ldots,x_p)$ may have come. Then, our problem is to find out the discrimination criterion to assign $\underset{\sim}{x}$ to a population of the two namely π_1 and π_2.

Since irrespective of the form of the population we know a general rule for discrimination that assign the individual $\underset{\sim}{x}$ to π_1 if

$$\{p_1f_1(x) - p_2f_2(x)\} \geq 0 ,$$

otherwise assign $\underset{\sim}{x}$ to π_2 [see for example, Anderson (1958)]. Here p_1 and p_2 are priori probabilities of π_1 and π_2. If we assume $p_1=p_2$, then the discrimination rule says to assign the individual to π_1 if $f_1(x) \geq f_2(x)$. Consequently, in the present case, we assign the observation x to π_1 if

$$g\{(x-\theta_1)^T\Lambda^{-1}(x-\theta_1)\} \geq g\{(x-\theta_2)^T\Lambda^{-1}(x-\theta_2)\} .$$

Since it is assumed that $g(\cdot)$ is a monotonic decreasing function of $(x-\theta)^T\Lambda^{-1}(x-\theta)$, therefore, the assignment of the observation x to π_1 is satisfied if

$$(x-\theta_1)^T\Lambda^{-1}(x-\theta_1) \leq (x-\theta_2)^T\Lambda^{-1}(x-\theta_2) \qquad (3.2)$$

After a simple algebra (3.2) reduces to

$$(\theta_1-\theta_2)^T\Lambda^{-1}x \geq \tfrac{1}{2}\{(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2)\} \quad .$$

Hence, we assign $\underset{\sim}{x}$ to π_1 if

$$(\theta_1-\theta_2)^T\Lambda^{-1}x \geq \tfrac{1}{2}\{(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2)\} \quad , \qquad (3.3)$$

otherwise assign $\underset{\sim}{x}$ to π_2. From (3.3), $(\theta_1-\theta_2)^T\Lambda^{-1}X$ is a linear function in $\underset{\sim}{X}$ and it is termed as linear discriminant function. The purpose of the following theorem is to show the distribution of this linear discriminant function.

Theorem 3.1

If $\underset{\sim}{X}$ has a distribution with the p.d.f.

$$f(x) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\theta)^T\Lambda^{-1}(x-\theta)\} \ ,$$

then the distribution of the linear discriminant function $(\theta_1-\theta_2)^T\Lambda^{-1}X = PX = U$ (say) has the distribution given by

$$f(u) = K\ |P\Lambda P^T|^{-\frac{1}{2}}g_1\{(u-P\theta)^T(P\Lambda P^T)^{-1}(u-P\theta)\} \ , \qquad (3.4)$$

where

$$P\theta = (\theta_1-\theta_2)^T\Lambda^{-1}\theta$$

$$P\Lambda P^T = (\theta_1-\theta_2)^T\Lambda^{-1}\Lambda\Lambda^{-1}(\theta_1-\theta_2)$$

$$= (\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2)$$

Proof:

The theorem directly follows from the theorem 2.2.

Mean and Variance of the Discriminant Function

Analogous to the corresponding multivariate normal set-up, we calculate the mean and variance of the quantity

$$X^T\Lambda^{-1}(\theta_1-\theta_2) - \tfrac{1}{2}(\theta_1+\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2) = U_1^* \quad \text{(say)} .$$

If X has come from the population π_1, then the p.d.f. of X is given by

$$f(x) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\theta_1)^T\Lambda^{-1}(x-\theta_1)\} .$$

Then $\Lambda^{-\frac{1}{2}}(x-\theta_1) = Y$ has the spherical distribution whose p.d.f. is given by

$$f(y) = K_p g(Y^TY).$$

If follows from the p.d.f. of Y that $E(Y) = \underset{\sim}{0}$ and covariance matrix of the form

$$E(YY^T) = \sigma^2\cdot I_p = \begin{bmatrix} \sigma^2 & 0 \;\ldots\ldots & 0 \\ 0 & \sigma^2 \;\ldots\ldots & 0 \\ \ldots & \ldots\ldots\ldots & \ldots \\ 0 & 0 \;\ldots\ldots & \sigma^2 \end{bmatrix} , \sigma^2 > 0 ,$$

then it follows that

$$E(X) = E(\Lambda^{\frac{1}{2}}Y+\theta_1)$$

$$= \theta_1 , \qquad (3.5)$$

and $\text{cov}(X) = E(X-\theta_1)(X-\theta_1)^T$

$$= E(\Lambda^{\frac{1}{2}}YY^T\Lambda^{\frac{1}{2}})$$

$$= \sigma^2\Lambda \tag{3.6}$$

Therefore,

$$E(U_1^*) = E\{X^T\Lambda^{-1}(\theta_1-\theta_2) - \tfrac{1}{2}(\theta_1+\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2)\}$$

$$= \theta_1^T\Lambda^{-1}(\theta_1-\theta_2) - \tfrac{1}{2}(\theta_1+\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2)$$

$$= (\theta_1^T-\tfrac{1}{2}\theta_1^T-\tfrac{1}{2}\theta_2^T)\Lambda^{-1}(\theta_1-\theta_2)$$

$$= \tfrac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2)$$

$$= \frac{\alpha}{2} \quad \text{(say)} ,$$

and

$$\text{cov}(U_1^*) = \text{cov}\{X^T\Lambda^{-1}(\theta_1-\theta_2)\}$$

$$= (\theta_1-\theta_2)^T\Lambda^{-1}\text{cov}(X)\Lambda^{-1}(\theta_1-\theta_2)$$

$$= \sigma^2(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2) \text{ , by (3.6)}$$

$$= \sigma^2\cdot\alpha$$

3.1.2 Probability of Misclassification

Consider the discrimination criterion (3.3) that classify X in π_1 if

$$(\theta_1-\theta_2)^T\Lambda^{-1}X - \tfrac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2) \geq 0 ,$$

and in π_2 otherwise. Let the probability of misclassifying an individual from π_1 into π_2 be e_1 and the probability of misclassifying an individual from π_2 in π_1 be e_2. If the distributions π_1 and π_2 are multivariate normal with parameters (θ_1,Σ) and (θ_2,Σ) respectively, where Σ is the common covariance matrix, then it is well-known that probabilities of misclassifications e_1 and e_2 are same [see for example Anderson (1958), Srivastava and Carter (1983)] and they are given by

$$e_1 = e_2 = \Phi\left[-\tfrac{1}{2}\{(\theta_1-\theta_2)^T\Sigma^{-1}(\theta_1-\theta_2)\}^{\frac{1}{2}}\right] ,$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$.

In the present case, when the underlying distributions are elliptically symmetric, the probabilities of misclassifications are given as in the following theorem.

Theorem 3.2

If $\underset{\sim}{X}$ (px1) has the p.d.f.

$$f_1(x) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\theta_1)^T\Lambda^{-1}(x-\theta_1)\} ,$$

when it comes from a population π_1 ; and if $\underset{\sim}{X}$ has the p.d.f. given by

$$f_2(x) = K_p|\Lambda|^{-\frac{1}{2}}g\{(x-\theta_2)^T\Lambda^{-1}(x-\theta_2)\} ,$$

when it comes from another population π_2 ; and if e_1, e_2 are the probability of misclassing an individual from π_1 into π_2 and π_2 into π_1 respectively, then e_1 and e_2 are same and given by

$$e_1 = e_2 = \int_{-\infty}^{-\frac{1}{2}D} Kg_1(z^T z)dz \quad ,$$

where $D^2 = (\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2)$

Proof:

By definition e_1 and e_2 are

$$e_1 = \Pr\left[\{(\theta_1-\theta_2)^T\Lambda^{-1}X - \tfrac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2)\}\le o \,|\, x\varepsilon\pi_1\right] ,$$

and

$$e_2 = \Pr\left[\{(\theta_1-\theta_2)^T\Lambda^{-1}X - \tfrac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2)\}< o \,|\, x\varepsilon\pi_2\right] .$$

Thus, $e_1 = \Pr\left[(\theta_1-\theta_2)^T\Lambda^{-1}X \le \tfrac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2) \,|\, x\varepsilon\pi_1\right]$.

Let $(\theta_1-\theta_2)^T\Lambda^{-1} = P$, then we write

$$e_1 = \Pr\left[PX \le \tfrac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2) \,|\, x\varepsilon\pi_1\right] .$$

Next by (3.4), this probability of misclassification, namely, e_1 can be written as

$$e_1 = \int_{-\infty}^{\frac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1+\theta_2)} K\,|P\Lambda P^T|^{-\frac{1}{2}}\, g_1\{(u-P\theta_1)^T(P\Lambda P^T)^{-1}(u-P\theta_1)\}du$$

Put $Z = (P\Lambda P^T)^{-\frac{1}{2}}(U-P\theta_1)$ in the last equation. Then it is readily seen that

$$e_1 = \int_{-\infty}^{-\frac{1}{2}D} Kg_1(z^Tz)dz \quad ,$$

where $D = \frac{1}{2}(\theta_1-\theta_2)^T\Lambda^{-1}(\theta_1-\theta_2)$. Similarly, it can be shown that

$$e_2 = \int_{\frac{1}{2}D}^{\infty} Kg_1(z^Tz)dz \quad .$$

Therefore, by symmetriness property of the distribution of Z, we obtain $e_1 = e_2$. Hence the theorem.

3.2 Classification of an observation into one of two multivariate t-populations with known parameters

Now we shall use the general procedure outlined above, as an example, to the case when the two underlying distributions are multivariate. Let $t_p(\theta_1,\Sigma,\nu)$ and $t_p(\theta_2,\Sigma,\nu)$ are the two multivariate t-populations, where $\theta_i = (\theta_{i1},\ldots,\theta_{ip})^T$ is the mean vector of the ith $(i = 1,2)$ population; Σ is the covariance matrix and ν is the degrees of freedom for both populations. The p.d.f. of the above described ith $(i=1,2)$ population is given by

$$f_i(x) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\Sigma|^{-\frac{1}{2}}\{(\nu-2) + (x-\theta_i)^T\Sigma^{-1}(x-\theta_i)\}^{-\frac{\nu+p}{2}} \quad , \qquad (3.7)$$

by (4.2). As before we name these two populations as π_1 and π_2 respectively.

Let x denote an observed sample of size 1. Since $\{\cdot\}^{-\frac{\nu+p}{2}}$ in (3.7) is a decreasing function of $(x-\theta_1)^T\Sigma^{-1}(x-\theta_1)$, therefore, it follows from (3.1) through (3.3) that $\underset{\sim}{x}$ should be assigned to π_1 if

$$(\theta_1-\theta_2)^T\Sigma^{-1}x \geq \tfrac{1}{2}\{(\theta_1-\theta_2)^T\Sigma^{-1}(\theta_1+\theta_2)\} , \tag{3.8}$$

otherwise we assign x to π_2.

<u>Distribution of the discriminant function</u>

Theorem 3.3

If $\underset{\sim}{X}$ is distributed according to $t_p(\theta_1,\Sigma,\nu)$, then $U_1^* = X^T\Sigma^{-1}(\theta_1-\theta_2) - \tfrac{1}{2}(\theta_1+\theta_2)^T\Sigma^{-1}(\theta_1-\theta_2)$ has the distribution of the form $t_1(\frac{\alpha}{2},\alpha,\nu)$, where $\alpha = (\theta_1-\theta_2)^T\Sigma^{-1}(\theta_1-\theta_2)$ for $\nu > 2$.

Proof:

Since $E(X) = \underset{\sim}{\theta}_1$, it is readily verified that $E(U_1^*) = \frac{\alpha}{2}$ and $\text{cov}(U_1^*) = \alpha$, where α is given by $\alpha = (\theta_1-\theta_2)^T\Sigma^{-1}(\theta_1-\theta_2)$. Now, the theorem follows directly from theorem 3.1.

<u>Probability of Misclassifications</u>

Theorem 3.4

Let $\underset{\sim}{X}$ has the multivariate t-distribution with the p.d.f.

$$f_i(x) = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\!\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\,\Gamma\!\left(\frac{\nu}{2}\right)} |\Sigma|^{-\frac{1}{2}}\{(\nu-2) + (x-\theta_i)^T\Sigma^{-1}(x-\theta_i)\}^{-\frac{\nu+p}{2}} ,$$

when $\underset{\sim}{X}$ comes from population π_i $(i=1,2)$. Also let e_1 and e_2 are the probability of misclassifying an individual from π_1 into π_2 and π_2 into π_1 respectively. Then e_1 and e_2 are given by

$$e_1 = e_2 = \int_{-\infty}^{-\frac{1}{2}D^*} \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{\frac{1}{2}}\,\Gamma\left(\frac{\nu}{2}\right)} \{(\nu-2) + z^2\}^{-\frac{\nu+1}{2}} dz , \tag{3.9}$$

where $D^{*2} = (\theta_1-\theta_2)^T\Sigma^{-1}(\theta_1-\theta_2)$.

Proof:

This follows from the theorem 3.2.

A Numerical Example on Probability of Misclassification:

Rao (1948a) considers three populations consisting of the Brahmin caste (π_1), the Artisan caste (π_2), and the Korwa caste (π_3) of India. The measurements for each individual of a caste are stature (x_1), sitting height (x_2), nasal depth (x_3), and nasal height (x_4). From Rao's example, we now consider two populations namely the Brahmin caste (π_1) and the Artisan caste (π_2). The means of the four variables in the two populations and the matrix of correlations among four variables for all the populations are given as follows:

	Means	
	Brahmin (π_1)	Artisan (π_2)
Stature (x_1)	164.51	160.53
Sitting height (x_2)	86.43	81.47
Nasal depth (x_3)	25.49	23.84
Nasal height (x_4)	51.24	48.62

and correlation matrix R (say)

$$R = \begin{pmatrix} 1.0000 & 0.5849 & 0.1774 & 0.1974 \\ & 1.0000 & 0.2094 & 0.2170 \\ & & 1.0000 & 0.2910 \\ & & & 1.0000 \end{pmatrix}$$

The standard deviations for four variables X_1, X_2, X_3 and X_4 are given as $\sigma_1 = 5.74$, $\sigma_2 = 3.20$, $\sigma_3 = 1.75$ and $\sigma_4 = 3.50$ respectively. Now using these standard deviations and the correlation matrix R, we have the common covariance matrix for all populations as

$$\Sigma = \begin{pmatrix} 32.9476 & 10.7434 & 1.7820 & 3.9569 \\ & 10.2400 & 1.1726 & 2.4304 \\ & & 3.0625 & 1.7824 \\ & & & 12.2500 \end{pmatrix}$$

Now we assume that π_1 refers to $t_p(\theta_1,\Sigma,\nu)$ and the second population π_2 refers to $t_p(\theta_2,\Sigma,\nu)$. Then $(\theta_1-\theta_2)^T = (3.98, 4.96, 1.65, 2.62)$. By using (3.9), we obtain the following probabilities of misclassifications for different degrees of freedom.

ν	5	10	15	30	∞ (Normal case)
e_1 or e_2	0.160	0.176	0.184	0.1901	0.1949

CHAPTER 4

MULTIVARIATE t-DISTRIBUTION

It has been mentioned earlier that the class of elliptic distributions of the continuous type could be a suitable generalization of the multivariate normal model. This class has the p.d.f. of the form given by

$$f(x) = K|\Lambda|^{-\frac{1}{2}} g\{(x-\theta)^T \Lambda^{-1}(x-\theta)\} \quad .$$

This elliptic class of distributions is a wide one and it is reasonable to expect that distributional results would usually depend on the form of the function $g(\cdot)$. Hence, for the purpose of inferences, hypothesis testing, computations of probabilities etc., it is appropriate to put further restrictions on the form of $g(\cdot)$. In particular, in the present thesis, we have restricted ourselves to the subclass, namely the multivariate t-distribution, as a generalization of the normal model in the hope of obtaining sharper results. The p-dimensional multivariate t-distribution of $\underset{\sim}{X}(p\times 1)$ has the p.d.f. given by

$$f(x) = \frac{\nu^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \cdot \pi^{\frac{p}{2}}} |\Lambda|^{-\frac{1}{2}} \{\nu + (x-\theta)^T \Lambda^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}} , \qquad (4.1)$$

[cf. Cornish (1954), Dunnett and Sobel (1954)] where θ is the location parameter, Λ is a scale matrix and ν is known as the degrees of freedom.

This subclass of multivariate t-distribution contains at one end the multivariate cauchy distribution when $\nu=1$ and at the other end the multivariate normal distribution when $\nu\to\infty$.

It is readily verified that [see Cornish (1954)] the multivariate t-distribution (4.1) possesses second order moments for $\nu>2$ and has the following mean vector and covariance matrix:

$$E(X) = \theta,\ \text{cov}(X) = \frac{\nu}{\nu-2}\Lambda = \Sigma\ \text{(say)}$$

We therefore reparametrize Λ in (4.1) by the covariance matrix $\Sigma = \frac{\nu}{\nu-2}\Lambda$ and obtain

$$f(x) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\Sigma|^{-\frac{1}{2}}\{(\nu-2)+(x-\theta)^T\Sigma^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}} \tag{4.2}$$

In what follows, we will denote this distribution (4.2) in short by $t_p(\theta,\Sigma,\nu)$, where θ will be referred to as the centre or mean, Σ the covariance matrix and ν the degrees of freedom of the t-distribution.

4.1 Marginal and conditional distributions

In notation of theorem 2.1, $g(\cdot)$ in (4.2) is given by $g(\cdot) = \{(\nu-2)+(x-\theta)^T\Sigma^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}}$. By direct integrations, it is readily verified that $g_m(\cdot)$ has the similar form as $g(\cdot)$ for the multivariate t-distribution. Therefore, we have the following theorem:

Theorem 4.1

If $X\sim t_p(\theta,\Sigma,\nu)$ and if X, θ and Σ are partitioned as

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},\quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix},\ \text{and}\ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where X_1 and X_2 are of order (rx1) and (p-rx1), θ_1, Σ_{11} of order (rx1) and (rxr) respectively, then $X_1 \sim t_r(\theta_1,\Sigma_{11},\nu)$.

Among others Cornish (1954, pp. 535-47), Raiffa and Schlaifer (1961, p. 528), Johnson and Kotz (1972) discuss the marginal and conditional t-distributions. Johnson and Kotz (1972) discuss the conditional multivariate t-distribution for the special case when $X \sim t_p(o,I_p,\nu)$. Cornish (1954) derives a conditional distribution function and uses that function to calculate the mean and the covariance matrix of a conditional t-distribution.

Raiffa and Schlaifer (1961) claim that if $X \sim t_p(\theta,\Sigma,\nu)$ and if X, θ, and Σ are partioned in a manner similar to that in theorem 4.1, then the conditional distribution of X_1 given $X_2=x_2$, is a multivariate t-distribution with appropriate mean, variance, and the degrees of freedom ν (same as the degrees of freedom of X). We now show that this claim is incorrect and the correct version is given in the following theorem 4.2.

Furthermore we note that the mean and covariance matrix of the conditional t-distribution [which were derived by Cornish (1954) from a conditional distribution function in a laborious way] directly follow from the form of the conditional distribution given in the theorem 4.2.

Theorem 4.2

If X is distributed according to $t_p(\theta,\Sigma,\nu)$ and if X, θ, Σ and Σ^{-1} are partitioned as

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \; \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \; \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \text{ and } \Sigma^{-1} = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}$$

where both X_1 and θ_1 are of order rx1 (say), and both of Σ_{11} and Σ^{11} are of order (rxr), then the distribution of X_1 given $X_2=x_2$ has the form

$$t_r\{\theta_1-(\Sigma^{11})^{-1}\Sigma^{12}(x_2-\theta_2),\ \frac{\nu-2+(x_2-\theta_2)^T\Sigma_{22}^{-1}(x_2-\theta_2)}{\nu-2+p-r}(\Sigma^{11})^{-1},(\nu+p-r)\}$$

Proof:

By definition, $f(x_1/x_2=x_2) = \dfrac{f(x_1,x_2)}{f(x_2)}$

For simplicity suppose $\underset{\sim}{\theta}=\underset{\sim}{0}$. Since by theorem 4.1, $X_2 \sim t_{p-r}(0,\Sigma_{22},\nu)$, therefore,

$$f(x_1/x_2) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{r}{2}}\,\Gamma\left(\frac{\nu+p-r}{2}\right)}|\Sigma|^{-\frac{1}{2}}|\Sigma_{22}|^{\frac{1}{2}}$$

$$\times\left[\{(\nu-2)+x_2^T\Sigma_{22}^{-1}x_2\}^{\frac{\nu+p-r}{2}}\{(\nu-2)+X^T\Sigma^{-1}x\}^{-\frac{\nu+p}{2}}\right] \quad (4.3)$$

In (4.3),

$$X^T\Sigma^{-1}X = (X_1\ X_2)^T\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}X_1\\ X_2\end{pmatrix}$$

$$= (X_1^T\Sigma^{11}X_1 + 2x_2^T\Sigma^{21}X_1 + x_2^T\Sigma^{22}x_2) \quad (4.4)$$

Since $\Sigma\Sigma^{-1}=I_p$, so $\Sigma_{22}^{-1} = \Sigma^{22} - (\Sigma^{21})^T(\Sigma^{11})^{-1}\Sigma^{12}$, provided Σ^{11} and Σ_{22} are non-singular [see for example Press (1972), p. 25-26]. Using this identity in (4.4), we get

$$X^T\Sigma^{-1}X = x_2^T\Sigma_{22}^{-1}x_2 + \{X_1 + (\Sigma^{11})^{-1}\Sigma^{12}x_2\}^T\Sigma^{11}\{X_1 + (\Sigma^{11})^{-1}\Sigma^{12}x_2\} \quad (4.5)$$

Substitute $X_1 + (\Sigma^{11})^{-1}\Sigma^{12}x_2 = W$. Then by (4.5), it follows from (4.3) that

$$f(w/x_2) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)|\Sigma^{11}|^{\frac{1}{2}}}{\pi^{\frac{r}{2}}\Gamma\left(\frac{\nu+p-r}{2}\right)}\{(\nu-2) + x_2^T\Sigma_{22}^{-1}x_2\}^{\frac{\nu+p-r}{2}}\{(\nu-2) + x_2^T\Sigma_{22}^{-1}x_2 + w^T\Sigma^{11}w\}^{-\frac{\nu+p}{2}}$$

This is because

$$|\Sigma|^{-\frac{1}{2}}|\Sigma_{22}|^{\frac{1}{2}} = |\Sigma^{-1}|^{\frac{1}{2}}|\Sigma_{22}|^{\frac{1}{2}}$$

$$= |\Sigma^{11}\Sigma^{22} - \Sigma^{12}\Sigma^{21}|^{\frac{1}{2}}\,|\Sigma^{22} - (\Sigma^{21})^T(\Sigma^{11})^{-1}\Sigma^{12}|^{-\frac{1}{2}}$$

$$= |\Sigma^{11}|^{\frac{1}{2}}$$

Now reparametrizing Σ^{11} by $\left\{\frac{(\nu-2) + x_2^T\Sigma_{22}^{-1}x_2}{\nu-2 + p-r}\right\}(\Sigma^{11})^*$, we obtain

$$f(w/x_2) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{r}{2}}\Gamma\left(\frac{\nu+p-r}{2}\right)}|\Sigma^{11^*}|^{\frac{1}{2}}(\nu-2+p-r)^{\frac{\nu+p-r}{2}}(\nu-2+p-r+w^T\Sigma^{11^*}w)^{-\frac{\nu+p}{2}} \qquad (4.6)$$

which is the p.d.f. of a multivariate t-distribution $t_r\{(0, (\Sigma^{11*})^{-1}, (\nu+p-r)\}$. Therefore, by linear transformation, we obtain the distribution of $X_1/X_2=x_2$ as in the theorem.

We note that it immediately follows from the above theorem:

$$E(X_1/X_2=x_2) = \theta_1 - (\Sigma^{11})^{-1}\Sigma^{12}(x_2-\theta_2)\text{ , and}$$

$$V(X_1/X_2=x_2) = (\Sigma^{11*})^{-1} = \frac{\{(\nu-2) + (x_2-\theta_2)^T\Sigma_{22}^{-1}(x_2-\theta_2)\}}{\nu+p - r-2}(\Sigma^{11})^{-1}\ .$$

4.2 On a conditional distribution considered by Raiffa and Schlaifer

Raiffa and Schlaifer (1961, p. 256) considered the student density function given by

$$f_s^{(r)}(z|m,H,\nu) \equiv \int_0^\infty f_N^{(r)}(z|m,hH)f_{\gamma 2}(h|1,\nu)dh \qquad (4.7)$$

$$= \frac{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+r-2}{2}\right)}{\pi^{\frac{r}{2}}\Gamma\left(\frac{\nu-2}{2}\right)}\{\nu + (z-m)^T H(z-m)\}^{-\frac{\nu+r}{2}}|H|^{\frac{1}{2}} ,$$

$$-\infty < z < \infty$$

$$-\infty < m < \infty$$

$$\nu > 0,\ H > 0$$

where $f_N^{(r)}(\cdot)$ is the p.d.f. of a r-dimensional multivariate normal distribution and $f_{\gamma 2}(\cdot)$ is the p.d.f. of gamma variate of a second type. They considered the following partitions

$$Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},\quad m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix},\quad H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix},$$

where z_1, m_1 are of order qx1 and H_{11} is of order (qxq). Then they have obtained the marginal distribution as

$$f_s^{(r-q)}(z_2|m_2, V_{22}^{-1},\nu)$$

$$= \int_0^\infty f_N(r-q)(z_2|m_2,h(H_{22}-H_{21}H_{11}^{-1}H_{12}))f_{\gamma 2}(h|1,\nu)dh \qquad (4.8)$$

$$= f_s^{(r-q)}(z_2|m_2,H_{22}-H_{21}H_{11}^{-1}H_{12},\nu) ,$$

and the conditional distribution of $Z_1/Z_2=z_2$ as

$$\begin{aligned} f_s^{(q)}&(z_1|m_1-H_{11}^{-1}H_{12}(z_2-m_2),H_{11},\nu) \\ &= \int_0^\infty f_N^{(q)}(z_1|m_1-H_{11}^{-1}H_{12}(z_2-m_2),hH_{11})f_{\gamma 2}(h|1,\nu)dh \\ &= f_s^{(q)}(z_1|m_1+V_{12}V_{22}^{-1}(z_2-m_2),\ (V_{11}-V_{12}V_{22}^{-1}V_{21})^{-1},\nu) \end{aligned} \tag{4.9}$$

where $V\equiv H^{-1}$, and V was partitioned as $V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$ with V_{22} as a positive definite matrix of order $(r-q)x(r-q)$

We now rewrite (4.7) as

$$\begin{aligned} f_s^{(r)}(z|m,H,\ ) &= \int_0^\infty f_N^{(r)}(z|m,hH)f_{\gamma 2}(h|1,\nu)dh \\ &= \int_0^\infty \Big[f_N^{(q)}(z_1\ m_1-H_{11}^{-1}H_{12}(z_2-m_2),hH_{11}) \\ &\quad x\ f_N^{(r-q)}(z_2|m_2,h(H_{22}-H_{22}H_{11}^{-1}H_{12})\cdot f_{\gamma 2}(h|1,\nu)\Big]dh \end{aligned} \tag{4.10}$$

Since the expression in (4.10) cannot be equal to the product of the two right hand expressions given by (4.8) and (4.9), therefore their method in obtaining conditional density is defective.

We also cite a simple example. Let us assume that in their derivations $H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, where Z is a two-dimensional variable. Then (4.8) yields

$$f(z_2) = \frac{\nu^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)} (\nu+z_2^2)^{-\frac{\nu+1}{2}}$$

and by (4.9) $f(z_1/z_2)$ is given by

$$f(z_1/z_2) = \frac{\nu^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)} (\nu+z_1^2)^{-\frac{\nu+1}{2}} .$$

Now because

$$f(z_2)f(z_1/z_2) \neq f(z_1,z_2) = \frac{\nu^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+1}{2}\right)}{\pi \Gamma\left(\frac{\nu}{2}\right)} (\nu+z_1^2+z_2^2)^{-\frac{\nu+2}{2}} ,$$

therefore their derivation for conditional density is defective.

4.3 Distribution of linear combinations

If $X \sim t_p(\theta,\Sigma,\nu)$, then it is readily verified that the standardized variable $\Sigma^{-\frac{1}{2}}(X-\theta) = Z$ is distributed according to $t_p(0,I_p\nu)$, so that the density function of Z is given by

$$f(z) = \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}} \Gamma\left(\frac{\nu}{2}\right)} \{(\nu-2) + \sum_{i=1}^{p} z_i^2\}^{-\frac{\nu+p}{2}} \qquad (4.11)$$

The distribution of Z given by (4.11) is spherically symmetric and Z is invariant under any orthogonal transformation. Thus $HZ \sim t_p(0,I_p,\nu)$, where H is any (pxp) orthogonal matrix. Conversely, if $Z \sim t_p(0,I_p,\nu)$ then $X = \theta + \Sigma^{\frac{1}{2}}Z \sim t_p(\theta,\Sigma,\nu)$, or, in general $X = \theta + BZ \sim t_p(\theta,BB^T,\nu)$, where B is a positive definite matrix of rank p. We now derive the distribution of linear combinations as in the following theorem.

Theorem 4.3 (standardization)

Let $X \sim t_p(\theta,\Sigma,\nu)$ and A be any qxp matrix (q≤p) of rank q, then

$$(A\Sigma A^T)^{-\frac{1}{2}}A(X-\theta) \sim t_q(0,I_q,\nu)$$

Proof:

Set $Z = \Sigma^{-\frac{1}{2}}(X-\theta)$. Then $Z \sim t_p(0,I_p,\nu)$. From theorems 2.2 and 4.1 it follows that

$$(LL^T)^{-\frac{1}{2}} LZ \sim t_q(0,I_q,\nu) ,$$

where $L=A\Sigma^{\frac{1}{2}}$. L is a matrix of order qxp with rank q, because by hypothesis A is a qxp matrix of rank q and Σ is a pxp non-singular matrix of rank p. Hence the theorem.

We note that from the above theorem it follows that the linear combination AX has the distribution of the form $AX \sim t_q(A\theta,A\Sigma A^T,\nu)$, where A is a qxp matrix of rank q.

4.4 Characteristic function

Fisher and Healy (1956) gave a polynomial expression for the characteristic function ϕ_ν of the univariate t-distribution with odd degrees of freedom only. They developed the function using recurrence relations. Ifram (1970) has given results for ϕ_ν of the t_ν-distribution but they are found incorrect by Pestana (1977).

In this section we obtain the characteristic function of the multivariate t-distribution and show that the c.f. has the pedagogical virtue of reducing the multivariate problem to the case of an analogous univariate problem. Furthermore we discuss some applications of the characteristic function.

4.4.1 The characteristic function of multivariate t-distribution

Theorem 4.4

If X has a p-dimensional multivariate t-distribution with the p.d.f.

$$f(x,\theta,\Sigma,\nu) = K_p|\Sigma|^{-\frac{1}{2}}\{(\nu-2) + (x-\theta)^T\Sigma^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}},$$

where $-\infty < X < \infty$, $\nu > 2$, and $K_p = \dfrac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}$; then the characteristic function of X is given by

(a) for odd degrees of freedom ν

$$\phi_X(t_1,\ldots,t_p) = \frac{\sqrt{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right) \cdot e^{it^T\theta-\sqrt{(\nu-2)t^T\Sigma t}}}{2^{\nu-1}\Gamma\left(\frac{\nu}{2}\right)} \sum_{r=1}^{m}\left[\binom{2m-1-r}{m-r}\frac{\{2\sqrt{(\nu-2)t^T\Sigma t}\}^{r-1}}{(r-1)!}\right],$$

where $\frac{\nu+1}{2} = m$.

(b) for <u>even</u> ν

$$\phi_X(t_1,\ldots,t_p) = \frac{(-1)^{m+1}\,\Gamma\!\left(\frac{\nu+1}{2}\right) e^{it^T\theta}}{\sqrt{\pi}\,\Gamma\!\left(\frac{\nu}{2}\right)\prod_{j=1}^{m}(m+\frac{1}{2}-j)} \sum_{n=0}^{\infty}\left[\left(\frac{\sqrt{(\nu-2)t^T\Sigma t}}{2}\right)^{2n}\right.$$

$$\left.\times\ \frac{1}{(n!)^2}\prod_{j=0}^{m-1}(n-j)\left\{\sum_{j=0}^{m-1}\frac{1}{n-j} + \log\left(\frac{(\nu-2)t^T\Sigma t}{4}\right) - \frac{\Gamma'(n+1)}{\Gamma(n+1)}\right\}\right]$$

where $m = \frac{\nu}{2}$ and $\frac{\Gamma'(n+1)}{\Gamma(n+1)}$ is the logarithmic derivation of the gamma function $\Gamma(n+1)$.

(c) for fractional ν

$$\phi_X(t_1,\ldots,t_p) = \frac{(\nu-2)^{\frac{\nu}{2}-m}\,\Gamma\!\left(\frac{\nu+1}{2}\right)(-1)^m}{\Gamma\!\left(\frac{\nu}{2}\right)\cdot 2^{\xi}\,\Gamma(\xi+\frac{1}{2})\prod_{j=1}^{m}\left(\frac{\nu+1}{2}-j\right)} \cdot \frac{\pi}{\sin\xi\pi} \cdot e^{it^T\theta}$$

$$\times \sum_{n=0}^{\infty}\left[\left(\frac{\sqrt{(\nu-2)t^T\Sigma t}}{2}\right)^{2n} \cdot \frac{1}{n!}\right.$$

$$\left.\left\{\frac{2^{\xi}\prod_{j=0}^{m-1}(n-\xi-j)}{\nu^{\xi}\,\Gamma(n+1-\xi)} - \frac{(t^T\Sigma t)^{\xi}\prod_{j=0}^{m-1}(n-j)}{2^{\xi}\,\Gamma(n+1+\xi)}\right\}\right]$$

where $\xi = \frac{\nu}{2} - \left[\frac{\nu+1}{2}\right] = \frac{\nu}{2} - m$, with $[X]$ as an integral value of x.

Proof:

Given $f(x,\theta,\Sigma,\nu) = K_p|\Sigma|^{-\frac{1}{2}}\{(\nu-2)+(x-\theta)^T\Sigma^{-1}(x-\theta)\}^{-\frac{\nu+P}{2}}$, where

$K_p = \dfrac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+P}{2}\right)}{\pi^{\frac{P}{2}}\,\Gamma\left(\frac{\nu}{2}\right)}$, $-\infty < x < \infty$, $\nu > 0$. We wish to derive

$\phi_X(t_1,t_2,\ldots,t_p) = \mathrm{Exp}(it^Tx)$, where $t^T = (t_1,t_2,\ldots,t_p)$.

Let us make the transformation $\Sigma^{-\frac{1}{2}}(x-\theta) = Y$, then

$g(\underset{\sim}{y}) = K_p\{(\nu-2)+\sum_{i=1}^{p} y_i^2\}^{-\frac{\nu+p}{2}}$ and the c.f. of Y is given by

$\phi_Y(u_1,\ldots,u_p) = K_p\int e^{iu^Ty}\{(\nu-2)+\sum_{i=1}^{p} y_i^2\}^{-\frac{\nu+P}{2}}dy$. To evaluate the last integral, we make an orthogonal transformation $Y_{Px1} = \Gamma_{PxP}\, Z_{Px1}$, with first column of the orthogonal matrix Γ_{PxP} as the vector

$\frac{u_1}{|u|}, \frac{u_2}{|u|}, \ldots, \frac{u_p}{|u|}$, where $|u| = (u^Tu)^{\frac{1}{2}}$. Then

$\phi_Z(u_1,u_2,\ldots,u_p) = K_p\int e^{i|u|z_1}((\nu-2)+\sum_{i=1}^{p} z_i^2)^{-\frac{\nu+p}{2}}\,d\underset{\sim}{z}$, which again can be written as

$$\phi(u_1,\ldots,u_p) = K_p\int e^{i\,u\cdot z_1}dz_1\int\cdots\int\left[c_p+z_p^2\right]^{-\frac{\nu+P}{2}}dz_2\ldots dz_p, \qquad (4.12)$$

where $c_p = (\nu-2) + \sum_{i=1}^{p-1} z_i^2$

Let $I_p \equiv \int \frac{dz_p}{(c_p+z_p^2)^{\frac{\nu+p}{2}}}$. On substituting $\sqrt{c_p}\tan\theta$ for z_p, we get

after a little algebra,

$$I_p = \frac{1}{(c_p)^{\frac{\nu+p-1}{2}}} \cdot \frac{\pi^{\frac{1}{2}} \Gamma\left(\frac{\nu+p-1}{2}\right)}{\Gamma\left(\frac{\nu+p}{2}\right)} = \frac{1}{c_p^{\frac{\nu+p-1}{2}}} \cdot \ell_p \quad \text{(say)}$$

Now from (4.12) it immediately follows that

$$\phi_Z(u_1,\ldots,u_p) = K_p\cdot\ell_p\cdot\ell_{p-1}\cdot \;\ldots\; \cdot\ell_3\cdot\ell_2\cdot\int_{-\infty}^{\infty} e^{i|u|z_1}((\nu-2)+z_1^2)^{-\frac{\nu+1}{2}} dz_1$$

$$= K_p\cdot\ell_p\cdot\ell_{p-1}\cdot \;\ldots\; \cdot\ell_3\cdot\ell_2\cdot J_1 \quad \text{(say)} \qquad (4.13)$$

Case (a) <u>when ν is odd</u>

For odd ν we evaluate J_1 as follows: Let us consider a contour consisting of the segment $[-R,R]$ of the real axis and the upper semi-circle $|w| = R$, $I_m w \geq 0$, and also consider the function

$f(w) = \frac{e^{i|u|w}}{((\nu-2)+w^2)^{\frac{\nu+1}{2}}}$. Now, for $R > (\nu-2)$, the contour encloses the pole of $f(w)$ at $w_0 = i\sqrt{(\nu-2)}$, this pole is of mth order, where $\frac{\nu+1}{2} = m$ and ν is odd, the corresponding residue being

$$C_{-1} = \frac{1}{(m-1)!} \lim_{w \to i\sqrt{(\nu-2)}} \left[\frac{\partial^{m-1}}{\partial w^{m-1}} \{(w-i\sqrt{(\nu-2)})^m \right.$$

$$\left. \times \frac{e^{i|u|w}}{((w+i\sqrt{(\nu-2)})(w-i\sqrt{(\nu-2)}))^m}\}_{w=i\sqrt{\nu-2}}\right]$$

Differentiating the function (in 2nd bracket) (m-1) times and putting $w = i\sqrt{(\nu-2)}$, we get

$$C_{-1} = \frac{e^{-|u|\sqrt{(\nu-2)}}}{(i)\left[(m-1)!\right]^2} \sum_{r=1}^{m} \frac{(2m-r-1)!}{(2\sqrt{(\nu-2)})^{2m-r}} \cdot \binom{m-1}{r-1} (|u|)^{r-1}$$

Since, by an elementary residue theorem

$$\int_{-R}^{R} f(w)dw \equiv \int_{-\infty}^{\infty} \frac{e^{-i|u|z_1}}{(z_1^2+\sqrt{(\nu-2)})^m} dz_1 = 2\pi i. \quad C_{-1}, \text{ as } R\to\infty, \text{ therefore,}$$

we get

$$J_1 = \frac{2\, e^{-|u|\sqrt{(\nu-2)}}}{(2\sqrt{(\nu-2)})^{2m-1}} \sum_{r=1}^{m} \binom{2m-1-r}{m-r} \frac{(2|u|\sqrt{(\nu-2)})^{r-1}}{(r-1)!}$$

Now, using J_1 in (4.13), we obtain

$$\phi_Y(\cdot) = \phi_Z(u_1,\ldots,u_p) = \frac{\sqrt{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)}{2^{\nu-1}\,\Gamma\left(\frac{\nu}{2}\right)} e^{-|u|\sqrt{(\nu-2)}} \sum_{r=1}^{m} \binom{2m-1-r}{m-r} \frac{(2|u|\sqrt{(\nu-2)})^{r-1}}{(r-1)!} \tag{4.14}$$

Next substituting back $X = \theta + \Sigma^{\frac{1}{2}}Y$, we obtain

$$\phi_X(t_1,\ldots,t_p) = Ee^{it^TX}$$

$$= e^{it^T\theta}Ee^{it^T\Sigma^{\frac{1}{2}}Y}$$

$$\phi_X(t_1,\ldots,t_p) = \frac{\sqrt{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)}{2^{\nu-1}\,\Gamma\left(\frac{\nu}{2}\right)} e^{it^T\theta-\sqrt{(\nu-2)}\sqrt{t^T\Sigma t}} \sum_{r=1}^{m} \binom{2m-1-r}{m-r} \frac{(2\sqrt{(\nu-2)t^T\Sigma t})^{r-1}}{(r-1)!}. \tag{4.15}$$

which is the characteristic function of the multivariate t-distribution when ν is odd. Hence the case (a) of the theorem is proved. Note that this characteristic function is of the form $e^{it^T\theta}\Psi_1(t^T\Sigma t)$, where

$$\Psi_1(x) = \frac{\sqrt{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)}{2^{\nu-1}\,\Gamma\left(\frac{\nu}{2}\right)} e^{-\sqrt{(\nu-2)x}} \sum_{r=1}^{m} \binom{2m-1-r}{m-r} \frac{(2\sqrt{(\nu-2)x})^{r-1}}{(r-1)!} \qquad (4.16)$$

Case (b) when ν is even

Whether ν is an even or a fraction, $\frac{\nu+1}{2}$ is not an integer quantity. Let $\frac{\nu+1}{2} = m+q$, where m is the integral part and q is the fractional part. In this case, the pole of the fraction f(w) considered in case (a) is of fractional order. We evaluate the integral $J_1 = \int_{-\infty}^{\infty} f(z_1)dz_1$ as follows:

Suppose $f(z_1)$ has pole at $\pm i\sqrt{a(\nu-2)}$, where 'a' is a variable such that $1-h \leq a \leq 1+h$ and h is very small positive quantity. Also, suppose $J_2 = \int_{-\infty}^{\infty} e^{i|u|z_1}(z_1^2+a(\nu-2))^{-\frac{\nu+1}{2}} dz_1$, and J_2 is differentiable with respect to 'a'. Then, J_2 can be demonstrated very easily as

$$J_2 = \frac{1}{(\nu-2)^{m+q}} \frac{\partial^m}{\partial a^m} \int_{-\infty}^{\infty} e^{i|u|z_1} \int \cdots \int \{a+z_1^2 / (\nu-2)\}^{-m-q}\, da..da dz_1$$

$$J_2 = \frac{(-1)^m}{(\nu-2)^m \prod_{j=1}^{m} (m+q-j)} \frac{\partial^m}{\partial a^m} \int_{-\infty}^{\infty} e^{i|u|z_1} \{z_1^2+a(\nu-2)\}^{-q} dz_1 \qquad (4.17)$$

Next let $q = \xi + \frac{1}{2}$ so that $|\xi| < \frac{1}{2}$, we immediately obtain

$$J_2 = \frac{(-1)^m}{(\nu-2)^m \prod_{j=1}^{m} (m+q-j)} \frac{\partial^m}{\partial a^m} \frac{\sqrt{\pi}\, |u|^{\xi} \mathcal{K}_{\xi}(\sqrt{a(\nu-2)}\,|u|)}{2^{\xi-1} (\sqrt{a(\nu-2)})\, \Gamma(\xi+\frac{1}{2})}, \tag{4.18}$$

because $\int_{-\infty}^{\infty} \frac{e^{izx}}{(x^2+a^2)^{\nu+\frac{1}{2}}} dx = \frac{\sqrt{\pi}\, z^{\nu} \mathcal{K}_{\xi}(az)}{2^{\nu-1} a^{\nu}\, \Gamma(\nu+\frac{1}{2})}$, $a > 0$, $z > 0$, $|\nu| < \frac{1}{2}$

(see q 24, p. 63, Gröbner and Hofreiter, 1961), where $\mathcal{K}_{\xi}(\cdot)$ is a modified Bessel function of the second kind, and $\mathcal{K}_{\xi}(\sqrt{a(\nu-2)}\,|u|)$ is given

by $$\mathcal{K}_{\xi}(\sqrt{a(\nu-2)}\,|u|) = (-1) I_0(\sqrt{a(\nu-2)}\,|u|) \log\left(\frac{\sqrt{a(\nu-2)}\,|u|}{2}\right) + \sum_{n=0}^{\infty} \frac{\Psi(n+1)}{(n!)^2} \left(\frac{\sqrt{a(\nu-2)}\,|u|}{2}\right)^{2n},$$

for $\xi = 0$, where

$$I_0(\sqrt{a(\nu-2)}\,|u|) = \sum_{n=0}^{\infty} \frac{1}{(n!)^2} \left(\frac{\sqrt{a(\nu-2)}\,|u|}{2}\right)^{2n},$$

and $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the logarithmic derivative of the gamma function.

Note that $\xi=0$ is our interest now, because $\xi=0$ implies $q=\frac{1}{2}$, which in turn implies ν is even and is given by $\nu=2m$, where m is an integer quantity.

Putting $\xi=0$, (4.18) can be re-written as

$$J_2 = \frac{(-1)^m}{(\nu-2)^m \prod_{j=1}^{m} (m+q-j)} \frac{\partial^m}{\partial a^m} \left\{\frac{\mathcal{K}_0(\sqrt{a(\nu-2)}\,|u|)}{2^{-1}}\right\} \tag{4.19}$$

Now, we need

$$\frac{\partial^m}{\partial a^m}\{\mathcal{K}_o(\sqrt{a(\nu-2)}|u|)\}$$

$$= \frac{\partial^m}{\partial a^m}\ (-1)\sum_{n=0}^{\infty}\frac{1}{(n!)^2}\left(\frac{\sqrt{a(\nu-2)}|u|}{2}\right)^{2n}\cdot\log\left(\frac{\sqrt{a(\nu-2)}|u|}{2}\right)$$

$$+\sum_{n=0}^{\infty}\frac{\Gamma'(n+1)}{\Gamma(n+1)}\cdot\frac{1}{(n!)^2}\left(\frac{\sqrt{a(\nu-2)}|u|}{2}\right)^{2n}$$

$$= (-1)\frac{\partial^m}{\partial a^m}\left[\sum_{n=0}^{\infty}\left(\frac{\sqrt{(\nu-2)}|u|}{2}\right)^{2n}\frac{1}{(n!)^2}\right.$$

$$\left.\left\{(a)^n\cdot\log\left(\frac{\sqrt{a(\nu-2)}|u|}{2}\right)-\frac{\Gamma'(n+1)}{\Gamma(n+1)}(a)^n\right\}\right]$$

$$= (-1)\frac{\partial^m}{\partial a^m}\left[\sum_{n=0}^{\infty}\left(\frac{\sqrt{(\nu-2)}|u|}{2}\right)^{2n}\frac{1}{(n!)^2}\right.$$

$$\left.\{a^n\log\sqrt{a}+\log\frac{\sqrt{(\nu-2)}|u|}{2}\quad-\frac{\Gamma'(n+1)}{\Gamma(n+1)}(a)^n\}\right]$$

Now

$$\frac{\partial^m}{\partial a^m}(a^n \log\sqrt{a})$$

$$= \tfrac{1}{2}\frac{\partial^m}{\partial a^m}(a^n \log a)$$

$$= \tfrac{1}{2}\left[\prod_{j\neq 0}^{m-1}(n-j) + \prod_{j\neq 1}^{m-1}(n-j) + \dots + \prod_{j\neq (m-1)}^{m-1}(n-j)\right]a^{n-m}$$

$$+ \prod_{j=0}^{m-1}(n-j)\log a\cdot a^{n-m} \qquad (4.20)$$

and

$$\frac{\partial^m}{\partial a^m}(a^n) = n(n-1)(n-2)\dots(n-m+1)a^{n-m}$$

$$= \prod_{j=0}^{m-1}(n-j)a^{n-m}$$

Therefore, using (4.20) in (4.19), we obtain

$$J_1 = \frac{\partial^m}{\partial a^m}\{J_2\}_{a=1}$$

$$= \frac{(-1)^{m+1}\cdot 2}{(\nu-2)^m \prod_{m=1}^{m}(m+q-j)}\left[\sum_{n=0}^{\infty}\left(\frac{\sqrt{(\nu-2)}|u|}{2}\right)^{2n}\frac{1}{(n!)^2}\{\tfrac{1}{2}\cdot\prod_{j=0}^{m-1}(n-j)\left(\sum_{j=0}^{m-1}\frac{1}{n-j}\right)\right.$$

$$\left. + \tfrac{1}{2}\prod_{j=0}^{m-1}(n-j)\cdot\log\left(\frac{(\nu-2)|u|^2}{4}\right) - \frac{\Gamma'(n+1)}{\Gamma(n+1)}\prod_{n=0}^{m-1}(n-j)\}\right]$$

$$= \frac{(-1)^{m+1}}{(\nu-2)^m \prod_{j=1}^{m} (m+q-j)} \sum_{n=0}^{\infty} \left[\frac{\sqrt{(\nu-2)}\,|u|}{2}^{2n} \cdot \frac{1}{n!^2} \cdot \prod_{j=0}^{m-1} (n-j) \right.$$

$$\left. \left\{ \sum_{j=0}^{m-1} \frac{1}{n-j} + \log\left(\frac{(\nu-2)\ u^2}{4}\right) - \frac{\Gamma'(n+1)}{\Gamma(n+1)} \right\} \right]$$

Hence, by (4.13), we obtain

$$\phi_Z(u_1,\ldots,u_P) = K_P \cdot \ell_P \cdot \ \ldots\ \cdot \ell_2 \cdot J_1$$

$$= \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+P}{2}\right)}{\pi^{\frac{P}{2}} \Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{\pi^{\frac{1}{2}} \Gamma\left(\frac{\nu+P-1}{2}\right)}{\Gamma\left(\frac{\nu+P}{2}\right)} \cdot \ldots \cdot \frac{\pi^{\frac{1}{2}} \Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu+2}{2}\right)} \cdot J_1$$

$$= \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)} \cdot J_1 = \frac{(\nu-2)^m \Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)} \cdot J_1$$

i.e. $\phi_Y(u_1,\ldots,u_P) = \phi_Z(u_1,\ldots,u_P)$.

$$= \frac{(-1)^{m+1} \Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi} \prod_{j=1}^{m} (m+\frac{1}{2}-j)} \sum_{n=0}^{\infty} \left[\left(\frac{\sqrt{(\nu-2)}\,|u|}{2}\right)^{2n} \cdot \frac{1}{(n!)^2} \prod_{j=0}^{m-1} (n-j) \left\{ \sum_{j=0}^{m-1} \frac{1}{n-j} \right.\right.$$

$$\left.\left. + \log \frac{(\nu-2)|u|^2}{4} - \frac{\Gamma'(n+1)}{\Gamma(n+1)} \right\} \right], \qquad (4.21)$$

because $q=\frac{1}{2}$ for even ν.

Next as in case (a) substituting back $X = \theta+\Sigma^{\frac{1}{2}}Y$, we obtain $\phi_X(t_1,\ldots,t_P)$ as in case (b) of the theorem. Note that for even ν, the characteristic function is of the form

$$e^{i\underset{\sim}{t}^T\underset{\sim}{\theta}}\psi_2(t^T\Sigma t) \ ,$$

$$\text{where } \Psi_2(x) = \frac{(-1)^{m+1}\,\Gamma\!\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\,\Gamma\!\left(\frac{\nu}{2}\right)\prod_{j=1}^{m}(m+\tfrac{1}{2}+j)} \sum_{n=0}^{\infty}\left[\left(\frac{\sqrt{(\nu-2)x}}{2}\right)^{2n}\cdot\frac{1}{(n!)^2}\prod_{j=0}^{m-1}(n-j)\right.$$

$$\left. \times\left\{\sum_{j=0}^{m-1}\frac{1}{n-j} + \log\frac{(\nu-2)x}{4} - \frac{\Gamma'(n+1)}{\Gamma(n+1)}\right\}\right]$$

Case (c) <u>when ν is a fraction</u>

We set $\frac{\nu+1}{2} = m+q = m+\xi+\frac{1}{2}$, where $|\xi|<\frac{1}{2}$, and m is the integral part of $\frac{\nu+1}{2}$. Now if ν is a fractional quantity, then $\xi\neq 0$ but $|\xi|<\frac{1}{2}$. In this case, when $\xi\neq 0$, $\mathfrak{K}_\xi(\cdot)$ in (4.18) is given by

$$\mathfrak{K}_\xi(\sqrt{a(\nu-2)}\,|u|) = \frac{\pi}{2\sin\xi\pi}\left[I_{-\xi}(\sqrt{a(\nu-2)}\,|u|) - I_\xi(\sqrt{a(\nu-2)}\,|u|)\right] \ ,$$

where $I_\xi(\sqrt{a(\nu-2)}\,|u|) = \sum_{n=0}^{\infty}\frac{1}{n!}\frac{1}{\Gamma(n+\xi+1)}\left(\frac{\sqrt{a(\nu-2)}\,|u|}{2}\right)^{\xi+2n}$, is a modified Bessel function of the first kind (cf. Grobner and Hofreiter, 1961, pp. 192-193). Therefore, by (4.18),

$$J_2 = \frac{(-1)^m\sqrt{\pi}\,|u|^\xi}{(\nu-2)^m\prod_{j=1}^{m}(m+q-j)\cdot 2^{\xi-1}(\sqrt{(\nu-2)})^\xi\,\Gamma(\xi+\frac{1}{2})}\frac{\partial^m}{\partial a^m}\left\{\frac{\mathfrak{K}_\xi(\sqrt{a(\nu-2)}\,|u|)}{(\sqrt{a})^\xi}\right\} \quad (4.23)$$

We now simplify

$$\frac{\partial^m}{\partial a^m}\left\{\frac{\mathfrak{K}_\xi(\sqrt{a(\nu-2)}\,|u|)}{(\sqrt{a})^\xi}\right\}$$

$$= \frac{\partial^m}{\partial a^m}\left\{\frac{I_{-\xi}(\sqrt{a(\nu-2)}|u|)}{(\sqrt{a})^{\xi}} - \frac{I_{\xi}(\sqrt{a(\nu-2)}|u|)}{(\sqrt{a})^{\xi}}\right\}\cdot\frac{\pi}{2\sin\xi\pi}$$

$$= \frac{\pi}{2\sin\xi\pi}\frac{\partial^m}{\partial a^m}\left[\sum_{n=0}^{\infty}\frac{1}{n!\,\Gamma(n+\xi+1)}\frac{1}{(\sqrt{a})^{\xi}}\left(\frac{\sqrt{a(\nu-2)}|a|}{2}\right)^{-\xi+2n}\right.$$

$$\left. - \sum_{n=0}^{\infty}\frac{1}{n!\,\Gamma(n+\xi+1)}\frac{1}{(\sqrt{a})^{\xi}}\frac{\sqrt{a(\nu-2)}|u|}{2}^{\xi+2n}\right]$$

$$= \frac{\pi}{2\sin\xi\pi}\sum_{n=0}^{\infty}\{\frac{1}{n!\,\Gamma(n-\xi+1)}\cdot\left(\frac{\sqrt{(\nu-2)}|u|}{2}\right)^{-\xi+2n}\frac{\partial^m}{\partial a^m}a^{n-\xi}$$

$$- \frac{1}{n!\,\Gamma(n+\xi+1)}\left(\frac{\sqrt{(\nu-2)}|u|}{2}\right)^{\xi+2n}\frac{\partial^m}{\partial a^m}(a^n)\}$$

Since $\frac{\partial^m}{\partial a^m}a^{n-\xi}$

$$= (n-\xi)(n-\xi-1)\;\ldots\;(n-\xi-m+1)a^{n-\xi-m+1}$$

and $\frac{\partial^m}{\partial a^m}a^n = n(n-1)\;\ldots\;(n-m+1)a^{n-m+1}$, therefore the last equation can be written as

$$\frac{\pi}{2\sin\xi\pi}\sum_{n=0}^{\infty}\left(\frac{\sqrt{(\nu-2)}|u|}{2}\right)^{2n}\cdot\frac{1}{n!}\{\frac{2^{\xi}(\sqrt{(\nu-2)}|u|)}{\Gamma(n-\xi+1)}\prod_{j=0}^{m-1}(n-\xi-j)\cdot a^{n-\xi-m+1}$$

$$- \frac{(|u|\sqrt{(\nu-2)})^{\xi}}{2^{\xi}\,\Gamma(n+\xi+1)}\prod_{j=0}^{m-1}(n-j)a^{n-m+1}\}$$

Therefore, using this equation in (4.21), we get

$$J_2 = \frac{(-1)^m\sqrt{\pi}}{(\nu-2)^m \prod_{j=1}^{m}(m+q-j)\cdot 2^{\xi}\,\Gamma(\xi+\frac{1}{2})}\cdot\frac{\pi}{\operatorname{Sin}n\pi}\sum_{n=0}^{\infty}\left[\left(\frac{|u|\sqrt{(\nu-2)}}{2}\right)^{2n}\cdot\frac{1}{n!}\right.$$

$$\left.\times\left\{\frac{2^{\xi}\prod_{j=0}^{m-1}(n-\xi-j)a^{n-\xi-m+1}}{(\nu-2)^{\xi}\,\Gamma(n-\xi+1)} - \frac{|u|^{2\xi}\prod_{j=0}^{m-1}(n-j)}{2^{\xi}\,\Gamma(n+1+\xi)}\right\}\right]$$

Now $J_1 = J_2\Big]_{a=1}$. Putting J_1 in (4.11), we finally obtain

$$\phi_Z(u_1,\ldots,u_p) = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{\frac{1}{2}}\,\Gamma\left(\frac{\nu}{2}\right)}\cdot J_1$$

$$= \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+1}{2}\right)(-1)^m}{\Gamma\left(\frac{\nu}{2}\right)\cdot 2^{\xi}\,\Gamma(\xi+\frac{1}{2})\prod_{j=1}^{m}(m+q-j)}\cdot\frac{\pi}{\operatorname{Sin}\xi\pi}$$

$$\times\sum_{n=0}^{\infty}\left[\left(\frac{\sqrt{(\nu-2)}\,|u|}{2}\right)^{2n}\cdot\frac{1}{n!}\left\{\frac{2^{\xi}\prod_{j=0}^{m-1}(n-\xi-j)}{(\nu-2)^{\xi}\,\Gamma(n+1-\xi)}\right.\right.$$

$$\left.\left. - \frac{|u|^{2\xi}\prod_{j=0}^{m-1}(n-j)}{2^{\xi}\,\Gamma(n+1+\xi)}\right\}\right] \tag{4.24}$$

Next substituting back $X = \theta+\Sigma^{\frac{1}{2}}Y$, we get the c.f. $\phi_X(t_1,\ldots,t_p)$ as in case (c) of the theorem. This $\phi_X(t_1,\ldots,t_p)$ can be written as

$$e^{it^T\theta}\Psi_3(t^T\Sigma t)\ ;$$

where
$$\Psi_3(x) = \frac{(\nu-2)^{\frac{\nu}{2}-m}\,\Gamma\!\left(\frac{\nu+1}{2}\right)(-1)^m}{\Gamma\!\left(\frac{\nu}{2}\right)\cdot 2^{\xi}\,\Gamma(\xi+\tfrac{1}{2})\prod_{j=1}^{m}(m+q-j)} \cdot \frac{\pi}{\sin\xi\pi}$$

$$\times \sum_{n=0}^{\infty}\left[\left(\frac{\sqrt{(\nu-2)x}}{2}\right)^{2n}\cdot\frac{1}{n!}\left\{\frac{2^{\xi}\prod_{j=0}^{m-1}(n-\xi-j)}{(\nu-2)^{\xi}\,\Gamma(n+1-\xi)} - \frac{(\sqrt{x})^{2\xi}\prod_{j=0}^{m-1}(n-j)}{2^{\xi}\,\Gamma(n+1+\xi)}\right\}\right]$$

4.4.2 Special case - the characteristic function of the univariate t-distribution

Putting $u_1=u$, and $u_2=u_3=\ldots=u_p=0$ in (4.14), (4.21) and (4.24), we obtain the characteristic function of the univariate t distribution as follows:

(i) For odd ν

$$\phi_X(u) = Ee^{iux} = \frac{\sqrt{\pi}\,\Gamma\!\left(\frac{\nu+1}{2}\right)\cdot e^{-u\sqrt{(\nu-2)}}}{2^{\nu-1}\,\Gamma\!\left(\frac{\nu}{2}\right)}\cdot\sum_{r=1}^{m}\binom{2m-1-r}{m-r}\frac{(2u\sqrt{\nu-2})^{r-1}}{(r-1)!}$$

where $(\nu+1)/2 = m$,

(ii) for even ν

$$\phi_X(u) = Ee^{iux} = \frac{(-1)^{m+1}\,\Gamma\!\left(\frac{\nu+1}{2}\right)}{\Gamma\!\left(\frac{\nu}{2}\right)\sqrt{\pi}\prod_{j=0}^{m}(m+\tfrac{1}{2}-j)}\sum_{n=0}^{\infty}\left[\left(\frac{u\sqrt{(\nu-2)}}{2}\right)^{2n}\cdot\frac{1}{(n!)^2}\prod_{j=0}^{m-1}(n-j)\right.$$

$$\left.\times\left\{\sum_{j=0}^{m-1}\frac{1}{n-j} + \log\left(\frac{(\nu-2)u^2}{4}\right) - \frac{\Gamma'(n+1)}{\Gamma(n+1)}\right\}\right]$$

where $m = \nu/2$, and

(iii) for fractional ν

$$\phi_X(u) = Ee^{iux} = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) 2^{\xi}\,\Gamma\left(\xi+\tfrac{1}{2}\right)\prod_{j=0}^{m}(m+q-j)} \cdot \frac{\pi}{\sin\xi\pi}$$

$$\times \sum_{n=0}^{\infty}\left[\left(\frac{u\sqrt{\nu-2}}{2}\right)^{2n} \cdot \frac{1}{n!}\left\{\frac{2^{\xi}\prod_{j=0}^{m-1}(n-\xi-j)}{(\nu-2)^{\xi}\,\Gamma(n+1-\xi)} - \frac{u^{2\xi}\prod_{j=0}^{m-1}(n-j)}{2^{\xi}\,\Gamma(n+1-\xi)}\right\}\right]$$

Moments of t-distribution using characteristic function

All odd moments of multivariate t-distribution are zero. If, in (4.2), $\theta=0$ and $\Sigma=I$, identity matrix, then the p.d.f. of X is

$$f(x) = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\,\Gamma\left(\frac{\nu}{2}\right)}\left\{(\nu-2) + \sum_{i=1}^{p} x_i^2\right\}^{-\frac{\nu+p}{2}}$$

It follows from Johnson and Kotz (1972, p. 136) that the even moments of the above spherical t-distribution, if they exist, is given by

$$\mu_{2r_1,2r_2,\ldots,2r_p} = \frac{(\nu-2)^{\sum_{i=1}^{p} r_i}\,\Gamma\left(\frac{\nu}{2} - \sum_{i=1}^{p} r_i\right)}{(\sqrt{\pi})\,\Gamma\left(\frac{\nu}{2}\right)} \prod_{i=1}^{p}\Gamma\left(r_i+\tfrac{1}{2}\right);$$ they exist when

$2\sum_{i=1}^{p} r_i < \nu$. This result can be verified using any of the equations (4.14), (4.21) and (4.24). For simplicity, we prove $\mu_{2,0,\ldots,0}=1$ for all types of ν. This moment namely $\mu_{2,\ldots,0}$ is in fact the second moment of the univariate student's t-distribution.

Case 1) ν odd

To prove $\mu_{2,0,\ldots,0}=1$ in this case, we put $u_1=u$, $u_2=u_3=\ldots=u_p=0$ in (4.14) and then collect the coefficient of $(iu_1)^2/2$. From (4.14), it follows that

$$\phi(u,0,\ldots,0) = \frac{\sqrt{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)}{2^{\nu-1}\,\Gamma\left(\frac{\nu}{2}\right)}\left\{e^{-u\sqrt{(\nu-2)}}\sum_{r=1}^{m}\binom{2m-1-r}{m-r}\frac{(2u\sqrt{\nu-2})^{r-1}}{(r-1)!}\right\}$$

$$= \frac{\sqrt{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)}{2^{\nu-1}\,\Gamma\left(\frac{\nu}{2}\right)}\left\{\left(1-u\sqrt{\nu-2}+\frac{u^2(\nu-2)}{2!}-\frac{u^3(\nu-2)^{\frac{3}{2}}}{3!}+\ldots\right)\right\}$$

$$\times\left\{\binom{2m-2}{m-1}+\binom{2m-3}{m-2}(2u\sqrt{\nu-2})^1\right.$$

$$\left.+\binom{2m-4}{m-3}\frac{(2u\sqrt{\nu-2})^2}{2!}+\ldots\right\}$$

Now, coefficient of u^2 from the last expression is given by

$$\frac{\sqrt{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)}{2^{\nu-1}\,\Gamma\left(\frac{\nu}{2}\right)}\left\{\binom{2m-4}{m-3}\cdot 2(\nu-2)-\binom{2m-3}{m-2}2(\nu-2)+\frac{\nu-2}{2}\binom{2m-2}{m-1}\right\},$$

where $(\nu+1)/2 = m$. After a simple algebra, this coefficient of u^2 reduces to $-\frac{1}{2}$. Thus, coefficient of $\frac{(iu)^2}{2} = \mu_{2,0,\ldots,0} = 1.$

Case 2) ν <u>even</u>

For even ν, from (4.21) it follows that

$$\phi(u,0,\ldots,0) = \frac{(-1)^{m+1}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}\prod_{j=0}^{m}(m+\frac{1}{2}-j)} \sum_{n=0}^{m}\left[\left(\frac{u\sqrt{\nu-2}}{4}\right)^{2n} \cdot \frac{1}{(n!)^2}\right.$$

$$\left. \times \prod_{j=0}^{m-1}(n-j)\left\{\sum_{j=0}^{m-1}\frac{1}{(n-j)} + \log\left(\frac{u(\nu-2)}{2}\right)^2 - \frac{\Gamma'(n+1)}{\Gamma(n+1)}\right\}\right]$$

Thus, the coefficient of u^2 is given by

$$\frac{(-1)^{m+1}\,\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)\prod_{j=1}^{m}(m+\frac{1}{2}-j)}\left\{\frac{(\nu-2)}{4}\,(n(n-2)(n-3)\ldots.(n-m+1))_{n=1}\right\},$$

because $\prod_{j=0}^{m-1}(n-j) = 0$ for n=1. This quantity reduces to $-\frac{1}{2}$, proving that $\mu_{2,\ldots,0} = 1$ for even ν.

Case 3) when ν is a <u>fractional</u> quantity

Now we put $u_1=u$, $u_2=u_3=\ldots=u_p=0$ in (4.24) and obtain

$$\phi(u,0,\ldots.,0) = \frac{(\nu-2)^{\frac{\nu}{2}-m}\,\Gamma\left(\frac{\nu+1}{2}\right)(-1)^m}{2^{\xi}\,\Gamma\left(\frac{\nu}{2}\right)\Gamma(\xi+\frac{1}{2})\prod_{j=1}^{m}(m+q-j)} \cdot \frac{\pi}{\sin\xi\pi}$$

$$\times \sum_{n=0}^{\infty}\left[\left(\frac{u\sqrt{\nu-2}}{2}\right)^{2n} \cdot \frac{1}{n!}\left\{\frac{2^{\xi}\prod_{j=0}^{m-1}(n-\xi+j)}{(\nu-2)^{\xi}\,\Gamma(n+1-\xi)} - \frac{u^{2\xi}\prod_{j=0}^{m-1}(n-j)}{2^{\xi}\,\Gamma(n+1+\xi)}\right\}\right]$$

Now, because

$$\prod_{j=0}^{m-1} (n-j) = 0 \text{ for } n = 1,$$

$$m+\xi+\tfrac{1}{2} = (\nu+1)/2 = m+q, \text{ i.e. } m = \frac{\nu}{2} - \xi$$

the coefficient of u^2 from the last expression for $\phi(u,0,\ldots,0)$ is given by

$$\frac{(\nu-2)^{\xi}\,\Gamma\!\left(\frac{\nu+1}{2}\right)(-1)^{\frac{\nu}{2}-\xi}}{(\nu-2)^{\xi}\,\Gamma\!\left(\frac{\nu}{2}\right)\Gamma(\xi+\tfrac{1}{2})\prod_{j=1}^{m}\left(\frac{\nu+1}{2}-j\right)} \cdot \frac{\pi}{\sin\xi\pi} \cdot \frac{(\nu-2)}{4} \cdot \frac{\prod_{j=0}^{\frac{\nu}{2}-\xi-1}(1-\xi-j)}{\Gamma(2-\xi)}$$

It can be shown that in the above expression

$$\prod_{j=1}^{m}\left(\frac{\nu+1}{2}-j\right) = \frac{\Gamma\!\left(\frac{\nu+1}{2}\right)}{\Gamma(\xi+\tfrac{1}{2})},$$

$$\prod_{j=0}^{\frac{\nu}{2}-\xi-j}(1-\xi-j) = \frac{(-1)^{\frac{\nu}{2}-\xi}\,\Gamma\!\left(\frac{\nu}{2}-1\right)}{\Gamma(\xi-1)},$$

$$\frac{\pi}{\sin\xi\pi} = \frac{\Gamma(\xi)\,\Gamma(1-\xi)}{\Gamma(1)} \text{ for } -\tfrac{1}{2}<\xi<\tfrac{1}{2}.$$

Using these results we simplify the coefficient of u^2 as $-\tfrac{1}{2}$ which implies that also for fractional ν, $\mu_{2,0,\ldots,0} = 1$.

4.4.3 A Remark on C.F. by Awad (1980)

Awad (1980) has made a remark on the characteristic function of the univariate F-distribution and mentioned similar arguments for the c.f. of the univariate Student's t-distribution. But as examined in the sequel, Awad (1980) is in error.

The series for the c.f. of the univariate F-distribution, in equation (2) of Awad's (1980) paper is not uniformly convergent. The series can be written as

$$\frac{1}{B(\frac{m}{2}, \frac{n}{2})} \sum_{r=0}^{\infty} \frac{(it)^r}{r!(r-\frac{n}{2})!} \sum_{j=0}^{\infty} \frac{(j+1)(j+2)\ldots(j+r-\frac{n}{2})}{(j+r+\frac{m}{2})}$$

for $r > \frac{n}{2}$. Now the later sum converges to ∞, because for large j, $(j+r-\frac{n}{2})/(j+r+\frac{m}{2})$ is of unit order. Hence the characteristic function of the univariate F and t-distribution explained by Awad (1980) are defective. However, the c.f. for the univariate F-distribution has already appeared in the literature due to Phillips (1982).

4.4.4 Another application of the c.f. of multivariate t-distribution

Theorem 4.5

Let $X_1, X_2, \ldots, X_p$ be p scalar random variables and $Y_1, Y_2, \ldots Y_n$ be n linear combination of the X's related as follows:

$$\underset{\sim}{Y} = A\underset{\sim}{X} ,$$

where A is a matrix of order nxp. Now if the joint distribution of the Ys' is n-dimensional multivariate student t, then the joint distribution of the p-scalar random variable xs' will be p-dimensional multivariate student t with appropriate parameters.

Proof:

Let us consider the linear combination $Y_j = a_{j1}X_1+a_{j2}X_2+\ldots+a_{jp}X_p$ for at least one $a_{jk}\neq 0$ ($j=1,\ldots,n$; $k=1,\ldots,p$), where $\underset{\sim}{a}_j^T = (a_{j1},\ldots,a_{jp})$ is the jth row of the matrix A in the theorem. Then the characteristic function of Y_j is given by

$$\phi_{Y_j}(t^*) = Ee^{it^* a_j^T X}$$

$$= Ee^{it^*(a_{j1}X_1+\ldots+a_{jp}X_p)} \tag{4.26}$$

As it is given that the joint distribution of Ys' is MVSt, therefore, by the theorem 4.1 and 4.4, we obtain the c.f. of Y_j in the form

$$e^{it^T\theta_1}\,\Psi(t^t\Sigma_{11}t) \,, \tag{4.27}$$

as in (a) or (b) or (c) of the theorem 4.4.

Now if $X = (X_1,\ldots,X_p)$ has mean μ and covariance matrix Σ, then by (4.26) and (4.27),

$$\phi_{a_j^T X}(t^*) = e^{it^* \underset{\sim}{a}_j^T \mu}\,\Psi(t^{*2} a_j^T \Sigma a_j) \,.$$

Putting $t^* = 1$, we then obtain

$$\phi_{a_j^T X}(1) = e^{i\underset{\sim}{a}_j^T \mu}\,\Psi(a_j^T \Sigma a_j) \,.$$

which in turn is equal to $\phi_X(a_1,\ldots,a_p)$. Therefore, $X \sim t_p(\mu,\Sigma,\nu)$.

CHAPTER 5

SAMPLING THEORY FOR MULTIVARIATE t-DISTRIBUTION

5.1 A proposed sampling set-up from a multivariate t-population

Consider $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ as n samples, where each of $\underset{\sim}{X}_j$ (j=1,...,n) is a p-dimensional vector random variable defined as $\underset{\sim}{X}_j=(X_{1j},\ldots,X_{pj})^T$. We assume that $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ are pairwise uncorrelated with p.d.f.

$$f(\underset{\sim}{x}_1^T,\ldots,\underset{\sim}{x}_n^T)=\frac{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\Lambda|^{-\frac{n}{2}}\{\nu+\sum_{j=1}^{n}(\underset{\sim}{x}_j-\underset{\sim}{\theta})^T\Lambda^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\theta})\}^{-\frac{\nu+np}{2}}, \tag{5.1}$$

where $\underset{\sim}{\theta}$ is a location parameter and Λ is a scale parameter.

The joint distribution of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ considered in (5.1) is elliptically symmetric belonging to the class of distributions given by (2.7) and is termed multivariate t-distribution [cf. Cornish (1954), Dunnett and Sobel (1954)].

The univariate version of this model has been considered by Zellner (1976). The p.d.f. in (5.1) is a direct multivariate generalization of Zellner's model as shown below.

Let

$$E(X_{ij}) = \theta_i \text{ for all } i,j$$

$$E(X_{ij}-\theta_i)^2 = \sigma^2\Lambda_{ii} \text{ for all } j, \text{ and } i=1,\ldots,p$$

$$E(X_{ij}-\theta_i)(X_{\ell j}-\theta_\ell) = \sigma^2\Lambda_{i\ell} \text{ for all } j \text{ and } i,\ell=1,\ldots,p$$

and

$E(X_{ij}-\theta_i)(X_{\ell j'}-\theta_\ell) = 0$ for all i and ℓ and $j \neq j'$, where Λ_{ij} are unknown parameters.

Furthermore following Zellner (1976) we assume that for a given σ, the component variables $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ are independently and normally distributed and the distribution of $\underset{\sim}{X}_j=(X_{1j},\ldots,X_{pj})^T$ is $N_p(\theta,\sigma^2\Lambda)$ for all $j=1,\ldots,n$; while σ is assumed to be a random variable having an inverted gamma distribution with probability density function given by

$$g(\sigma) = \frac{2\cdot(\nu/2\sigma^2)^{\frac{\nu+1}{2}}}{(\nu/2)^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)} \cdot \text{Exp}\left(-\tfrac{1}{2}\frac{\nu}{\sigma^2}\right), \quad \sigma > 0,\ \nu > 0,$$

where ν is an unknown parameter.

Clearly, the distribution of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ is then given by

$$f(x_1^T,\ldots,x_n^T) = \int_0^\infty f(x_1^T,\ldots,x_n^T\,;\,\sigma)\,g(\sigma)\,d\sigma$$

$$= K_{np}\int_0^\infty \text{Exp}\left[-\frac{1}{2\sigma^2}\{\nu+\sum_{j=1}^n(\underset{\sim}{x}_j-\underset{\sim}{\theta})^T\Lambda^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\theta})\}\right]$$

$$\times\ \sigma^{-(\nu+np+1)}\,d\sigma, \tag{5.2}$$

where K_{np} is the normalizing constant. Upon direct integration, we obtain the p.d.f. given by (5.1). For the model (5.1), when $\nu>2$, it is readily verified that

$$E(\underset{\sim}{X}_j) = \theta,\ E(\underset{\sim}{X}_j-\theta)(x_j-\theta)^T = \frac{\nu}{\nu-2}\Lambda$$

for all j, and $E(X_j-\theta)(X_\ell-\theta)^T = 0$ for $j\neq\ell, j,\ell=1,\ldots,n$ [see Cornish (1954)]. Hence, by reparametrization of Λ in (5.1) by the covariance matrix $\Sigma = \frac{\nu}{\nu-2}\Lambda$, we finally obtain

$$f(x_1^T,\ldots,x_n^T) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\Sigma|^{-\frac{n}{2}}\{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}} \tag{5.3}$$

which we propose as the joint distribution of the samples $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$. We remark that

$$E(X_j) = \theta$$

$$E(X_j-\theta)(X_j-\theta)^T = \Sigma ,$$

and

$$E(\underset{\sim}{X}_j-\theta)(\underset{\sim}{X}_\ell-\theta)^T = 0,\ j\neq\ell, j,\ell=1,\ldots,n ;$$

so that it is appropriate to call θ as the location (or mean) and Σ as the scale (or covariance) parameters of the model. ν in (5.3) is referred as the degrees of freedom of the distribution. Also, it is seen that $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ are pairwise uncorrelated while they are not necessarily independent. Independence implies that $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ are normally distributed (c.f. Kelker, 1970). Finally we remark that when $\nu\to\infty$, the limiting form (5.3) is multivariate normal, so that the proposed model accommodates the usual case when the samples $X_1,\ldots,X_n$ are assumed to be independently and identically distributed according to $N_p(\theta,\Sigma)$.

5.2 Distribution of the sample mean vector

The following theorem is a direct consequence of theorems 2.3 and 4.1.

Theorem 5.1

Let the p.d.f. of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ be

$$f(x_1^T,\ldots,x_j^T,\ldots,x_n^T) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\Sigma|^{-\frac{n}{2}}\{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}}$$

Then the p.d.f. of $\underset{\sim}{\bar{X}} = (\bar{X}_1,\ldots,\bar{X}_i,\ldots,\bar{X}_p)^T$ is given by

$$f(\bar{x},\theta,\Sigma,\nu) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}|\Sigma/n|^{-\frac{1}{2}}\{(\nu-2)+(x-\theta)^T\left(\frac{\Sigma}{n}\right)^{-1}(x-\theta)\}^{-\frac{\nu+p}{2}} . \qquad (5.4)$$

This result follows from theorems 2.3 and 4.1 by letting

$$K_{np} = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\Gamma\left(\frac{\nu}{2}\right)} , \text{ and}$$

$$g\{\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\} = \{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}} ,$$

in (2.9).

5.3 Distribution of the sample sum of product matrix

The generalization of the Wishart distribution in (2.29) was based on the assumption that the underlying distribution belongs to the class of the elliptic distributions of the continuous type. In the following theorem this result is specialized to the case when the underlying distribution is assumed to belong to the subclass of multivariate t-distributions. It is finally shown that the usual Wishart matrix distribution is a special case of this theorem.

Theorem 5.2

If $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_j, \ldots, \underset{\sim}{X}_n$ has the p.d.f.

$$f(x_1^T, \ldots, x_n^T) = \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+np}{2}\right) |\Sigma|^{-\frac{n}{2}}}{\pi^{\frac{np}{2}} \Gamma\left(\frac{\nu}{2}\right)} \{(\nu-2)+\sum_{j=1}^{n} (x_j-\theta)^T \Sigma^{-1} (x_j-\theta)\}^{-\frac{\nu+np}{2}}$$

then the p.d.f. of the elements of

$$V = \sum_{j=1}^{n} (\underset{\sim}{x}_j-\bar{x})(x_j-\bar{x})^T$$

$$= ((\sum_{j=1}^{n} (x_{ij}-\bar{x}_i)(x_{kj}-\bar{x}_k)))$$

$$= ((V_{ik})) \text{ say,}$$

is given by

$$f(v_{ik}/i\leq k,\ i=1,\dots,p\ ;\ k=1,\dots,p)$$

$$=\frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+n'p}{2}\right)|\Sigma|^{-\frac{n'}{2}}|V|^{\frac{n'-p-1}{2}}}{\pi^{\frac{p(p-1)}{4}}\Gamma\left(\frac{\nu}{2}\right)\prod_{i=1}^{p}\Gamma\left(\frac{n'-i+1}{2}\right)}\{(\nu-2)+\text{trace }\Sigma^{-1}V\}^{-\frac{\nu+n'p}{2}} \quad (5.5)$$

where $\underset{\sim}{X}_j = (X_{1j},\dots,X_{pj})^T$, $n' = n-1$.

Proof:

Substitute $K_{np} = \dfrac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\Gamma\left(\frac{\nu}{2}\right)}$, and

$g(\cdot) = \{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}}$ in the p.d.f. of $\underset{\sim}{X}_1,\dots,\underset{\sim}{X}_j,\dots,\underset{\sim}{X}_n$ of the theorem 2.5. From (2.29) and the marginal property of t-distribution, it then follows that

$$f(v_{ik}/i\leq k,\ i=1,\dots,p\ ;\ k=1,\dots,p)$$

$$= C_{n'p}|\Sigma|^{-\frac{n'}{2}}|V|^{\frac{n'-p-1}{2}}\{(\nu-2)+\text{trace }\Sigma^{-1}V\}^{-\frac{\nu+n'p}{2}} \quad (5.6)$$

where $C_{n'p}$ is a normalizing constant such that $\int_V f(v)dv=1$.

We evaluate C_{np} first. In doing this we write the analogous form to the equation (2.23) as

$$C_{np}\{(\nu-2)+\sum_{i=1}^{p}\sum_{r=1}^{i-1}b_{ir}^{*2}+\sum_{i=1}^{p}R_i\}^{-\frac{\nu+np}{2}}\prod_{i=1}^{p}R_i^{\frac{n-(i-1)}{2}-1}dR_i\prod_{i=1}^{p}\prod_{r=1}^{i-1}db_{ir}^{*}$$

(5.7)

This is because in the present case

$$g(\cdot) = \{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}}$$

We note that in (5.7), $b_{ir}^* \sim t(0,1)$ whereas in (2.23) b_{ir}^* was distributed spherically around the centre $\underset{\sim}{0}$.

Now we integrate over b_{ir}^* and R_i's of (5.7) to evaluate C_{np} such that

$$C_{np}\int\cdots\int\{(\nu-2)+\sum_{i=1}^{p}\sum_{r=1}^{i-1}b_{ir}^{*2}+\sum_{i=1}^{p}R_i\}^{-\frac{\nu+np}{2}}\prod_{i=1}^{p}R_i^{\frac{n-(i-1)}{2}-1}dR_i$$

$$\times\prod_{i=1}^{p}\prod_{r=1}^{i-1}db_{ir}^* = 1.$$

We integrate out $R_p, R_{p-1},\ldots,R_1$ one by one. Let $a = \{(\nu-2)+\prod_{i=1}^{p}\prod_{r=1}^{i-1}b_{ir}^{*2}\}$. Then the above equation can be written as

$$C_{np}\int\frac{\prod_{i=1}^{p-1}(R_i)^{\frac{n-(i-1)}{2}-1}dR_i}{(a+\sum_{i=1}^{p-1}R_i)^{\frac{\nu+np}{2}}}\prod_{i=1}^{p}\prod_{r=1}^{i-1}db_{ir}^*\int_0^\infty\frac{R_p^{\frac{n-p+1}{2}-1}dR_p}{\left\{1+\frac{R_p}{(a+\sum_{i=1}^{p-1}R_i)}\right\}^{\frac{\nu+np}{2}}} = 1$$

Putting $R_p/\{a+\sum_{i=1}^{p-1}R_i\} = u$ and performing the Beta-integration of the first kind, we get

$$C_{np}B\left(\frac{\nu+np-n+p-1}{2},\frac{n-p+1}{2}\right)\int\cdots\int\frac{\prod_{i=1}^{p-1}(R_i)^{\frac{n-i+1}{2}-1}dR_i\prod_{i=1}^{p}\prod_{r=1}^{i-1}db_{ir}^*}{\left\{a+\sum_{i=1}^{p-1}R_i\right\}^{\frac{\nu+np-n+p-1}{2}}} = 1$$

Making similar substitution and performing the integration over $R_{p-1}, R_{p-2}, \ldots, R_1$, we obtain

$$C_{np}\,\beta\left(\frac{\nu+np-n+p-1}{2}, \frac{n-p+1}{2}\right)\beta\left(\frac{\nu+np-2n+2p-3}{2}, \frac{n-p+2}{2}\right)\ldots\ldots$$

$$\ldots\ldots x\ \beta\left(\frac{\nu+P^2-(P^2+P)/2}{2}, \frac{n-p+p}{2}\right)\int \frac{\prod\prod db^*_{ir}}{\{(\nu-2)+\sum_{i=1}^{p}\sum_{r=1}^{i-1} b^{*2}_{ir}\}^{\frac{\nu+p(P-1)}{2}}} = 1 \tag{5.8}$$

Since $b^*_{ir} \sim t(0,1)$, it then follows that

$$\int \{(\nu-2)+\prod_{i=1}^{p}\prod_{r=1}^{i-1} b^{*2}_{ir}\}^{-\frac{\nu+p(P-1)}{2}} db_{21}\ldots db_{P(P-1)} = \frac{\Gamma\left(\frac{\nu}{2}\right)\pi^{\frac{p(P-1)}{4}}}{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+p(P-1)/2}{2}\right)}$$

Using this integration result in (5.8), a simple algebric calculation yields the C_{np} as

$$C_{np} = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{p(P-1)}{4}}\,\Gamma\left(\frac{\nu}{2}\right)\prod_{i=1}^{p}\Gamma\left(\frac{n-(p-i)}{2}\right)} \tag{5.9}$$

C_{np} given by (5.9) is a normalising constant in the p.d.f. of the elements of $M = \sum_{j=1}^{n} \underset{\sim}{x}_j \underset{\sim}{x}_j^T$, where $\underset{\sim}{X}_j(p\times1) \sim t_p(0, I_p)$ and $\underset{\sim}{X}_j, \underset{\sim}{X}_s$ $(j\neq s)$ are uncorrelated.

Next by (2.29), the normalising constant in the p.d.f. of the elements of $V = \sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T$ would be $C_{n'p}$, where $n'=n-1$ and C_{np} is given by (5.9). Thus changing n by n' in (5.9), we obtain

$$C_{n'p} = \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+n'p}{2}\right)}{\pi^{\frac{p(P-1)}{4}} \Gamma\left(\frac{\nu}{2}\right) \prod_{i=1}^{p} \Gamma\left(\frac{n'-(p-i)}{2}\right)}$$

Hence, from (5.6) the theorem follows.

We note that when $\nu \to \infty$, the p.d.f. given by (5.5) converges to the form

$$\frac{|\Sigma|^{-\frac{n'}{2}} |V|^{\frac{n'-p-1}{2}} \operatorname{Exp}\{-\frac{1}{2} (\operatorname{trace} \Sigma^{-1} V)\}}{2^{\frac{n'p}{2}} \pi^{\frac{p(P-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(n'-i-1)\right)}$$

which is the p.d.f. of the usual central Wishart matrix distribution.

5.3.1 Moments of the sum of product matrix

In this section we derive the first and second moments of the sum of product matrix, namely, $V = \sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T$, whose p.d.f. is given by (5.5)

Consider $\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_j,\ldots,\underset{\sim}{Z}_{n'}$ with p.d.f. given by

$$f(z_1,\ldots,z_{n'}) = \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+n'p}{2}\right)}{\pi^{\frac{n'p}{2}} \Gamma\left(\frac{\nu}{2}\right)} |\Sigma|^{-\frac{n'}{2}} \{(\nu-2)+\sum_{j=1}^{n'} z_j^T \Sigma^{-1} z_j\}^{-\frac{\nu+n'p}{2}} \tag{5.10}$$

so that each of $Z_j \sim t_p(0,\Sigma,\nu)$ for $j=1,\ldots,n'$; and Z_j, Z_ℓ are uncorrelated when $j \neq \ell$ for $j,\ell=1,\ldots,n'$. Then from (2.26) it follows that the p.d.f. of the elements of the matrix $V_1 = \sum_{j=1}^{n'} z_j z_j^T$ is the same as that of the p.d.f. of $V = \sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T$ given by (5.5).

Therefore, finding the moments of $V = \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T$ is equivalent to find the moments of $V_1 = \sum_{j=1}^{n'} z_j z_j^T$, where $Z_1, \ldots, Z_{n'}$ has the p.d.f. given by (5.10). For simplicity of calculations, we derive the first and second moments of V_1, which are also first and second moments of V, as in the following:

First moment of V

$$E(V) = E(V_1) = n'\Sigma,$$

where V and V_1 are defined as in the above. This result can be shown as follows:

Let $V_{i\alpha}$ be the (i,α)th element of the matrix V_1 for $i,\alpha = 1,\ldots,p$. Now, because from (5.10) it follows that $E(Z_j Z_j^T) = \Sigma$, therefore we have

$$E(V_{i\alpha}) = E\{\sum_{j=1}^{n'} z_{ij} z_{\alpha j}^T\}$$

$$= n' E(z_{ij} z_{\alpha j}^T)$$

$$= n' \cdot \sigma_{i\alpha}, \qquad (5.11)$$

where $\sigma_{i\alpha}$ is the (i,α)th element of Σ matrix. Thus $E(V_1) = n'\Sigma$

Second moments of V

We simplify second moments of V, namely, $\mathrm{cov}(V_{i\alpha}, V_{k\beta})$ and then specialize the result to obtain the well-known second moments of the Wishart distribution.

$$\mathrm{Cov}(V_{i\alpha},V_{k\beta}) = E\{(V_{i\alpha}-n'\sigma_i)(V_{k\beta}-n'\sigma_{k\beta})\}$$

$$= E(V_{i\alpha}V_{k\beta}) - n'^2\sigma_{i\alpha}\sigma_{k\beta} \quad , \tag{5.11}$$

because by (5.10) $E(V_{i\alpha}) = n'\sigma_{i\alpha}$. We now calculate $E(V_{i\alpha}V_{k\beta})$. By definition,

$$E(V_{i\alpha}V_{k\beta}) = E\{\sum_{j=1}^{n'}\sum_{r=1}^{n''} z_{ij}z_{\alpha j}z_{kr}z_{\beta r}\}$$

$$= E\{\sum_{j=1}^{n'} z_{ij}z_{\alpha j}z_{kj}z_{\beta j} + \sum_{j\neq r}^{n'} z_{ij}z_{\alpha j}z_{kr}z_{\beta r}\} \tag{5.12}$$

Substitute $\Sigma^{-\frac{1}{2}}Z_j=Y_j$ in (5.10) for all $j=1,\ldots,n'$. Then it follows that the marginal density of Y_j is given by

$$f(y_j) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}\{(\nu-2)+\sum_{i=1}^{p} y_{ij}^2\}^{-\frac{\nu+p}{2}} ,$$

hence from Johnson and Kotz (1972), it now follows that

$$E(y_{1j}^{2r_1}\cdots y_{pj}^{2r_p})$$

$$= \mu_{2r_1,\ldots,2r_i,\ldots,2r_p}$$

$$= \frac{(\nu-2)^{\sum_{i=1}^{p} r_i}\Gamma\left(\frac{\nu}{2}-\sum_{i=1}^{p} r_i\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}\prod_{i=1}^{p}\Gamma(r_i+\tfrac{1}{2}) \quad . \tag{5.13}$$

Let $\Sigma^{\frac{1}{2}} = ((m_{iu}))$, such that $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}=\Sigma$, for $i,u=1,\ldots,p$. Because from the substitution we have $Z_j = \Sigma^{\frac{1}{2}}Y_j$, therefore we can write

$$z_{ij} = \sum_{u=1}^{p} m_{iu} y_{uj}$$

Now by (5.14) the first term ~~in~~ (5.12) can be simplified as

$$E(z_{ij} z_{\alpha j} z_{kj} z_{\beta j})$$

$$= E\left(\sum_{u=1}^{p} m_{iu} y_{uj} \sum_{u=1}^{p} m_{\alpha u} y_{uj} \sum_{u=1}^{p} m_{ku} y_{uj} \sum_{u=1}^{p} m_{\beta u} y_{uj} \right)$$

$$= E\Big[\Big\{ \sum_{u=1}^{p} m_{iu} m_{\alpha u} y_{uj}^2 + \sum_{u<v}\sum (m_{iu} m_{\alpha v} + m_{iv} m_{\alpha u}) y_{uj} y_{vj}$$

$$\times \sum_{u=1}^{p} m_{ku} m_{\beta u} y_{uj}^2 + \sum_{u<v}\sum (m_{ku} m_{\beta v} + m_{kv} m_{\beta u}) y_{uj} y_{vj} \Big\} \Big].$$

$$= E\Big[\sum_{u=1}^{p} m_{iu} m_{\alpha u} m_{ku} m_{\beta u} y_{uj}^4$$

$$+ \sum_{u<v}\sum \{ (m_{iu} m_{\alpha u} m_{kv} m_{\beta v} + m_{iv} m_{\alpha v} m_{ku} m_{\beta u}) y_{uj}^2 y_{vj}^2 \}$$

$$+ \sum_{u<v}\sum \{ (m_{iu} m_{\alpha v} + m_{iv} m_{\alpha u})(m_{ku} m_{\beta v} + m_{kv} m_{\beta u}) y_{uj}^2 y_{vj}^2 \}$$

$$+ \{\text{other terms containing odd powers of } y_{uj}\} \Big] \qquad (5.15)$$

Any odd moment of y_{uj} is zero and in addition it follows from (5.13) that

$$E(y_{uj}^4) = 3\left(\frac{\nu-2}{\nu-4}\right) ;$$

$$E(y_{uj}^2 y_{vj}^2) = \frac{\nu-2}{\nu-4} , \text{ and } E(y_{uj}^2) = 1.$$

Using these results in (5.15), we obtain

$$E(z_{ij}z_{\alpha j}z_{kj}z_{\beta j})$$

$$= \frac{\nu-2}{\nu-4}\Big[3\sum_{u=1}^{p} m_{iu}m_{\alpha u}m_{ku}m_{\beta u}$$

$$+ \{\sum_{u<v}\sum(m_{iu}m_{\alpha u}m_{ku}m_{\beta v}+m_{iv}m_{\alpha v}m_{ku}m_{\beta u})$$

$$+ \sum_{u<v}\sum(m_{iu}m_{\alpha v}+m_{iv}m_{\alpha u})(m_{ku}m_{\beta v}+m_{kv}m_{\beta u})\}\Big] \qquad (5.16)$$

Next we evaluate the expectation of the second term in (5.12) as follows:

$$E\sum_{j\neq r}^{n'} z_{ij}z_{\alpha j}z_{kr}z_{\beta r}$$

$$= \sum_{j\neq r}^{n'} E\Big[\{\sum_{u=1}^{p} m_{iu}y_{uj}\sum_{u=1}^{p} m_{\alpha u}y_{uj}\}\{\sum_{u=1}^{p} m_{ku}y_{ur}\sum_{u=1}^{p} m_{\beta u}y_{ur}\}\Big]$$

$$= n'(n'-1)E\Big[\{\sum_{u=1}^{p} m_{iu}m_{\alpha u}y_{uj}^2 + \sum_{u<v}(m_{iu}m_{\alpha v}+m_{iv}m_{\alpha u})y_{uj}y_{vj}\}$$

$$\times \{\sum_{u=1}^{p} m_{ku}m_{\beta u}y_{ur}^2 + \sum_{u<v}(m_{ku}m_{\beta u}+m_{kv}m_{\beta u})y_{ur}y_{vr}\}$$

$$= n'(n'-1)E\Big[\sum_{u=1}^{p} m_{iu}m_{\alpha u}m_{ku}m_{\beta u}y_{uj}^2y_{ur}^2$$

$$+ \sum_{u<v}\{(m_{iu}m_{\alpha u}m_{kv}m_{\beta v}+m_{iv}m_{\alpha v}m_{ku}m_{\beta u})y_{uj}^2y_{ur}^2\}$$

$$+ \{\text{other terms having odd powers of } y_{rj}\}\Big] \qquad (5.17)$$

It can be shown that

$$E(y_{ij}^2 y_{vr}^2) = \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+2}{2}\right)}{\pi \Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{y_{ij}^2 y_{vr}^2}{\{(\nu-2)+y_{ij}^2+y_{vr}^2\}^{\frac{\nu+2}{2}}} dy_{ij} dy_{vr} = \frac{\nu-2}{\nu-4} .$$

Similarly $E(u_{uj}^2 y_{ur}^2) = \frac{\nu-2}{\nu-4}$ for $j \neq r$. Using these results in (5.17), we obtain

$$E \sum_{j \neq r}^{n'} z_{ij} z_{\alpha j} z_{kr} z_{\beta r}$$

$$= n'(n'-1) \cdot \frac{\nu-2}{\nu-4} \Big[\sum_{u=1}^{p} m_{iu} m_{\alpha u} m_{ku} m_{\beta u}$$

$$+ \sum_{u<v} (m_{iu} m_{\alpha u} m_{kv} m_{\beta v} + m_{iv} m_{\alpha v} m_{ku} m_{\beta u}) \qquad (5.18)$$

First we use (5.16) and (5.18) in (5.12), then by (5.11), we finally obtain

$$\mathrm{Cov}(V_{i\alpha}, V_{k\beta})$$

$$= E(V_{i\alpha} V_{k\beta}) - n'^{2} \sigma_{i\alpha} \sigma_{k\beta}$$

$$= \frac{\nu-2}{\nu-4} \Big[\Big(\sum_{u=1}^{p} m_{iu} m_{\alpha u} m_{ku} m_{\beta u} \Big) \{3n' + n'(n'-1)\}$$

$$+ (n')^2 \{ \sum_{u<v} (m_{iu} m_{\alpha u} m_{kv} m_{\beta v} + m_{iv} m_{\alpha v} m_{ku} m_{\beta u}) \}$$

$$+ n' \{ \sum_{u<v} (m_{iu} m_{\alpha v} + m_{iv} m_{\alpha u})(m_{ku} m_{\beta v} + m_{kv} m_{\beta u}) \} \Big]$$

$$- n'^{2} \Big(\sum_{u=1}^{p} m_{iu} m_{\alpha u} \Big) \Big(\sum_{u=1}^{p} m_{ku} m_{\beta u} \Big) \qquad (5.19)$$

$\underline{Var(V_{i\alpha})}$

If i=k, $\alpha=\beta$ in (5.19), we get the variance of $V_{i\alpha}$ which is given by

$$E(V_{i\alpha}-n'\sigma_{i\alpha})^2$$

$$= \frac{\nu-2}{\nu-4}\left[n'^2\left(\sum_{u=1}^{p} m_{iu}m_{\alpha u}\right)^2 + 2n'\sum_{u=1}^{p} m_{iu}^2 m_{\alpha u}^2\right.$$

$$\left. + n'\sum_{u<v}(m_{iu}m_{\alpha v}+m_{iv}m_{\alpha u})^2\right] - n'^2\left(\sum_{u=1}^{p} m_{iu}m_{\alpha u}\right)^2 \qquad (5.20)$$

Limiting case when $\nu\to\infty$

Under this situation when $\nu\to\infty$, (5.20) reduces to

$$Var(V_{i\alpha})$$

$$= n'\left[2\sum_{u=1}^{p} m_{iu}^2 m_{\alpha u}^2 + \sum_{u<v}(m_{iu}m_{\alpha v}+m_{iv}m_{\alpha u})^2\right]$$

$$= n'\left[2\sum_{u=1}^{p} m_{iu}^2 m_{\alpha u}^2 + \sum_{u<v} m_{iu}^2 m_{\alpha v}^2 + \sum_{u<v} m_{iv}^2 m_{\alpha u}^2\right.$$

$$\left. + 2\sum_{u<v} m_{iu}m_{\alpha v}m_{\alpha u}m_{iv}\right]$$

$$= n'\left[\left(\sum_{u=1}^{p} m_{iu}m_{\alpha u}\right)^2 + \sum_{u=1}^{p} m_{iu}^2 m_{\alpha u}^2 + \sum_{u<v} m_{iu}^2 m_{\alpha v}^2 + \sum_{u<v} m_{iu}^2 m_{\alpha u}^2\right]$$

$$= n'\left[\left(\sum_{u=1}^{p} m_{iu}m_{\alpha u}\right)^2 + \sum_{u=1}^{p} m_{iu}^2 \sum_{u=1}^{p} m_{\alpha u}^2\right]$$

Now because $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$ i.e. $((m_{ij}))((m_{\alpha j})) = ((\sigma_{i\alpha}))$, therefore the last equation can be written as

$$Var(V_{i\alpha}) = n'\left[\sigma_{i\alpha}^2 + \sigma_{ii}\,\sigma_{\alpha\alpha}\right], \qquad (5.21)$$

which is the variance of the (i,j)th element of the usual Wishart matrix [cf. Anderson, 1958, p. 161].

5.4 Distribution of a Quadratic form

For the normal case, it is well known that the distribution of the quadratic form is non-central chi-square. In the following theorem we generalize this result by assuming the underlying distribution to have a suitable multivariate t-distribution.

Theorem 5.3

Let $Y=(y_1,\ldots,y_j,\ldots,y_n)$, where $y_j=(y_{1j},\ldots,y_{pj})^T$. Let Y^* be the stacked random variable corresponding to Y so that $Y^*=(y_{11},\ldots,y_{p1},y_{12},\ldots,y_{p2},\ldots,y_{1n},\ldots,y_{pn})^T$. If $Y^* \sim t_{np}(\underset{\sim}{1}_n \otimes \mu, I_{np}, \nu)$ where $1_{n\times 1} = (1,1,\ldots,1)^T$, then $W = \frac{\nu}{\nu-2} Y^{*T}Y^* = \frac{\nu}{\nu-2} \sum_{i=1}^{p} \sum_{j=1}^{n} Y_{ij}^2$ has the distribution given by

$$f(w) = \frac{\nu^{\frac{\nu}{2}} w^{\frac{np-1}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{\nu+np}{2}+2j\right)}{\Gamma\left(\frac{np}{2}+j\right)\Gamma(j+1)} (\lambda w)^j (\lambda+w+\nu)^{-\frac{\nu+np}{2}-2j} \tag{5.22}$$

where $\lambda = \frac{\nu}{\nu-2} \sum_{i=1}^{p} \mu_i^2$.

In the special case when $\lambda=0$, W/np is distributed according to the usual F-distribution with degrees of freedom np and ν.

Proof:

By hypothesis

$$f(y_{11},y_{12},\ldots.y_{np})$$

$$= \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\!\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\,\Gamma\!\left(\frac{\nu}{2}\right)} \left\{(\nu-2) + \sum_{i=1}^{p}\sum_{j=1}^{n}(y_{ij}-\mu_i)^2\right\}^{-\frac{\nu+np}{2}}$$

Let us set $(\nu-2)^{\frac{1}{2}}U_{ij} = \nu^{\frac{1}{2}}Y_{ij}$ and $(\nu-2)^{\frac{1}{2}}m_i = \nu^{\frac{1}{2}}\mu_i$ so that $(\nu-2)^{\frac{1}{2}}(U_{ij}-m_i) = \nu^{\frac{1}{2}}(Y_{ij}-\mu_i)$. It follows that

$$f(u_{11},\ldots,u_{np}) = \frac{\nu^{\frac{\nu}{2}}\,\Gamma\!\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\,\Gamma\!\left(\frac{\nu}{2}\right)} \left\{\nu + \sum_{i=1}^{p}\sum_{j=1}^{n}(u_{ij}-m_i)^2\right\}^{-\frac{\nu+np}{2}}$$

Then $W = \frac{\nu}{\nu-2}\sum_{i}^{p}\sum_{j=1}^{n}Y_{ij}^2 = \sum_{i}^{p}\sum_{j=1}^{n}U_{ij}^2$ has the distribution function $F(w)$ given by

$$F(w) = \int\limits_{\sum\sum_{ij} u_{ij}^2 \le w} f(u_{11},\ldots,u_{np}) \prod_{ij} du_{ij} \tag{5.23}$$

It is readily verified by direct integration over z that

$$\int_0^{\infty} \pi^{-\frac{np}{2}} z^{-np}\, \mathrm{Exp}\left\{-\frac{1}{2z^2}\sum_{i=1}^{p}\sum_{j=1}^{n}(u_{ij}-m_i)^2\right\}$$

$$\times\ 2\left(\Gamma\!\left(\frac{\nu}{2}\right)\right)^{-1}\left(\frac{\nu}{2}\right)^{-\frac{1}{2}}\left(\frac{\nu}{2z^2}\right)^{\frac{+1}{2}} \mathrm{Exp}\left(-\frac{\nu}{2z^2}\right)dz$$

$$= \int_0^\infty G(u_{ij}, m_i, z)\,dz \quad \text{(say)}$$

$$= f(u_{11}, u_{12}, \ldots, u_{np})$$

Substituting the above integral in (5.23) for $f(u_{11}, \ldots, u_{np})$ and changing the order of integration we then obtain

$$F(w) = \int_0^\infty \int_{\sum\sum_{ij} u_{ij}^2 \le w} G(u_{ij}, m_i, z)\,dz \prod_{ij} du_{ij} .$$

The simple transformation $u_{ij} = zv_{ij}$ then yields

$$F(w) = \int_0^\infty (\Gamma\frac{\nu}{2})^{-1} (\frac{\nu}{2})^{-\frac{1}{2}} \cdot 2(\frac{\nu}{2z^2})^{\frac{\nu+1}{2}} \text{Exp}(-\frac{\nu}{2z^2}) \cdot H(\frac{w}{z^2})\,dz, \qquad (5.24)$$

where

$$H(x) = \int_{\sum\sum_{ij}^{pn} v_{ij}^2 \le x} \pi^{-\frac{np}{2}} \text{Exp}\{-\tfrac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{n} (v_{ij} - \frac{mi}{z})^2\} \prod_{ij} dv_{ij},$$

which is the distribution function of the usual noncentral x^2 with np degrees of freedom and noncentrality parameter

$\sum_{i=1}^{p} m_i^2/z^2 = \nu(\nu-2)^{-1} (\sum_{i=1}^{p} \mu_i^2)/z^2 = \lambda/z^2$ where λ is defined in the theorem.

Hence from (5.24), it follows that

$$f(w) = \frac{dF(w)}{dw} = \int_0^\infty 2\left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}\left(\frac{\nu}{2}\right)^{-\frac{1}{2}}\left(\frac{\nu}{2z^2}\right)^{\frac{\nu}{2}+1} \operatorname{Exp}\left(-\frac{\nu}{2z^2}\right)\cdot\frac{1}{z^2}\, h\left(\frac{w}{z^2}\right) dz,$$

where

$$h(x) = \frac{dH(x)}{dx} = 2^{-\frac{np}{2}} \operatorname{Exp} -\tfrac{1}{2}(x+\lambda/z^2) \times \sum_{j=0}^{\infty} \frac{(x)^{\frac{np}{2}+j-1}\,(\lambda/z^2)^j}{\Gamma\left(\frac{np}{2}+j\right)\cdot 2^{2j}\,\Gamma(j+1)}$$

[See Johnson and Kotz (1970)], so that

$$f(w) = \nu^{\frac{\nu}{2}}\, w^{\frac{np}{2}-1}\left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1} 2^{-\frac{\nu+np}{2}+1}$$

$$\times \sum_{j=0}^{\infty} \frac{(w)^j}{\Gamma\left(\frac{np}{2}+j\right)\Gamma(j+1)\cdot 2^{2j}} \int_0^\infty (z^2)^{-\frac{\nu+np}{2}-2j} \operatorname{Exp} -\frac{1}{2z^2}(\lambda+w+\nu)\, dz$$

Upon direct integration over z, we finally obtain the p.d.f. of W given by (5.22). This proves the first part of the theorem. For the special case when $\lambda=0$, we have from (5.22)

$$f(w) = \frac{\nu^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+np}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{np}{2}\right)} \cdot w^{\frac{np}{2}-1}\,(\nu+w)^{-\frac{\nu+np}{2}}$$

$$= \frac{(np/\nu)^{\frac{np}{2}}}{B\left(\frac{\nu}{2}, \frac{np}{2}\right)} \cdot w^{\frac{np}{2}-1}\left(1+\frac{np}{\nu}w\right)^{-\frac{\nu+np}{2}},$$

which is the probability density of the usual central F-distribution with degrees of freedom np and ν.

5.5 Joint distribution of $\overline{X}$ and V

From (5.3), we have the p.d.f. of $X_1,\ldots,X_j,\ldots,X_n$ as

$$f(x_1^T,\ldots,x_j^T,\ldots,x_n^T)\alpha|\Sigma|^{-\frac{n}{2}}\{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}}dX\ , \qquad (5.25)$$

where X denotes a matrix now and is given by

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}$$

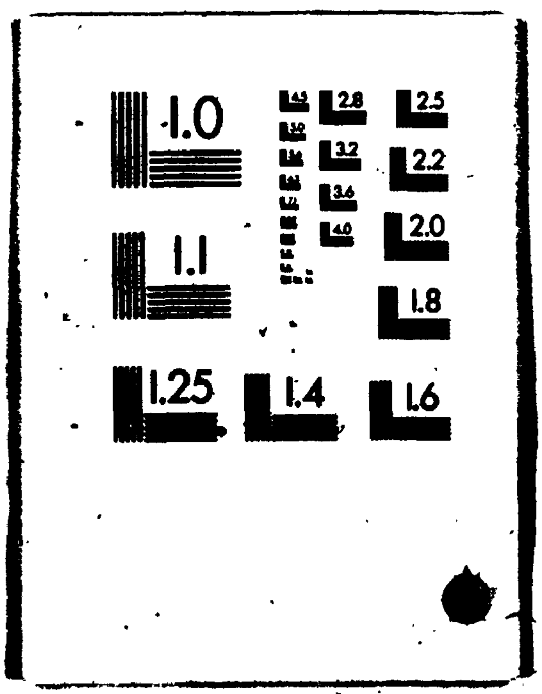
Let $X_1 = x\Gamma$,

where Γ is an orthogonal matrix given by

$$\Gamma_{nxn} = \begin{bmatrix} 1/\sqrt{n} & \gamma_{12} & \cdots & \gamma_{1n} \\ 1/\sqrt{n} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & & \vdots \\ 1/\sqrt{n} & \gamma_{n1} & \cdots & \gamma_{nn} \end{bmatrix} .$$

$$\text{Then, } X_1 = \left(\begin{array}{c|c} \sqrt{n}\ \ \overline{x}_1 & \\ \sqrt{n}\ \ \overline{x}_2 & \\ \vdots & Y : px(n-1) \\ \sqrt{n}\ \ \overline{x}_p & \end{array}\right) \text{ (say)} \qquad (5.26)$$

2 OF/DE 2



Now $XX^T = X\Gamma\Gamma^T X^T$

$$= X_1 X_1^T$$

$$= n\bar{X}\bar{X}^T + YY^T \text{ by } (5.26)$$

Therefore, $YY^T = XX^T - n\bar{X}\bar{X}^T$. Jacobian of the transformation is given by

$$J(X \to X_1) = J(X \to \sqrt{n}\,\bar{X},\ Y) = 1$$

Since in (5.25),

$$\sum_{j=1}^{n} (x_j-\theta)^T \Sigma^{-1} (x_j-\theta) = \text{trace} \sum_{j=1}^{n} (x_j-\theta)^T \Sigma^{-1} (x_j-\theta)$$

$$= \text{trace}\ \Sigma^{-1} \sum_{j=1}^{n} (x_j-\theta)(x_j-\theta)^T$$

$$= \text{trace}\ \Sigma^{-1}\{\sum_{j=1}^{n} (x_j-\bar{x}+\bar{x}-\theta)(x_j-\bar{x}+\bar{x}-\theta)^T\}$$

$$= \text{trace}\ \Sigma^{-1}\{\sum_{j=1}^{n} (x_j-\bar{x})(x_j-\bar{x})^T + n(\bar{x}-\theta)(\bar{x}-\theta)^T\}$$

$$= \text{trace}\ \Sigma^{-1}(YY^T) + \text{trace}\ (\bar{x}-\theta)^T (\frac{\Sigma}{n})^{-1} (\bar{x}-\theta)$$

$$= \text{trace}\ \Sigma^{-1}(YY^T) + (\bar{x}-\theta)^T (\frac{\Sigma}{n})^{-1} (\bar{x}-\theta) ,$$

from (5.25) it follows that

$$f(\bar{x},y) = K|\Sigma|^{-\frac{n}{2}}\{(\nu-2)+(\bar{x}-\theta)^T(\frac{\Sigma}{n})^{-1}(\bar{x}-\theta) + \text{trace}\ \Sigma^{-1}(yy^T)\}^{-\frac{\nu+np}{2}},$$

where K is a normalising constant such that

$$\int\int f(\bar{x},y)\,d\bar{x}\,dy = 1 \ .$$

We now apply theorem 5.2 and obtain $f(\bar{x},v)$ from $f(\bar{x},y)$. After some algebra, it can be shown that the associated normalising constant is given by

$$K = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\!\left(\frac{\nu+np}{2}\right)\,|v|^{\frac{n-p-2}{2}}\,n^{\frac{p}{2}}}{\Gamma\!\left(\frac{\nu}{2}\right)\prod\limits_{i=1}^{p}\Gamma\!\left(\tfrac{1}{2}(n-i)\right)\cdot\pi^{\frac{p(P+1)}{4}}}$$

Hence, the joint density of $\bar{X}$ and V is given by

$$f(\bar{x},v) = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\!\left(\frac{\nu+np}{2}\right)\,|V|^{\frac{n-p-2}{2}}\,|\Sigma|^{-\frac{n}{2}}}{\Gamma\!\left(\frac{\nu}{2}\right)\pi^{\frac{p(P+1)}{4}}\prod\limits_{i=1}^{p}\Gamma\!\left(\tfrac{1}{2}(n-i)\right)}$$

$$\times\ \{(\nu-2) + (\bar{x}-\theta)^{T}\left(\frac{\Sigma}{n}\right)^{-1}(\bar{x}-\theta) + \mathrm{trace}\ \Sigma^{-1}V\}^{-\frac{\nu+np}{2}} \qquad (5.27)$$

CHAPTER 6

ESTIMATION OF PARAMETERS OF MULTIVARIATE t-DISTRIBUTION

6.1 Estimation of Parameters

Consider the proposed model (5.3) given by

$$f(x_1^T,\ldots,x_n^T) = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\,\Gamma\left(\frac{\nu}{2}\right)}\,|\Sigma|^{-\frac{n}{2}}\{(\nu-2)+\sum_{j=1}^{n}(x_j-\theta)^T\Sigma^{-1}(x_j-\theta)\}^{-\frac{\nu+np}{2}} \tag{6.1}$$

where $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_j,\ldots,\underset{\sim}{X}_n$ as before are n samples uncorrelated to each other. We recall that for $\nu > 2$, $E(\underset{\sim}{X}_j) = \underset{\sim}{\theta}$, $E(\underset{\sim}{X}_j-\theta)(\underset{\sim}{X}_j-\theta)^T = \Sigma$, and $E(X_j-\theta)(X_\ell-\theta) = 0$ for $j\neq\ell$; $j,\ell = 1,2,\ldots,n$. In (6.1), θ is a location parameter, Σ scale matrix for X_j $(j=1,\ldots,n)$ and ν is termed as the degrees of freedom of the multivariate t-distribution.

We exploit the so-called method of moments for estimation of the parameters θ, Σ, and ν. We propose the sample mean $\bar{X}$ and the sample covariance matrix V/n as estimators for θ and Σ respectively, where

$$\left.\begin{aligned}\underset{\sim}{\bar{X}} &= (\bar{X}_1,\ldots,\bar{X}_i,\ldots,\bar{X}_p)^T\\ V/n &= \frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X})(X_j-\bar{X})^T\end{aligned}\right| \tag{6.2}$$

with $\underset{\sim}{X}_j = (X_{1j},\ldots,X_{ij},\ldots,X_{pj})^T$, $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n}X_{ij}$; and V a sum of product matrix.

We assume $\nu > 4$ so that fourth moment of X_{ij} exist. We then propose moment estimator for ν based on the fourth moment of X_{ij} given by

$$E(X_{ij}-\theta_i)^4 = 3\left(\frac{\nu-2}{\nu-4}\right)\sigma_{ii}^2 , \tag{6.3}$$

where σ_{ii} is the ith diagonal element of the covariance matrix Σ.

The fourth moment of X_{ij} is easily computed as follows: We make the transformation $(X_j-\theta) = \Sigma^{\frac{1}{2}}U_j$ for $j=1,\ldots,n$ in (6.1) and then integrate out $u_1,\ldots,u_{j-1},u_{j+1},\ldots,u_n$ so that we obtain

$$f(u_j) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}\{(\nu-2) + \sum_{i=1}^{p} u_{ij}^2\}^{-\frac{\nu+p}{2}} .$$

It follows then from Johnson and Kotz (1972, pp. 135-36) that

$$\begin{aligned} &E(u_{ij}) = 0,\ E(u_{ij}u_{rj}) = 0,\ E(u_{ij}^3u_{rj}) = 0 , \\ &E(u_{ij}u_{rj}^3) = 0,\ E(u_{ij}^4) = 3(\nu-2)/(\nu-4) , \\ &E(u_{ij}^2u_{rj}^2) = (\nu-2)/(\nu-4) , \end{aligned} \tag{6.4}$$

for all $i=1,\ldots,p$; $j=1,\ldots,n$ and $i\neq r$.

Now let $\Sigma^{\frac{1}{2}} = ((a_{is}))$ so that

$$\left.\begin{aligned} &\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma = ((\sigma_{is})) , \\ &\text{and} \\ &(X_{ij}-\theta_i) = \sum_{s=1}^{p} a_{is}u_{sj} \end{aligned}\right| \tag{6.5}$$

Hence, $E(X_{ij}-\theta_i)^4 = E\{\sum_{s=1}^{p} a_{is}u_{sj}\}^4$

$$= E\{\sum_{s=1}^{p} a_{is}^4 u_{sj}^4 + 4\sum_{s<r}^{p} a_{is}^3 u_{sj}^3 a_{ir} u_{rj}$$

$$+ 6\sum_{s<r} a_{is}^2 u_{sj}^2 a_{ir}^2 u_{rj}^2 + 4\sum_{s<r} a_{is} u_{sj} a_{ir}^3 u_{rj}^3\}$$

$$= 3\left(\frac{\nu-2}{\nu-4}\right)\left(\sum_{s=1}^{p} a_{is}^2\right)^2, \text{ by (6.4).}$$

Since by (6.5) $\left(\sum_{s=1}^{p} a_{is}^2\right)^2 = \sigma_{ii}^2$, it is now readily seen that

$$E(x_{ij}-\theta_i)^4 = 3\left(\frac{\nu-2}{\nu-4}\right)\sigma_{ii}^2 \text{ , which is the fourth moment of } X_{ij}.$$

Hence, we have

$$E\sum_{j=1}^{n}(x_{ij}-\theta_i)^4 = 3n\left(\frac{\nu-2}{\nu-4}\right)\sigma_{ii}^2 \text{ and finally summing over } i=1,\ldots,p,$$

we get

$$E\sum_{i=1}^{p}\sum_{j=1}^{n}(x_{ij}-\theta_i)^4 = 3n\left(\frac{\nu-2}{\nu-4}\right)\sum_{i=1}^{p}\sigma_{ii}^2 \quad (6.6)$$

Thus, we propose a moment estimator for ν as the solution for ν of the equation

$$\{(\nu-2)/(\nu-4)\} = \frac{1}{3n}\{\sum_{i=1}^{p}\sum_{j=1}^{n}(x_{ij}-\hat{\theta}_i)^4\}/\{\sum_{i=1}^{p}\hat{\sigma}_{ii}^2\}, \quad (6.7)$$

where $\hat{\theta}_i$ and $\hat{\sigma}_{ii}$ are given by (6.2).

Finally, we have

$$\hat{\nu} = \frac{2\{3\sum_{i=1}^{p}\hat{\sigma}_{ii}^{2} - \frac{2}{n}\sum_{i}\sum_{j}(x_{ij}-\hat{\theta}_{i})^{4}\}}{\{3\sum_{i=1}^{p}\hat{\sigma}_{ii}^{2} - \frac{1}{n}\sum_{i}\sum_{j}(x_{ij}-\hat{\theta}_{i})^{4}\}} \tag{6.8}$$

6:2 Statistical properties of the estimators

From theorem 5.1, if $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ has the p.d.f. given by (6.1), then $\bar{\underset{\sim}{X}} \sim t_p(\theta,\Sigma/n,\nu)$. This implies that $\bar{X}$, the estimator of θ given by (6.2) is an unbiased estimator of θ. Also when $n\to\infty$, $\Sigma/n\to 0$, therefore, $\bar{X}$ is a consistent estimator of θ.

Next, from (5.11) it is known that $E(V)=n'\Sigma=(n-1)\Sigma$. Therefore, by (6.2), it follows that

$$E(\hat{\Sigma}) = (1-\frac{1}{n})\Sigma ,$$

which implies that the sample covariance matrix is a biased estimator of Σ (the covariance matrix of $\underset{\sim}{X}_j$, for $j=1,\ldots,n$). However for large n, $\hat{\Sigma}$ is unbiased for Σ. Furthermore, as in the following, we show that $\hat{\Sigma}$ is a consistent estimator of Σ too.

In order to show the consistency of $\hat{\Sigma}$ we express the (i,r)th element of $\hat{\Sigma}$, namely, $\hat{\sigma}_{ir}$ as

$$\hat{\sigma}_{ir} = \frac{1}{n}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)(x_{rj}-\bar{x}_r) \quad , \tag{6.9}$$

by (6.2). Now (6.9) can be rewritten as

$$\hat{\sigma}_{ir} = \frac{1}{n}\sum_{j=1}^{n}\{(x_{ij}-\theta_i) - (\bar{x}_i-\theta_i)\}\{(x_{rj}-\theta_r) - (\bar{x}_r-\theta_r)\}$$

$$= \frac{1}{n}\sum_{j=1}^{n}(x_{ij}-\theta_i)(x_{rj}-\theta_r) - (\bar{x}_i-\theta_i)(\bar{x}_r-\theta_r),$$

so that

$$\underset{n\to\infty}{\text{Plim}}\ \hat{\sigma}_{ir} = \underset{n\to\infty}{\text{Plim}}\ \underline{\frac{1}{n}\sum_{j=1}^{n}(x_{ij}-\theta_i)(x_{rj}-\theta_r)}$$

$$- \underset{n\to\infty}{\text{Plim}}\ (\bar{x}_i-\theta_i)\cdot \underset{n\to\infty}{\text{Plim}}\ (\bar{x}_r-\theta_r) \qquad (6.10)$$

Now, $\text{Plim}\ \frac{1}{n}\sum_{j=1}^{n}(x_{ij}-\theta_i)(x_{rj}-\theta_r) = \sigma_{ir}$, because $E(x_{ij}-\theta_i)(x_{rj}-\theta_r) = \sigma_{ir}$, the (i,r)th element of Σ, for all j=1,....,n. The last term in (6.10) is zero, because $\bar{X}$ is consistent for θ implies that $\bar{x}_i$ is consistent for θ_i, also $\bar{x}_r$ is consistent for θ_r. Hence it follows from (6.10) that

$$\underset{n\to\infty}{\text{Plim}}\ \hat{\sigma}_{ir} = \sigma_{ir} \quad , \qquad (6.11)$$

which implies that $\hat{\Sigma}$ is a consistent estimator of Σ.

Finally we show that $\hat{\nu}$ given by (6.8) is a consistent estimator of ν. In order to prove this, we examine $\underset{n\to\infty}{\text{Plim}}\ \frac{1}{n}\sum_i\sum_j(x_{ij}-\hat{\theta}_i)^4$.

$$\operatorname*{Plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij}-\hat{\theta}_i)^4$$

$$= \operatorname*{Plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} \{(x_{ij}-\theta_i) - (\hat{\theta}_i-\theta_i)\}^4$$

$$= \operatorname*{Plim}_{n\to\infty} \left[\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij}-\theta_i)^4 - \frac{4}{n} \sum_i \sum_j (x_{ij}-\theta_i)^3 (\bar{x}_i-\theta_i) \right.$$

$$+ \frac{6}{n} \sum_i \sum_j (x_{ij}-\theta_i)^2 (\bar{x}_i-\theta_i)^2 - \frac{4}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij}-\theta_i)(\bar{x}_i-\theta_i)^3$$

$$\left. + \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} (\bar{x}_i-\theta_i)^4 \right] \tag{6.12}$$

Now $\operatorname*{Plim}_{n\to\infty} \frac{1}{n} \sum_i \sum_j (x_{ij}-\theta_i)^4 = 3\left(\frac{\nu-2}{\nu-4}\right) \sum_{i=1}^{p} \sigma_{ii}^2$, by (6.6), while the probability limit of each of the remaining terms in (6.12) is seen to be equal to zero.

Consider for example

$$\operatorname*{Plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij}-\theta_i)^2 (\bar{x}_i-\theta_i)^2$$

$$= \left[\sum_{i=1}^{p} \operatorname*{Plim}_{n\to\infty} |\bar{x}_i-\theta_i|^2 \cdot \operatorname*{Plim}_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (x_{ij}-\theta_i)^2 \right].$$

Since $\operatorname*{Plim}_{n\to\infty} \bar{x}_i = \theta_i$ and $\operatorname*{Plim}_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (x_{ij}-\theta_i)^2 = \sigma_{ii}$, an application of Slutsky's theorem (cf. Cramer, H., 1946) readily shows that the right hand side of the last equation reduces to 0 (zero).

We then apply Slutsky's theorem to (6.8), and obtain the following result for $\operatorname*{Plim}_{n\to\infty} \hat{\nu}$, by (6.11) and (6.12):

$$\underset{n\to\infty}{\text{Plim}}\ \hat{\nu} = \frac{2\{3\ \underset{n\to\infty}{\text{Plim}} \sum_{i=1}^{p} \hat{\sigma}_{ii}^2 - 2\ \underset{n\to\infty}{\text{Plim}} \frac{1}{n} \sum_{i}^{p} \sum_{j}^{n} (x_{ij}-\bar{x}_i)^4\}}{\{3\ \underset{n\to\infty}{\text{Plim}} \sum_{i=1}^{p} \hat{\sigma}_{ii}^2 - \underset{n\to\infty}{\text{Plim}} \frac{1}{n} \sum_{i} \sum_{j} (x_{ij}-\bar{x}_i)^4\}}$$

$$= \frac{2\{3 \sum_{i=1}^{p} \sigma_{ii}^2 - 2\cdot 3(\frac{\nu-2}{\nu-4}) \sum_{i=1}^{p} \sigma_{ii}^2\}}{3 \sum_{i=1}^{p} \sigma_{ii}^2 - 3(\frac{\nu-2}{\nu-4}) \sum_{i=1}^{p} \sigma_{ii}^2}$$

$$= \nu \qquad (6.13)$$

Therefore $\hat{\nu}$ given by (6.8) is a consistent estimator of ν.

CHAPTER 7

FACTOR ANALYSIS WITH MULTIVARIATE t-ERRORS AND UNOBSERVABLE FACTORS

7.1 The orthogonal factor model

Let $Y_j=(y_{1j},\ldots,y_{ij},\ldots,y_{pj})^T$ be the jth $(j=1,\ldots,n)$ observation vector, where each element of Y_j may be thought of as having been generated by a linear combination of orthogonal unobservable factors (f_j) upon which some error (ε_j) has been superimposed. Let $f_j=(f_{1j},\ldots,f_{ri},\ldots,f_{mj})$ be the vector random variables of the m factors and $\varepsilon_j=(\varepsilon_{1j},\ldots,\varepsilon_{pj})$ be the error vector variable with following distributional assumptions:

Assumptions

(i) $f_j \sim t_m(0,I_m,\nu)$, $m\leq p$

(ii) $\varepsilon_j \sim t_p(0,\Psi^2,\nu)$, where $\Psi^2=\mathrm{Diag}(\Psi_1^{\,2},\ldots,\Psi_p^{\,2})$

(iii) ε_j and f_j are uncorrelated, having the joint p.d.f. given by

$$f(\varepsilon_j,f_j)= K\begin{vmatrix}\Psi^2 & 0\\ 0 & I_m\end{vmatrix}^{-\frac{1}{2}}\{(\nu-2) + f_j^{\,T}f_j + \varepsilon_j^{\,T}(\Psi^2)^{-1}\varepsilon_j\}^{-\frac{\nu+(m+p)}{2}}$$

Also let $\theta=(\theta_1,\ldots,\theta_p)^T$ be a vector of unknown parameters usually called location parameter and $Q=((q_{ir}))$ be a pxm matrix of coefficient parameters, usually called as the factor loading matrix. Taking into account all the above considerations, we now formulate and hence deal with the following linear model

$$Y_j = \theta + Qf_j + \varepsilon_j$$

$$j = 1,\ldots,n \qquad (7.2)$$

This model (7.2) is an orthogonal factor model as it has been assumed in (i) of (7.1) that $f_j(m\times 1) \sim t_m(0, I_m, \nu)$, which implies that factors $f_1,\ldots,f_m$ are orthogonal to each other. Also by assumption (i) it follows from the theorem 4.3 that $Qf_j \sim t_p(0, QQ^T, \nu)$. Hence, by assumptions (ii) and (iii), it now follows that

$$Y_j \sim t_p(\theta, \Sigma, \nu) \quad , \qquad (7.3)$$

where $\Sigma = QQ^T + \Psi^2$. Furthermore, in accordance with (5.3), we assume that $\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_n$ are pairwise uncorrelated and they have the joint p.d.f. given by

$$f(y_1^T,\ldots,y_n^T) = \frac{(\nu-2)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu+np}{2}\right) |QQ^T+\Psi^2|^{-\frac{n}{2}}}{\Gamma\left(\frac{\nu}{2}\right) \pi^{\frac{np}{2}}}$$

$$\times \left\{(\nu-2) + \sum_{j=1}^{n} (y_j-\theta)^T (QQ^T+\Psi^2)^{-1} (y_j-\theta)\right\}^{-\frac{\nu+np}{2}} \qquad (7.4)$$

We note that in (7.4), for $\nu > 2$

$$E(Y_j) = \underset{\sim}{\theta}$$

$$\mathrm{Var}(Y_j) = QQ^T + \Psi^2$$

$$E(Y_j-\theta)(Y_\ell-\theta) = 0 \qquad \text{for } j \neq \ell,\ j, \ell = 1,\ldots,n.$$

7.2 Estimation of the parameters

As the p.d.f. of $\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_j,\ldots,\underset{\sim}{Y}_n$ given by (7.4) is similar to the p.d.f. in (6.1), it follows from equations (6.2) and (6.8) that for the present case

$$\left.\begin{aligned}
&\hat{\theta} = \bar{y}\,, \\
&\hat{\Sigma} = \hat{Q}\hat{Q}^T + \hat{\Psi}^2 = \frac{V}{n} = \frac{1}{n}\sum_{j=1}^{n}(y_j-\bar{y})(y_j-\bar{y})^T\,, \\
&\text{and} \\
&\hat{\nu} = \frac{2\{3\sum_{i=1}^{p}\hat{\sigma}_{ii}^2 - \frac{2}{n}\sum_i\sum_j(y_{ij}-\hat{\theta}_i)^4\}}{\{3\sum_{i=1}^{p}\hat{\sigma}_{ii}^2 - \frac{1}{n}\sum_i\sum_j(y_{ij}-\hat{\theta}_i)^4\}}
\end{aligned}\right| \qquad (7.5)$$

It then remains to estimate the factor loading matrix Q, and the dispersion matrix Ψ^2 (of the error random variable) such that the second equation in (7.5) is satisfied, namely, $\hat{Q}\hat{Q}^T+\hat{\Psi}^2 = \hat{\Sigma} = V/n$.

Substituting $QQ^T+\Psi^2$ for Σ, it follows from (5.27) that the joint distribution of $\bar{Y}$ and V is given by

$$f(\bar{y},v) = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\!\left(\frac{\nu+np}{2}\right)|V|^{\frac{n-p-2}{2}}\,|QQ^T+\Psi^2|^{-\frac{n}{2}}}{\Gamma\!\left(\frac{\nu}{2}\right)\prod_{i=1}^{p}\Gamma\!\left(\tfrac{1}{2}(n-i)\right)\pi^{\frac{p(p+1)}{4}}}$$

$$\times\left\{(\nu-2) + \operatorname{trace}\,(QQ^T+\Psi^2)^{-1}V + (\bar{y}-\theta)^T\left(\frac{QQ^T+\Psi^2}{n}\right)^{-1}(\bar{y}-\theta)\right\}^{-\frac{\nu+np}{2}} \qquad (7.6)$$

The last term in the second parenthesis in (7.6) may be set equal to zero by noting that it vanishes if $\hat{\theta}=\bar{y}$. Moreover ν in (7.6) is estimated by

$\hat{V}$ as in (7.5). For the estimation of Q and Ψ^2, we propose the generalized least square method due to Jöreskog and Goldberger (1972).

In this method, ½ trace $\left[I_p - S^{-1}(QQ^T + \Psi^2)\right]^2$ is minimized in obtaining the estimates for Q and Ψ^2, where $S = V/n = \frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})(y_j - \bar{y})^T$.

Let $R = (r_{iu})$ be a correlation matrix, where $r_{iu} = s_{iu}/(s_{ii}^{1/2} s_{uu}^{1/2})$, and s_{iu} is the element of S matrix in the ith row and uth column. In many applications both the origin and the unit in the scales of measurement are arbitrary or irrelevent and then only the correlation matrix R is of any interest (see p. 126, Enslein, Ralston, and Wilf, 1977). In such cases, we take S to be a correlation matrix R in what follows. For computational purpose, we summarize the generalized least square estimation technique due to Joreskog from Enslein et al. (pp. 127-136) as in the following:

(1) Set $\Psi_i = +\sqrt{e^{\alpha i}}$ $\quad i=1,2,\ldots,p$

(2) The starting point $\alpha_i^{(1)}$ is chosen as

$$\alpha_i^{(1)} = \log\left[(1-m/2p)/s^{ii}\right],$$

where s^{ii} is the ith diagonal element of S^{-1}.

(3) Using (2) in (1) we have $\Psi_i^{(1)}$. Check the matrix $\Psi_i^{(1)} = \text{Diag}\,(\Psi_1^{(1)}, \Psi_2^{(1)}, \ldots \Psi_p^{(1)})$, whether it is non-singular or not. We consider the case when Ψ_i is non-singular only as it happens mostly in practical situations. Then we find out the eigenvalues

$\gamma_1<\gamma_2<\gamma_3<\cdots<\gamma_m<\gamma_{m+1}<\cdots<\gamma_p$ of the matrix $\Psi S^{-1}\Psi$. Next we calculate $w_1, w_2, \ldots, w_m, w_{m+1}, \ldots, w_p$ the corresponding orthonormal eigenvectors of the matrix $\Psi S^{-1}\Psi$. Let $w_i=(w_{i1},\ldots,w_{ik},\ldots,w_{ip})^T$.

(4) Now calculate

$$h(i) = \frac{\partial g}{\partial \alpha_i} = \sum_{k=m+1}^{p} (\gamma_k^2-\gamma_k) w_{ik}^2 \quad ,$$

and

$$H_{ij} = \delta_{ij} \cdot \frac{\partial g}{\partial \alpha_i} + \sum_{k=m+1}^{p} \gamma_k w_{ik} w_{jk} \left\{ \sum_{\ell=1}^{m} \gamma_\ell \frac{\gamma_k+\gamma_\ell-2}{\gamma_k-\gamma_\ell} w_{i\ell} w_{j\ell} + s^{ij}\psi_i\psi_j \right\} ,$$

where $\delta_{ij} = \sum_{k=1}^{p} w_{ik} w_{jk}$

and

$$g(\Psi) = \tfrac{1}{2} \sum_{k=m+1}^{p} (\gamma_k - 1)^2 \quad .$$

(5) Let $h^{(1)}(i)$ and $H_{ij}^{(1)}$ be the values of $h(i)$ and H_{ij} respectively depending on $\alpha_i^{(1)}$. Now the Newton-Raphson iteration procedure is

$$H_{ij}^{(1)}\delta^{(1)} = h^{(1)}$$

$\alpha^{(2)} = \alpha^{(1)} - \delta^{(1)}$, where $\delta^{(1)}$ is a column vector of corrections.

Once we obtain $\alpha^{(2)}$ we go from (1) through (5) to get $\alpha^{(3)}$ and so on. The convergence criterion is that the largest absolute correction be less than a prescribed small number ε. Thus, Ψ_i is estimated by

$+\sqrt{e^{\hat{\alpha}_i}}$, i.e. $\hat{\psi}_i = \sqrt{e^{\hat{\alpha}_i}}$.

Finally Q is estimated using the equation $\hat{Q}=\hat{\Psi}\Omega_1(\Gamma_1^{-1}-I_m)^{\frac{1}{2}}$, (see pp. 129–130, Enslein et al.) where

$$\Gamma_1 = \text{diag}\ (\gamma_1,\gamma_2,\ldots,\gamma_m)$$

and

$\Omega = (w_1,\ldots,w_m,\ldots,w_p)$ be partitioned as $\Omega = \begin{bmatrix}\Omega_1\Omega_2\end{bmatrix}$ with Ω_1 consists of the first m vectors. Thus we obtain the estimates $\hat{\Psi}$ and $\hat{Q}$ for Ψ and Q respectively.

Once we obtain the estimate of the principal factor loading matrix Q, our next effort is to find significance in patterns detected in the factor loading matrix. In doing so, more often, the primary estimate $\hat{Q}$ for Q is not sufficient to interpret the factors. In such cases, often the elements of factors in $\hat{Q}$ are rotated orthogonally (see Press (1972), Morrison (1967)). But, due to these rotations, often the ambiguity arises in factor analysis solutions. However among various procedures proposed for eliminating the ambiguity, the varimax rotation procedure of Kaiser (1958) is widely used. Kaiser's (1958) suggestion is to select the elements of the orthogonal matrix so that 's', the normalized total simplicity is maximized; where

$$s = \sum_{j=1}^{m} s_j \tag{7.7}$$

with $s_j = \frac{1}{p}\sum_{i=1}^{p} q_{ij}^{*4} - \frac{1}{p^2}\left(\sum_{i=1}^{p} q_{ij}^{*2}\right)^2$

and $q_{ij}^{*} = q_{ij}\Big/\sqrt{\sum_{j=1}^{m} q_{ij}^2}$,

q_{ij} is the (i,j)th element of the factor loading matrix Q.

Upon considering a single pair of factors at a time, factors are rotated until all ½m(m-1) pairs of factors have been rotated. The iterative solution for the rotation proceeds in the following fashion:

Consider rth and sth factors as 1st and 2nd factors in general. Now, the first and 2nd factors are rotated by an angle determined from the expression

$$\tan 4\phi = \Big[2\{2p \sum_{i=1}^{p} (q_{ir}^{*2} - q_{is}^{*2}) q_{ir}^{*} q_{is}^{*} - \sum_{i=1}^{p} (q_{ir}^{*2} - q_{is}^{*2})\} \times (2 \sum_{i=1}^{p} q_{ir}^{*} q_{is}^{*})\} \Big] \Big/ \Big[p \sum_{i=1}^{p} \{(q_{ir}^{*2} - q_{is}^{*2}) - (2 q_{ir}^{*} q_{is}^{*})^2\} - \{(\sum_{i=1}^{p} (q_{ir}^{*2} - q_{is}^{*2}))^2 - 2(\Sigma q_{ir}^{*} q_{is}^{*})^2\} \Big] \tag{7.8}$$

and the following table

		Sign of numerator in (7.8)	
		+	-
Sign of denominator in (7.8)	+	I: $0° \le 4\phi < 90°$	IV: $-90° \le 4\phi < 0°$
	-	II: $90° \le 4\phi < 180°$	III: $-180° \le 4\phi < -90°$

[see Morrison (1967, p. 285)]. The new first factor is rotated with the original third factor, and so on, until all ½m(m-1) pairs of factors have been rotated. In this way, finally we obtain a rotated matrix, say, $Q^{**} = (q_{ij}^{**})$ for m factors each factor containing p elements.

CHAPTER 8

REGRESSION ANALYSIS WITH MULTIVARIATE t-ERROR VARIABLE

A regression model with a vector response variable under the assumption that the error has a multivariate t-distribution was considered by Zellner (1976) for a single response variable. This chapter is a direct multidimensional generalization of the above-mentioned model to the case of several response variables. The parameters of the model namely the regression parameters as well as the scale parameters and the degrees of freedom of the error variable are estimated and the estimation procedure is illustrated by an actual stock market data taken from the New York Stock Exchange. Also the properties of the estimators and tests for the regression parameters are discussed.

8.1 A stock market problem and the regression model

An important problem in the area of stock market analysis is the study of the performance of all stocks of some selected firms relative to the overall performance of all stocks trading on a particular (or several) stock exchange(s). For example, consider the price change data in Table 1 (included at the end of the chapter) for the stocks of four selected firms -

1. General Electric,
2. Standard Oil,

3. I.B.M., and

4. Sears ;

trading on the New York Stock Exchange, in relation to the performance of the New York Stock Exchange as a whole (or perhaps in conjunction with several other stock exchanges).

Let Y_{ij} denote the monthly return on $100 of capital, invested on the ith stock during the jth month, $Y_{ij} = 100\{(Q_{ij}-P_{ij}) + R_{ij}\}/P_{ij}$, where P_{ij} is the price of the ith stock at the beginning of the jth month, Q_{ij} the price at the end of the jth month and R_{ij} the dividends earned during the jth month. Let m_j denote the weighted average of these returns during the jth month for the aggregate of all stocks trading on the New York Stock Exchange, called MARKET for short. The primary objective is to study the linear regression of the joint monthly returns of the selected stocks on the corresponding monthly return of the MARKET as a whole.

For example, an appropriate regression set-up for the data shown in Table 1, of Y_{ij} and m_j of the 4 firms over 20 monthly periods would be:

$$Y_{ij} = \alpha_i + \beta_i m_j + \varepsilon_{ij},$$

$$i = 1,2,\ldots,4;\ j = 1,2,\ldots,20,$$

and we would be primarily interested in obtaining estimates of the parameters $(\alpha_1, \beta_1),\ldots,(\alpha_4, \beta_4)$ under suitable distributional assumptions on the error variable ε_{ij}.

A straightforward generalization of the above model would be the case when there are k markets (say NYSE, AMEX, TSE) instead of one and we have data for Y_{ij}, m_{rj} for p firms, $i=1,\ldots,p$ and k MARKETS $r=1,\ldots,k$ and n periods $j=1,\ldots,n$.

The resulting regression set-up would then be

$$Y_{ij} = \alpha_i + \sum_{r=1}^{k} \beta_{ir} m_{rj} + \varepsilon_{ij} , \qquad (8.1)$$

for i=1,2,....,p and j=1,2,....,n, the regression parameters of interest being α_i and β_{i1}, β_{i2}, ..., β_{ik} for i=1,2,....,p.

We therefore consider the more general model (employing matrix notation) give by:

$$Y_{(Pxn)} = \theta_{(Pxk)} X_{(kxn)} + \varepsilon_{(pxn)} \qquad (8.2)$$

where θ is the parameter to be estimated, X is the design matrix and ε, an error matrix with suitable distributional assumption to be discussed later.

We remark that this model accommodates (8.1) with $\theta=(\alpha,\beta)$ and $x^T = [\underset{\sim}{1}_n, M^T]$, where $\alpha=(\alpha_1,\alpha_2,....,\alpha_p)^T$, $\beta_{(Pxk)}=((\beta_{i\gamma}))$, $\underset{\sim}{1}_n=(1,1,....,1)^T$, and $M_{(kxn)}=((m_{rj}))$, where X^T stands for the transpose of X.

In accordance with Zellner (1976), we now make the following assumptions on the error variable $\varepsilon=((\varepsilon_{ij}))$:

$E(\varepsilon_{ij}) = 0$, for all i, j

$E(\varepsilon_{ij}^2) = \sigma^2\Lambda_{ii}$ for all j, and i = 1,2,....,P,

$E(\varepsilon_{ij}\varepsilon_{\ell j}) = \sigma^2\Lambda_{i\ell}$ for all j, and i, = 1,2,....,P,

and $E(\varepsilon_{ij}\varepsilon_{\ell j'}) = 0$ for all i and ℓ and $j \neq j'$

where Λ_{ij} are unknown parameters.

Further we assume that for a given σ, the errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independently and normally distributed, the distribution of $\underset{\sim}{\varepsilon}_j=(\varepsilon_{1j},\ldots,\varepsilon_{pj})^T$ being $N_p(0,\sigma^2\Lambda)$ for $j=1,2,\ldots,n$; while σ is assumed to be a random variable having an inverted gamma distribution with probability density function given by

$$g(\sigma) = \frac{2}{\Gamma\left(\frac{\nu}{2}\right)\left(\frac{\nu}{2}\right)^{\frac{1}{2}}} \operatorname{Exp}\{-\tfrac{1}{2}\left(\frac{\nu}{\sigma^2}\right)\} \cdot \left(\frac{\nu}{2\sigma^2}\right)^{\frac{\nu+1}{2}}, \quad \sigma > 0, \nu > 0 ,$$

where ν is an unknown parameter.

Then similar calculations to (5.2) through (5.3) yields

$$f(\varepsilon_1^T,\ldots,\varepsilon_n^T) = \int_0^\infty f(\varepsilon_1^T,\ldots,\varepsilon_n^T,)g(\sigma)d\sigma$$

$$= \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\,\Gamma\left(\frac{\nu}{2}\right)} |\Sigma|^{-\frac{n}{2}} \{(\nu-2) + \sum_{j=1}^{n} \varepsilon_j^T \Sigma^{-1} \varepsilon_j\}^{-\frac{\nu+np}{2}} \tag{8.3}$$

where $\Sigma = \frac{\nu}{\nu-2}\Lambda$.

We propose (8.3) as the distribution of the error variable of the regression model (8.2). It follows then [see for example Cornish (1954)] that

$$E(\varepsilon_j) = 0$$

$$E(\varepsilon_j\varepsilon_j^T) = \Sigma$$

and

$$E(\varepsilon_j\varepsilon_s^T) = 0, \quad j\neq s, \; j,s=1,2,\ldots,n.$$

In this model, it is to be noted that while $\varepsilon_1,\ldots,\varepsilon_n$ are pairwise uncorrelated, they are not necessarily independent. Independence implies that $\varepsilon_1,\ldots,\varepsilon_n$ are normally distributed [c.f. Kelker, p. 423, 1970]. Finally we remark that the proposed model accommodates the usual case when the errors $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n$ are assumed to be independently and identically distributed according to $N_p(0,\Sigma)$ by letting $\nu\to\infty$ in (8.3).

8.2 Estimation of the parameters

Consider the proposed model (8.2) given by:

$$Y = \theta X + \varepsilon, \tag{8.4}$$

where Y as before is (Pxn) response variable, θ a (Pxk) matrix of unknown parameters to be estimated, X a known design matrix of order kxn and $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_j,\ldots,\varepsilon_n)$ an error variable where $\varepsilon_j=(\varepsilon_{1j},\ldots,\varepsilon_{pj})^T$, and p.d.f. of ε is given by the multivariate t-distribution:

$$f(\varepsilon) = \frac{(\nu-2)^{\frac{\nu}{2}}\,\Gamma\!\left(\frac{\nu+np}{2}\right)}{\pi^{\frac{np}{2}}\,\Gamma\!\left(\frac{\nu}{2}\right)}\,|\Sigma|^{-\frac{n}{2}}\{(\nu-2) + \sum_{j=1}^{n} \varepsilon_j^T\Sigma^{-1}\varepsilon_j\}^{-\frac{\nu+np}{2}} \tag{8.5}$$

We recall that for $\nu > 2$, $E(\varepsilon_j)=0$, $E(\varepsilon_j\varepsilon_j^T)=\Sigma$, and $E(\varepsilon_j\varepsilon_s^T)=0$ for $j\neq s$; $j,s=1,2,\ldots,n$.

To estimate θ we minimize $\sum_{j=1}^{n}(y_j-\theta x_j)^T(y_j-\theta x_j)$, where $y_j=(y_{1j},\ldots,y_{pj})^T$ and $x_j=(x_{1j},\ldots,x_{kj})^T$, with respect to θ and obtain the following least square estimate of θ:

$$\hat{\theta}^T = (xx^T)^{-1}xy^T \tag{8.6}$$

Next we propose the following moment estimator for the covariance matrix :

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^{n} (y_j - \hat{\theta} x_j)(y_j - \hat{\theta} x_j)^T \tag{8.7}$$

where $\hat{\theta}$ is determined by (8.6). In order to estimate ν by the method of moments we now consider the fourth moment of ε_{ij}. Similar calculations to (6.3) through (6.7), yields the equation

$$\{(\nu-2)/(\nu-4)\} = \frac{1}{3n} \left[\sum_i \sum_j \left(y_{ij} - \sum_{r=1}^{k} \hat{\theta}_{ir} x_{rj} \right)^4 \right] \Big/ \left[\sum_{i=1}^{p} \hat{\sigma}_{ii}^2 \right] , \tag{8.8}$$

where $\hat{\theta}_{ir}$ and $\hat{\sigma}_{ii}$ are given by (8.6) and (8.7) respectively. In what follows we will denote the so-called 'estimated residual' by $\hat{\varepsilon}_{ij}$ where

$$\hat{\varepsilon}_{ij} = y_{ij} - \sum_{r=1}^{k} \hat{\theta}_{ir} x_{rj} , \tag{8.9}$$

in analogy with $\varepsilon_{ij} = (y_{ij} - \sum_{r=1}^{k} \theta_{ir} x_{rj})$ by virtue of (8.4). Then, from (8.8), we obtain the estimate of ν as

$$\hat{\nu} = 2\left(3 \sum_{i=1}^{p} \hat{\sigma}_{ii}^2 - \frac{2}{n} \sum_i \sum_j \hat{\varepsilon}_{ij}^4\right) \Big/ \left(3 \sum_{i=1}^{p} \hat{\sigma}_{ii}^2 - \frac{1}{n} \sum_i \sum_j \hat{\varepsilon}_{ij}^4\right) . \tag{8.10}$$

<u>A Numerical Example</u>

Let us consider the stock market problem of section 8.1 relating to the four selected firms, having the regression model

$$Y_{ij} = \alpha_i + \beta_i m_j + \varepsilon_{ij}, \quad i=1,2,\ldots,4; \; j=1,2,\ldots,20. \tag{8.11}$$

where the error variable is assumed to have a t-distribution of the form given by (8.3) with the unknown parameters ν and Σ.

We illustrate the estimation of all unknown parameters of this model namely α_i, β_i for i=1,2,3,4; Σ and ν.

$$\text{Put} \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix} = \theta^T \text{ and set}$$

$$X = \begin{pmatrix} 1 & 1 & & 1 \\ m_1 & m_2 & & m_{20} \end{pmatrix}, \; Y=(y_1,\ldots.y_4)^T, \text{ where } y_i=(y_{i1},\ldots.,y_{i20}) \text{ for}$$

i=1,....,4. Then (8.11) can be rewritten as $Y = \theta X + \varepsilon$, which reduces to the model (8.4). The values for m_1, m_2, ..., m_{20} and the Y_{ij} are listed in Table 1. Since the error variable in (8.11) is assumed to have a t-distribution of the form given by (8.3), therefore direct application of the equations (8.6), (8.7) and (8.10) yields the following estimates for θ, Σ and ν.

	G.E. (1)	STAND (2)	I.B.M. (3)	SEARS (4)
$\hat{\alpha}_i$	-0.27489	-0.89038	0.21585	-2.00951
$\hat{\beta}_i$	1.18155	1.01322	0.95128	1.11962

$$\hat{\Sigma} = \begin{bmatrix} 7.8015 & 6.0205 & -4.0967 & -0.8160 \\ & 12.3545 & -1.1775 & -3.6845 \\ & & 13.6160 & 9.5150 \\ & & & 22.6845 \end{bmatrix} \text{ and } \hat{\nu} = 21.32$$

8.3 Properties of the estimators

In this section, it is shown that the estimates $\hat{\theta}$, $\hat{\Sigma}$ and $\hat{\nu}$ given by (8.6), (8.7), and (8.10) are consistent estimators (as $n\to\infty$) for the respective parameters. It is well known that $\hat{\theta}$ in (8.6) is an unbiased estimator for θ. This is because by (8.4) $\hat{\theta}^T$ can be expressed as

$$\hat{\theta}^T = (XX^T)^{-1}XY^T = (XX^T)^{-1}X\,(X^T\theta^T + \varepsilon^T)$$

so that

$$\hat{\theta}^T = \theta^T + (XX^T)^{-1}X\varepsilon^T \tag{8.12}$$

but $E(\varepsilon^T)=0$. However, $\hat{\Sigma}$ and $\hat{\nu}$ are not unbiased and no attempt has been made to remove the bias. It is of great interest to study the convergence of the estimators to the corresponding population parameters. We recall that X is a design matrix of order (kxn). Then, as n becomes infinite so does the order of X. Therefore, in order to examine the asymptotic behaviour (as $n\to\infty$) of the estimates we now assume that

$\lim_{n\to\infty} (\frac{1}{n}XX^T)$ exists so that $\lim_{n\to\infty} (\frac{1}{n}XX^T)^{-1}$ also exists.

Consistency of $\hat{\theta}$:

It follows from (8.12) that $\hat{\theta}^T-\theta^T = (\frac{1}{n}XX^T)^{-1}\frac{1}{n}X\varepsilon^T$. Next by (8.5) it is readily verified that $cov(\frac{1}{n}X\varepsilon_i^T)=(\sigma_{ii}/n)\otimes(\frac{1}{n}XX^T)$ so that $cov(\frac{1}{n}X\varepsilon^T) = (\frac{1}{n}\Sigma)\otimes(\frac{1}{n}XX^T)$, where $\varepsilon_i=(\varepsilon_{i1},\ldots,\varepsilon_{in})$. Therefore, $cov(\hat{\theta}^T) = (\frac{1}{n}XX^T)^{-1}\otimes(\Sigma/n)$. Since by assumption $\lim_{n\to\infty} (\frac{1}{n}XX^T)^{-1}$ exists, and Σ is a finite dimensional with bounded elements, it follows that $\lim_{n\to\infty} cov(\hat{\theta}^T)=0$. Therefore $\hat{\theta}$ is a consistent estimator of θ.

Consistency of $\hat{\Sigma}$:

By (8.7), the (i,r)th element of $\hat{\Sigma}$, namely $\hat{\sigma}_{ir}$ can be expressed as

$$\hat{\sigma}_{ir} = \frac{1}{n}\hat{\varepsilon}_i\hat{\varepsilon}_r^T \quad , \tag{8.13}$$

where $\hat{\varepsilon}_i = (\hat{\varepsilon}_{i1},\ldots,\hat{\varepsilon}_{in})$.

Also by (8.9) we have $\hat{\varepsilon}_i=(Y_i-\hat{\theta}_iX)$, which in turn by (8.12) noting that $Y_i-\theta_{ir}=\varepsilon_i$, can be written as

$$\hat{\varepsilon}_i = \varepsilon_i[I-X^T(XX^T)^{-1}X] \tag{8.14}$$

Using (8.14) in (8.13), a simple calculation shows that

$$\hat{\sigma}_{ir} = \frac{1}{n}\varepsilon_i\varepsilon_r^T - \frac{1}{n}\varepsilon_iX^T(XX^T)^{-1}(X\varepsilon_r^T).,$$

so that

$$\underset{n\to\infty}{\text{plim}}\ \hat{\sigma}_{ir} = \underset{n\to\infty}{\text{plim}}\left(\frac{1}{n}\varepsilon_i\varepsilon_r^T\right) - \{\underset{n\to\infty}{\text{plim}}\left(\frac{1}{n}\varepsilon_iX^T\right)$$

$$\lim_{n\to\infty}\left(\frac{1}{n}XX^T\right)^{-1}\underset{n\to\infty}{\text{plim}}\left(\frac{1}{n}X\varepsilon_r^T\right) \tag{8.15}$$

Now $\underset{n\to\infty}{\text{plim}}\left(\frac{1}{n}\varepsilon_i\varepsilon_r^T\right) = \sigma_{ir}$, because $E\left(\frac{1}{n}\varepsilon_i\varepsilon_r^T\right) = \sigma_{ir}$.

The last term in (8.15) is zero, because $E(\varepsilon_i) = 0$ and $\lim_{n\to\infty}\left(\frac{1}{n}XX^T\right)^{-1}$ exists by assumption.

Hence, $\underset{n\to\infty}{\text{plim}}\ \hat{\sigma}_{ir} = \sigma_{ir}$, (8.16)

which implies that $\hat{\Sigma}$ is a consistent estimator of Σ.

Consistency of $\hat{\nu}$:

From (8.9) and (8.4) it is seen that $\hat{\varepsilon}_{ij}-\varepsilon_{ij} = -\sum_{r=1}^{k}(\hat{\theta}_{ir}-\theta_{ir})x_{rj}$

so that $\frac{1}{n}\sum\sum_{ij}\hat{\varepsilon}_{ij}^4 = \frac{1}{n}\sum\sum_{ij}\{\varepsilon_{ij} - \sum_{r=1}^{k}(\hat{\theta}_{ir}-\theta_{ir})x_{rj}\}^4$

$$= \frac{1}{n}\sum\sum_{ij}\varepsilon_{ij}^4 - 4\sum\sum_{ij}\frac{1}{n}\varepsilon_{ij}^3\{\sum_{r=1}^{k}(\hat{\theta}_{ir}-\theta_{ir})x_{rj}\}$$

$$+ 6\sum\sum\frac{1}{n}\varepsilon_{ij}^2\{\sum_{r=1}^{k}(\hat{\theta}_{ir}-\theta_{ir})x_{rj}\}^2$$

$$- 4\sum\sum_{ij}\frac{1}{n}\varepsilon_{ij}\{\sum_{r=1}^{k}(\hat{\theta}_{ir}-\theta_{ir})x_{rj}\}^3$$

$$+ \sum\sum_{ij}\frac{1}{n}\{\sum_{r=1}^{k}(\hat{\theta}_{ir}-\theta_{ir})x_{rj}\}^4 \qquad (8.17)$$

Now it follows from (5.3) that

$E(\varepsilon_{ij}^4) = 3(\frac{\nu-2}{\nu-4})\sigma_{ii}^2$, therefore

$$\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n}\varepsilon_{ij}^4 = 3(\frac{\nu-2}{\nu-4})\sum_{i=1}^{p}\sigma_{ii}^2,$$

while the probability limit under the assumption that $|x_{rj}| \le M$ for all r and j, of each of the remaining terms in (8.17) is seen to be equal to zero. Consider for example

$$\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum\sum_{ij}\varepsilon_{ij}^2\{\sum_{r=1}^{k}(\hat{\theta}_{ir}-\theta_{ir})x_{rj}\}^2$$

$$\le \underset{n\to\infty}{\text{plim}}\sum\sum_{ij}\frac{1}{n}\varepsilon_{ij}^2\{\sum_{r=1}^{k}|(\hat{\theta}_{ir}-\theta_{ir})x_{rj}|\}^2$$

$$\leq M^2 \sum_{i=1}^{p} \left[\underset{n\to\infty}{\text{plim}} \{ \sum_{r=1}^{k} |\hat{\theta}_{ir} - \theta_{ir}| \}^2 \underset{n\to\infty}{\text{plim}} \sum_{j=1}^{n} \frac{1}{n} \varepsilon_{ij}^2 \right] = 0$$

since $\underset{n\to\infty}{\text{plim}}\ \hat{\theta}_{ir} = \theta_{ir}$

Applying Slutsky's theorem (cf. Cramer, H., 1946) to (8.10), from (8.16) and (8.17) it follows that $\underset{n\to\infty}{\text{plim}}\ \hat{\nu} = \nu$.

8.4 Hypothesis testing

Consider the model $Y=\theta x+\varepsilon$ given by (8.4), where Y is a (pxn) response variable, θ a (pxk) matrix of unknown parameters estimated by (8.6), x a (kxn) known design matrix of rank $k \leq n$ and ε, a (pxn) error matrix which has a multivariate t-distribution given by (8.5). The model can be rewritten as:

$$Y^* = (I_p \otimes x^T)\ \theta^* + \varepsilon^* \quad , \tag{8.18}$$

where $\varepsilon^* \sim t_{np}(0,\Sigma \otimes I_n,\nu)$ by (3.2). In (8.18), $Y^*=(y_1,\ldots,y_i,\ldots,y_p)^T$, $\theta^*=(\theta_1,\ldots,\theta_p)^T$, $\varepsilon^*=(\varepsilon_1,\ldots,\varepsilon_p)^T$, where $y_i=(y_{i1},\ldots,y_{in})$, $\theta_i=(\theta_{i1},\ldots,\theta_{ik})$, and $\varepsilon_i=(\varepsilon_{i1},\ldots,\varepsilon_{in})$ for $i=1,2,\ldots,p$.

We wish to test the hypothesis $H:\theta=\theta_0$ (or equivalently $\theta^*=\theta_o^*$) versus the alternative hypothesis $H_1:\theta\neq\theta_o$ (or $\theta^*\neq\theta_o^*$) for the cases when ν, Σ are known and when ν, Σ are unknown but n is sufficiently large.

Case 1 ν and Σ known

In notation of (8.18), from (8.6) we obtain

$$\hat{\theta}^* = [I_p \otimes (xx^T)^{-1}x]Y^* \tag{8.19}$$

Therefore, by (8.12) we can write

$$\hat{\theta}^* - \theta^* = [I_p \otimes (xx^T)^{-1}x]\varepsilon^*, \tag{8.20}$$

where $\varepsilon^* \sim t_{np}(0, \Sigma \otimes I_n, \nu)$. Now, let $\beta = \Sigma \otimes (xx^T)^{-1}$ and $U = \beta^{-\frac{1}{2}}(\hat{\theta}^* - \theta_o^*)$, where $\hat{\theta}^*$ is given by (8.19) and θ_o^* is the value of θ^* under the null hypothesis H_o. We now propose (in analogy with the corresponding problem when the error is assumed to have multivariate normal set up) the test statistic D given by

$$D = \frac{\nu}{\nu-2} U^T U$$

$$= \frac{\nu}{\nu-2} (\hat{\theta}^* - \theta_o^*)^T \beta^{-1} (\hat{\theta}^* - \theta_o^*)$$

$$= \frac{\nu}{\nu-2} (\hat{\theta}^* - \theta_o^*)^T \{\Sigma \otimes (xx^T)^{-1}\}^{-1} (\hat{\theta}^* - \theta_o^*) \tag{8.21}$$

Obviously lower values of the test statistic D will favor the H_o while the higher values will direct the rejection of H_o.

Distribution of the test statistic D:

By assumption Rank $(x) = k \leq n$, so that the rank of $[I_p \otimes (xx^T)^{-1}x]$ is kp, whereas the rank of $(\Sigma \otimes I_n)$ is np. Since $\varepsilon^* \sim t_{np}(0, \Sigma \otimes I_n, \nu)$ while from (8.20) $(\hat{\theta}^* - \theta^*)$ is a linear function of ε^*, straightforward application of theorem 4.3 shows that $\beta^{-\frac{1}{2}}(\hat{\theta}^* - \theta^*) \sim t_{kp}(0, I_{kp}, \nu)$, where

$\beta = \Sigma \otimes (xx^T)^{-1}$. It now follows that

$$U = \beta^{-\frac{1}{2}}(\hat{\theta}^* - \theta_o^*) = \beta^{-\frac{1}{2}}\left[(\hat{\theta}^* - \theta^*) + (\theta^* - \theta_o^*)\right] \sim t_{kp}(\beta^{-\frac{1}{2}}(\theta^* - \theta_o^*), 1_{kp}, \nu)$$

Hence from theorem 5.3, it follows that $D = \frac{\nu}{\nu-2} U^T U$ has the density given by

$$f(d) = \frac{\nu^{\frac{\nu}{2}} d^{\frac{kp}{2}-1}}{\Gamma\left(\frac{\nu}{2}\right)} \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{\nu+kp}{2}+2j\right)}{\Gamma\left(\frac{kp}{2}+j\right)\Gamma(j+1)} \frac{(\lambda d)^j}{(\lambda+\nu+d)^{\frac{\nu+kp}{2}+2j}} \qquad (8.22)$$

where $\lambda = \frac{\nu}{\nu-2}(\theta^* - \theta_o^*)^T \beta^{-1} (\theta^* - \theta_o^*)$.

We note that under $H_o: \theta^* = \theta_o^*$ so that $\lambda = 0$, D/kp has the usual central F-distribution with degrees of freedom ν and kp by virtue of theorem 5.3.

Case 2 ν, Σ are unknown, n is large

We recall from Section 8.3 that $\hat{\Sigma}$ and $\hat{\nu}$ as determined by (8.7) and (8.10) are consistent estimators for Σ and ν respectively. Hence the F-test based upon $D = \frac{\hat{\nu}}{\hat{\nu}-2} U^T U$, where $U = \beta^{-\frac{1}{2}}(\hat{\theta}^* - \theta_o^*)$ may still be approximately valid with $\hat{\nu}$ in place of ν.

Table 1: A Monthly Return Data

Month j	MARKET (m_j)	FIRM: 1. General Electric (y_{1j})	2. Standard Oil (y_{2j})	3. IBM (y_{3j})	4. Sears (y_{4j})
1	-3.9597910	-4.9438201	-6.2893093	-1.3882671	-9.4202876
2	-1.6576912	-4.4917259	-5.3243849	1.1807449	2.0000000
3	-1.0816958	-1.8316831	-2.6315790	0.1811394	-4.3650795
4	0.4112013	9.6692085	5.4054055	-6.4195275	-4.7717843
5	-1.2405984	0.6960556	-1.3519812	-4.1545894	-3.3551198
6	5.1096663	5.1612902	2.8708134	7.5356424	5.4545455
7	-1.5521422	-3.7610620	-0.6976746	1.5625000	0.4310343
8	-1.4268629	-1.8390805	-8.1498802	1.0256410	7.6909840
9	0.0679950	-2.9508196	-1.5503876	-2.7480204	-3.2128513
10	-3.9339226	-2.1951221	0.2624672	-1.4367815	-3.7344400
11	4.2175182	1.0972567	-2.0418849	3.5957240	4.7068965
12	0.5467556	-0.4987530	7.3170722	3.5984848	-7.0539415
13	-5.7218745	-8.2706749	-7.0707083	-2.9250458	-9.8214269
14	-1.2142301	-2.1857925	-2.2826087	-4.2824857	-2.2574261
15	3.1896032	5.4189943	3.3898305	-6.2686563	-7.7319562
16	8.3155274	12.0643440	12.2950790	12.9511650	15.6424580
17	1.8877115	1.1961721	0.6326035	-1.8308274	-4.2318844
18	-1.3075344	-3.2624114	-5.3921569	-0.3872217	-5.6122448
19	5.6815811	6.1881188	3.6269430	9.2322648	3.2432433
20	3.7603755	1.1655010	-0.6000001	4.8950177	-2.4921466

Source: The Monthly Stock Returns file of Centre for Research in Security Prices, University of Chicago, Graduate School of Business, 1101 E. 58th Street, Chicago, IL 60637

REFERENCES

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

AWAD, A.M. (1980), "Remark on the characteristic function of the F-distribution," Sankhyā, Ser. A, 42, 128-129.

BARTLETT, M.S. (1934), "The vector representation of a sample," Proc. Camb. Phil. Soc. 30, 327-340.

BOX, G.E.P. (1952), "Multi-factor designs of first order," Biometrika 39, 49-57.

BOX, G.E.P. (1953), "Spherical distributions," (Abstract), Ann. Math. Statist. 24, 687-688.

CHMIELEWSKI, M.A. (1980a), "Invariant scale matrix hypothesis tests under elliptical symmetry," J. Mult. Anal. 10, 343-350.

CHMIELEWSKI, M.A. (1980b), "Tests for block diagonal structure under elliptical symmetry," Unpublished manuscript.

CHMIELEWSKI, M.A. (1981), "Elliptically Symmetric Distributions: A Review and Bibliography," International Statistical Review, 49, p. 67-74.

CORNISH, E.A. (1954), "The Multivariate t-Distribution Associated with a Set of Normal Sample Deviates," Australian Journal of Physics, Vol. 7, 531-542.

CRAMER, H. (1946), "Mathematical Methods of Statistics," Princeton University Press.

DAWID, A.P. (1977), "Spherical matrix distributions and a multivariate model," J. R. Statistical Society, B, 39, 254-261.

DUNNETT, C.W., and SOBEL, M. (1954), "A Bivariate Generalization of Student's t-distribution, with Tables for Certain Special Cases," Biometrika, Vol. 41, 153-169.

EFRON, B. (1969), "Student's t-test under symmetry conditions," J. Am. Statist. Assoc., 64, 1278-1302.

ENSLEIN, K., RALSTON, A., and WILF, H.S. (1977), "Statistical Methods for Digital Computers," Vol. III, John Wiley and Sons Inc., 125-153.

FISHER, R.A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenies, Vol. 7, pp. 179-188.

FISHER, R.A. and HEALHY, M.J.R. (1956), "New tables of Behrens' test of significance," J. R. Stats. Soc., Ser. B, 18, 212.

FRASER, D.A.S. (1951), "Generalized Hit Probabilities with a Gaussian Target," Annals of Math. Stats., V-22, 248-255.

GEISSER, S. (1964), "Posterior Odds for Multivariate Normal Classifications," Journal of the Royal Statistical Society, B, Vol. 26, pp. 69-76.

GILLILAND, D.C. (1968), "On maximizations of the integral of a bell-shaped function over a symmetric set," Naval Res. Logistic Quart., 15, 507-515.

GRÖBNER, W. and HOFREITER, N. (1961), "Integraltafel zweiter teil Bestimmte Integrale heransgegeben," Wien Springer-Verlag.

HARTMAN, P. and WINTNER, A. (1940), "On the spherical approach to the normal distribution law," Am. J. Math. 62, 759-779.

IFRAM, A.F. (1970), "On the characteristic function of the F- and t-distributions," Sankhya, Ser. A, 32, 350-352.

JEFFREYS, H. (1957), "Scientific Inference," 2nd ed., Cambridge, England: Cambridge University Press.

JOHNSON, N.L. and KOTZ, S. (1972), "Distributions in Statistics: Continuous Multivariate Distributions," John Wiley and Sons Inc.

JOHNSON, N.L. and KOTZ, S. (1970), "Distributions in Statistics: Continuous Univariate Distributions - 2," John Wiley and Sons Inc.

JÖRESKOG, K., and GOLDBERGER, A. (1972), "Factor Analysis by Generalized Least Squares," Psychometrika, 37, 243-260.

KAISER, H.F. (1958), "The Varimax criterion for analytic rotation in factor analysis," Psychometrika, 23, 187-200.

KARIYA, T. (1981), "A Robustness Property of Hotelling's T^2-Test," The Annals of Statistics, Vol. 9, No. 1, 211-214.

KARIYA, T. (1981), "Robustness of Multivariate Tests," The Annals of Statistics, Vol. 9, No. 6, 1267-1275.

KARIYA, T. and EATON, M.L. (1977), "Robust tests for spherical symmetry," Ann. Statistics, 5, 206-215.

KELKER, D. (1970), "Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization," Sankhyā, Series A, 32, 419-430.

KULLBACK, S. (1959), "Information Theory and Statistics," New York: John Wiley and Sons.

LORD, R. D. (1954), "The Use of the Hankel Transform in Statistics. I. General theory and examples, Biometrika, 41, 44-55.

MANDELBROT, B. (1963), "The Variation of Certain Speculative Prices," Journal of Búsiness, 36, 394-419.

MARDIA, K.V., KENT, J.T., and BIBBY, J.M. (1979), "Multivariate Analysis," Academic Press.

MATHAI, A.M. and PEDERJOLI, G. (1977), "Characterization of the normal probability law," John Wiley, New Delhi.

MAXWELL, J.C. (1860), "Illustration of the dynamical theory of gases - Part I. On the motions and collisions of perfectly elastic bodies," Taylor's Phil. Mag., 19, 19-32.

MEHTA, M.M. (1967), "Random Matrices and the Statistical Theory of Energy Levels," New York: Academic Press.

MORRISON, D.F. (1967), "Multivariate Statistical Methods," McGraw-Hill.

MUIRHEAD, R.J. (1982), "Aspects of Multivariate Statistical Theory," John Wiley and Sons, Inc.

NASH, D. and KLAMKIN, M.S. (1976), "A spherical characterization of the normal distribution," J. Mult. Anal. 55, 156-158.

NEUDECKER, H. (1969), "Some theorems on Matrix Differentiation with special Reference to Kronecker Products," JASA, V. 64.

NEWCOMB, S. (June 1886), "A Generalized Theory of the Combination of Observations so as to Obtain the Best Results," American Journal of Mathematics, 343-366.

PESTANA, D. (1977), "Note on a paper of Ifram," Sankhyā, Ser. A, 39, 396-397.

PHILLIPS, P.C.B. (1982), "The true characteristic function of the F-distribution," Biometrika, 69, 1, pp. 261-264.

PRAETZ, P.D. (1972), "The distribution of share price changes," Journal of Business, 45, 49-55.

PRESS, S.J. (1972), "Applied Multivariate Analysis," Holt, Rinehart and Winston, Inc.

RIAFFA, H. and SCHLAIFER, R. (1961), "Applied Statistical Decision Theory," Boston: Harvard University Press.

RAO, C.R. (1948a), "The utilization of multiple measurements in problems of biological classification," J.R.S.S., B, 10, pp. 159-193.

SRIVASTAVA, M.S. and CARTER, E.M. (1983), "An Introduction to Applied Multivariate Statistics," North-Holland.

ZELLNER, A. (1976), "Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Term," Journal of the American Statistical Association, Vol. 71, 400-405.

END

24 03 85

FIN