UNIVERSITY OF CAPE TOWN

DEPARTMENT OF MATHEMATICAL STATISTICS

****

ASPECTS

OF

NON-CENTRAL MULTIVARIATE t DISTRIBUTIONS

by

June M. Juritz

****

A thesis prepared under the supervision of Professor C.G. Troskie in fulfilment of the requirements for the degree of Doctor of Philosophy in Mathematical Statistics

****

To: John, Robert, Jennifer and David

ACKNOWLEDGEMENTS

My warmest thanks go to Professor C.G. Troskie for his encouragement and guidance not only during the preparation of this thesis but throughout my post-graduate statistical career. In particular I appreciate the time he devoted to discussing the first draft as it was during his well-earned sabbatical leave. Since he did not read the final draft any errors and omissions are entirely my own responsibility.

Secondly, I should like to thank my colleagues in the Department of Mathematical Statistics and in particular Dr. Arthur Money, for the many stimulating discussions we have had over the years.

I thank Mrs. C. Barends, Mrs. M.I. Cousins, Mrs. B. Prosser and Miss L. Jennings for their skilful typing of the manuscript. I am especially grateful to the first three mentioned above who, in spite of other heavy commitments, took over the job when Miss Jennings was prevented from continuing because of serious illness. Any variations in the style of the typescript must not be attributed to lack of expertise on their part but to the idiosyncrasies of their particular typewriters.

To my husband, John, go my thanks for his practical help with the proof reading and to him and my children my appreciation of their forbearance and cheerfulness over the last few months when the completion of this thesis was uppermost in my thoughts. I also wish to thank my maid, Miss Theresa Booysen, who by keeping our household running smoothly, contributed in no small way to the completion of this task.

Finally I wish to acknowledge the debt I owe to the University of South Africa and in particular to Professor H.S. Steyn and the Department of Mathematical Statistics. Ten years ago, with no thought of ever becoming a statistician, I registered for a course with U.N.I.S.A. I found a subject which has become an endless source of fascination to me. Anything that is worthwhile in this thesis sprang from the seeds that were sown in those early days.

CONTENTS

PREFACE

CHAPTER 1 INTRODUCTION

1.1 The univariate Student's t distribution 1.1

1.2 Multivariate generalisations of the t distribution 1.3

CHAPTER 2 FOLKLORE

2.1 Matrices and determinants 2.1

2.2 The multivariate gamma function 2.5

2.3 The generalised hypergeometric function and zonal polynomials 2.7

2.4 Some univariate probability distributions 2.11

2.5 Some multivariate probability distributions 2.16

CHAPTER 3 THE MULTIVARIATE t DISTRIBUTION AND THE INVERTED t DISTRIBUTION

3.1 Introduction 3.1

3.2 Four cases of the multivariate t distribution 3.3

3.3 The inverted multivariate t distribution 3.21

3.4 An application of the lower non-central multivariate t distribution 3.29

TABLE OF PERCENTAGE POINTS OF THE UNIVARIATE LOWER NON-CENTRAL t DISTRIBUTION following page 3.36

CHAPTER 4 QUADRATIC FORMS OF t VARIABLES

4.1 Introduction 4.1

4.2 Quadratic forms of normal variables 4.2

4.3 The joint distribution of quadratic forms of t variables 4.5

4.4 Quadratic forms in a partitioned t vector 4.11

4.5 The general distribution of t'At 4.16

4.6 The distribution of W 4.22

CHAPTER 5 STATISTICS DERIVED FROM THE MULTIVARIATE t

5.1 Introduction 5.1

5.2 The central distribution of $\tilde{T}$ 5.5

5.3 Three non-central distributions of $\tilde{T}$ 5.10

5.4 The distribution of the roots of the central $\tilde{T}$ 5.22

CHAPTER 6 THE MATRIX T DISTRIBUTION

6.1 Introduction 6.1

6.2 The central matrix T distribution 6.5

6.3 Non-central matrix T distributions 6.10

6.4 The inverted matrix T distribution 6.22

CHAPTER 7 THE MATRIX T AND MULTIVARIATE REGRESSION
7.1 Introduction 7.1
7.2 The distribution of the matrix of regression coefficients 7.2
7.3 The non-central distribution of the matrix of regression coefficients 7.8
7.4 The canonical form of the non-central distribution of the matrix of regression coefficients 7.10
7.5 The distribution of two-stage least-squares estimators 7.21

TABLE OF THE DISTRIBUTION OF THE 2SLS ESTIMATOR following page 7.41

CHAPTER 8 A BIVARIATE t DISTRIBUTION AND PARTITIONED HOTELLING'S T^2 TESTS
8.1 Introduction 8.1
8.2 Partitioned T^2 tests 8.3

TABLE OF PERCENTAGE POINTS OF HOTELLING'S T^2 DISTRIBUTION following page 8.11

BIBLIOGRAPHY

P R E F A C E

In accordance with the regulations for the Degree of Ph.D. from the University of Cape Town, the candidate presents a summary of the contents of the thesis indicating in what way they constitute a contribution to knowledge.

Chapter 1 reviews univariate non-central t distributions and three of the multivariate extensions that have been defined in the literature. No new results are given.

Chapter 2 comprises a number of well-known results that are required for developments in subsequent chapters. These have been included to avoid unnecessary deviations in proofs given later.

Chapter 3 discusses three non-central versions of the multivariate t distribution defined by Cornish (1954a) and Dunnett and Sobel (1954), namely the upper non-central t (Khsirsagar, 1960), the lower non-central t (Miller, 1968) and the doubly non-central t (Patil and Kovner 1969). We show that the multivariate lower non-central t distribution can be represented as an infinite series of central multivariate t densities each weighted by a term from the Poisson distribution. The doubly non-central t distribution is an extension of the bivariate distribution defined by Patil and Kovner (1969). Expressions for some moments are given in each case. Three non-central versions of the inverted multivariate t distribution, Raiffa and Schlaiffer (1961) are derived. In the final section we give a method for comparing a number of treatments with a control when the samples are non-homogeneous.

Chapter 4 discusses some quadratic forms of multivariate t distributions and we show that many of the familiar results on normal quadratic forms have their counterparts in quadratic forms of t variables, with the F distribution replacing the chi-squared distribution of the normal form. However the independence property associated with certain sets of normal forms no longer holds for the corresponding sets of t variables. The joint distribution of these forms is given. We also demonstrate the relationship between certain quadratic forms of t variables and the F tests of the Analysis of Variance. Most of the above results rest on the idempotency of certain matrices. In the last sections we consider the distribution of a quadratic form of t variables when the matrices are not idempotent.

Chapter 5 discusses the distribution of a "Wishart" matrix of t variables, $\tilde{T}$, and extends results given by Cornish (1954b) for the central case and Khsirshagar (1960) for the linear non-central case. The moments of the determinant $|\tilde{T}|$ are given and the distribution of the characteristic roots of $\tilde{T}$ in the central case.

Chapter 6 discusses the matrix T and inverted matrix T distribution as defined by Dickey (1967). A lower non-central matrix T distribution is defined and some of its properties are investigated. The distribution of the upper non-central matrix T and the non-central inverted matrix T are given in integral form.

Chapter 7. The matrix of sample regression coefficients $B = A_{11}^{-1}A_{12}$, where A is Wishart's matrix, has a matrix T distribution (Kshirsagar, 1960). We show that the matrix of partial regression coefficients $B_3 = A_{11.3}^{-1}A_{12.3}$ also has a matrix T distribution with degrees of freedom reduced by the number of variables held fixed. We next consider the distribution of B when A has a non-central Wishart distribution and show, for the linear case, that B has a lower non-central matrix T distribution if A_{11} has a non-central Wishart distribution, and another type of distribution otherwise. In the final section we show that in some cases the distribution of two-stage least squares estimators is lower non-central matrix T.

Chapter 8 compares two simultaneous multivariate procedures using Hotelling's T^2 statistics.

CHAPTER 1

INTRODUCTION

Tests based on the Student t-distribution must be among the most widely used (and misused) of all statistical procedures. As Rupert G. Miller remarks in his book, Simultaneous Statistical Inference, "Everyone is born knowing the t distribution." In spite of this, we begin with a brief review of Student's t and its three non-central forms as a background to the main topic of this thesis - multivariate non-central t distributions. There is no unique generalisation (central or non-central) of Student's t to the multivariate case. The different generalisations that have appeared in the literature are discussed in Johnson and Kotz (1972). In this chapter we review only the three generalisations which will be the subjects of further study in subsequent chapters.

1.1 THE UNIVARIATE STUDENT'S t DISTRIBUTION

If X is a normal variable with unit variance, is independent of the variable Z, which has a chi-squared distribution with n degrees of freedom, then the ratio

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X \qquad 1.1.1$$

has a t distribution with n degrees of freedom. If X has zero mean, and Z has a central chi-squared distribution, then the ratio 1.1.1 has a central Student's t distribution. The distribution was derived by

W.S. Gosset (1908) writing under the pen-name "Student" and its probability density function is

$$f(t) = \frac{\Gamma(\frac{1}{2}(n+1))}{(n\pi)^{\frac{1}{2}}\Gamma(\frac{1}{2}n)} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} \qquad -\infty < t < \infty$$

Since the synthetic representation of t given in 1.1.1 is a ratio, we see that there are various ways in which a non-central t distribution can occur. The non-centrality can enter the distribution through the numerator, if X has a non-central distribution; through the denominator, if Z has a non-central distribution, or through the numerator and denominator if both X and Z have non-central distributions.

If X has a non-zero mean, then the ratio 1.1.1 is usually said to have a non-central t distribution. We shall however call a distribution of this type an *upper non-central t distribution* since the non-centrality occurs in the numerator. An excellent review article on the upper non-central t distribution and its applications is given by Owen (1968). Tables of percentage points are given by Resnikoff and Lieberman (1957).

If X has zero mean and Z has a non-central chi-squared distribution, then we shall say that the ratio 1.1.1 has a *lower non-central t distribution,* since the non-centrality occurs in the denominator. The lower non-central t distribution was derived by Marakathavalli (1954) who showed that the density function can be expressed as an infinite series of central t densities each weighted by a term from the Poisson distribution.

In addition he provided a table of the upper 2,5% points for values of the non-centrality parameter $\lambda = 2(2)20$ and degrees of freedom $n = 6\ (1)20,30,40,\infty$. The lower non-central t distribution can be used to test the significance of the difference of two means when the samples are non-homogeneous. We shall consider a multivariate extension of this test in Chapter 3.

If both X and Z have non-central distributions, then the ratio 1.1.1 has a *doubly non-central distribution.* This distribution was derived by Robbins (1948). A doubly non-central t distribution arises when sampling from mixtures of normal populations or from a single population in which the mean exhibits a secular trend. Krishnan (1967) derived the moments of the doubly non-central t and gave two approximations to the density function based on moments. The approximations can be used to obtain percentage points of the distribution.

1.2 MULTIVARIATE GENERALISATIONS OF THE t DISTRIBUTION

Essentially a multivariate generalisation of the t distribution (central or non-central) would entail finding the joint distribution of a number of correlated variables each of which has a univariate t distribution. However a glance at the synthetic representation, $t = (\frac{Z}{n})^{-\frac{1}{2}}X$, shows us that several generalisations are possible. Since both X and Z can be replaced by some multivariate counterpart either separately or together. Even then all possible generalisations are not exhausted since if we think of a t statistic, apart from the constant term, as being the ratio of a mean, $\overline{X}$ say, to some estimator of a standard deviation S, i.e. $t \propto \overline{X}/S$, it can happen that S^2 is not a "chi-squared" estimator of the variance. The joint distribution of a number of such variables would have a type of multivariate t distribution but the marginal distribution would not be Student's t. Such a generalisation was considered by Patil and Liao (1970). As mentioned before the various generalisations are discussed in Johnson and Kotz (1972).

We shall only discuss three types of multivariate t distributions, two of which are equivalent although they arise from different synthetic representations and a third which is a multivariate t distribution that can be constructed from a sample from a multivariate normal population.

GENERALISATION 1

This multivariate t distribution is one which is obtained by replacing the univariate standard normal variable X in 1.1.1 by a $p\times 1$ vector of correlated normal variables, each with unit variance. The synthetic representation of this multivariate t is

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X \qquad 1.2.1$$

where t is now a $p\times 1$ vector. Z is a scalar random variable which has a chi-squared distribution with n degrees of freedom (abbreviated $Z \sim \chi^2_n$) and is independent of the $p\times 1$ vector X. X has a multivariate normal distribution with mean vector zero and correlation matrix R, where R is the $p\times p$ positive definite correlation matrix (abbreviated $X \sim N_p(0,R)$). In the central case, the parameters of the distribution are the $\frac{1}{2}p(p-1)$ correlation coefficients in R and the degrees of freedom of the common chi-squared variable Z. The marginal distributions of the elements of t are univariate Student t distributions in the central case.

This distribution was first derived by Cornish (1954a) and independently by Dunnett and Sobel (1954). A special case of the distribution which occurs when $R = I_p$ (i.e. all the correlations are zero) is called the multivariate t distribution with independent numerators. If in addition Z has only one degree of freedom, each element of t is the ratio of two independent standard normal variables and the joint distribution of $t_1,\ldots,t_p$ is the multivariate Cauchy distribution, a distribution which is remarkable because it has no moments of any order.

If Z has a non-central chi-squared distribution the random vector t defined in 1.2.1 has a lower non-central multivariate t, as defined by Miller (1968). If X has a non-zero mean, μ, say, we then have the upper non-central multivariate t of Kshirsagar (1960), who obtained the distribution of X/S where $X \sim N_p(\mu,\sigma^2 R)$ and S is such that $nS^2/\sigma^2 \sim \chi_n^2$. If both X and Z have non-central distributions, we have a doubly non-central multivariate t distribution. A bivariate distribution of this type was derived by Patil and Kovner (1969) and we extend this to the p variate case in Chapter 3.

A method for evaluating the probability integral for any dimension of the central case has been given by John (1961) but the actual evaluation for $p > 2$ is complicated, not only because of the difficulties of evaluating a multi-dimensional integral but also because the integral depends on the $\frac{1}{2}p(p-1)$ correlations, which makes the provision of tables of percentage points for any configuration of the correlation coefficients impracticable. Most tabulations have been confined to percentage points of the studentised maximum value

$$V = \max t_i ,$$

the studentised maximum modulus

$$U = \max |t_i| ,$$

and the studentised minimum modulus

$$W = \min |t_i|$$

where $t_1 \ldots t_p$ have a multivariate t-distribution with all the correlations equal. References to available tabulations are given in Dunn and Massey (1965) and more exactly in Johnson and Kotz (1972) p. 137-42. Steffens (1969) tabulated some powers of the studentised maximum modulus and studentised minimum modulus tests in the bivariate case.

The multivariate t distribution can be used to construct simultaneous confidence intervals for several types of problem. These include

(i) Mean values of normal populations, John (1961), Miller (1966), Dunn and Massey (1965).

(ii) Regression coefficients, John (1961), the regression equation at k points, Hahn and Hendrickson (1971), bands for the regression line over a finite range, Dunn (1968), uniform confidence bands for a quadratic regression line, Trout and Chow (1972).

(iii) Certain interactions in the two-way layout, Dunn and Massey (1965).

These specialised confidence intervals are shorter than those given by the more general S method of Scheffé (1953). Other applications include multiple comparisons of a treatment with a control, Dunnett (1955); the ranking of a number of populations according to their means, Bechofer, Dunnett and Sobel (1954); the construction of prediction intervals, Hahn (1969, 1970) and confidence bounds for a normal cumulative distribution, Kanofsky (1968).

We see that this generalisation of Student's t arises naturally in Regression and Analysis of Variance problems, since we can identify X with a vector of mean values or regression coefficients which have a multivariate normal distribution with common variance σ^2 and Z with nS^2/σ^2 where S^2 is an unbiased estimate of σ^2 with n degrees of freedom. In these applications the construction of the vector t will ensure that the central distribution is independent of the unknown variance σ^2, but of course still dependent upon the correlation structure of the underlying normal distribution.

GENERALISATION 2

A second generalisation of t occurs when the chi-squared variable Z is replaced by a $p \times p$ matrix U which has a Wishart distribution and X is a $p \times 1$ vector of standard normal variables. Thus the $p \times 1$ vector t has the synthetic representation

$$t = \sqrt{n}\ (U^{\frac{1}{2}'})^{-1}X \qquad 1.2.2$$

where $X \sim N_p(0,I)$ independently of U which has a Wishart distribution with covariance matrix $P = R^{-1}$ and $U^{\frac{1}{2}}$ is the symmetric square root of U. In Chapter 6 we shall show in detail that the density function of 1.2.2 is the same as that of Generalisation 1.

This representation of t arises in the Bayesian analysis of the multivariate normal distribution, Ando and Kaufmann (1965) Geisser and Cornfield (1963) and in the Bayesian analysis of the multivariate regression model, Taio and Zellner (1964).

Among others, the works of De Groot (1970) and Press (1972) discuss statistical analysis from a Bayesian viewpoint. Briefly, in the Bayesian analysis of a sample from a multivariate normal population, both the observations and the parameters μ and Σ^{-1} are assumed to be random variables. (Σ^{-1} is the inverse of the covariance matrix, and is called the precision matrix.) The experimenter's knowledge of the possible values of μ and Σ^{-1} before any observations are made, is expressed in the joint prior distribution of μ and Σ^{-1}. A sample is then drawn and the sample values are then combined with his prior knowledge (through the likelihood function and the prior distribution) to obtain the posterior distribution of μ and Σ^{-1}. The posterior distribution which expresses the experimenter's belief in the possible values of μ and Σ^{-1} in the light of the additional information yielded by the sample, is used to make inferences about μ and Σ^{-1}. If the prior joint distribution of μ and Σ^{-1} is Normal-Wishart, then both the marginal prior and posterior distributions of μ are multivariate t distributions (with different parameters), Ando and Kauffman (1965), De Groot (1970) page 179, Press (1972) page 169. If μ and Σ^{-1} have a diffuse or non-informative prior distribution, Geisser and Cornfield (1963), Press (1972), the marginal posterior distribution of μ is also multivariate t. In the Bayesian analysis of an m-equation multivariate regression model when the set of regression coefficients has a non-informative prior distribution, the marginal posterior distribution of the regression coefficient vector for any equation is multivariate t, Taio and Zellner (1964).

A natural extension of 1.2.2, is one which is obtained by replacing the vector X, by a $p \times q$ matrix, the rows of which are independently distributed as $N_q(0,Q)$. This is the matrix T distribution which was defined by Dickey (1967), and will be the main topic of Chapters 6 and 7, where we consider some non-central distributions.

GENERALISATION 3

In this generalisation, the numerators of the components of t are correlated normal variables and the denominators are the square roots of correlated Chi-squared variables. The numerators and denominators are independent.

Thus

$$t_i = \left(\frac{Z_i}{n}\right)^{-\frac{1}{2}} X_i \qquad i = 1,\ldots,p$$

where $X_1,\ldots,X_p$ have a joint normal distribution with mean $\mu_1,\ldots,\mu_p$. The variance of X_i is σ^2, $i = 1,\ldots,p$ and the correlation between X_i and X_j is ρ, $i,j = 1,\ldots,p$, $i \neq j$. $Z_1,\ldots,Z_p$ have a multivariate chi-squared distribution with n degrees of freedom and correlation ρ. The bivariate distribution of $t' = (t_1 t_2)'$ has been studied by Siddiqui (1967), Miller (1968) and Krishnan (1970, 1972). The distribution for $p > 2$ becomes complicated and appears to be unknown.

Multivariate t distributions of this type can be constructed from a sample from a multivariate normal population. Let (X_{1i},X_{2i}) $i = 1,\ldots,N$, be a random sample from a bivariate normal population with parameters

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Let
$$\bar{X}_1 = \frac{1}{N}\Sigma X_{1i} \qquad \bar{X}_2 = \frac{1}{N}\Sigma X_{2i}$$

$$S_1^2 = \frac{1}{N}\Sigma(X_{1i}-\bar{X}_1)^2 \qquad S_2^2 = \frac{1}{N}\Sigma(X_{2i}-\bar{X}_2)^2$$

and
$$t_1 = \frac{(N-1)^{\frac{1}{2}}\bar{X}_1}{S_1} \qquad t_2 = \frac{(N-1)^{\frac{1}{2}}\bar{X}_2}{S_2}$$

Then $t' = (t_1,t_2)'$ has the bivariate t distribution described above. Siddiqui (1967) obtained the joint distribution of t_1 and t_2 for $N = 2$ and an asymptotic expression for large N. Krishnan (1972) considered the upper-noncentral distribution and the doubly-noncentral distribution, Krishnan (1970). It is clear that the marginal distributions of t_1 and t_2 are univariate Student t (as are the marginal distributions of all the multivariate t distributions we have discussed). Jensen (1973) considered a slightly different generalisation of a t distribution of this type, where the univariate t statistics are replaced by Hotellings T^2 statistics, and used it to construct some tests of significance for partitioned normal mean vectors. These tests will be the topic of Chapter 8.

CHAPTER 2

FOLKLORE

This chapter comprises a number of well-known definitions and theorems that are used in subsequent chapters. For the most part the theorems are stated without proof.

2.1 MATRICES AND DETERMINANTS

Definition 2.1.1

Let A be a p×p matrix, then

A' is the transpose of A

A^{-1} is the inverse of A

$|A|$ is the determinant of A

$A^{\frac{1}{2}}$ is the symmetric square root of A

$\mathrm{tr}A$ is the trace of A

$\mathrm{etr}A = e^{\mathrm{tr}A}$

Theorem 2.1.1 *(Graybill (1969))*

Let A be a p×p real non-singular matrix, partitioned as follows:

$$A \quad \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{ij} *is a* $p_i \times p_j$ *matrix* $\quad i,j = 1,2$

and $p_1 + p_2 = p$

Let $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$

$A_{22\cdot 1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$

Then

$$|A| = |A_{11}|\ |A_{22\cdot 1}| = |A_{22}|\ |A_{11\cdot 2}|$$

and

$$A^{-1} = \begin{pmatrix} A_{11\cdot 2}^{-1} & -A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} & A_{22\cdot 1}^{-1} \end{pmatrix}$$

When considering certain integrals involving zonal polynomials it will be necessary to ensure that the matrices involved are symmetric. Hence we require:

Definition 2.1.2

The p×p matrix A is symmetric if A = A'.

Theorem 2.1.2

If B is any m×n matrix then the following matrices are symmetric,

$$BB';\quad B'B;\quad I_n \pm B'B;\quad I_m \pm BB';\quad S \pm BRB'$$

where I_p *is the identity matrix of order p*

and *R and S are both symmetric.*

Definition 2.1.3

Let A be a real, symmetric matrix, then A is positive definite (A>0) if and only if |A|>0 and all principal minors of A are positive.

It is well-known that a symmetric matrix can be represented as a linear combination of idempotent matrices; the coefficients of the combination being the distinct characteristic roots of the matrix. (Rao (1965) p.36) This representation is called the spectral decomposition of the matrix. The next two theorems show that certain non-symmetric matrices also have a spectral decomposition.

Theorem 2.1.3 (Rao 1965)

If A and B are p×p real symmetric matrices and B is positive definite, then there exists a non-singular matrix R such that

$$A = R'\Lambda R$$
$$B = R'R$$

where

$$\Lambda = diag(\lambda_1,\ldots\lambda_p)$$

and

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$$

are the roots of

$$|A-\lambda B| = 0$$

Since B is positive definite, it is also non-singular ($|B|>0$); hence using Theorem 2.1.3 we can write

$$AB^{-1} = R'\Lambda R(R'R)^{-1}$$
$$= R'\Lambda R'^{-1}$$

Let $a_1, a_2, \ldots a_r$ be the r distinct values of $\lambda_1,\ldots\lambda_p$. If C_j is a p×p diagonal matrix which has elements 1 where Λ has elements a_j and 0 otherwise then

$$\Lambda = \sum_{j=1}^{r} a_j C_j$$

and $$AB^{-1} = \sum_{j=1}^{r} a_j R'C_j R'^{-1}$$

$$= \sum_{j=1}^{r} a_j E_j$$

where $E_j = R'C_j R'^{-1}$

Now $E_j^2 = (R'C_j R'^{-1})(R'C_j R'^{-1}) = R'C_j R'^{-1} = E_j$

So E_j is idempotent, $j = 1 \ldots r$

Also $E_i E_j = 0$, $i \neq j$ since $C_i C_j = 0$.

Further, the rank of E_j = rank of $C_j = r_j$ where r_j is the multiplicity of a_j.

Letting $B^{-1} = V$, we can summarise the above in a theorem which will define the spectral decomposition of the non-symmetric matrix $W = AV$.

Theorem 2.1.4 (*Baldessari 1967*)

Let A and V ($V = B^{-1}$) *be two real symmetric p×p matrices where V is positive definite, then the matrix AV has the spectral decomposition*

$$AV = \sum_{j=1}^{r} a_j E_j$$

where

(*i*) $a_1, a_2 \ldots a_r$ *are the distinct roots of* $|AV - \lambda I| = 0$

(ii) E_j *is idempotent* $j = 1, \ldots\ldots r$

(iii) $E_i E_j = 0 \quad i \neq j$

(iv) *the rank of* E_j *is* r_j *where* r_j *is the multiplicity of* a_j .

Definition 2.1.4 The Symmetric Square Root of a Matrix.

Let A be a p×p symmetric positive definite matrix and let P be an orthogonal matrix such that

$$A = P\Lambda P'$$

where

$$\Lambda = \text{diag}\ (\lambda_1, \ldots \lambda_p)$$

Let $$\Lambda^{\frac{1}{2}} = \text{diag}\ (\lambda_1^{\frac{1}{2}}, \ldots \lambda_p^{\frac{1}{2}})$$

Then

$$A^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P'$$

is the symmetric square root of A.

2.2 THE MULTIVARIATE GAMMA FUNCTION

Definition 2.2.1

We define the univariate gamma function as

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx \qquad n > 0$$

A simple change of variable in definition 2.2.1 gives the useful integral,

Theorem 2.2.1

For $Q > 0$ and $n > 0$

$$\int_0^\infty x^{n-1} e^{-Qx} dx = \Gamma(n)/Q^n$$

Definition 2.2.2 The Multivariate Gamma Function.

The multivariate gamma function is defined as

$$\Gamma_p(\lambda) = \pi^{\frac{1}{4}p(p-1)} \Gamma(\lambda)\Gamma(\lambda - \tfrac{1}{2}) \ldots\ldots \Gamma(\lambda - \tfrac{p}{2} + \tfrac{1}{2})$$

$$= \pi^{\frac{1}{4}p(p-1)} \prod_{i=1}^{p} \Gamma(\lambda - \tfrac{i}{2} + \tfrac{1}{2})$$

Theorem 2.2.2

If $p = q + r$ and $n > p$ then

$$\frac{\Gamma_q(\frac{1}{2}(n-r))}{\Gamma_p(\frac{1}{2}n)} = \frac{1}{\pi^{\frac{1}{2}rq}\Gamma_r(\frac{1}{2}n)}$$

Proof

From the definition of the multivariate gamma function

$$\frac{\Gamma_q(\frac{1}{2}(n-r))}{\Gamma_p(\frac{1}{2}n)} = \frac{\pi^{\frac{1}{4}q(q-1)}}{\pi^{\frac{1}{4}p(p-1)}} \times \frac{\prod_{i=1}^{p} \Gamma(\frac{1}{2}(n-r) - \frac{1}{2}i + \frac{1}{2})}{\prod_{i=1}^{p} \Gamma(\frac{1}{2}n - \frac{1}{2}i + \frac{1}{2})}$$

Consider the denominator of the right-hand side:

Since $p = q + r$, $\pi^{\frac{1}{4}p(p-1)} = \pi^{\frac{1}{2}rq}\ \pi^{\frac{1}{4}r(r-1)}\ \pi^{\frac{1}{4}q(q-1)}$

and

$$\prod_{i=1}^{p} \Gamma(\tfrac{1}{2}n - \tfrac{1}{2}i + \tfrac{1}{2}) = \prod_{i=1}^{r} \Gamma(\tfrac{1}{2}n - \tfrac{1}{2}i + \tfrac{1}{2}) \prod_{i=r+1}^{q+r} \Gamma(\tfrac{1}{2}n - \tfrac{1}{2}i + \tfrac{1}{2})$$

$$= \prod_{i=1}^{r} \Gamma(\tfrac{1}{2}n - \tfrac{1}{2}i + \tfrac{1}{2}) \prod_{j=1}^{q} \Gamma(\tfrac{1}{2}n - \tfrac{1}{2}(r+j) + \tfrac{1}{2})$$

$$= \prod_{i=1}^{r} \Gamma(\tfrac{1}{2}n - \tfrac{1}{2}i + \tfrac{1}{2}) \prod_{j=1}^{q} \Gamma(\tfrac{1}{2}(n-r) - \tfrac{1}{2}j + \tfrac{1}{2})$$

Hence $\quad \Gamma_p(\tfrac{1}{2}n) = \pi^{\frac{1}{2}rq}\Gamma_r(\tfrac{1}{2}n)\Gamma_q(\tfrac{1}{2}(n-r))$

and the result follows.

2.3 THE GENERALISED HYPERGEOMETRIC FUNCTION AND ZONAL POLYNOMIALS

The generalised hypergeometric function plays an important role in the theory of non-central distributions. Many univariate non-central distributions, among them the non-central χ^2 and F distributions involve Bessel functions and hypergeometric functions which can be written as special cases of the generalised hypergeometric function.

Definition 2.3.1 (Rainville 1960)

The generalised hypergeometric function with scalar argument is defined as

$${}_pF_q(a_1,a_2,\ldots a_p;\ b_1,b_2,\ldots b_q;\ x) = \sum_{k=0}^{\infty} \frac{(a_1)_k\ldots(a_p)k}{(b_1)_k\ldots(b_q)k} \frac{x^k}{k!}$$

where $\quad (a)_k = a(a+1)\ldots.(a+k-1)$

$\quad\quad (a)_0 = 1$

${}_pF_q$ is a function of the real or complex variable x, depending on the real or complex numbers $a_1 \dots a_p$, $b_1 \dots b_q$. Throughout this thesis x and the parameters a_i and b_j will always be real.

We note the following results:

(i) No denominator parameter, b_j, is allowed to be zero or a negative integer.

(ii) If any numerator parameter a_i is zero or a negative integer the series terminates.

(iii) If $p \leq q$ the series converges for all finite x.

(iv) If $p=q+1$ the series converges for all $|x|<1$ and diverges for $|x|>1$.

(v) If $p>q+1$ the series diverges for $x \neq 0$.

Multivariate non-central distributions involve a generalisation of this function to the case where the variable x is replaced by a symmetric matrix S and ${}_pF_q$ is a real- or complex-valued symmetric function of the latent roots of S. The hypergeometric functions which appear in the non-central distributions of matrix variables were defined by Constantine (1963).

Definition 2.3.2 (Constantine 1963)

If S is an m×m symmetric matrix, the generalised hypergeometric function with matrix argument is

$${}_pF_q(a_1,\dots a_p;\, b_1 \dots b_q;\, S) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa \dots (a_p)_\kappa}{(b_1)_\kappa \dots (b_q)_\kappa} \frac{C_\kappa(S)}{k!}$$

where

$$(a)_\kappa = \prod_{i=1}^{m} (a - \tfrac{1}{2}(i-1))_{k_i}$$

$$(a)_{k_i} = a(a+1) \ldots (a + k_i - 1)$$

a_i and b_j are arbitrary complex numbers, subject only to the restriction that b_j is not an integer or half-integer less than or equal to $\frac{1}{2}(m-1)$.

$\kappa = (k_1, \ldots k_m)$ is a partition of k into not more than m parts such that $\Sigma k_i = k$ and $k_1 \geq k_2 \geq \ldots \geq k_m \geq 0$.

$C_\kappa(S)$ is a certain homogeneous symmetric polynomial known as a zonal polynomial (see Definition 2.3.3).

Note also that:

(i) If one of the a_i is a negative integer ($a_i = -n$, say) then for $k \geq mn + 1$ all the coefficients vanish, so the function reduces to a polynomial of degree mn.

(ii) If $p \leq q$ the series converges for all S.

(iii) If $p = q+1$ the series converges for $||S||<1$ where $||S||$ is the maximum of the absolute values of the characteristic roots of S.

(iv) If $p > q+r$ the series diverges for $||S|| \neq 0$.

(v) If a is such that the gamma functions exist then

$$(a)_\kappa = \Gamma_m(a,\kappa)/\Gamma_m(a)$$

where

(vi) $$\Gamma_m(a,\kappa) = \pi^{\frac{1}{4}m(m-1)} \prod_{i=1}^{m} \Gamma(a + k_i - \tfrac{1}{2}(i-1))$$

Using zonal polynomials the non-central distributions of many random variables can be expressed in a compact form. Zonal polynomials were first applied to statistical theory by James (1961, 1964) and Constantine (1963), and have become an important tool for research into the non-central distributions of statistics derived from an underlying normal population. We define a zonal polynomial and list only the properties we shall need in the sequel. For a fuller discussion the reader is referred to the articles cited above.

Let S be an m×m symmetric matrix and $\phi(S)$ a polynomial in the $\frac{1}{2}(m)(m+1)$ different elements of S.

Let V_k be the vector space of all homogeneous polynomials $\phi(S)$ of degree k.

V_k can be decomposed into a direct sum of irreducible invariant subspaces V_κ corresponding to each partition κ of k into not more than m parts, i.e.

$$V_k = \oplus V_\kappa$$

The polynomial $(\mathrm{tr}S)^k \varepsilon V_k$ then has the unique decomposition

$$(\mathrm{tr}S)^k = \sum_\kappa C_\kappa(S)$$

into polynomials $C_\kappa(S)\varepsilon V_\kappa$, belonging to the respective invariant subspaces.

<u>Definition 2.3.3</u>

The Zonal polynomial $C_\kappa(S)$ is defined as the component of $(\mathrm{tr}S)^k$ in the subspace V_κ.

We note that

(i) $C_\kappa(S)$ is a symmetric homogeneous polynomial of degree k in the latent roots of S;

(ii) If S is symmetric, and R positive definite then RS and $R^{\frac{1}{2}}SR^{\frac{1}{2}}$ have the same roots so

$$C_\kappa(RS) = C_\kappa(R^{\frac{1}{2}}SR^{\frac{1}{2}}) \quad ;$$

(iii) If a is any scalar, it follows from (i) that

$$C_\kappa(aS) = a^k C_\kappa(S) \quad .$$

The values of many integrals involving zonal polynomials have been given in the literature. We shall need only the following:

Theorem 2.3.1 *(Constantine 1963)*

Let R be a complex symmetric matrix whose real part is positive definite and let T be an arbitrary complex symmetric matrix. Then

$$\int_{S>0} etr(-RS)|S|^{t-\frac{1}{2}(m+1)} C_\kappa(TS)\,dS = \Gamma_m(t,\kappa)|R|^{-t} C_\kappa(TR^{-1})$$

the integration being over the space of positive definite m×m matrices, and valid for all complex numbers t satisfying $R(t) > \frac{1}{2}(m-1)$.

2.4 SOME UNIVARIATE PROBABILITY DISTRIBUTIONS

Definition 2.4.1 The Central χ^2 Distribution

Z has a central χ^2 distribution with n degrees of freedom (written $Z \sim \chi^2_n$) if the probability density function of Z is

$$f(z) = \frac{1}{2^{\frac{1}{2}n}\,\Gamma(\frac{1}{2}n)} z^{\frac{1}{2}n-1} e^{-\frac{1}{2}z} \qquad z>0$$

Theorem 2.4.1 *The Moments of the Central χ^2 Distribution*

If $Z \sim \chi^2_n$ then the r^{th} moment of Z is

$$E(Z^r) = \frac{2^r\Gamma(\frac{1}{2}n+r)}{\Gamma(\frac{1}{2}n)}$$

We shall be particularly interested in the "inverted" moments of the χ^2 distribution.

Definition 2.4.2

If X is any random variable, the k^{th} inverted moment of X is defined as

$$E(X^{-k}) \qquad k>o$$

provided the expectation exists.

Theorem 2.4.2 *The Inverted Moments of the Central χ^2 Distribution*

If $Z \sim \chi^2_n$ then the k^{th} inverted moment of Z is

$$E(Z^{-k}) = \frac{\Gamma(\frac{1}{2}n-k)}{2^k\Gamma(\frac{1}{2}n)} \qquad 2k<n$$

Proof

$$E(Z^{-k}) = \frac{1}{2(\frac{1}{2}n)\Gamma(\frac{1}{2}n)}\int_0^\infty z^{\frac{1}{2}(n-2k)-1}e^{-\frac{1}{2}z}dz$$

$$= \frac{2^{\frac{1}{2}(n-2k)}}{2^{\frac{1}{2}n}}\frac{\Gamma(\frac{1}{2}(n-2k))}{\Gamma(\frac{1}{2}n)} \qquad 2k<n$$

$$= \frac{1}{2^k}\frac{\Gamma(\frac{1}{2}n-k)}{\Gamma(\frac{1}{2}n)}$$

In particular if $k = \frac{1}{2}$ or $k = 1$ we have

Corollary 2.4.1

$$E(Z^{-\frac{1}{2}}) = \frac{\Gamma(\frac{1}{2}(n-1))}{2^{\frac{1}{2}}\Gamma(\frac{1}{2}n)}$$

$$E(Z^{-1}) = \frac{\Gamma(\frac{1}{2}(n-2))}{2\Gamma(\frac{1}{2}n)} = \frac{1}{n-2}$$

Definition 2.4.3 The Non-Central χ^2 Distribution

If Z has a non-central χ^2 distribution with n degrees of freedom and non-centrality parameter λ, (written $Z \sim \chi^2_n(\lambda)$), then

$$f(z) = \sum_{i=o}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{z^{\frac{1}{2}(n+2i)-1}e^{-\frac{1}{2}z}}{2^{\frac{1}{2}(n+2i)}\Gamma(\frac{1}{2}(n+2i))} \qquad z>o$$

We note

(i) f(z) can be represented as an infinite series of central χ^2 densities, each weighted by a term from the Poisson distribution. This follows because

$$\frac{e^{-\lambda}\lambda^i}{i!} = P(X=i)$$

where X has a Poisson distribution, and

$$\frac{z^{\frac{1}{2}(n+2i)-1}e^{-\frac{1}{2}z}}{2^{\frac{1}{2}(n+2i)}\Gamma(\frac{1}{2}(n+2i))}$$

is the density function of a χ^2_{n+2i} variable.

(ii) If $X_1, X_2, \ldots X_n$ are independent normal variables with unit variance and means $\mu_1, \mu_2, \ldots \mu_n$ respectively, then

$$Z = \sum_{i=1}^{n} X^2_i \sim \chi^2_n(\lambda)$$

where

$$\lambda = \tfrac{1}{2}\sum_{i=1}^{n} \mu^2_i$$

Theorem 2.4.3 *The Inverted Moments of the Non-Central χ^2 Distribution*

If $Z \sim \chi^2_n(\lambda)$ then

$$E(Z^{-k}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{\Gamma(\frac{1}{2}(n+2i-2k))}{2^k\Gamma(\frac{1}{2}(n+2i))} \qquad 2k<n$$

Proof:

$$E(Z^{-k}) = \int_0^{\infty} z^{-k}f(z)dz$$

Integrating term by term we obtain the result.

In particular if $k = \frac{1}{2}$, $k = 1$ we have

Corollary 2.4.2

If $Z \sim \chi^2_n(\lambda)$

$$E(Z^{-\frac{1}{2}}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{\Gamma(\frac{1}{2}(n+2i-1))}{2^{\frac{1}{2}}\Gamma(\frac{1}{2}(n+2i))}$$

$$E(Z^{-1}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{\Gamma(\frac{1}{2}(n+2i)-1)}{\Gamma(\frac{1}{2}(n+2i))}$$

$$= \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{1}{(n+2i-2)}$$

Definition 2.4.4 The Doubly Non-Central F Distribution

The random variable W has a doubly non-central F distribution if it is the ratio of two independent non-central χ^2 variables each divided by their degrees of freedom, i.e.

$$W = \frac{s}{r} \frac{\chi_r^2(\phi)}{\chi^2{}_s(\lambda)}$$

We refer to this distribution by $W \sim F(r,s,\phi,\lambda)$ and note

(i) If both non-centrality parameters are zero ($\lambda = o$ and $\phi = o$) then W has a central F distribution, written as

$$W \sim F(r,s)$$

(ii) If the denominator non-centrality parameter is zero ($x = o$) then W has an upper non-central F distribution, written as

$$W \sim F(r,s,\phi,o)$$

(This distribution is usually called the non-central F distribution).

(iii) If the numerator parameter is zero, then W has a lower non-central F distribution, written as

$$W \sim F(r,s,o,\lambda)$$

2.5 SOME MULTIVARIATE PROBABILITY DISTRIBUTIONS

Definition 2.5.1 **The Multivariate F (Inverted Dirichlet) Distribution (Press 1972)**

Let $Z_o, Z_1, \ldots Z_p$ be independent scalar random variables

where $\quad Z_i \sim \chi^2_{r_i} \qquad i = o \ldots p$

Let $\quad Y' = [Y_1, \ldots Y_p]'$

where $\quad Y_i = \dfrac{Z_i}{Z_o} \qquad i = 1 \ldots p$

Then the distribution of the random vector Y is multivariate F with density function

$$f(Y) = \frac{\Gamma(\frac{1}{2}(r_o+r+\ldots r_p))}{\Gamma(\frac{1}{2}r_o)\Gamma(\frac{1}{2}r\ )\ldots\Gamma(\frac{1}{2}r_p)} \; \frac{\prod_{i=1}^{p} y_i^{\frac{1}{2}r_i-1}}{\left(1+\sum_{1}^{p} y_i\right)^{\frac{1}{2}r}}$$

where $\quad r = \sum_{i=o}^{p} r_i$.

The multivariate F distribution is a special case of the inverted Dirichlet distribution (Tiao and Guttman 1965). The moments of Y are obtained by a straightforward integration.

Theorem 2.5.1 ***The Moments of Y***

If Y has a multivariate F distribution, then

$$E\left(y_1^{k_1} \, y_2^{k_2} \ldots \, y_p^{k_p}\right) = \frac{\Gamma(\frac{1}{2}r_o-k)\Gamma(\frac{1}{2}r_1+k_1)\ldots\Gamma(\frac{1}{2}r_p+k_p)}{\Gamma(\frac{1}{2}r_o)\Gamma(\frac{1}{2}r_1)\ldots\Gamma(\frac{1}{2}r_p)}$$

where $k = \sum_{i=1}^{p} k_i$ *and* $r_o > 2k$.

Corollary 2.5.1

$$E(Y_i) = \frac{r_i}{r_o - 1} \qquad i = 1 \ldots p$$

$$Var(Y_i) = \frac{2r_i(r_i + r_o - 2)}{(r_o - 2)^2(r_o - 4)} \qquad i = 1 \ldots p$$

$$Cov(Y_iY_j) = \frac{2r_ir_j}{(r_o - 2)^2(r_o - 4)} \qquad i,j = 1 \ldots p \quad i \neq j$$

Definition 2.5.2 The Multivariate Normal Distribution

The $p \times 1$ random vector X has a multivariate normal distribution with mean vector μ and covariance matrix $\Sigma > 0$ (written $X \sim N_p(\mu, \Sigma)$) if

$$f(X) = \frac{1}{(2\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\right)$$

where $-\infty < X < \infty$ $\quad -\infty < \mu < \infty$

Definition 2.5.3 The Matrix Normal Distribution

The $p \times n$ random matrix X has a matrix normal distribution if the columns of X are independent $N_p(\mu^{(i)}, \Sigma)$ vectors $i = 1 \ldots n$. The density function of X is

$$f(X) = \frac{1}{(2\pi)^{\frac{1}{2}pn}|\Sigma|^{\frac{1}{2}n}} \operatorname{etr}\left(-\tfrac{1}{2}\Sigma^{-1}(X-M)(X-M)'\right)$$

where $M = \left(\mu^{(1)} \ldots \mu^{(n)}\right)$ is the p×n matrix of mean vectors.

Definition 2.5.4 The Central Wishart Distribution

If the p×n matrix X has a matrix normal distribution with M = 0 and covariance matrix Σ, then the p×p matrix

$$A = XX'$$

has a central Wishart distribution with n degrees of freedom. (Written $A \sim W_p(\Sigma, n)$). The density function of A is

$$f(A) = \frac{|A|^{\frac{1}{2}(n-p-1)} \operatorname{etr}(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}np}\Gamma_p(\frac{1}{2}n)|\Sigma|^{\frac{1}{2}n}} \qquad A>0 \qquad \Sigma>0$$

Definition 2.5.5 The Non-Central Wishart Distribution (Constantine 1963)

If the p×n matrix X has a matrix normal distribution with M ≠ 0 and covariance matrix Σ, then the p×p matrix

$$A = XX'$$

has a non-central Wishart distribution with n degrees of freedom and covariance matrix Σ and non-centrality parameter matrix

$$\Omega = \tfrac{1}{2}MM'\Sigma^{-1}$$

This distribution is denoted by $A \sim W_p(\Sigma, n, \Omega)$ and the density function of A is

$$f(A) = \frac{\operatorname{etr}(-\Omega)|A|^{\frac{1}{2}(n-p-1)} \operatorname{etr}(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}np}\Gamma_p(\frac{1}{2}n)|\Sigma|^{\frac{1}{2}n}} \, {}_0F_1(\tfrac{1}{2}n;\ \tfrac{1}{2}\Sigma^{-1}\Omega A) \qquad \begin{matrix} A>0 \\ \Sigma>0 \end{matrix}$$

If $\Omega = 0$, then A has a central Wishart distribution. The rank of Ω determines the form of the non-central Wishart distribution more explicitly. If the rank of Ω is one, then A has a linear non-central Wishart distribution, so called because the n mean vectors of the underlying normal variables lie on a line in p dimensional space.

Definition 2.5.6 — The Canonical Form of the Linear Non-Central Wishart Distribution (Anderson 1946)

If A has a non-central Wishart distribution with n degrees of freedom, $\Sigma = I$ and the rank of Ω is one, then

$$f(A) = \frac{e^{-\frac{1}{2}\kappa^2}|A|^{\frac{1}{2}(n-p-1)}etr(-\frac{1}{2}A)}{2^{\frac{1}{2}np}\Gamma_p(\frac{1}{2}n)} \sum_{i=o}^{\infty} \frac{(\frac{1}{2}\kappa^2)^i a_{11}^i \Gamma(\frac{1}{2}n)}{i!2^i\Gamma(\frac{1}{2}n+i)}$$

where a_{11} is the (1,1)th element of A and κ^2 is the non-centrality parameter.

From the fact that $\int f(A)dA = 1$ we obtain the useful integral identity in Σ:

Theorem 2.5.2

$$\int_{A>0} |A|^{\frac{1}{2}(n-p-1)} etr(-\tfrac{1}{2}\Sigma^{-1}A)\ dA = 2^{\frac{1}{2}np}\Gamma_p(\tfrac{1}{2}n)|\Sigma|^{\frac{1}{2}n}$$

NOTATION

Throughout this thesis we shall have occasion to consider zero's, null vectors and matrices. In general

$\lambda = o$ means that the scalar λ is zero;

$a = 0$ means that all the components of the vector a are zero;

$A = 0$ means that all the elements of A are zero.

CHAPTER 3

THE MULTIVARIATE t DISTRIBUTION AND THE INVERTED t DISTRIBUTION

3.1 INTRODUCTION

In this chapter we consider in some detail the distribution of the p-dimensional random vector, with the synthetic representation

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X. \qquad 3.1$$

X has a p-dimensional multivariate normal distribution and is independent of the scalar Z, which has either a central or non-central chi-squared distribution with n degrees of freedom. This is the multivariate generalisation of Student's t discussed in Chapter 1 under the heading Generalisation 1 and is often referred to in the literature as "the" multivariate t distribution or the multivariate t distribution with common denominator. As remarked in Chapter 1, this distribution arises naturally when considering the problem of simultaneous statistical inference on sets of mean values or regression coefficients. In these applications, the covariance structure of X is I, R or $\sigma^2 R$ where R is the correlation matrix. A slightly more general definition of t, which includes the above-mentioned instances as special cases, is one that allows X to have any covariance structure Σ. In this case $t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X$ where $X \sim N_p(\mu,\Sigma)$ independently of $Z \sim \chi_n^2(\lambda)$. Although this distribution loses the convenient property of being independent of σ^2 in the central case, (except when $\Sigma = I$, R or $\sigma^2 R$) it will be the definition of t we shall adopt throughout the thesis, because in studying some general aspects of the t distribution, we feel that it is not necessary to restrict ourselves to a particular covariance structure for X. We do however require that Σ be positive definite so that the inverse is well defined.

We consider four cases of this distribution:

(i) the central distribution, Cornish (1954a), Dunnett and Sobel (1954), where

$$X \sim N_p(0,\Sigma),\ \Sigma > 0,\ \text{independently of } Z \sim \chi_n^2$$

(ii) the upper non-central distribution, Kshirsagar (1960), where

$$X \sim N_p(\mu,\Sigma),\ \Sigma > 0,\ \text{independently of } Z \sim \chi_n^2$$

(iii) the lower non-central distribution, Miller (1968), where

$$X \sim N_p(0,\Sigma),\ \Sigma > 0,\ \text{independently of } Z \sim \chi_n^2(\lambda)$$

(iv) the doubly non-central distribution, where

$$X \sim N_p(\mu,\Sigma),\ \Sigma > 0,\ \text{independently of } Z \sim \chi_n^2(\lambda).$$

A bivariate version of (iv) was discussed by Patil and Kovner (1969) and we give the distribution for the general case.

Although each of the distributions (i)-(iii) can be obtained from (iv) by setting the appropriate non-centrality parameters equal to zero, we derive each of the distributions separately for ease of reading. In addition, we also show that the marginal distribution of any sub-vector is a multivariate t distribution of the same kind and also give expressions for the mean vector and the expected value of the p×p matrix tt'.

In Section 3, we derive three non-central distributions of the Inverted Multivariate t distribution. Raiffa and Schlaiffer (1961) defined the central Inverted Multivariate t distribution as the distribution of the p×1 vector $r = 1 + t'\left(\frac{\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}} t$ where t has a central multivariate t distribution. We apply this transformation to the three non-central t distributions given in Section 2. In Section 4 we discuss an application of the lower non-central multivariate t distribution.

3.2 FOUR CASES OF THE MULTIVARIATE t-DISTRIBUTION

Theorem 3.2.1. The Central t Distribution (Cornish (1954a) Dunnett and Sobel (1954)).

If the p×1 random vector $X \sim N_p(0,\Sigma)$, $\Sigma>0$, independently of the scalar $Z \sim \chi^2_n$, then the density function of

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X$$

is given by

$$f(t) = \frac{\Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}\Gamma(\frac{1}{2}n)|\Sigma|^{\frac{1}{2}}} \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)} \qquad 3.2.1$$

where $-\infty<t<\infty$.

<u>Proof</u>: By definitions 2.5.2 and 2.4.1

$$f(x) = \frac{\exp(-\frac{1}{2}X'\Sigma^{-1}X)}{(2\pi)^{\frac{1}{2}p} \quad |\Sigma|^{\frac{1}{2}}}$$

$$f(z) = \frac{z^{\frac{1}{2}n-1} \; e^{-\frac{1}{2}z}}{2^{\frac{1}{2}n} \; \Gamma(\frac{1}{2}n)}$$

Since X and Z are independent,

$$f(X,z) \propto z^{\frac{1}{2}n-1} \exp(-\tfrac{1}{2}[z+X'\Sigma^{-1}X]) \qquad 3.2.2$$

where the constant of proportionality is

$$(2^{\frac{1}{2}(n+p)}\pi^{\frac{1}{2}p}\Gamma(\tfrac{1}{2}n) \; |\Sigma|^{\frac{1}{2}})^{-1} \qquad 3.2.3$$

Transforming from X to t in (3.2.2) by

$$X = \left(\frac{Z}{n}\right)^{\frac{1}{2}} t$$

with Jacobian, $J(X\to t) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p}$,

$$f(t,z) \propto n^{\frac{1}{2}p} \; z^{\frac{1}{2}(n+p)-1} \; \exp\left[-\tfrac{1}{2}\left(1+\frac{t'\Sigma^{-1}t}{n}\right) z\right] \qquad 3.2.4$$

To obtain f(t), 3.2.4 is integrated over z, and using Theorem 2.2.1,

$$\int_0^\infty f(t,z)dz \propto \frac{n^{-\frac{1}{2}p} \; 2^{\frac{1}{2}(n+p)} \; \Gamma(\frac{1}{2}(n+p))}{\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}(n+p)}} \qquad 3.2.5$$

Evaluating the constant term from 3.2.3 and 3.2.5, gives the density function 3.2.1.

Definition 3.2.1. If the p×1 vector t has the density function 3.2.1, then t has a multivariate central distribution with n degrees of freedom and parameter matrix Σ. This distribution will be denoted by $t_p(n;\Sigma)$.

To obtain the marginal distribution of any sub-vector of t, $t^{(1)}$ say, we recall a fundamental property of the normal distribution.

Suppose $X \sim N_p(\mu,\Sigma)$ and is partitioned into two sets of components

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} \qquad 3.2.6$$

where $X^{(1)}$ is q×1 and $X^{(2)}$ is r×1, q+r = p

$$\text{Let} \quad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad 3.2.7$$

be partitioned accordingly, so that $\mu^{(1)}$ is q×1, $\mu^{(2)}$ is r×1, Σ_{12} is q×r and Σ_{22} is r×r.

Then

$$X^{(1)} \sim N_q(\mu^{(1)}, \Sigma_{11}) \qquad 3.2.8$$

Let $t = \begin{bmatrix} t^{(1)} \\ t^{(2)} \end{bmatrix}$ where $t^{(1)}$ is q×1, $t^{(2)}$ is r×1

$$\text{Then} \quad t^{(1)} = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X^{(1)} \qquad 3.2.9$$

where $X^{(1)} \sim N_q(0,\Sigma_{11})$ and $Z \sim \chi_n^2$ independently of $X^{(1)}$.

Applying the same derivation as in Theorem 3.2.1 to $t^{(1)}$ gives

Theorem 3.2.2. *Cornish (1954a).*

Let $t = \begin{bmatrix} t^{(1)} \\ t^{(2)} \end{bmatrix} \begin{matrix} q\times 1 \\ r\times 1 \end{matrix}$ *be distributed as* $t_p(n,\Sigma)$.

Then the marginal distribution of $t^{(1)}$ *is*

$$f(t^{(1)}) = \frac{\Gamma(\frac{1}{2}(n+q))}{(n\pi)^{\frac{1}{2}q}\ \Gamma(\frac{1}{2}n)\,|\Sigma_{11}|^{\frac{1}{2}n}} \cdot\cdot \left\{1+t^{(1)'}\Sigma_{11}^{-1}t^{(1)}\right\}^{-\frac{1}{2}(n+q)}$$

3.2.10

Since the labelling of the components of t is arbitrary the marginal distribution of any sub-vector of t is again multivariate t. If q = 1 and $\Sigma_{11} = \sigma_{11} = 1$, then the marginal distribution of the single element t_1 is univariate Student's t. It can also be shown (Cornish 1954a) that the conditional distribution of $t^{(1)}$, given that $t^{(2)} = a$, is also multivariate t.

The moments of t can be found from the synthetic representation of t. In particular, the mean and covariance matrix of t are given by the next theorem.

Theorem 3.2.3. *Cornish (1954a)*

If $t \sim t_p(n,\Sigma)$ *then the mean vector of t is*

$$E(t) = 0 \qquad 3.2.12$$

and the covariance matrix of t is

$$E(tt') = \frac{n}{n-2}\,\Sigma\,. \qquad 3.2.13$$

Proof: $E(t) = E\left[\left(\frac{Z}{n}\right)^{-\frac{1}{2}} X\right]$

$= E\left[\left(\frac{Z}{n}\right)^{-\frac{1}{2}}\right] E(X)$ since X and Z are independent

$= 0$ since $E(X) = 0$

Since $E(t) = 0$ the covariance matrix of t is

$$E(tt') = E\left[\left(\frac{Z}{n}\right)^{-1}\right] E(XX')$$

$$= n\, E(Z^{-1})\, E(XX')$$

By Corollary 2.4.1, $E(Z^{-1}) = (n-2)^{-1}$.

Therefore $E(tt') = \frac{n}{n-2}\,\Sigma$

If $\Sigma = I$ in theorem 3.2.3, all the covariance terms are zero and so the elements of t are uncorrelated. However this does not imply that the elements of t are independent, since setting $\Sigma = I$ in the density function of t, we see that it does not factorize into a product of the marginal densities. So the multivariate t distribution is an example of a distribution in which zero correlation does not imply independence. The reason for this is, of course, the fact that the elements of t are bound together by the common Z.

Theorem 3.2.4. If $t \sim t_p(n,\Sigma)$ then the mean and modal vectors coincide.

Proof: The mode of t is the point at which the density function is a maximum. From Theorem 3.2.1,

$$f(t) \propto \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)}$$

Hence f(t) is maximised, when $(1 + \frac{t'\Sigma^{-1}t}{n})$ is a minimum. Since Σ^{-1} is positive definite, $t'\Sigma^{-1}t \geq 0$ for all t and is only zero if t is the null vector. But t = 0 is also E(t) - hence the mean and mode coincide.

Corollary 3.2.1. *E(t) satisfies the set of partial differential equations*

$$\frac{\partial f(t)}{\partial t_i} = 0 \qquad i = 1 \cdots p$$

where f(t) is the density function of t.

Corollary 3.2.2. *E(t) satisfies the set of partial differential equations*

$$\frac{\partial Q}{\partial t_i} = 0 \qquad i = 1 \cdots p$$

where $Q = t'\Sigma^{-1}t$.

If the mean vector of X is not zero, then t has an upper non-central multivariate t distribution. We derive this distribution in the next theorem.

Theorem 3.2.5. *The Upper Non-Central t Distribution Kshirsagar (1960).*

If the p×1 vector $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, *independently of* $Z \sim \chi^2_n$ *then the density function of*

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X$$

is given by

$$f(t) = \frac{\exp(-\frac{1}{2}\mu'\Sigma^{-1}\mu)}{(n\pi)^{\frac{1}{2}p}\Gamma(\frac{1}{2}n)|\Sigma|^{\frac{1}{2}}} \sum_{k=0}^{\infty} \frac{2^{\frac{1}{2}k}\Gamma(\frac{1}{2}(n+p+k))}{n^{\frac{1}{2}k}\, k!} \frac{(t'\Sigma^{-1}\mu)^k}{\left(1+\frac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}(n+p+k)}} \qquad 3.2.14$$

where $-\infty < t < \infty$.

<u>Proof</u>: By definition 2.5.2

$$f(X) = \frac{1}{(2\pi)^{\frac{1}{2}p}\ |\Sigma|^{\frac{1}{2}}}\ \exp\ (-\tfrac{1}{2}[(X-\mu)'\Sigma^{-1}(X-\mu)])$$

Expanding the quadratic form and rearranging

$$f(X) = \frac{\exp(-\frac{1}{2}\mu'\Sigma^{-1}\mu)}{(2\pi)^{\frac{1}{2}p}\ |\Sigma|^{\frac{1}{2}}}\ \exp(-\tfrac{1}{2}[(X'\Sigma^{-1}X-2X'\Sigma^{-1}\mu]) \qquad 3.2.15$$

From Definition 2.4.1

$$f(z) = \frac{z^{\frac{1}{2}n-1}\ e^{-\frac{1}{2}z}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} \qquad 3.2.16$$

So

$$f(x,z) \propto z^{\frac{1}{2}n-1}\ \exp(-\tfrac{1}{2}[z+X'\Sigma^{-1}X - 2X'\Sigma^{-1}\mu]) \qquad 3.2.17$$

where the constant of proportionality is

$$\frac{\exp(-\frac{1}{2}\mu'\Sigma^{-1}\mu)}{2^{\frac{1}{2}(n+p)}\ \pi^{\frac{1}{2}p}\ \Gamma(\frac{1}{2}n)\ |\Sigma|^{\frac{1}{2}}} \qquad 3.2.18$$

Transform from X to t in 3.2.17 by

$$X = \left(\frac{Z}{n}\right)^{\frac{1}{2}} t \quad \text{with} \quad J(X\to t) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p}$$

Then
$$f(t,z) \propto \frac{z^{\frac{1}{2}(n+p)-1}}{n^{\frac{1}{2}p}}\ \exp\left(-\tfrac{1}{2}\left[z+\frac{z}{n}\ t'\Sigma^{-1}t\ -\ 2\left(\frac{z}{n}\right)^{\frac{1}{2}} t'\Sigma^{-1}\mu\right]\right) \qquad 3.2.19$$

The exponent in 3.2.19 can be written as

$$\exp(-\tfrac{1}{2}\left[1\ +\ \frac{t'\Sigma^{-1}t}{n}\right]\ z)\ \exp\left[\left(\frac{z}{n}\right)^{\frac{1}{2}} t'\Sigma^{-1}\mu\right] \qquad 3.2.20$$

The second term of 3.2.20 can be expanded in a power series as

$$\exp\left\{\left(\frac{z}{n}\right)^{\frac{1}{2}} t'\Sigma^{-1}\mu\right\} = \sum_{k=0}^{\infty} \frac{(t'\Sigma^{-1}\mu)^k}{n^{\frac{1}{2}k}} \quad \frac{z^{\frac{1}{2}k}}{k!} \qquad 3.2.21$$

So 3.2.19 becomes

$$f(t,z) \propto \sum_{k=0}^{\infty} \frac{(t'\Sigma^{-1}\mu)^k}{n^{\frac{1}{2}(p+k)}\ k!} \quad z^{\frac{1}{2}(n+p+k)-1} \exp\left(-\tfrac{1}{2}\left\{1 + \frac{t'\Sigma^{-1}t}{n}\right\}z\right)$$

3.2.22

To obtain f(t) we integrate 3.2.22 termwise over z. By Theorem 2.2.1,

$$\int_0^{\infty} z^{\frac{1}{2}(n+p+k)-1} \quad \exp\left(-\tfrac{1}{2}\left\{1 + \frac{t'\Sigma^{-1}t}{n}\right\} z\right)dz$$

$$= \frac{2^{\frac{1}{2}(n+p+k)} \quad \Gamma(\frac{1}{2}(n+p+k))}{\left\{1 + \frac{t'\Sigma^{-1}t}{n}\right\}^{\frac{1}{2}(n+p+k)}} \qquad 3.2.23$$

Evaluating the constant terms from 3.2.18 and 3.2.23, we obtain the density function given in 3.2.14.

<u>Definition 3.2.2.</u> If the p×1 vector t has the density function 3.2.14, then t has a multivariate upper non-central t distribution with n degrees of freedeom, parameter matrix Σ and non-centrality parameter μ. This distribution will be denoted by $t \sim t_p(n,\Sigma,\mu)$.

Kshirsagar (1960) found the density function by considering the distribution of $\underset{\sim}{t} = X/s$ where $X \sim N_k\ (\mu, \sigma^2 R)$ (R is the correlation matrix) and $u = fs^2/\sigma'^2 \sim \chi^2_f$ independently of X. If in 3.2.14 we make the substitions $p = k$, $n = f$, $\Sigma = \sigma^2 R$, $k = \alpha$ and change the variable t to

Proof:

$$E(t) = n^{\frac{1}{2}} E(Z^{-\frac{1}{2}})E(X).$$

By Corollary 2.4.1, $E(Z^{-\frac{1}{2}}) = \dfrac{\Gamma(\frac{1}{2}(n-1))}{2^{\frac{1}{2}}\,\Gamma(\frac{1}{2}n)}$

and $E(X) = \mu$. Hence Eq. 3.2.24 follows.

$$E(tt') = n\,E(Z^{-1})E(XX')$$

By Corollary 2.4.1, $E(Z^{-1}) = (n-2)^{-1}$ and $E(XX') = \Sigma+\mu\mu'$. Hence Eq. 3.2.25 follows.

If Z has a non-central chi-squared distribution and X has zero mean, then t has a lower non-central multivariate t distribution. The density function is given in the next theorem.

Theorem 3.2.8. The Lower Non-Central t Distribution Miller (1968).

If the p×1 vector $X \sim N_p(0,\Sigma), \Sigma > 0$, independently of $Z \sim \chi_n^2(\lambda)$, then the density function of

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X$$

is given by

$$f(t) = \frac{e^{-\lambda}}{(n\pi)^{\frac{1}{2}p}\,|\Sigma|^{\frac{1}{2}}} \sum_{i=0}^{\infty} \frac{\lambda^i\,\Gamma(\frac{1}{2}(n+2i+p))}{i!\,\Gamma(\frac{1}{2}(n+2i))} \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+2i+p)} \qquad 3.2.26$$

Proof: From definition 2.5.2

$$f(X) = \frac{1}{(2\pi)^{\frac{1}{2}p}\,|\Sigma|^{\frac{1}{2}}} \exp(-\tfrac{1}{2}X'\Sigma^{-1}X)$$

and from definition 2.4.3

$$f(z) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} \frac{z^{\frac{1}{2}(n+2i)-1} \quad e^{-\frac{1}{2}z}}{2^{\frac{1}{2}(n+2i)} \quad \Gamma(\frac{1}{2}(n+2i))}$$

$$f(x,z) \propto \sum_{i=0}^{\infty} \frac{\lambda^{i}}{i!} \frac{z^{\frac{1}{2}(n+2i)-1} \quad \exp(-\frac{1}{2}[z+X'\Sigma^{-1}X])}{2^{\frac{1}{2}(n+2i)} \quad \Gamma(\frac{1}{2}(n+2i))} \qquad 3.2.27$$

where the constant of proportionality is

$$\frac{e^{-\lambda}}{(2\pi)^{\frac{1}{2}p}\,|\Sigma|^{\frac{1}{2}}} \qquad 3.2.28$$

Transform from X to t in 3.2.27 by

$$X = \left(\frac{Z}{n}\right)^{\frac{1}{2}} t \quad \text{with} \quad J(X\to t) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p}$$

Then

$$f(t,z) \propto \sum_{i=0}^{\infty} \frac{\lambda^{i} \quad z^{\frac{1}{2}(n+p+2i)-1} \quad \exp(-\frac{1}{2}[1+\frac{t'\Sigma^{-1}t}{n}]z)}{i! \quad 2^{\frac{1}{2}(n+2i)} \quad n^{\frac{1}{2}p} \quad \Gamma(\frac{1}{2}(n+2i))}$$

3.2.29

To obtain f(t) we integrate 3.2.29 termwise over z by Theorem 2.2.1.

$$\int_0^{\infty} z^{\frac{1}{2}(n+p+2i)-1} \exp\left\{-\frac{1}{2}\left(1+\frac{t'\Sigma^{-1}t}{n}\right) z\right\} dz$$

$$= \frac{2^{\frac{1}{2}(n+p+2i)} \quad \Gamma(\frac{1}{2}(n+2i+p))}{\left(1+\frac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}(n+2i+p)}}$$

and so

$$f(t) \propto \sum_{i=0}^{\infty} \left(\frac{2}{n}\right)^{\frac{1}{2}p} \lambda^{i} \frac{\Gamma(\frac{1}{2}(n+2i+p))}{\Gamma(\frac{1}{2}n)} \left(1+\frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+2i+p)}$$

3.2.30

Evaluating the constant terms from 3.2.28 and 3.2.30 we obtain the density function given by Eq. 3.2.26.

The density function of the lower non-central t can be written in terms of the hypergeometric function (see definition 2.3.1). Writing $f(t) = \dfrac{\Gamma(\frac{1}{2}(n+p))\ \Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n)\ \ \ \Gamma(\frac{1}{2}(n+p))}\ f(t)$ and noting that

$$\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+2i+p)} = \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)} \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-i},$$

Eq. 3.2.26 can be expressed as

$$f(t) = \frac{e^{-\lambda}\ \ \Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{1}{2}n)} \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)}$$

3.2.31

$$\times \sum_{i=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n+p)+i)\ \Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}(n+p))\ \Gamma(\frac{1}{2}n+i)} \left\{\lambda\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-1}\right\}^{i} / i!$$

The infinite series in 3.2.31 is ${}_1F_1\left(\frac{1}{2}(n+p);\frac{1}{2}n;\lambda\left(1+\frac{t'\Sigma^{-1}t}{n}\right)^{-1}\right)$ and so

$$f(t) = \frac{e^{-\lambda}\ \ \Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{1}{2}n)} \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)}$$

$$\times\ {}_1F_1\left(\tfrac{1}{2}(n+p);\ \tfrac{1}{2}n;\ \lambda\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-1}\right)$$

3.2.32

<u>Definition 3.2.3.</u> If the $p\times 1$ vector t has the density function 3.2.26 or 3.2.32 then t has a multivariate lower non-central t distribution with n degrees of freedom, parameter matrix Σ and non-centrality parameter λ. This distribution will be denoted by $t \sim t_p(n,\Sigma,0,\lambda)$.

We noted in definition 2.4.3 that the non-central χ^2 density can be expressed as an infinite series of central χ^2 densities each weighted by a term from the Poisson distribution. In the next theorem we shall show that an analogous result holds for the lower non-central multivariate t distribution.

<u>*Theorem 3.2.9.*</u> *If* $t \sim t_p(n,\Sigma,0,\lambda)$ *then*

$$f(t) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} f(t;\ n+2i,\ \frac{n}{n+2i}\Sigma\ ) \qquad 3.2.33$$

where $f(t;\ m,\phi)$ *is the density function of the central multivariate t with m degrees of freedom and parameter matrix* ϕ.

<u>Proof:</u> From Eq. 3.2.26, the density function of the lower non-central t is

$$f(t) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i} \left[\left(\frac{\Gamma(\frac{1}{2}(n+2i+p))}{(n\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{1}{2}(n+2i))} \right) \left(1+ \frac{t'\Sigma^{-1}t}{n} \right)^{-\frac{1}{2}(n+2i+p)} \right]$$

The term in square brackets can be written as

$$\frac{\Gamma(\frac{1}{2}(n+2i+p))}{((n+2i)\pi)^{\frac{1}{2}p}\left(\frac{n}{n+2i}\right)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{1}{2}(n+2i))} \left[1 + \left(\frac{n+2i}{n}\right) \frac{t'\Sigma' t}{(n+2i)} \right]^{-\frac{1}{2}(n+2i+p)}$$

3.2.34

Noting that $\left(\frac{n}{n+2i}\right)^{\frac{1}{2}p} |\Sigma|^{\frac{1}{2}} = \left| \frac{n}{n+2i}\Sigma \right|^{\frac{1}{2}}$

we recognise 3.2.34 as the density function of a central multivariate t distribution with n+2i degrees of freedom and parameter matrix $\frac{n}{n+2i}\Sigma$. Hence the density function of $t_p(n,\Sigma,0,\lambda)$ can be written as Eq. 3.2.33.

As in Theorems 3.2.2 and 3.2.6, we can deduce that the marginal distribution of any subvector of t, $t^{(1)}$ say, is again lower non-central multivariate t. The density function of $t^{(1)}$ can be obtained from Eq. 3.2.26 by replacing Σ by Σ_{11}, and p by q, where q is the dimension of $t^{(1)}$. When q = 1, the distribution reduces to a univariate lower non-central t, as studied by Marakathavalli (1954).

The moments of the distribution can be found either from the synthetic representation of t or using the form of the density function given in Theorem 3.2.9. In particular we have:

Theorem 3.2.10. *If $t \sim t_p(n,\Sigma,0,\lambda)$ then*

$$E(t) = 0$$

$$\text{and} \quad E(tt') = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{n}{(n+2i-2)} \Sigma \qquad 3.2.35$$

Proof: If $t \sim t_p(n,\Sigma,0,\lambda)$ then

$$E(t) = E\left\{\left(\frac{Z}{n}\right)^{-\frac{1}{2}}\right\} E(X)$$

where $X \sim N(0,\Sigma)$. Hence $E(t) = 0$

$$E(tt') = E\left\{\left(\frac{Z}{n}\right)^{-1}\right\} E(XX').$$

From Corollary 2.4.2, $E\left\{\left(\frac{Z}{n}\right)^{-1}\right\} = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{n}{(n+2i-2)}$

and $E(XX') = \Sigma$ since $E(X) = 0$.

Hence $E(tt') = \sum_{i=o}^{\infty} \frac{e^{-\lambda} \lambda^{i}}{i} \frac{n}{(n+2i-2)} \Sigma.$

Alternatively, by Theorem 3.2.9, f(t) can be written as an infinite series of central t densities, each weighted by a term from the Poisson distribution. Hence expected values of t can be expressed as an infinite series of expected values of central t densities each weighted by a term from the Poisson distribution. By Theorem 3.2.3, if $t \sim t_p(n+2i, \frac{n}{n+2i} \Sigma)$ then

$$E(tt') = \frac{n+2i}{n+2i-2} \left(\frac{n}{n+2i} \Sigma \right) = \frac{n}{n+2i-2} \Sigma .$$

So, for the lower non-central t,

$$E(tt') = \sum_{i=o}^{\infty} \frac{e^{-\lambda} \lambda^{i}}{i!} \frac{n}{n+2i-2} \Sigma .$$

We now consider the doubly non-central multivariate t distribution which arises if both X and Z have non-central distributions. Patil and Kovner (1969) gave the density function for p = 2 and we now derive the distribution for the general case.

Theorem 3.2.11. *The Doubly-Non-Central t Distribution*

If the p×1 vector $X \sim N_p(\mu,\Sigma)$, $\Sigma > 0$, independently of $Z \sim \chi^2_n(\lambda)$ then the density function of

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X$$

is given by

$$f(t) = \frac{exp(-\frac{1}{2}[\mu'\Sigma^{-1}\mu+2\lambda])}{(n\pi)^{\frac{1}{2}p} \quad |\Sigma|^{\frac{1}{2}}} \sum_{i=o}^{\infty} \sum_{k=o}^{\infty} \frac{\lambda^{i}}{i!} \left(\frac{2}{n}\right)^{\frac{1}{2}k} \frac{\Gamma(\frac{1}{2}(n+2i+p+k))}{\Gamma(\frac{1}{2}(n+2i)}$$

$$\times \quad \frac{(t'\Sigma^{-1}\mu)^{k}}{k!} \quad \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+2i+p+k)} \qquad 3.2.36$$

Proof: From definition 2.5.2 and Eq. 3.2.15

$$f(X) = \frac{\exp(-\frac{1}{2}\mu'\Sigma^{-1}\mu)}{(2\pi)^{\frac{1}{2}p} \quad |\Sigma|^{\frac{1}{2}}} \quad \exp\left(-\frac{1}{2}[X'\Sigma^{-1}X - 2X'\Sigma^{-1}\mu]\right)$$

3.2.37

and from definition 2.4.3

$$f(z) = \sum_{i=o}^{\infty} \frac{e^{-\lambda} \lambda^{i}}{i!} \frac{z^{\frac{1}{2}(n+2i)-1} e^{-\frac{1}{2}z}}{2^{\frac{1}{2}(n+2i)} \Gamma(\frac{1}{2}(n+2i))}$$

So $f(X,z) \propto \sum_{i=o}^{\infty} \phi(i) \quad z^{\frac{1}{2}(n+2i)-1} \quad \exp(-\frac{1}{2}[z+X'\Sigma^{-1}X])$

$$\times \exp(X'\Sigma^{-1}\mu) \qquad 3.2.38$$

where the constant of proportionality is

$$\frac{\exp(-\frac{1}{2}\mu'\Sigma^{-1}\mu)}{(2\pi)^{\frac{1}{2}p} \quad |\Sigma|^{\frac{1}{2}}} \qquad 3.2.39$$

and

$$\phi(i) = \frac{e^{-\lambda} \lambda^{i}}{i! \; 2^{\frac{1}{2}(n+2i)} \; \Gamma(\frac{1}{2}(n+2i)} \qquad 3.2.40$$

Transform from X to t in 3.2.38 by

$$X = \left(\frac{Z}{n}\right)^{\frac{1}{2}} t \quad \text{with} \quad J(X\to t) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p}$$

Then

$$f(t,z) \propto \sum_{i=o}^{\infty} \phi(i)\ z^{\frac{1}{2}(n+2i+p)-1} \exp\left[-\tfrac{1}{2}\left(1+ \frac{t'\Sigma^{-1}t}{n}\right) z\right]$$

$$\times \exp\left\{\left(\frac{z}{n}\right)^{\frac{1}{2}} t'\Sigma^{-1}\mu\right\} \qquad 3.2.41$$

The second exponential term in 3.2.41 can be expanded as the power series

$$\exp\left[\left(\frac{z}{n}\right)^{\frac{1}{2}} t'\Sigma^{-1}\mu\right] = \sum_{k=o}^{\infty} \frac{(t'\Sigma^{-1}\mu)^k}{n^{\frac{1}{2}k}} \frac{z^{\frac{1}{2}k}}{k!} \qquad 3.2.42$$

Substituting in 3.2.41 we have

$$f(t,z) \propto \sum_{i=o}^{\infty} \sum_{k=o}^{\infty} \phi(i) \frac{(t'\Sigma^{-1}\mu)^k}{n^{\frac{1}{2}k}k!}\ z^{\frac{1}{2}(n+2i+p+k)-1} \exp\left[-\tfrac{1}{2}\left(1+ \frac{t'\Sigma^{-1}t}{n}\right) z\right]$$

To obtain f(t) we integrate 3.2.43 termwise over z.

By Theorem 2.2.1

$$\int_0^{\infty} z^{\frac{1}{2}(n+2i+p+k)-1} \exp\left[-\tfrac{1}{2}\left(1+\frac{t'\Sigma^{-1}t}{n}\right)z\right] dz$$

$$= \frac{2^{\frac{1}{2}(n+2i+p+k)}\ \Gamma(\frac{1}{2}(n+2i+p+k))}{\left(1+\frac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}(n+2i+p+k)}} \qquad 3.2.44$$

and

$$f(t) \propto \sum_{i=o}^{\infty} \sum_{k=o}^{\infty} \phi(i) \frac{(t'\Sigma^{-1}\mu)^k\, 2^{\frac{1}{2}(n+2i+p+k)}\ \Gamma(\frac{1}{2}(n+2i+p+k))}{n^{\frac{1}{2}k}k! \left(1+\frac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}(n+2i+p+k)}}$$

3.2.45

Substituting for $\phi(i)$ from Eq. 3.2.40 and the constant of proportionality given in Eq. 3.2.39 we obtain the density function given in Eq. 3.2.36.

Definition 3.2.4. If the $p\times 1$ vector t has the density function 3.2.36, then t has a multivariate doubly non-central t distribution with n degrees of freedom, parameter matrix Σ and noncentrality parameters μ and λ. The distribution will be referred to as $t_p(n,\Sigma,\mu,\lambda)$.

The density function 3.2.36 reduces to that of the multivariate

(i)	central t if $\mu = 0$ and $\lambda = o$	(Theorem 3.2.1)
(ii)	upper non-central t if $\lambda = o$	(Theorem 3.2.5)
(iii)	lower non-central t if $\mu = 0$	(Theorem 3.2.8)

As in the other cases, the marginal distribution of any subvector of t, $t^{(1)}$ say, can be deduced from the synthetic representation. The density function has the same form as 3.2.36 with μ replaced by $\mu^{(1)}$, Σ by Σ_{11} and p by q where q is the dimension of $t^{(1)}$. If $q = 1$ and $\Sigma_{11} = \sigma_{11} = 1$, the marginal distribution of t_1 is doubly non-central univariate Student's t (Krishnan (1967)).

The moments of the distribution can be deduced from the synthetic representation. In particular, we have

Theorem 3.2.12.

If $t \sim t_p(n,\Sigma,\mu,\lambda)$ *then*

$$E(t) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{\Gamma(\frac{1}{2}(n+2i-1))}{\Gamma(\frac{1}{2}(n+2i))} \left(\frac{n}{2}\right)^{\frac{1}{2}} \mu \qquad 3.2.46$$

$$E(tt') = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{n}{(n+2i-2)} [\Sigma+\mu\mu'] \qquad 3.2.47$$

Proof: Since X and Z are independent

$$E(t) = E\left(\left(\frac{Z}{n}\right)^{-\frac{1}{2}}\right) E(X)$$

By Corollary 2.4.2, $E(Z^{-\frac{1}{2}}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i \, \Gamma(\frac{1}{2}(n+2i-1))}{i! \;\; 2^{\frac{1}{2}} \, \Gamma(\frac{1}{2}(n+2i))}$

and $E(X) = \mu$

Hence Eq. 3.2.46 follows

$$E(tt') = E\left(\left(\frac{Z}{n}\right)^{-1}\right) E(XX')$$

By Corollary 2.4.2, $E(Z^{-1}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{1}{(n+2i-2)}$

and $E(XX') = \Sigma+\mu\mu'$.

Hence Eq. 3.2.47 follows.

3.3 THE INVERTED MULTIVARIATE t-DISTRIBUTION

Raiffa and Schlaiffer (1961) page 259, defined the inverted multivariate t distribution as the distribution of the $p\times 1$ vector

$$r = \frac{t}{\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}}} \qquad 3.3.1$$

where t has a central multivariate t distribution with parameter matrix Σ, $\Sigma > 0$, and n degrees of freedom. The elements of r all lie between -1 and 1, and r may be considered to be a multivariate generalisation of the scalar sample correlation from two independent univariate normal populations. The distribution also plays an important role in the Bayesian approach to experimental design (preposterior analysis, see Raiffa and Schlaiffer (1961)). If preposterior analysis is applied to the multivariate normal distribution when both the mean vector and precision matrix are unknown, and have a joint Normal-Wishart prior distribution, then the preposterior distribution of the sample mean vector is inverted multivariate t. (Ando and Kaufman (1965)).

In this section, we consider the distribution of r when the transformation 3.3.1 is applied to the non-central t distributions derived in section 3.2.

Theorem 3.3.1. *The Central Inverted Multivariate t distribution*

(Raiffa and Schlaiffer (1961))

If $t \sim t_p(n,\Sigma)$ *and*

$$r = \frac{t}{\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}}}$$

then $$f(r) = \frac{\Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}\,\Gamma(\frac{1}{2}n)\,|\Sigma|^{\frac{1}{2}}}\left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n-1} \qquad 3.3.2$$

where $-1 < r < 1$ *and* $r'\Sigma^{-1}r < n$.

<u>Proof</u>: If $t \sim t_p(n,\Sigma)$, then

$$f(t) = \frac{\Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}\ \Gamma(\frac{1}{2}n)\ |\Sigma|^{\frac{1}{2}}} \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)} \qquad 3.3.3$$

If $r = \dfrac{t}{\left(1 + \dfrac{t'\Sigma^{-1}t}{n}\right)^{\frac{1}{2}}}$

then

$$t = \frac{r}{\left(1 - \dfrac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}}} \qquad 3.3.4$$

and $$\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-1} = \left(1 - \frac{r'\Sigma^{-1}r}{n}\right) \qquad 3.3.5$$

The Jacobian of the transformation from t to r is (Raiffa ans Schlaiffer (1961) page 260)

$$J(t \to r) = \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{-\frac{1}{2}p-1} \qquad 3.3.6$$

Transforming f(t) using the above equations gives f(r) as stated in 3.3.2. The condition $r'\Sigma^{-1}r < n$ ensures that the density function is non-negative.

<u>*Theorem 3.3.2.*</u> <u>*The Inverted Upper Non-Central Multivariate t.*</u>

If the p×1 vector $t \sim t_p(n,\Sigma,\mu)$ and r is as defined in Eq. 3.3.1 then

$$f(r) = \frac{exp(-\frac{1}{2}\mu'\Sigma^{-1}\mu)}{(n\pi)^{\frac{1}{2}p}\Gamma(\frac{1}{2}n)|\Sigma|^{\frac{1}{2}}} \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n-1} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n+p+k)}{k!} \left(\frac{2^{\frac{1}{2}} r'\Sigma^{-1}\mu}{n^{\frac{1}{2}}}\right)^{k}$$

for $r'\Sigma^{-1}r < n$ and $-1 < r < -1$.

<u>Proof</u>: From Theorem 3.2.5 and Eq. 3.2.14

$$f(t) \propto \sum_{k=o}^{\infty} \phi(k)\ (t'\Sigma^{-1}\mu)^k \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p+k)} \qquad 3.3.7$$

where the constant of proportionality is

$$\frac{\exp(-\frac{1}{2}\mu'\Sigma^{-1}\mu)}{(n\pi)^{\frac{1}{2}p}\ \Gamma(\frac{1}{2}n)\ |\Sigma|^{\frac{1}{2}}} \qquad 3.3.8$$

and

$$\phi(k) = \left(\frac{2}{n}\right)^{\frac{1}{2}k} \frac{\Gamma(\frac{1}{2}(n+p+k))}{k!} \qquad 3.3.9$$

Transforming in f(t) using Eq. 3.3.4 and 3.3.5 we find that

$$(t'\Sigma^{-1}\mu)^k \text{ becomes } (r'\Sigma^{-1}\mu)^k \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{-\frac{1}{2}k}$$

$$\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p+k)} \text{ becomes } \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}(n+p+k)}$$

and the Jacobian is given by 3.3.6.

Hence

$$f(r) \propto \sum_{k=o}^{\infty} \phi(k)\ (r'\Sigma^{-1}\mu)^k \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n-1} \qquad 3.3.10$$

The Ratio test shows that the series converges, since taking the limit of the ratio of the (k+1)th and k^{th} term as $k \to \infty$ we have

$$\lim_{k\to\infty} \frac{\phi(k+1)}{\phi(k)} \frac{(r'\Sigma^{-1}\mu)^{k+1}}{(r'\Sigma^{-1}\mu)^k}$$

$$= \left(\frac{2}{n}\right)^{\frac{1}{2}} r'\Sigma^{-1}\mu \lim_{k\to\infty} \frac{\Gamma(\frac{1}{2}(n+p+k)+\frac{1}{2})}{\Gamma(\frac{1}{2}(n+p+k))} \frac{1}{k+1} \qquad 3.3.11$$

Using Stirling's approximation, it is easily shown that for large N, $\Gamma(N+\frac{1}{2})/\Gamma(N)$ is approximately $\sqrt{N}$. Hence the limit of 3.3.11 is zero and the series is absolutely convergent and thus convergent.

Substituting for $\phi(k)$ and the constant of proportionality gives the density function 3.3.6.

Theorem 3.3.3. *The Inverted Lower Non-Central Multivariate t*

Let $t \sim t_p(n,\Sigma,0,\lambda)$ and r be defined as in Eq. 3.3.1 then

$$f(r) = \frac{e^{-\lambda}}{(n\pi)^{\frac{1}{2}p}\;|\Sigma|^{\frac{1}{2}}} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{\Gamma(\frac{1}{2}(n+p+2i))}{\Gamma(\frac{1}{2}(n+2i))} \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n+i-1} \qquad 3.3.12$$

where $r'\Sigma^{-1}r < n$ and $-1 < r < 1$.

Proof: Since t has a lower non-central t distribution

$$f(t) \propto \sum_{i=0}^{\infty} \phi(i) \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+2i+p)} \qquad 3.3.13$$

$$\text{where } \phi(i) = \frac{\lambda^i}{i!} \frac{\Gamma(\frac{1}{2}(n+2i+p))}{\Gamma(\frac{1}{2}(n+2i))} \qquad 3.3.14$$

and the constant of proportionality is

$$\frac{e^{-\lambda}}{(2\pi)^{\frac{1}{2}p}\;|\Sigma|^{\frac{1}{2}}} \qquad 3.3.15$$

Transforming in f(t) using Eqs. 3.3.4 and 3.3.5 with Jacobian given by 3.3.6, we find

$$f(r) \propto \sum_{i=0}^{\infty} \phi(i) \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n+i-1}$$

Substituting for the constant term and $\phi(i)$ from Eqs. 3.3.14 and 3.3.15 gives the density function given in 3.3.12.

With some rearrangement, the density function of the lower non-central inverted t can be written in terms of the hypergeometric series. Rewriting 3.3.12, we have

$$f(r) = \frac{e^{-\lambda}}{(n\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}} \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n-1} \sum_{i=o}^{\infty} \frac{\Gamma(\frac{1}{2}(n+p)+i)}{\Gamma(\frac{1}{2}n+i)} \left\{\lambda\left(1-\frac{r'\Sigma^{-1}r}{n}\right)\right\}^{i} \frac{1}{i!}$$

3.3.16

The infinite series in 3.3.16 can be written as

$$\frac{\Gamma(\frac{1}{2}(n+p))}{\Gamma(\frac{1}{2}n)} \sum_{i=o}^{\infty} \frac{\Gamma(\frac{1}{2}(n+p)+i)}{\Gamma(\frac{1}{2}(n+p))} \quad \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n+i)} \left\{\lambda \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)\right\}^{i} \frac{1}{i!}$$

3.3.17

The infinite series in 3.3.16 is ${}_1F_1\left\{\frac{1}{2}(n+p);\ \frac{1}{2}n;\ \lambda\left(1- \frac{r'\Sigma^{-1}r}{n}\right)\right\}$

and so

$$f(r) = \frac{e^{-\lambda}}{(n\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}} \frac{\Gamma(\frac{1}{2}(n+p))}{\Gamma(\frac{1}{2}n)} \left(1- \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n-1} {}_1F_1\left\{\frac{1}{2}(n\ p);\frac{1}{2}n;\lambda\left(1-\frac{r'\Sigma^{-1}r}{n}\right)\right\}$$

3.3.18

As with the lower non-central t, the density function of the inverted lower non-central t can be written as an infinite series of inverted central t densities, each weighted by a term from the Poisson distribution.

Theorem 3.3.4.

If the p×1 vector r has an inverted lower non-central multivariate t distribution then

$$f(r) = \sum_{i=o}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} g(r;\ n+2i,\ \frac{n}{n+2i}\Sigma) \qquad 3.3.19$$

where g(r; m, ϕ) is the density function of a central inverted multivariate t, with m degrees of freedom and parameter matrix ϕ.

Proof:

From Theorem 3.3.1, if r has central inverted multivariate t distribution with (n+2i) degrees of freedom and parameter ϕ, then

$$g(r;n+2i,\phi) = \frac{\Gamma(\frac{1}{2}(n+2i+p))}{((n+2i)\pi)^{\frac{1}{2}p}\ \Gamma(\frac{1}{2}(n+2i))|\phi|^{\frac{1}{2}}} \left(1 - \frac{r'\phi^{-1}r}{n+2i}\right)^{\frac{1}{2}(n+2i)-1}$$

3.3.20

From Theorem 3.3.3,

$$f(r) = \sum_{i=o}^{\infty} \frac{e^{-\lambda}\lambda^{i}\ \Gamma(\frac{1}{2}(n+2i+p)}{i!\ (n\pi)^{\frac{1}{2}p}\ (\frac{1}{2}(n+2i))|\Sigma|^{\frac{1}{2}}} \left(1 - \frac{r'\Sigma^{-1}r}{r}\right)^{\frac{1}{2}(n+2i)-1}$$

3.3.21

$$\text{Now } (n\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}} = ((n+2i)\pi)^{\frac{1}{2}p} \left(\frac{n}{n+2i}\right)^{\frac{1}{2}p} |\Sigma|^{\frac{1}{2}}$$

$$= ((n+2i)\pi)^{\frac{1}{2}p} \left|\frac{n}{n+2i}\Sigma\right|^{\frac{1}{2}} \qquad 3.3.22$$

Also $\frac{r'\Sigma^{-1}r}{n} = \left(\frac{n+2i}{n}\right) \frac{r'\Sigma^{-1}r}{n+2i}$ 3.3.23

Substituting 3.3.22 and 3.3.23 in (3.3.21) and comparing with 3.3.20, we see that

$$f(r) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} g(r; n+2i; \frac{n}{n+2i}\Sigma)$$

Theorem 3.3.5. The Doubly Non-Central Inverted Multivariate t.

If the p×1 vector $t \sim t_p(n,\Sigma,\mu,\lambda)$ and r is as defined in Eq. 3.1, then

$$f(r) = \frac{\exp(-\frac{1}{2}(\mu'\Sigma^{-1}\mu+2\lambda))}{(n\pi)^{\frac{1}{2}p} \quad |\Sigma|^{\frac{1}{2}}} \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \frac{\lambda^i}{i!} \frac{\Gamma(\frac{1}{2}(n+2i+p+k))}{\Gamma(\frac{1}{2}(n+2i))}$$

3.3.24

$$\times \left(\frac{2}{n}\right)^{\frac{1}{2}k} \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n+i-1} \frac{(r'\Sigma^{-1}\mu)^k}{k!}$$

where $-1 < r < 1$ and $r'\Sigma^{-1}r < n$.

Proof: From Theorem 3.2.11, if t has a doubly non-central t distribution then

$$f(t) \propto \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \phi(i,k) \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+2i+p+k)} \frac{(t'\Sigma^{-1}\mu)^k}{k!}$$

3.3.25

where

$$\phi(i,k) = \frac{\lambda^i}{i!} \left(\frac{2}{n}\right)^{\frac{1}{2}k} \frac{\Gamma(\frac{1}{2}(n+p+2i+k))}{\Gamma(\frac{1}{2}(n+2i))}$$ 3.3.26

and the constant of proportionality is

$$\frac{\exp(-\frac{1}{2}[\mu'\Sigma^{-1}\mu+2\lambda])}{(n\pi)^{\frac{1}{2}p}\qquad|\Sigma|^{\frac{1}{2}}} \qquad 3.3.27$$

Transforming in f(t) by using Eqs. 3.3.4 and 3.3.5

$$(t'\Sigma^{-1}\mu)^k \text{ becomes } (r'\Sigma^{-1}\mu)^k \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{-\frac{1}{2}k}$$

$$\left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+2i+p+k)} \text{ becomes } \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}(n+2i+p+k)}$$

and the Jacobian is given by 3.3.6.

Hence

$$f(r) \propto \sum_{i=o}^{\infty} \sum_{k=o}^{\infty} \phi(i,k)(r'\Sigma^{-1}\mu)^k \left(1 - \frac{r'\Sigma^{-1}r}{n}\right)^{\frac{1}{2}n+i-1}$$

Substituting for $\phi(i)$ and the constant term gives the density function 3.3.24.

3.4 AN APPLICATION OF THE LOWER NON-CENTRAL MULTIVARIATE t DISTRIBUTION

Suppose an experiment is conducted to compare p treatment means with a single control mean. Independent random samples, each of size n, are drawn from the treatments and the control and an appropriate measurement is made on each unit. The measurements are assumed to be normally distributed and all measurements have a common variance σ^2. The usual procedure that the experimenter follows is that given by Dunnett (1955,1964) and Miller (1966) page 76.

Let Y_{ij} , $i = 0,1, \cdots p$, $j = 1,2 \cdots n$ be the observations from the p+1 samples, where i = o labels the observations on the central and i = 1 ·· p the observations on the treatments. The p hypotheses to be tested are

$$H_o: \ \mu_i - \mu_o = 0 \quad i = 1 \cdots p \qquad 3.4.1$$

against the alternative

$$H: \ \mu_i \neq \mu_o \qquad i = 1 \cdots p$$

where μ_i i = 1 ·· p denotes the i^{th} population mean and μ_o denotes the control mean.

The population means are estimated by the sample means $\overline{Y}_i$, i = 0,1, ··· p and the variance σ^2 is estimated by

$$s^2 = \frac{1}{(p+1)(n-1)} \sum_{ij} (Y_{ij} - \overline{Y}_i)^2 \qquad 3.4.2$$

where

$$s^2 \sim \frac{\sigma^2 \chi^2{}_\nu}{\nu} \quad \text{with } \nu = (p+1)(n-1).$$

The natural statistic for comparing the i^{th} treatment mean with the control is

$$t_i = \frac{(\overline{Y}_i - \overline{Y}_o) - (\mu_i - \mu_o)}{s\sqrt{2/n}} \qquad i = 1,2, \cdots p \qquad 3.4.3$$

Individually each of these statistics has a t distribution with ν degrees of freedom. It is easily shown that their joint distribution is central multivariate t under H_o.

Let

$$X_i = \frac{(\bar{Y}_i - \bar{Y}_o) - (\mu_i - \mu_o)}{\sqrt{2/n}} \qquad i = 1 \cdots p \qquad 3.4.4$$

Then $X_i \sim N(0,\sigma^2)$. The correlation between any pair X_i, X_j is $\rho_{ij} = \frac{1}{2}$. Thus $X' = (X_1, \cdots X_p) \sim N_p(0,\sigma^2 R)$ where

$$R = \begin{bmatrix} 1 & & & \frac{1}{2} \\ & \cdot & & \\ & & \cdot & \\ \frac{1}{2} & & & 1 \end{bmatrix} \qquad 3.4.5$$

Then $t_i = \frac{X_i}{\sigma} \Big/ \frac{\sigma}{s} = \frac{X_i}{s}$ and

$t' = (t_1, \cdots t_p)'$ has a multivariate t distribution with ν degrees of freedom and parameter matrix R.

To compare a single treatment mean with the control the critical point for H_o would be $t_{\frac{1}{2}\alpha,\nu}$ - the upper $100\alpha/2$ percentage point of the univariate t distribution with ν degrees of freedom. However, if the p comparisons are grouped into a family, the critical point must be increased to achieve a Type I probability error rate of α. (A Type I probability error rate of α means that the probability of erroneously rejecting one or more of the hypotheses when all of them are true is α). The appropriate critical point in this case is the upper 100α percentage point of the distribution of

$$|d|_{p,\nu} = \max_{1 \leq i \leq p} |t_i| \qquad 3.4.6$$

where $t = (t_1 \cdots t_p) \sim t_p(R,\nu)$.

Tabulations of the critical points are given by Dunnett (1964). The appropriate confidence intervals for the difference $\mu_i - \mu_o$ are

$$\overline{Y}_i - \overline{Y}_o \pm |d|_{\alpha,p,\nu} \quad s\left(\frac{2}{n}\right)^{\frac{1}{2}} \qquad 3.4.7$$

The above is the standard procedure for comparing a number of treatment means with a control. Let us now suppose that two measuring instruments are used to obtain the observations and in each sample m measurements are made with the first instrument and the remaining (n-m) with the second. Suppose it is found by some independent means that the second instrument has a systematic bias, δ , which affects the true mean of the observations. The variance is unaffected by the bias and is still σ^2 for all the observations. Thus the i^{th} sample contains m observations with mean μ_i and (n-m) with mean $(\mu_i+\delta)$. If the experimenter has noted the instrument used to obtain each observation, he could separate the observations and analyse each set separately.

Suppose however he does not do this and only knows that each of the p+1 samples contains m measurements made with the first instrument and (n-m) with the second. Unfortunately, in this case, since his samples are not homogeneous, the technique outlined above cannot be used to compare the treatments with the control. However, if the experimenter knows the ratio δ^2/σ^2 , an attack can be made upon the problem.

Consider the i^{th} sample. Since the expected values of the observations are independent of the labelling, we can assume without loss of generality that the first m observations were made with the first instrument and the remaining (n-m) with the second, although this fact will be unknown to the experimenter. Thus if Y_{ij} $j = 1 \cdots n$ are the observations in the i^{th} sample,

$$Y_{ij} \sim N_1(\mu_i, \sigma^2) \quad \text{for } j = 1 \cdots m$$

$$\sim N_1(\mu_i + \delta, \sigma^2) \quad j = m+1, \cdots n.$$

Then $$E(\bar{Y}_i) = \frac{1}{n} \sum_{j=1}^{n} Y_{ij}$$

$$= \frac{1}{n} (m\mu_i + (n-m)(\mu_i + \delta))$$

$$= \mu_i + \frac{(n-m)}{n} \delta \quad i = 1, \cdots p \qquad 3.4.8$$

$$Var(\bar{Y}_i) = \frac{\sigma^2}{n} \qquad 3.4.9$$

since the variance is not affected by the bias.

Let $(Y_{ij} - \bar{Y}_i)$ be the deviation of the j^{th} observation from the mean.

Then, for $j = 1, \cdots m$;

$$E(Y_{ij} - \bar{Y}_i) = \mu_i - \left[\mu_i + \frac{(n-m)}{n}\delta\right]$$

$$= - \frac{(n-m)}{n} \delta \qquad 3.4.10$$

and for $j = m+1, \cdots n$

$$E(Y_{ij} - \bar{Y}_i) = \mu_i + \delta - \left[\mu_i + \frac{n-m}{n} \delta\right]$$

$$= \frac{m}{n} \delta \qquad 3.4.11$$

Let $$s_i^2 = \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_i)^2$$

Then from 3.4.10 and 3.4.11,

$$s_i^2 \sim \sigma^2 \chi^2_{n-1} (\lambda_i) \qquad 3.4.12$$

The non-centrality parameter is given by

$$\lambda_i = \frac{\delta^2}{2\sigma^2} \left[m \left(\frac{n-m}{n}\right)^2 + (n-m) \left(\frac{m}{n}\right)^2 \right]$$

$$= \frac{\delta^2}{2\sigma^2} \frac{m(n-m)}{n} \qquad 3.4.13$$

The unknown variance σ^2 is estimated by

$$s^2 = \frac{1}{(p+1)(n-1)} \sum_{i=o}^{p} \sum_{j=o}^{n} (Y_{ij} - \overline{Y}_i)^2$$

$$= \frac{1}{(p+1)(n-1)} \sum_{i=o}^{p} s_i^2 . \qquad 3.4.14$$

Since the (p+1) samples are independent,

$$\frac{\nu s^2}{\sigma^2} \sim \chi^2_\nu (\lambda) \qquad 3.4.15$$

where $\nu = (p+1)(n-1)$ 3.4.16

and

$$\lambda = \sum_{i=o}^{p} \lambda_i = \frac{(p+1)\delta^2 \; m(n-m)}{2\sigma^2 n} \qquad 3.4.17$$

Consider now the statistic used for comparing the i^{th} treatment with the control, namely,

$$t_i = \frac{(\overline{Y}_i - \overline{Y}_o) - (\mu_i - \mu_o)}{s\sqrt{2/n}} \qquad i = 1 \cdots p$$

From 3.4.8, it follows that

$$E(\overline{Y}_i - \overline{Y}_o) = \left(\mu_i + \frac{(n-m)}{n} \delta\right) - \left(\mu_o + \frac{(n-m)}{n} \delta\right)$$

$$= \mu_i - \mu_o$$

and we see that the numerator of t_i is not affected by the biased measurements. From 3.4.9 and the independence of the samples

$$\mathrm{Var}(\bar{Y}_i - \bar{Y}_o) = \frac{2\sigma^2}{n} \qquad i = 1 \cdots p.$$

From 3.4.15

$$\frac{s^2}{\sigma^2} \sim \frac{\chi^2{}_\nu(\lambda)}{\nu}$$

and we see that s^2 is affected by the biased measurements through the non-centrality parameter λ. Hence t_i has a univariate lower non-central t distribution (Marakathavalli (1954)) with ν degrees of freedom and non-centrality parameter λ as defined in 3.4.17. The joint distribution of $t' = (t_1, \cdots t_p)'$ is lower non-central multivariate t with non-centrality parameter λ and parameter matrix R as defined in 3.4.5. Thus the critical point for the family of p comparisons with the control is $|d'|_{\alpha,p,\nu}$ which is the upper 100α per cent point of the distribution of

$$|d'|_{p,\nu} = \max_{1 \leq i \leq p} |t_i|_\nu$$

where $t' = (t_1, \cdots t_p) \sim t_p(\nu, R, 0, \lambda)$.

Unfortunately, this distribution has not been tabulated. However, simultaneous confidence intervals for $\mu_i - \mu_o$ could be obtained using the Bonferroni Inequality. In this case, the intervals would be

$$\bar{Y}_i - \bar{Y}_o \pm t^* s \left(\frac{2}{n}\right)^{\frac{1}{2}} \qquad i = 1 \cdots p \qquad 3.4.18$$

where t^* is upper $100\left(\frac{\alpha}{2p}\right)$ per cent point of the univariate lower non-central t distribution with ν degrees of freedom and non-centrality parameter λ. If $p = 1$, the single interval is that given by Marakathavalli (1954) for comparing two means when the samples are non-homogeneous.

A tabulation of the 97,5% points of t^* is given in Marakathavalli (1954) and we include them here as the source is not readily obtainable. The tabulation is for $\nu = 6(1)20, 30, 40, \infty$ and $2\lambda = 0(2)20$ (Marakathavalli's λ is twice ours). An inspection of the table shows that the intervals given by 3.4.18 are shorter than those that would be obtained if we ignored the fact that the samples were non-homogeneous and used the central t distribution. The reason for this is clear, since to apply 3.4.18, we need to know the value of the non-centrality parameter λ, which is a function of δ/σ. Thus when using 3.4.18 we have more information about the samples than when we use the central t distribution and this is reflected by the shorter confidence intervals.

Since λ depends on δ/σ, this means that the experimenter must know the ratio of the bias to the true standard deviation. If this is not known precisely, a discrepancy enters the value of λ. However, in the experimental situation outlined above, it is quite likely that the degrees of freedom $\nu = (p+1)(n-1)$ will be fairly large and an inspection of the table shows that the critical point becomes less sensitive to λ as the degrees of freedom increase.

The 97,5 percentile points of the univariate lower non-central t distribution. Reproduced from Marakathavalli (1954).

TABLE IV

The 5% points of the t_1 distribution with n degrees of freedom and parameter λ

$n \backslash \lambda$	0	2	4	6	8	10	12	14	16	18	20
6.	2·447	2·099	1·844	1·652	1·503	1·385	1·288	1·208	1·141	1·083	1·033
7.	2·365	2·072	1·850	1·678	1·541	1·431	1·338	1·261	1·195	1·138	1·083
8.	2·306	2·053	1·856	1·700	1·574	1·469	1·381	1·307	1·242	1·186	1·137
9.	2·262	2·039	1·862	1·719	1·601	1·502	1·418	1·346	1·283	1·228	1·179
10.	2·228	2·029	1·867	1·735	1·624	1·530	1·450	1·380	1·319	1 265	1·217
11.	2·201	2·020	1·872	1·749	1·645	1·555	1·478	1·411	1·352	1 299	1·251
12.	2·179	2·014	1·877	1·761	1·663	1·578	1·503	1·438	1·380	1·329	1·282
13.	2·160	2·009	1·881	1·772	1·679	1·597	1·526	1·463	1·406	1·356	1·310
14.	2·145	2·004	1·885	1·782	1·693	1·615	1·546	1·485	1·430	1·381	1·336
15.	2·132	2·000	1·888	1·791	1·706	1·631	1·565	1·505	1·452	1·403	1·360
16.	2·120	1·997	1·891	1·799	1·717	1·646	1·581	1·524	1·472	1·424	1·381
17.	2·110	1·994	1·894	1·806	1·728	1·659	1·597	1·541	1·490	1·440	1·401
18.	2·101	1·992	1·897	1·813	1·738	1·671	1·611	1·556	1·507	1·461	1·420
19.	2·093	1·990	1·899	1·819	1·747	1·682	1·624	1·571	1·522	1·478	1·437
20.	2·086	1·988	1·901	1·824	1·755	1·692	1·636	1·584	1·537	1 494	1·454
30.	2·042	1·977	1·917	1·862	1·812	1·764	1·721	1·680	1·642	1 606	1·573
40.	2·021	1·972	1·927	1·834	1·844	1·806	1·770	1·736	1·704	1 674	1·645
∞	1·960	1·960	1·960	1·960	1·960	1·960	1·960	1·960	1·960	1 960	1·960

CHAPTER 4

QUADRATIC FORMS OF t VARIABLES

4.1 INTRODUCTION

In this chapter we consider the distribution of the quadratic form.

$$Q = \frac{1}{c} t'At$$

where t is a random vector with a multivariate t distribution as defined in Chapter 3; A is a symmetric matrix and c is a constant. Basically we shall show that many of the well-known properties of quadratic forms of normal variables have their counterparts in quadratic forms of t variables with the F distribution replacing the chi-squared distribution of the normal quadratic form. However, one major difference appears - the independence property associated with certain sets of normal quadratic forms does not hold for the corresponding sets of t-quadratic forms.

In Section 4.2 we review some results on normal quadratic forms which will be used to derive similar results for t variables. In Section 4.3 we consider the joint distribution of the k quadratic forms $\frac{1}{r_i} t'A_it$ obtained from the expression

$$t'At = \sum_{i=1}^{k} t'A_it .$$

In Section 4.4 we discuss the distribution of quadratic forms obtained from a partitioned t vector and show that these are equivalent to the familiar F statistics used for testing in the general linear hypothesis. All the results in the above sections depend on the idempotency of the matrix $A\Sigma$. In Section 4.5 we consider the distribution of $t'At$ when $A\Sigma$ is not idempotent. Cornish (1954a) gave the distribution for the central case in integral form. We shall show that, in general, the distribution of $t'At$ can be expressed as a linear combination of doubly non-central F distributions. The determination of an explicit expression for the density function is a difficult problem. In Section 4.6 we derive a special case of the central density function and give an expression for the moments in the general central case.

4.2 QUADRATIC FORMS OF NORMAL VARIABLES

Definition 4.2.1

If x is a p×1 vector and A a p×p symmetric matrix then

$$Q = \sum x_i x_j a_{ij} = x'Ax$$

is a quadratic form in the variables $x = (x_1, x_2 \ldots x_p)$. The matrix of the quadratic form is always assumed to be symmetric since, if it is not, it can be made so by defining $B = \frac{1}{2}(A+A')$ and considering $x'Bx$.

We first state a fundamental theorem on the independence and "chi-squared-ness" of certain quadratic forms in normal variables, which was proved by Graybill and Marsaglia (1957). It is a generalisation of Cochran's Theorem (Cochran, 1934) to the case where $X \sim N_p(\mu, \Sigma)$ and Σ is positive definite.

Theorem 4.2.1 *(Graybill and Marsaglia, 1957)*

Let $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, *and let*

$$X'AX = \sum_{i=1}^{k} X'A_iX$$

where the rank of A_i *is* r_i *and the rank of* A *is* r.

Then any one of the conditions $C_1, C_2, \ldots C_6$, *is a necessary and sufficient condition that* $X'A_iX$ *be independently distributed as* $\chi^2_{r_i}(\phi_i)$ *where* $\phi_i = \frac{1}{2}\mu'A_i\mu$

C_1 $A\Sigma$ *be idempotent and* $(r_1 + r_2 \ldots + r_k) = r$

C_2 $A\Sigma$ *and each* $A_i\Sigma$ *be idempotent.*

C_3 $A\Sigma$ *be idempotent and* $A_i\Sigma A_j = 0$ *for all* $i \neq j$

C_4 $X'AX$ *be distributed as* $\chi^2_r(\phi)$ *and* $(r_1 + r_2 \ldots r_k) = r$

C_5 $X'AX$ *be distributed as* $\chi^2_r(\phi)$ *and* $A_i\Sigma$ *be idempotent.*

C_6 $X'AX$ *be distributed as* $\chi^2_r(\phi)$ *and* $A_i\Sigma A_j = 0$ *for all* $i \neq j$.

In C_4 C_5 *and* C_6, $\phi = \frac{1}{2}\mu'A\mu$.

The condition that Σ be positive definite is important, since this ensures that X has a non-singular normal distribution. If the covariance matrix is singular the conditions for "chi-squared-ness" of normal quadratic forms changes (See Rayner and Livingstone (1965), Rayner and Niven (1970), Styan (1970)). A theorem similar to the one given above holds for quadratic forms of singular normal variables (Styan (1970) or Searle (1971) page 71). However, the theorem stated above is sufficient for our purposes since we wish to develop results for quadratic forms of multivariate t variables, and in the multivariate t distribution, the covariance matrix of the underlying normal distribution is positive definite.

A number of useful corollaries arise from Theorem 4.2.1

Corollary 4.2.1 (Cochran (1934))

If $X \sim N(0,I)$ and A_i is symmetric with rank r_i for $i = 1, ..k$, then $X'A_iX$ is independently distributed as $\chi^2_{r_i}$ if and only if $(r_1+r_2...r_k) = r$

Proof Let $\mu = 0$ and $\Sigma = I_p = A$ in Theorem 4.2.1.

Corollary 4.2.2

If $X \sim N_p(\mu,\Sigma), \Sigma > 0$, then $X'\Sigma^{-1}X \sim \chi^2_p(\phi)$

where $\phi = \frac{1}{2}\mu'\Sigma^{-1}\mu$.

Proof Let k = 1 and $A = \Sigma^{-1}$ in Theorem 4.2.1.

Corollary 4.2.3.

If $X \sim N_p(\mu,\Sigma)$, $\Sigma > 0$, then a necessary and sufficient condition that $X'AX \sim \chi^2_r(\phi)$, where r is the rank of A and $\phi = \frac{1}{2}\mu'A\mu$, is that $A\Sigma$ be idempotent.

Proof Let k = 1 in Theorem 4.2.1

If $A\Sigma$ is not idempotent, then $X'AX$ no longer has a chi-squared distribution. We shall consider this case in more detail in Section 4.4.

4.3. THE JOINT DISTRIBUTION OF QUADRATIC FORMS OF t VARIABLES.

We recall the definition of the multivariate t-distribution given in Chapter 3.

If $X \sim Np(\mu,\Sigma)$, $\Sigma>0$, and $Z \sim \chi_n^2(\lambda)$ independently of X then

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X$$

has a doubly non-central multivariate t distribution, i.e. $t \sim t_p(n,\Sigma,\mu,\lambda)$

By exploiting the independence of X and Z, the distribution of many quadratic forms of t variables can be found from Theorem 4.2.1 and its corollaries.

Theorem 4.3.1 *If* $t \sim t_p(n,\Sigma,\mu,\lambda)$, *then*

$$\frac{1}{p}\, t'\Sigma^{-1}t \sim F(p,n,\phi,\lambda),$$

where $F(p,n,\phi,\lambda)$ *is the doubly non-central F distribution with p and n degrees of freedom and non-centrality parameters* $\phi = \frac{1}{2}\mu' A\mu$ *and* λ. *(See definition 2.4.4)*

Proof From the definition of t it follows that

$$\frac{1}{p}\, t'\Sigma^{-1}t = \frac{n}{p}\,\frac{X'\Sigma^{-1}X}{Z}$$

where X and Z are independent and $Z \sim \chi_n^2(\lambda)$. By corollary 4.2.2 $X'\Sigma^{-1}X \sim \chi_p^2(\phi)$

$$\text{So } \frac{1}{p}\, t'\Sigma^{-1}t = \frac{n\ \chi_p^2(\phi)}{p\ \chi_n^2(\lambda)} = F(p,n,\phi,\lambda)$$

This result was proved by Cornish (1954a) for $\mu=0$ and $\lambda=o$. The quadratic form then has a central F distribution. If $\lambda=o$ the quadratic form has an upper non-central F distribution. From Theorem 4.3.1 it is easily seen that if $\lambda=o$ then

$$\frac{1}{p}(t-\mu)'\Sigma^{-1}(t-\mu) \sim F(p,n).$$

This result can be used to determine sample sizes in the Bayesian analysis of the multivariate normal distribution (see De Groot (1970) page 189).

<u>*Theorem 4.3.2*</u> *If $t \sim t_p(n,\Sigma,\mu,\lambda)$ then*

$$\frac{1}{r}t'At \sim F(r,n,\phi,\lambda)$$

where $\phi=\frac{1}{2}\mu'A\mu$, if and only if $A\Sigma$ is idempotent of rank r.

<u>Proof</u> $\frac{1}{r}t'At = \frac{n}{r}\frac{X'AX}{Z}$

Where $Z \sim \chi_n^2(\lambda)$ independently of $X'AX$. By corollary 4.2.3, $X'AX \sim \chi_r^2(\phi)$ if and only if $A\Sigma$ is idempotent. The result now follows from the definition of $F(r,n,\phi,\lambda)$.

Cornish (1954a) gives the distribution of $t'At$ for the central case in an integral form and remarks that a necessary and sufficient condition for $\frac{1}{r}t'At$ to have a central F distribution is that the non-zero roots of $A\Sigma$ each be equal to one. Since $\Sigma>0$, this is equivalent to the condition that $A\Sigma$ be idempotent. If Σ were singular, Cornish's condition would not hold, since then it is possible for $A\Sigma$ to have characteristic roots 0 and 1 but not be idempotent (see Styan (1970)). However this situation would not arise with quadratic forms of t variables, since the covariance matrix is always positive definite and hence non-singular. We shall consider the distribution of $t'At$ when $A\Sigma$ is not idempotent in Section 4.5.

We now derive a theorem, which is analogous to Theorem 4.2.1, in that it gives necessary and sufficient conditions for quadratic forms of t variables to have an F distribution. However there is one important difference - the quadratic forms of t variables are not independent as are their normal counterparts. As we shall show later, they are not even uncorrelated.

<u>*Theorem 4.3.3.*</u> *Let* $t \sim t_p(n, \Sigma, \mu, \lambda)$, $\Sigma > 0$, *and let*

$$t'At = \sum_{i=1}^{k} t'A_i t .$$

where the rank of A_i *is* r_i *and the rank of* A *is* r. *Then any one of the conditions* $C_1, C_2, \cdots C_6$ *is necessary and sufficient for* $\frac{1}{r_i} t'At$ *to be distributed as* $F(r_i, n, \phi_i, \lambda)$ *where* $\phi_i = \frac{1}{2}\mu' A_i \mu$.

C_1 $A\Sigma$ *be idempotent and* $(r_1 + r_2 \cdots + r_k) = r$

C_2 $A\Sigma$ *be idempotent and each* $A_i\Sigma$ *be idempotent*

C_3 $A\Sigma$ *be idempotent and* $A_i \Sigma A_j = 0$ *for* $i \neq j$

C_4 $\frac{1}{r} t'At$ *be distributed as* $F(r, n, \phi, \lambda)$ *and* $(r_1 + r_2 \cdots r_k) = r$

C_5 $\frac{1}{r} t'At$ *be distributed as* $F(r, n, \phi, \lambda)$ *and* $A_i\Sigma$ *be idempotent*

C_6 $\frac{1}{r} t'At$ *be distributed as* $F(r, n, \phi, \lambda)$ *and* $A_i \Sigma A_j = 0$ *for* $i \neq j$.

where $\phi = \frac{1}{2}\mu' A \mu$.

Proof:

$$t'At = \sum_{i=1}^{k} t'A_i t = \sum_{i=1}^{r} r_i \left(\frac{1}{r_i} t'A_i t\right)$$

But from the definition of t $t'A_i t = \frac{nX'A_iX}{Z}$ where $X \sim N_p(\mu,\Sigma)$ and $Z \sim \chi^2_n(\lambda)$

Hence

$$t'At = \sum_{i=1}^{r} r_i \left(\frac{n}{r_i} \frac{X'A_iX}{Z}\right)$$

Suppose any one of the conditions C_1, C_2 or C_3 holds. From Theorem 4.2.1, any one of these conditions is necessary and sufficient for $X'A_iX$ to be independently distributed as $\chi^2_{r_i}(\phi_i)$, where $\phi_i = \frac{1}{2}\mu'A_i\mu$. Since $Z \sim \chi^2_n(\lambda)$ independently of $X'A_iX$, it follows that C_1, C_2 or C_3 is necessary and sufficient for

$$\frac{1}{r_i} t'A_i t \sim F(r_i,n,\phi_i,\lambda).$$

Suppose that any one of the conditions C_4, C_5 or C_6 holds. Then

$$\frac{1}{r_i} t'At \sim F(r,n,\phi,\lambda)$$

But $F(r,n,\phi,\lambda) = \frac{n}{r} \frac{\chi^2_r(\phi)}{\chi^2_n(\lambda)}$ where the χ^2's are independent. So it follows that

$$\frac{1}{r} t'At = \frac{n}{r} \frac{\chi^2_r(\phi)}{\chi^2_n(\lambda)}$$

But from the definition of t we know also that

$$\frac{1}{r} t'At = \frac{n}{r} \frac{X'AX}{Z}$$

where $Z \sim \chi^2_n(\lambda)$.

$$\text{Hence} \quad \frac{1}{r}\, t'At \;=\; \frac{n}{r}\,\frac{X'AX}{Z} \;=\; \frac{n}{r}\,\frac{\chi^2_r(\phi)}{\chi^2_n(\lambda)} .$$

So $X'AX \sim \chi^2_r(\phi)$, and one of the conditions C_4, C_5 or C_6 of Theorem 4.2.1 must hold, which is necessary and sufficient for each of the quadratic forms $X'A_iX$ to be independently distributed as $\chi^2_{r_i}(\phi_i)$. Therefore it follows that each of the quadratic forms

$$\frac{n}{r}\,\frac{X'A_iX}{Z} \;=\; \frac{1}{r_i}\, t'A_it \sim F(r_i, n, \phi_i, \lambda).$$

We note that Theorem 4.3.1 and 4.3.2 can be derived as corollaries to Theorem 4.3.3. It is clear from their synthetic forms that the $t'A_it$ are not independent since they all share the same denominator Z. Let us assume that $\phi = 0$ and $\lambda = o$ in Theorem 4.3.3; then all the F distributions are central. Under the conditions of the theorem

$$t'A_it \;=\; n\,\frac{X'A_iX}{Z} \;=\; n\,\frac{\chi^2_{r_i}}{\chi^2_n} \qquad i = 1\cdot\cdot k$$

Hence

$$\frac{1}{n}\, t'A_it \;=\; \frac{X'A_iX}{Z} \;=\; \frac{\chi^2_{r_i}}{\chi^2_n} \qquad i = 1\cdot\cdot k .$$

Any one of the conditions C_1, C_2 ·· C_6 in Theorem 4.3.3 implies that the corresponding condition in Theorem 4.2.1 holds, and the $X'A_iX$ are also independent. Hence we can write

$$Y_i \;=\; \frac{1}{n}\, t'A_it \;=\; \frac{U_i}{U_o} \qquad i = 1\cdot\cdot k$$

where U_o, $U_1 \cdot\cdot U_k$ are independent χ^2 variables with $U_o \sim \chi^2_n$ and $U_i \sim \chi^2_{r_i}$. From definition 2.5.1 we see that the joint distribution of $Y \cdot\cdot Y_k$ is multivariate F (Inverted Dirichlet). We state this result as a theorem.

<u>*Theorem 4.3.4.*</u> *If $t \sim t_p(n, \Sigma)$ and any one of the conditions of Theorem 4.3.3 is satisfied, then the joint distribution of $Y_i = \frac{1}{n} t'A_i t \quad i = 1 \cdot\cdot k$ is multivariate F with parameters $(\frac{1}{2}r_1, \frac{1}{2}r_2 \cdots \frac{1}{2}r_k; \frac{1}{2}n)$, where r_i is the rank of A_i.*

By means of a change of variable in f(t), Taio and Guttman (1965) showed that if $t \sim t_k(n,I)$ then $\frac{1}{n} t^2 \quad i = 1 \cdot\cdot k$ is multivariate F with parameters $(\frac{1}{2}, \frac{1}{2}, \cdot\cdot \frac{1}{2}; n)$. This result can be proved using Theorems 4.3.3 and 4.3.4. In Theorem 4.3.3, let $p = k$, $\Sigma = I_k$ and $A = I_k$, and let A_i, $i = 1 \cdot\cdot k$ be a $k \times k$ matrix with the i^{th} diagonal element equal to one and all the other elements zero. Then

$$t'At = \sum_{i=1}^{k} t'A_i t$$

becomes

$$t'I_k t = \sum_{i=1}^{k} t_i^2$$

and the rank of I_k is k and that of A_i is 1. $i = 1 \cdot\cdot k$.

Clearly conditions C_1 C_2 and C_3 are satisfied and each one of them is necessary and sufficient for $t_i^2 \sim F(1,n)$.

Hence from Theorem 4.3.4 the joint distribution of $\frac{1}{n} \; t_i^2$ is multivariate F with degrees of freedom $(\frac{1}{2}, \frac{1}{2}; \cdot\cdot \frac{1}{2}; n)$.

The mean and variance of $t'A_i t$ and the correlation between $t'A_i t$ and $t'A_j t$ could be found from their synthetic representations using the facts that X and Z are independent and $A_i \Sigma$ is idempotent. However, in the central case, these quantities can be more easily found using the properties of the multivariate F given in Corollary 2.5.1.

Hence

$$E(t'A_it) = n\,E(Y_i) = \frac{n(\frac{1}{2}r_i)}{(\frac{1}{2}n-1)} = \frac{n}{n-2}\,r_i$$

$$Var(t'A_it) = n^2\,Var(Y_i) = \frac{n^2\frac{1}{2}r_i(\frac{1}{2}r_i+\frac{1}{2}n-1)}{(\frac{1}{2}n-1)^2(\frac{1}{2}n-2)}$$

$$= \frac{2n^2r_i(r_i+n-2)}{(n-2)^2(n-4)}$$

$$Cov(t'A_it,\ t'A_jt) = \frac{\frac{1}{4}n^2(r_ir_j)}{(\frac{1}{2}n-1)^2(\frac{1}{2}n-2)} = \frac{2n^2r_ir_j}{(n-2)^2(n-4)}$$

and

$$Corr(t'A_it,\ t'A_jt) = \left(\frac{r_ir_j}{(r_i+n-2)(r_j+n-2)}\right)^{\frac{1}{2}}$$

We note that the correlation is always positive.

4.4 QUADRATIC FORMS IN A PARTITIONED t VECTOR

We now consider the distribution of sets of quadratic forms which arises when t is partitioned into two sets of components. Suppose that $t \sim t_p(n,\Sigma,\mu,\lambda)$ and is partitioned into two sets of components $t^{(1)}$ and $t^{(2)}$, where $t^{(1)}$ is $q\times 1$ and $t^{(2)}$ is $r\times 1$. Let Σ,μ and X, the vector of the underlying normal distribution be partitioned accordingly. (See Chapter 3 Section 3.1 Eqs. 3.2.6 to 3.2.9). Consider the two quadratic forms

$$\frac{1}{q}\,t^{(1)'}\Sigma_{11}^{-1}\,t^{(1)} \qquad \text{and} \qquad \frac{1}{r}\,t^{(2)'}\Sigma_{22}^{-1}\,t^{(2)} \qquad 4.4.1$$

From the synthetic representation of t we have that

$$\frac{1}{q}\, t^{(1)'}\Sigma_{11}^{-1}\, t^{(1)} = \frac{n}{q}\,\frac{X^{(1)'}\Sigma_{11}^{-1}\, X^{(1)}}{Z}$$

and

$$\frac{1}{r}\, t^{(2)'}\Sigma_{22}^{-1}\, t^{(2)} = \frac{n}{r}\,\frac{X^{(2)'}\Sigma_{22}^{-1}\, X^{(2)}}{Z}$$

Since $X^{(1)} \sim N(\mu^{(1)}, \Sigma_{11})$ independently of $Z \sim \chi_n^2(\lambda)$, it follows from Corollary 4.2.2 that $X^{(1)'}\Sigma_{11}^{-1} X^{(1)} \sim \chi_n^2(\phi_1)$ where $\phi_1 = \frac{1}{2}\mu^{(1)'}\Sigma_{11}^{-1}\mu^{(1)}$ and so

$$\frac{1}{q}\, t^{(1)'}\Sigma_{11}^{-1}\, t^{(1)} \sim F(q,n,\phi_1,\lambda)$$

Similarly $\frac{1}{r}\, t^{(2)}\Sigma_{22}^{-1}\, t^{(2)} \sim F(r,n,\phi_2,\lambda)$ where $\phi_2 = \mu^{(2)'}\Sigma_{22}^{-1}\mu^{(2)}$.

If $\Sigma_{12} = 0$, then the chi-squared distributions in the numerators of the two quadratic forms are independent and hence the joint distribution of $\frac{1}{n}\, t^{(1)'}\Sigma_{11}^{-1}\, t^{(1)}$ and $\frac{1}{n}\, t^{(2)'}\Sigma_{22}^{-1}\, t^{(2)}$ in the central case is bivariate F with parameters $(\frac{1}{2}q, \frac{1}{2}r, \frac{1}{2}n)$. Wc state this result as

Theorem 4.4.1

Let $t \sim t_p(n,\Sigma,\mu,\lambda)$ be partitioned into two sets of components where $t^{(1)}$ is $q\times 1$ and $t^{(2)}$ is $r\times 1$, $q+r=p$. Let μ and Σ be partitioned accordingly. Then

$$\frac{1}{q}\, t^{(1)'}\Sigma_{11}^{-1}t^{(1)} \sim F(q,n,\phi_1,\lambda) \text{ where } \phi_1 = \tfrac{1}{2}\mu^{(1)'}\Sigma_{11}^{-1}\mu^{(1)}$$

and

$$\frac{1}{r}\, t^{(1)'}\Sigma_{11}^{-1}t^{(1)} \sim F(r,n,\phi_2,\lambda) \text{ where } \phi_2 = \tfrac{1}{2}\mu^{(2)'}\Sigma_{22}^{-1}\mu^{(2)}.$$

If $\lambda = o$, $\mu = 0$, $\Sigma_{12} = 0$, the joint distribution of $\frac{1}{n} t^{(1)'} \Sigma_{11}^{-1} t^{(1)}$ and $\frac{1}{n} t^{(2)'} \Sigma_{22}^{-1} t^{(2)}$ is bivariate F with parameters $(\frac{1}{2}q, \frac{1}{2}r; n)$.

Corollary 4.4.1. If $t \sim t_p(n,\Sigma,\mu,\lambda)$ and is partitioned into k sets of components where $t^{(i)}$ is $p_i \times 1$ and $\sum_{i=1}^{k} p_i = p$ and if μ and Σ are partitioned accordingly, then

$$\frac{1}{p_i} t^{(i)'} \Sigma_{ii}^{-1} t^{(1)} \sim F(p_i, n, \phi_i, \lambda)$$

where

$$\phi_i = \tfrac{1}{2}\mu^{(i)'}\Sigma_{ii}^{-1}\mu^{(i)} .$$

If $\lambda = o$, $\mu = 0$, and $\Sigma_{ij} = 0$ for all $i \neq j$, $i,j = 1 \cdot\cdot k$, then the joint distribution of $\frac{1}{n} t^{(i)'} \Sigma_{ii}^{-1} t^{(i)}$ $i = 1 \cdot\cdot k$, is k dimensional multivariate F with parameters $(\frac{1}{2}p_1, \ldots \frac{1}{2}p_k; n)$.

If $\Sigma_{12} \neq 0$, the chi-squared distributions in the numerators of the two quadratic forms are correlated, and the quadratic forms $\frac{1}{q} t^{(1)'} \Sigma_{11}^{-1} t^{(1)}$ and $\frac{1}{r} t^{(2)'} \Sigma_{22}^{-1} t^{(2)}$ would follow another type of multivariate F distribution which was derived by Jensen (1970).

If $V_1 = \frac{1}{q} t^{(1)'} \Sigma_{11}^{-1} t^{(1)}$ and $V_2 = \frac{1}{r} t^{(2)'} \Sigma_{22}^{-1} t^{(2)}$

then the joint density of V_1 and V_2 is (Jensen (1970))

$$f(V_1,V_2) = \sum_{k=o}^{\infty} G_K(\underline{\delta}) \sum_{i=o}^{k} \sum_{j=o}^{k} d_{jk}(k)$$

$$\times \frac{V_1^{\frac{1}{2}(q+2i-2)} \; V_2^{\frac{1}{2}(r+2j-2)}}{\left(1 + \frac{q}{n} V_1 + \frac{r}{n} V_2\right)^{\frac{1}{2}(q+r+2i+2j+n)}}$$

where $d_{ij}(k) = (-1)^{i+j} \binom{k}{i} \binom{k}{j}$

$$\times \quad \frac{q^{\frac{1}{2}(q+2i)} \, r^{\frac{1}{2}(r+2j)} \, \Gamma(\frac{1}{2}(q+r+2i+2j+n))}{n^{\frac{1}{2}(q+r+2i+2j)} \Gamma(\frac{1}{2}(q+2i)) \Gamma(\frac{1}{2}(r+2j)) \Gamma(\frac{1}{2}n)}$$

and $G(\underline{\delta})$ is a function of the squares of the canonical correlations between $X^{(1)}$ and $X^{(2)}$ in the underlying normal distribution and is

$$G_k(\underline{\delta}) = \sum_{j_1, \cdots j_s}^{k} a_{1j_1}, a_{2j_2} \cdots a_{sj_s}$$

with

$$a_{ij_i} = \delta_i^{j_i} \quad \Gamma(j_i + \tfrac{1}{2})/\Gamma(j_i+1)\Gamma(\tfrac{1}{2})$$

where δ_i is the square of the i^{th} canonical correlation between $X^{(1)}$ and $X^{(2)}$.

The distribution of our statistics follows directly from Jensen's results because in fact we are considering the same variables although they arise in a different context. We have been interested in deriving properties of the multivariate t-distribution; however we can show that, in fact, the quadratic forms in Theorem 4.4.1 are exactly the F statistics that arise in connection with tests of linear hypotheses in the Analysis of Variance.

Consider the general linear hypothesis of full rank

$$Y = X\beta + e$$

where Y is a n×1 vector of observations, X is a n×p matrix of known constants, β is an unknown p×1 vector of constants and e is a n×1 random vector such that E(e) = 0 and

$E(ee') = \sigma^2 I$. Suppose $Y \sim N(X\beta, \sigma^2 I)$.

Then following the standard procedure (see Graybill (1961)) the M.L.-L.S. estimator of β is,

$$\hat{\beta} = (X'X)^{-1}X'Y$$

and $\hat{\beta} \sim N(\beta, \sigma^2 M)$

where $M = (X'X)^{-1}$ (M is known)

σ^2 is estimated by

$$s^2 = (Y-X\hat{\beta})'(Y-X\hat{\beta})/(n-p)$$

and $\dfrac{(n-p)s^2}{\sigma^2} \sim \chi^2_{n-p}$.

Let $\beta = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$ and $M = \begin{pmatrix}M_{11} & M_{12}\\ M_{21} & M_{22}\end{pmatrix}$

If we wish to test the two hypotheses

$$H_o! \ \beta_1 = \beta_{10}$$

$$H_o! \ \beta_2 = \beta_{20}$$

we use the two statistics

$$V_1 = \frac{(\hat{\beta}_1 - \beta_{10})' M_{11}^{-1} (\hat{\beta}_1 - \beta_{10})}{qs^2}$$

$$V_2 = \frac{(\hat{\beta}_2 - \beta_{20})' M_{22}^{-1} (\hat{\beta}_2 - \beta_{20})}{rs^2}$$

Under the hypothesis $\beta = \beta_o$

$$\frac{1}{\sigma}(\hat{\beta} - \beta_o) \sim N_p(0,M) \text{ and } \frac{s^2}{\sigma^2} \sim \frac{\chi^2_{n-p}}{n-p}$$

So $t = \dfrac{\hat{\beta} - \beta_o}{s} \sim t_p(n-p,M)$

Partitioning t as β above, we find

$$t^{(1)} = \frac{(\hat{\beta}_1 - \beta_{10})}{s} \qquad t^{(2)} = \frac{(\hat{\beta}_2 - \beta_{20})}{s}$$

and

$$\frac{1}{q} t^{(1)'} M_{11}^{-1} t^{(1)} = \frac{(\hat{\beta}_1 - \beta_{10})' M_{11}^{-1} (\hat{\beta}_1 - \beta_{10})}{qs^2} = V_1 .$$

Similarly, $\frac{1}{r} t^{(2)'} M_{22}^{-1} t^{(2)} = V_2$.

So the F statistics for the general linear hypothesis of full rank can be expressed in terms of the quadratic forms of t variables given in Theorem 4.4.1.

4.5 THE GENERAL DISTRIBUTION OF t'At

In previous sections of this chapter we have used the idempotency of AΣ to deduce the distribution of the quadratic form t'At. What happens if AΣ is not idempotent? For $\mu = 0$ .. Cornish (1954a) gave the distribution function as an integral as follows:

If $t \sim t_p(n,\Sigma)$ and t'At is a quadratic form in t, then the distribution function of t'At is

$$\Pr(t'At \leq Q_o) = \int \cdots \int_{t'At \leq Q_o} f(t)dt$$

$$= \frac{\Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}\ \Gamma(\frac{1}{2}n)\ |\Sigma|^{\frac{1}{2}}} \int \cdots \int_{t'At \leq Q_o} \left(1+ \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)} dt \qquad 4.5.1$$

Let H be a non-singular matrix such that

$$H\Sigma^{-1}H' = I_p \qquad 4.5.2$$

and

$$HAH' = \Lambda \qquad 4.5.3$$

where $\Lambda = \text{diag}(\lambda_1, \cdots \lambda_p)$ and $\lambda_1 \cdots \lambda_p$ are the roots of $|A\Sigma - \lambda I| = 0$.

Let $t = Hx$; then, in view of 4.5.2, $J(t \to x) = |\Sigma|^{\frac{1}{2}}$ and 4.5.1 becomes

$$P(x'Ax \leq Q_o) = \frac{\Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}\,\Gamma(\frac{1}{2}n)} \int \cdots \int_{x'\Lambda x \leq Q_o} \left[1 + \frac{x'x}{n}\right]^{-\frac{1}{2}(n+p)} dx \qquad 4.5.4$$

If $\lambda_1, \lambda_2 \cdots \lambda_r$ are the non-zero characteristic roots of $|A\Sigma - \lambda I| = 0$, then $x'\Lambda x = \sum_{i=1}^{r} \lambda_i x_i^2$

Integrating over $(x_{r+1} \cdots x_p)$, 4.5.4 becomes

$$\frac{\Gamma(\frac{1}{2}(n+r))}{(n\pi)^{\frac{1}{2}p}\,\Gamma(\frac{1}{2}n)} \int \cdots \int \left[1 + \frac{x^{(1)'}\,x^{(1)}}{n}\right]^{-\frac{1}{2}(n+p)} dx^{(1)} \qquad 4.5.5$$

where $x^{(1)'} = (x_1 \cdots x_r)$ and the region of integration is $\sum_{i=1}^{r} \lambda_i x_i^2 \leq Q_o$

If the non-zero roots are all unity, then 4.5.5 reduces to the distribution function of $rF(r,n)$. (refer to Theorem 4.3.2). If the non-zero roots are not all unity, then this integral is difficult to evaluate. However, we can find out more about the distribution of $t'At$ in this case, and in the general case when $t \sim t_p(n,\Sigma,\mu,\lambda)$, using a result given by Baldessari (1967) for normal quadratic forms.

We have seen (Section 4.2) that the chi-squared distribution of the normal quadratic form $X'AX$ depends on the idempotency of $A\Sigma$. If $A\Sigma$ is not idempotent then the quadratic form is no longer distributed as χ^2. However, Baldessari (1967) showed that if $A\Sigma$ can be written in spectral form (see Theorem 2.1.4) then $X'AX$ is distributed as a certain linear combination of independent χ^2 variables (central or non-central).

More specifically, let $X \sim N_p(\mu,\Sigma)$ where $\Sigma > 0$ and let $X'AX$ be a quadratic form in X. Since A is symmetric and $\Sigma > 0$ the conditions of Theorem 2.1.4 hold, and $A\Sigma$ has the spectral decomposition given below.

Definition 4.5.1

$$A\Sigma = \sum_{j=1}^{r} a_j E_j \qquad 4.5.6$$

where (i) $a_1,\cdots a_r$ are the distinct characteristic roots of $|A\Sigma-\lambda I| = 0$

(ii) $E_j^2 = E_j$ for $j=1 \cdots r$

(iii) $E_i E_j = 0$ $i \neq j$.

(iv) Rank of $E_j = r_j$, where r_j is the multiplicity of a_j.

We now define a random variable V which is distributed as a linear combination of independent χ^2 variables.

Definition 4.5.2

Let V be a random variable such that

$$V = \sum_{j=1}^{r} a_j \chi^2_{r_j}(\tfrac{1}{2}\mu' L_j \mu)$$

where (i) the a_j are distinct

(ii) the χ^2's are independent

(iii) L_j is symmetric, positive semi-definite with rank r_j $j=1 \cdots r$.

Baldessari (1967) gives necessary and sufficient conditions for $X'AX$ to be distributed as V.

<u>*Theorem 4.5.1*</u> *(Baldessari (1967))*

If $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, then

$$X'AX \sim V$$

if and only if $A\Sigma$ has the spectral decomposition

$$A\Sigma = \sum_{j=1}^{r} a_j E_j .$$

The parameters of V are

(i) $\mu' L_j \mu = \mu' E_j \Sigma^{-1} \mu$

(ii) r_j *= rank of* E_j *where* $\sum_{1}^{r} r_j = p$

(iii) the distinct characteristic roots of $A\Sigma$, $a_1 \cdots a_r$.

We now use Theorem 4.5.1, to derive a corresponding result for a quadratic form of t variables, but first we need a lemma.

<u>*Lemma 4.5.1.*</u> *If $t \sim t_p(n, \Sigma, \mu, \lambda)$; and $Q = t'At$ is a quadratic form in t, then $A\Sigma$ always has a spectral decomposition.*

Proof: The parameter matrix, Σ, of the t distribution is by definition positive definite. The matrix A of the quadratic term Q can always be chosen to be symmetric. Hence the conditions of Theorem 2.5.1 apply and $A\Sigma$ has the required spectral decomposition.

Theorem 4.5.2. *Let $t \sim t_p(n,\Sigma,\mu,\lambda)$ and let $A\Sigma$ have the spectral decomposition guaranteed by lemma 4.5.1. Then*

$$\frac{1}{n} \; t'At \sim W \qquad 4.5.7$$

where

$$W = \sum_{j=1}^{r} \frac{a_j \, \chi^2_{r_j}(\tfrac{1}{2}\mu' L_j \mu)}{\chi^2_n(\lambda)} \qquad 4.5.8$$

with (i) $\chi^2_{r_j}(\cdot)$ *independent of* $\chi^2_n(\cdot)$ *for* $j = 1 \cdot\cdot r$.

(ii) $\tfrac{1}{2}\mu' L_j \mu = \tfrac{1}{2}\mu' E_j \Sigma^{-1} \mu$

(iii) a_j, E_j *and* r_j *defined by 4.5.6.*

Proof: If $t \sim t_p(n,\Sigma,\mu,\lambda)$ then

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X \quad \text{where } X \sim N_p(\mu,\Sigma),\ \Sigma > 0$$

and $Z \sim \chi^2_n(\lambda)$ independently of X. Hence

$$\frac{1}{n} \; t'At = \frac{X'AX}{Z}$$

From Theorem 4.5.1 $X'AX$ is distributed as V.

Hence $\frac{1}{n} \; t'AT \sim \frac{V}{Z} = W.$

It is easily seen that Theorems 4.3.1 and 4.3.2 are special cases of Theorem 4.5.2.

Corollary 4.5.1 $\frac{1}{p} t'\Sigma^{-1}t \sim F(p,n,\phi,\lambda)$ where $\phi = \frac{1}{2}\mu'\Sigma^{-1}\mu$

Proof: In Theorem 4.5.2, let $A = \Sigma^{-1}$. Then $A\Sigma^{-1} = I_p$ which is already in spectral form. I_p has a single distinct characteristic root, $a_1 = 1$, with multiplicity p. Hence $E_p = I_p$ with rank $r_1 = p$. Therefore $\frac{1}{2}\mu' L_j \mu = \frac{1}{2}\mu'\Sigma^{-1}\mu$.

So

$$\frac{1}{n} t'\Sigma^{-1}t = \frac{\chi_p^2(\frac{1}{2}\mu'\Sigma^{-1}\mu)}{\chi_n^2(\lambda)}$$

from which it follows that $\frac{1}{p} t'\Sigma^{-1}t \sim F(p,n,\phi,\lambda)$.

Corollary 4.5.2. $\frac{1}{r} t'At \sim F(r,n,\phi,\lambda)$ if and only if $A\Sigma$ is idempotent with rank r.

Proof: If $A\Sigma$ is idempotent with rank r then it has two distinct characteristic roots, $a_1 = 1$, with multiplicity r, and $a_2 = 0$ with multiplicity r-p.

Then $E_1 = A\Sigma$, $E_2 = (I-A\Sigma)$ and $E_1E_2 = 0$.

The spectral decomposition of $A\Sigma$ is

$$A\Sigma = A\Sigma + o(I-A\Sigma) = A\Sigma.$$

$$\tfrac{1}{2}\mu' L_1 \mu = \tfrac{1}{2}\mu' A\Sigma\Sigma^{-1}\mu = \tfrac{1}{2}\mu' A\mu$$

Hence $\frac{1}{n} t'At \sim \dfrac{\chi_r^2(\frac{1}{2}\mu' A\mu)}{\chi_n^2(\lambda)}$

From which it follows that $\frac{1}{r} t'At \sim F(r,n,\phi,\lambda)$.

From 4.5.7 and 4.5.8 it is easily seen that t'At can be written as a weighted sum of correlated doubly non-central F variables. Let $\phi_j = \frac{1}{2}\mu' L_j \mu$,

$$t'At = \sum_{j=1}^{r} a_j r_j \left(\frac{n\, \chi^2_{r_j}(\phi_j)}{r_j \chi^2_n(\lambda)} \right)$$

$$= \sum_{j=1} w_j \; F(r_j, n, \phi_j, \lambda) \qquad 4.5.9$$

The F variables are correlated because they all share the same χ^2 denominator. We summarise this result in the next theorem.

Theorem 4.5.3. A synthetic representation of t'At

If $t \sim t_p(n, \Sigma, \mu, \lambda)$ then

$$t'At = \sum_{j=1}^{r} w_j \; F(r_j, n, \phi_j, \lambda)$$

where (i) *the F distributions are correlated*

(ii) $w_j = a_j r_j$

(iii) $\phi_j = \frac{1}{2}\mu' E_j \Sigma^{-1} \mu$

(iv) a_j, r_j, E_j *and r are defined in 4.5.6.*

4.6 THE DISTRIBUTION OF W

Although we have been able to show that t'At is distributed as a linear combination of correlated F distributions, it is still difficult to determine the density function of t'At specifically, except in the special case where $A\Sigma$ is idempotent. From Theorem 4.5.2

$$\frac{1}{n}\, t'At = W = \sum_{j=1}^{r} \frac{a_j \; \chi^2_{r_j}(\phi_j)}{\chi^2_n(\lambda)} \qquad 4.6.1$$

$$\text{If } \phi = \lambda = o \quad \text{then } W = \sum_{j=1}^{r} a_j \frac{U_j}{U_o} \qquad 4.6.2$$

where $U_o, U_1, \cdots U_r$ are independent central χ^2 variables with degrees of freedom $n, r_1 \cdots r_r$. Let $Y_j = U_j/U_o$; then the joint distribution of $Y_1, \cdots Y_r$ is, by def.2.5.1. central multivariate F with parameters $(\frac{1}{2}r_1, \cdots \frac{1}{2}r_r; \frac{1}{2}n)$. Thus we must find the density function of

$$W = \sum_{j=1}^{r} a_j Y_j \qquad 4.6.3$$

Formally this could be found by a suitable transformation in the joint density of $(Y_1 \cdots Y_r)$. This is difficult to do in the general case. We shall consider a special case, where $r = 2$, and r_1, r_2 and n are even integers. To simplify notation let $\frac{1}{2}r_1 = k$, $\frac{1}{2}r_2 = \ell$ and $\frac{1}{2}n = m$, then k, ℓ and m are integers. The joint density of Y_1, Y_2 is then

$$f(Y_1, Y_2) = \frac{\Gamma(k+\ell+m) \qquad Y_1^{k-1}\; Y_2^{\ell-1}}{\Gamma(k)\;\Gamma(\ell)\;\Gamma(m)\;(1+Y_1+Y_2)^{k+\ell+m}} \qquad 4.6.4$$

where $0 < Y_1 < \infty$ and $0 < Y_2 < \infty$

Transform from Y_1 and Y_2 to θ_1 and θ_2 by

$$Y_1 = \theta_1 \qquad 4.6.5$$

$$Y_2 = \theta_2(1+\theta_1)$$

with Jacobian $J(Y_1, Y_2 \rightarrow \theta_1, \theta_2) = (1+\theta_1)$ 4.6.6

and noting that $(1+Y_1+Y_2) = (1+\theta_1)(1+\theta_2)$, 4.6.7

the density function of θ_1 and θ_2 is

$$f(\theta_1,\theta_2) \propto \frac{\theta_1^{k-1}\ \theta_2^{\ell-1}\ (1+\theta_1)^{\ell}}{\left\{(1+\theta_1)(1+\theta_2)\right\}^{k+\ell+m}} \qquad \begin{array}{l} 0 < \theta_1 < \infty \\ 0 < \theta_2 < \infty \end{array}$$

$$= \frac{\theta_1^{k-1}\ \theta_2^{\ell-1}}{(1+\theta_1)^{k+m}\ (1+\theta_2)^{k+\ell+m}} \qquad 4.6.8$$

where the constant of proportionality is

$$\frac{\Gamma(k+\ell+m)}{\Gamma(k)\Gamma(\ell)\Gamma(m)} \qquad 4.6.9$$

Now since $\Pr(W \leq w_o) = \Pr\left[a_1Y_1 + a_2Y_2 \leq w_o\right]$

$$= \Pr\left[a_1\theta_1 + a_2\theta_2(1+\theta_1) \leq w_o\right]$$

we can find the density function of W by transforming in 4.6.8. Since $W = a_1\theta_1 + a_2\theta_2(1+\theta_1)$, the required transformation is

$$\theta_1 = \frac{W - a_2\theta_2}{a_1+a_2\theta_2} \qquad 4.6.10$$

$$\theta_2 = \theta_2$$

and the Jacobian of the transformation is

$$J(\theta_1,\theta_2 \rightarrow W,\theta_2) = (a_1+a_2\theta_2)^{-1} \qquad 4.6.11$$

Noting that $1+\theta_1 = \dfrac{a_1 + W}{a_1 + a_2\theta_2}$ 4.6.12

the density function of W and θ_2 is

$$f(W,\theta_2) \propto \left(\frac{w-a_2\theta_2}{a_1+a_2\theta_2}\right)^{k-1} \cdot \left(\frac{a_1+a_2\theta_2}{a_1+w}\right)^{k+m} \cdot \frac{\theta_2^{\ell-1}}{(1+\theta_2)^{k+\ell+m}} \cdot \frac{1}{(a_1+a_2\theta_2)}$$

$$= \frac{(w-a_2\theta_2)^{k-1}(a_1+a_2\theta_2)^m\ \theta_2^{\ell-1}}{(a_1+w)^{k+m}\ (1+\theta_2)^{k+\ell+m}} \qquad 4.6.13$$

Now since k and m are integers we can expand the first two terms of the numerator of 4.6.13 using the binomial theorem and obtain

$$(w-a_2\theta_2)^{k-1} = \sum_{i=o}^{k-1} (-1)^i \binom{k-1}{i} w^{k-1-i}\ a_2^i\theta_2^i \qquad 4.6.14$$

$$(a_1+a_2\theta_2)^m = \sum_{j=o}^{m} \binom{m}{j} a_1^{m-j}\ a_2^j\ \theta_2^j \qquad 4.6.15$$

and so 4.6.13 becomes

$$f(W,\theta_2) \propto \sum_{i=o}^{k-1} \sum_{j=o}^{m} g(w;i,j)\ \frac{\theta_2^{i+j+\ell-1}}{(1+\theta_2)^{k+\ell+m}} \qquad 4.6.16$$

where

$$g(w;i,j)=(-1)^i \binom{k-1}{i}\binom{m}{j} a_1^{m-j}a_2^{i+j}w^{k-1-i}(a_1+w)^{-(k+m)} \qquad 4.6.17$$

To obtain f(w) we integrate 4.6.16 termwise with respect to θ_2. From the properties of the Beta function,

$$\int_o^\infty \frac{\theta_2^{i+j+\ell-1}}{(1+\theta_2)^{k+\ell+m}}\, d\theta_2 = B(\ell+i+j\ ;\ k+m-i-j)$$

$$= \frac{\Gamma(\ell+i+j)\Gamma(k+m-i-j)}{\Gamma(k+\ell+m)} \qquad 4.6.18$$

Hence

$$f(w) \propto \sum_{i=o}^{k-1} \sum_{j=o}^{m} \frac{\Gamma(\ell+i+j)\Gamma(k+m-i-j)g(w;i,j)}{\Gamma(k+\ell+m)} \qquad 4.6.19$$

Substituting for g(w,i,j) and the constant of proportionality from 4.6.9, we find

$$f(w) = \sum_{i=o}^{k-1} \sum_{j=o}^{m} \frac{(-1)^i \binom{k-1}{i}\binom{m}{j}\Gamma(\ell+i+j)\Gamma(k+m-i-j)a_1^{m-j}a_2^{i+j}w^{k-1-i}}{\Gamma(k)\Gamma(\ell)\Gamma(m) \quad (a_1+w)^{k+m}}$$

4.6.20

where $k = \frac{1}{2}r_1$, $\ell = \frac{1}{2}r_2$ $m = \frac{1}{2}n$.

We now consider the moments of W for the central case. From 4.6.3.

$$E(W^k) = E\left[\left(\sum_{j=1}^{r} a_j Y_j\right)^k\right] \qquad 4.6.21$$

where $(Y_1,\ldots Y_r)$ has a multivariate F distribution with parameters $(\frac{1}{2}r_1,\ldots\frac{1}{2}r_r; \frac{1}{2}n)$. Expanding the argument of 4.6.21 using the multinominal theorem, we have

$$E(W^k) = \sum_{\kappa} \frac{k!}{k_1!\,k_2!\cdots k_r!} a_1^{k_1}a_2^{k_2}\cdots a_r^{k_r}\; E(Y_1^{k_1}\cdots Y_r^{k_r})$$

4.6.22

where the summation is over all the partitions, κ , of k into not more than r parts.

By Theorem 2.5.1

$$E(Y_1^{k_1}\cdots Y_r^{k_r}) = \frac{\Gamma(\frac{1}{2}n-k)\Gamma(\frac{1}{2}r_1+k_1)\cdots\Gamma(\frac{1}{2}r_r+k_r)}{\Gamma(\frac{1}{2}n)\;\Gamma(\frac{1}{2}r_1)\;\cdots\;\Gamma(\frac{1}{2}r_r)}$$

4.6.23

where $\sum_{j=1}^{k} k_j = k$ and $n > 2k$. Hence

$$E(W^k) = \sum_{\kappa} \frac{k!\ a_1^{k_1} a_2^{k_2} \cdots a_r^{k_r}}{k_1!\ k_2! \cdots k_r!} \frac{\Gamma(\frac{1}{2}n-k)\Gamma(\frac{1}{2}r_1+k_1)\cdots\Gamma(\frac{1}{2}r_r+k_r)}{\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}r_1)\cdots\Gamma(\frac{1}{2}r_r)}$$

4.6.24

Now it was shown in Theorem 4.3.2 that if $A\Sigma$ is idempotent of rank r, $\frac{1}{r}\ t'At \sim F(r,n)$ when t has a central multivariate t distribution. In this case

$$E\left(\frac{1}{n}\ t'At\right) = \frac{r}{n-2} \qquad 4.6.25$$

We can also obtain this result from 4.2.24, since in this case $A\Sigma$ has a single non-zero root $a_1 = 1$ with multiplicity $r_1 = r$. Setting $k = k_1 = 1$ we have

$$E\left(\frac{t'At}{n}\right) = \frac{\Gamma(\frac{1}{2}n-1)\Gamma(\frac{1}{2}r+1)}{\Gamma(\frac{1}{2}n)\ \ \Gamma(\frac{1}{2}r)} = \frac{\frac{1}{2}r}{\frac{1}{2}n-1} = \frac{r}{n-2}\ .$$

CHAPTER 5

STATISTICS DERIVED FROM THE MULTIVARIATE t

5.1 INTRODUCTION

Let $Y_{(\alpha)}$ $\alpha = 1 \ldots K$ be a random sample from a p-dimensional normal population, with mean vector Θ and covariance matrix $\sigma^2 R$, where R is the correlation matrix. Suppose that R is known but σ^2 is unknown. Let S^2 be an unbiased estimate of σ^2 with n degrees of freedom which is independent of the $Y_{(\alpha)}$'s and is such that

$$Z = \frac{nS^2}{\sigma^2} \sim \chi^2_n \qquad 5.1.1$$

Consider the p×1 random vector

$$t_{(\alpha)} = \frac{1}{S} Y_{(\alpha)} \qquad 5.1.2$$

which represents an observation $Y_{(\alpha)}$ that has been "standardised" by dividing each component by the sample standard deviation. ("Standardised" is in inverted commas because the standardisation is obtained using a random variable S instead of the true standard deviation σ.)

We can write

$$t_{(\alpha)} = \left(\frac{\sigma}{S}\right)\left(\frac{Y_{(\alpha)}}{\sigma}\right) \qquad \alpha = 1,..K \qquad 5.1.3$$

Then $t_{(\alpha)}$ has the representation

$$t_{(\alpha)} = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X_{(\alpha)} \qquad \alpha = 1,..K \qquad 5.1.4$$

where

$$X_{(\alpha)} = \frac{Y_{(\alpha)}}{\sigma} \sim N(\mu, R) \qquad \alpha = 1 \ldots K \qquad 5.1.5$$

with $\mu = \dfrac{\Theta}{\sigma}$ 5.1.6

and $\left(\dfrac{Z}{n}\right)^{\frac{1}{2}} = \dfrac{S}{\sigma}$ 5.1.7

is independent of $X_{(\alpha)}$.

Hence $t_{(\alpha)}$ has an upper non-central multivariate t distribution with n degrees of freedom, parameter matrix R and non-centrality parameter μ.

Using the "standardised" variables, we can calculate the "standardised" mean vector

$$\bar{t} = \frac{1}{K}\sum_{\alpha=1}^{K} t_{(\alpha)} \qquad 5.1.8$$

$$= \frac{1}{S}\left(\frac{1}{K}\sum_{\alpha=1}^{K} Y_{(\alpha)}\right)$$

$$= \frac{1}{S}\bar{Y} \qquad 5.1.9$$

where $\bar{Y}$ is the mean vector of the $Y_{(\alpha)}$'s, and the matrix of sums of squares and cross-products

$$\tilde{T} = \sum_{\alpha=1}^{K} (t_{(\alpha)} - \bar{t})(t_{(\alpha)} - \bar{t})' \qquad 5.1.10$$

$$= \frac{1}{S^2} \sum_{\alpha=1}^{K} (Y_{(\alpha)} - \bar{Y})(Y_{(\alpha)} - \bar{Y})' \qquad 5.1.11$$

where the second term of 5.1.11 is the Wishart matrix constructed from the $Y_{(\alpha)}$'s.

$\tilde{T}$ is simply a Wishart matrix constructed from the "standardised" variables, instead of the original sample values. We resist the temptation to call this matrix the Twishart matrix, and shall simply refer to it as $\tilde{T}$. It is clear from 5.1.9 and 5.1.11 that, unlike their normal counterparts, $\bar{t}$ and $\tilde{T}$ are not independent, since they are bound together by S. Cornish (1954b) derived the joint distribution of $\bar{t}$ and $\tilde{T}$ when the underlying normal distribution has zero mean, and also obtained the marginal distributions of $\bar{t}$ and $\tilde{T}$. Kshirsagar (1960) derived the distribution of $\tilde{T}$ when the matrix

$$A = \sum_{\alpha=1}^{K} (Y_{(\alpha)}-\bar{Y})(Y_{(\alpha)}-\bar{Y})' \qquad 5.1.12$$

has a linear non-central Wishart distribution, with covariance matrix $\sigma^2 R$ and s^2 is a chi-squared estimator of some *other* variance σ'^2.

The distribution of $\tilde{T}$ when A has a non-central Wishart distribution with covariance matrix $\sigma^2 R$ and s^2 is an independent chi-squared estimator of the *same* variance σ^2, is extremely complicated for the following reason. If A has a non-central Wishart distribution, then the $Y_{(\alpha)}$'s are a random sample from a normal population in which the covariance matrix, $\sigma^2 R$, is the same for all observations but the means vary from observation to observation, i.e. $Y_{(\alpha)} \sim N_p(\Theta_{(\alpha)}, \sigma^2 R)$. Because the means vary in this way, any estimator of the common variance obtained from the sample will have a non-central chi-squared distribution (strictly speaking $Z = nS^2/\sigma^2$ is non-central chi-squared). The distribution of $\tilde{T}$ will be that of $\left(\frac{Z}{n}\right)^{-1} A$ where $A \sim W_p(k, \sigma^2 R, \Omega)$

and $Z \sim \chi^2_n(\lambda)$ independently of A. We shall indicate the method of deriving this distribution in Section 5.3, but will only derive the distribution explicitly for the linear case (see Theorem 5.3.7).

For the most part, we turn our attention to some other non-central forms of $\tilde{T}$, namely

(i) the "lower non-central $\tilde{T}$", when A has a central Wishart distribution and Z a non-central chi-squared distribution;

(ii) the "upper non-central $\tilde{T}$", which is the general case of Kshirsagar's T, when A has a non-central Wishart distribution of any rank and Z a central chi-squared distribution.

In addition we also discuss briefly the distribution of $\bar{t}$; some moments of the determinant of $\tilde{T}$ in the central case and lower non-central case, and finally, the distribution of the characteristic roots of $\tilde{T}$ in the central case.

Although in the foregoing discussion $\bar{t}$ and $\tilde{T}$ were statistics derived from an underlying normal population with covariance structure $\sigma^2 R$, in the subsequent derivation we shall consider a normal population with any covariance structure Σ, with the only proviso that Σ is positive definite, so that the inverse is well defined. We do so because we feel that in discussing the general aspects of those distributions it is not necessary to confine ourselves to a particular covariance structure. The abovementioned distributions can be obtained from ours by a simple change of parameter.

5.2 THE CENTRAL DISTRIBUTION OF $\tilde{T}$

Let $X_{(\alpha)}$, $\alpha = 1 \ldots K$, be a random sample from a $N_p(\mu,\Sigma)$ population. Let

$$\bar{X} = K^{-1} \sum_{\alpha=1}^{K} X_{(\alpha)} \qquad 5.2.1$$

be the sample mean vector and

$$A = \sum_{\alpha=1}^{K} (X_{(\alpha)}-\bar{X})(X_{(\alpha)}-\bar{X})' \qquad 5.2.2.$$

be the matrix of sums of squares and cross products. Then (Anderson (1958))

$$\bar{X} \sim N_p(\mu, K^{-1}\Sigma) \qquad 5.2.3$$

independently of

$$A \sim W_p(\Sigma, k) \qquad 5.2.4$$

where $k = K-1$

Let $Z \sim \chi^2_n$ independently of $X_{(\alpha)}$, $\alpha = 1 \ldots K$ and define

$$t_{(\alpha)} = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} X_{(\alpha)} \qquad \alpha = 1 \ldots K \qquad 5.2.5$$

Then $$\bar{t} = K^{-1} \sum_{\alpha=1}^{K} t_{(\alpha)} \qquad 5.2.6$$

$$= \left(\frac{Z}{n}\right)^{-\frac{1}{2}} \bar{X} \qquad 5.2.7$$

and

$$\tilde{T} = \sum_{\alpha=1}^{K} (t_{(\alpha)}-\bar{t})(t_{(\alpha)}-\bar{t})' \qquad 5.2.8$$

$$= \left(\frac{Z}{n}\right)^{-1} \sum_{\alpha=1}^{K} (X_{(\alpha)}-\overline{X})(X_{(\alpha)}-\overline{X})'$$

$$= \left(\frac{Z}{n}\right)^{-1} A \qquad 5.2.9$$

Then $t_{(\alpha)} \sim t_p(n,\Sigma,\mu) \qquad \alpha = 1...K \qquad 5.2.10$

i.e. $t_{(\alpha)}$ has an upper non-central multivariate t distribution. $\tilde{T}$ and $\overline{t}$ are dependent and if $\mu = 0$, their joint density function is (Cornish (1954b))

$$f(\tilde{T},\overline{t}) = C \frac{|\tilde{T}|^{\frac{1}{2}(k-p-1)}}{\left(1 + \frac{\overline{t}'(K^{-1}\Sigma)^{-1}\overline{t}+\mathrm{tr}\Sigma^{-1}\tilde{T}}{n}\right)^{\frac{1}{2}(n+pk)}} \qquad 5.2.11$$

where

$$C = \frac{\Gamma(\frac{1}{2}(n+pk))}{n^{\frac{1}{2}pk}\pi^{\frac{1}{2}p}|K^{-1}\Sigma|^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}k}\Gamma_p(\frac{1}{2}k)}$$

The marginal density functions of $\overline{t}$ and $\tilde{T}$ can be found by integration in 5.2.11, but they are more easily derived from the synthetic representations, 5.2.7 and 5.2.9. Clearly since $\overline{X} \sim N_p(\mu,K^{-1}\Sigma)$ independently of $Z \sim \chi^2_n$

$$\overline{t} \sim t_p(n,K^{-1}\Sigma,\mu) \qquad 5.2.12$$

If $\mu = 0$, then $\overline{t} \sim t_p(n,K^{-1}\Sigma)$.
The density function of $\tilde{T}$ is given in the next theorem.

Theorem 5.2.1 (Cornish 1954b)

Let $A \sim W_p(\Sigma,k)$ independently of $Z \sim \chi^2_n$. If

$$\tilde{T} = \left(\frac{Z}{n}\right)^{-1} A,$$

then

$$f(\tilde{T}) = \frac{\Gamma(\frac{1}{2}(n+pk))}{\Gamma(\frac{1}{2}n)\Gamma_p(\frac{1}{2}k)|n\Sigma|^{\frac{1}{2}k}} \frac{|\tilde{T}|^{\frac{1}{2}(k-p-1)}}{(1+tr(n\Sigma)^{-1}\tilde{T})^{\frac{1}{2}(n+pk)}} \qquad 5.2.13$$

for $\tilde{T} > 0$.

Proof: From definition 2.5.4

$$f(A) = \frac{|A|^{\frac{1}{2}(k-p-1)}etr(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}kp}\Gamma_p(\frac{1}{2}k)|\Sigma|^{\frac{1}{2}k}} \qquad 5.2.14$$

and from definition 2.4.1

$$f(z) = \frac{z^{\frac{1}{2}n-1}e^{-\frac{1}{2}z}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} \qquad 5.2.15$$

Since A and Z are independent,

$$f(A,z) \propto z^{\frac{1}{2}n-1}|A|^{\frac{1}{2}(k-p-1)}\exp(-\tfrac{1}{2}[z+tr\Sigma^{-1}A]) \qquad 5.2.16$$

where the constant of proportionality is

$$\left(2^{\frac{1}{2}(n+pk)}\Gamma(\tfrac{1}{2}n)\Gamma_p(\tfrac{1}{2}k)|\Sigma|^{\frac{1}{2}k}\right)^{-1} \qquad 5.2.17$$

Transforming from A to $\tilde{T}$ in 5.2.16 by

$$A = \left(\frac{Z}{n}\right)\tilde{T} \qquad 5.2.18$$

with Jacobian

$$J(A \to \tilde{T}) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p(p+1)} \qquad 5.2.19$$

and noting that

$$|A|^{\frac{1}{2}(k-p-1)} = \left|\frac{Z}{n}\tilde{T}\right|^{\frac{1}{2}(k-p-1)}$$

$$= \left(\frac{Z}{n}\right)^{\frac{1}{2}p(k-p-1)} |\tilde{T}|^{\frac{1}{2}(k-p-1)} \qquad 5.2.20$$

$$\mathrm{tr}\frac{\Sigma^{-1}\tilde{T}}{n} = \mathrm{tr}(n\Sigma)^{-1}\tilde{T} \qquad 5.2.21$$

$$f(\tilde{T},z) \propto n^{-\frac{1}{2}pk}|\tilde{T}|^{\frac{1}{2}(k-p-1)} z^{\frac{1}{2}(n+pk)-1} \exp(-\tfrac{1}{2}[1+\mathrm{tr}(n\Sigma)^{-1}\tilde{T}]z) \qquad 5.2.22$$

Integrating 5.2.21 with respect to z, using Theorem 2.2.1, gives

$$f(\tilde{T}) \propto n^{-\frac{1}{2}pk}|\tilde{T}|^{\frac{1}{2}(k-p-1)} 2^{\frac{1}{2}(n+pk)} \Gamma(\tfrac{1}{2}(n+pk))(1+\mathrm{tr}(n\Sigma)^{-1}\tilde{T})^{-\frac{1}{2}(n+pk)} \qquad 5.2.23$$

Evaluating the constant terms from 5.2.17 and 5.2.22 and noting in the evaluation that

$$n^{\frac{1}{2}pk}|\Sigma|^{\frac{1}{2}k} = |n\Sigma|^{\frac{1}{2}k} \qquad 5.2.24$$

gives $$f(\tilde{T}) = \frac{\Gamma(\frac{1}{2}(n+pk))}{\Gamma(\frac{1}{2}n)\Gamma_p(\frac{1}{2}k)|n\Sigma|^{\frac{1}{2}k}} \; \frac{|\tilde{T}|^{\frac{1}{2}(k-p-1)}}{(1+\mathrm{tr}(n\Sigma)^{-1}\tilde{T})^{\frac{1}{2}(n+pk)}}$$

as stated in the theorem.

It is clear that as $n \to \infty$, the limiting distribution of $\tilde{T}$ is $W_p(\Sigma,k)$.

The parameters of the distribution are

(i) p - the dimension of $\tilde{T}$

(ii) k - the degrees of freedom of A

(iii) Σ - the covariance matrix of A

(iv) n - the degrees of freedom of Z.

Definition 5.2.1

The central distribution of $\tilde{T}$ will be denoted by $TW(p,k,n,\Sigma)$.

Some other properties of $\tilde{T}$ which can be deduced from the synthetic representation are given in the next two theorems.

Theorem 5.2.2

If $\tilde{T} \sim TW(p,k,n,\Sigma)$, then

$$E(\tilde{T}) = \frac{nk}{n-2}\Sigma \qquad 5.2.25$$

Proof:

$$E(\tilde{T}) = nE(Z^{-1})E(A)$$

From Anderson (1958), page 53, $E(A) = k\Sigma$.
By Corollary 2.4.1, $E(Z^{-1}) = (n-2)^{-1}$, and the result follows.

Theorem 5.2.3

If $|\tilde{T}|$ is the determinant of $\tilde{T}$, then

$$E\left(|\tilde{T}|^h\right) = \frac{|n\Sigma|^h\Gamma(\frac{1}{2}(n-2ph))}{\Gamma(\frac{1}{2}n)} \prod_{i=1}^{p} \frac{\Gamma(\frac{1}{2}(K-i)+h)}{\Gamma(\frac{1}{2}(K-i))} \qquad 5.2.26$$

where $K = k+1$ and $n > 2ph$.

Proof:

$$E\left(|\tilde{T}|^h\right) = E\left(\left|\left(\frac{Z}{n}\right)^{-1}A\right|^h\right)$$

$$= n^{ph}E(Z^{-ph})E(|A|) \qquad 5.2.27$$

From Anderson (1958), page 171

$$E(|A|) = 2^{hp}|\Sigma|^h \prod_{i=1}^{p} \frac{\Gamma(\frac{1}{2}(K-i)+h)}{\Gamma(\frac{1}{2}(K-i))} \qquad 5.2.28$$

By Theorem 2.4.2

$$E(Z^{-ph}) = \frac{\Gamma(\frac{1}{2}(n-2ph))}{2^{hp}\Gamma(\frac{1}{2}n)} \qquad 5.2.29$$

Substituting 5.2.28 and 5.2.29 in 5.2.27 and noting that $n^{ph}|\Sigma|^h = |n\Sigma|^h$, the result follows.

5.3 THREE NON-CENTRAL DISTRIBUTIONS OF $\tilde{T}$

We first consider the distribution of $\bar{t}$ and $\tilde{T}$ when Z has a non-central chi-squared distribution.

Let $\bar{X}$ and A be as defined in section 5.2. Clearly, if $Z \sim \chi^2_n(\lambda)$ and $\bar{X} \sim N_p(\mu, K^{-1}\Sigma)$ independently of Z, then

$$\bar{t} = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} \bar{X}$$

has (i) a doubly non-central multivariate t distribution if $\mu \neq 0$.

i.e. $\bar{t} \sim t_p(n, K^{-1}\Sigma, \mu, \lambda)$

(ii) a lower non-central multivariate t distribution if $\mu = 0$.

i.e. $\bar{t} \sim t_p(n, K^{-1}\Sigma, 0, \lambda)$.

The distribution of $\tilde{T}$ does not depend on μ and is the same in both cases.

Theorem 5.3.1

Let $A \sim W_p(\Sigma, k)$ *and* $Z \sim \chi^2_n(\lambda)$, *independently of A.*

If $\tilde{T} = \left(\frac{Z}{n}\right)^{-1} A$, *then*

$f(\tilde{T})$ =

$$\sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}\Gamma(\frac{1}{2}(n+pk+2i))}{i!\Gamma(\frac{1}{2}(n+pk)\Gamma_p(\frac{1}{2}k)|n\Sigma|^{\frac{1}{2}k}} |\tilde{T}|^{\frac{1}{2}(k-p-1)} (1+tr(n\Sigma)^{-1}\tilde{T})^{-\frac{1}{2}(n+2i+pk)} \qquad 5.3.1$$

for $\tilde{T} > 0$.

Proof:

From definition 2.5.4

$$f(A) = \frac{|A|^{\frac{1}{2}(k-p-1)}etr(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}kp}\Gamma_p(\frac{1}{2}k)|\Sigma|^{\frac{1}{2}k}}$$

From definition 2.4.3

$$f(z) = \sum_{i=o}^{\infty} \frac{e^{-\lambda}\lambda^{i}z^{\frac{1}{2}(n+2i)-1}e^{-\frac{1}{2}z}}{i!2^{\frac{1}{2}(n+2i)}\Gamma(\frac{1}{2}(n+2i))}$$

$$f(z,A) \propto \sum_{i=o}^{\infty} \phi(i)z^{\frac{1}{2}(n+2i)-1}|A|^{\frac{1}{2}(k-p-1)}\exp(-\tfrac{1}{2}[z+tr\Sigma^{-1}A]) \qquad 5.3.2$$

where the constant of proportionality is

$$(2^{\frac{1}{2}kp}|\Sigma|^{\frac{1}{2}k}\Gamma_p(\tfrac{1}{2}k))^{-1} \qquad 5.3.3$$

and

$$\phi(i) = \frac{e^{-\lambda}\lambda^{i}}{2^{\frac{1}{2}(n+2i)}\Gamma(\frac{1}{2}(n+2i))} \qquad 5.3.4$$

Transforming from A to $\tilde{T}$ in 5.3.2 by

$$A = \left(\frac{Z}{n}\right)\tilde{T}$$

with Jacobian

$$J(A \to \tilde{T}) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p(p+1)}$$

and using 5.2.20 and 5.2.21 gives

$f(z,\tilde{T}) \propto$

$$\sum_{i=0}^{\infty} \phi(i) n^{-\frac{1}{2}pk} |\tilde{T}|^{\frac{1}{2}(k-p-1)} z^{\frac{1}{2}(n+2i+pk)-1} \exp(-\tfrac{1}{2}[1+tr(n\Sigma)^{-1}\tilde{T}]z) \qquad 5.3.5$$

Integrating 5.3.5 termwise over z, using Theorem 2.2.1, shows that the i^{th} integral has the value

$$\frac{2^{\frac{1}{2}(n+2i+pk)}\Gamma(\frac{1}{2}(n+2i+pk))}{(1+tr(n\Sigma)^{-1}\tilde{T})^{\frac{1}{2}(n+2i+pk)}} \qquad 5.3.6$$

Evaluating the constant terms from 5.3.3 to 5.3.6, noting that $n^{\frac{1}{2}pk}|\Sigma|^{\frac{1}{2}k} = |n\Sigma|^{\frac{1}{2}k}$, gives $f(\tilde{T})$ as stated in the theorem.

We could call this distribution the lower non-central $\tilde{T}$ distribution, since the non-centrality occurs in the denominator of the synthetic representation.

Definition 5.3.1

The lower non-central distribution of $\tilde{T}$ will be denoted by $TW(p,k,n,\Sigma,\lambda)$.

If $X_{(1)},\ldots X_{(K)}$ is a sample from a $N_p(\mu,\sigma^2 R)$ population, and A is the Wishart matrix of the sample then $A \sim W_p(k,\sigma^2 R)$.

Suppose that σ^2 is unknown, and an estimate of σ^2, s^2 say, with n degrees of freedom, based on another sample, is used. If, in fact, this sample

(i) has a different variance, σ'^2 say, and

(ii) is a non-homogeneous sample - i.e. the true mean-values of the observations vary from observation to observation -

then $$Z = \frac{ns^2}{\sigma'^2} \sim \chi^2_n(\lambda) .$$

Consider $$\tilde{T} = \frac{1}{s^2} A = \left(\frac{Z}{n}\right)^{-1} \left(\frac{1}{\sigma'^2}\right) A .$$

Since $$\frac{1}{\sigma'^2} A \sim W_p(k, \left(\frac{\sigma}{\sigma'}\right)^2 R)$$

it follows from Theorem 5.3.1 that

$$\tilde{T} \sim TW(p,k,n,\left(\frac{\sigma}{\sigma'}\right)^2 R,\lambda) .$$

If the second sample was non-homogeneous, but with the same true variance σ^2, then

$$\tilde{T} = \frac{1}{s^2}A = \left(\frac{Z}{n}\right)^{-1} \left(\frac{1}{\sigma^2}\right) A$$

and $\frac{1}{\sigma^2}A \sim W(k,R)$. Hence, in this case,

$$\tilde{T} \sim TW(p,k,n,R,\lambda) .$$

Like the lower non-central multivariate t distribution, the density of $\tilde{T}$ can be written as an infinite series of central $\tilde{T}$ densities, each weighted by a term from the Poisson distribution. In both cases this property results from the analogous property of the underlying non-central chi-squared distribution.

Theorem 5.3.2

If $\tilde{T} \sim TW(p,k,n,\Sigma,\lambda)$ then

$$f(\tilde{T}) = \sum_{i=o}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} g(\tilde{T};\ p,k,n+2i,\frac{n}{n+2i}\Sigma) \qquad 5.3.7$$

where $g(\ )$ is the density function of the central $TW(p,k,n+2i,\frac{n}{n+2i}\Sigma)$ distribution.

Proof

$f(\tilde{T}) =$

$$\sum_{i=o}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} \left\{ \frac{\Gamma(\frac{1}{2}(n+2i+pk))|\tilde{T}|^{\frac{1}{2}(k-p-1)}}{\Gamma(\frac{1}{2}(n+2i))\Gamma_p(\frac{1}{2}k)|n\Sigma|^{\frac{1}{2}k}} (1+tr(n\Sigma)^{-1}\tilde{T})^{-\frac{1}{2}(n+2i+pk)} \right\}$$

Let $\phi = \frac{n}{n+2i}\Sigma$;

then the terms in brackets can be written as

$$\frac{\Gamma(\frac{1}{2}(n+2i+pk))|\tilde{T}|^{\frac{1}{2}(k-p-1)}}{\Gamma(\frac{1}{2}(n+2i))\Gamma_p(\frac{1}{2}k)|(n+2i)\phi|^{\frac{1}{2}k}}(1+tr((n+2i)\phi)^{-1}\tilde{T})^{-\frac{1}{2}(n+2i+pk)}$$

which from Theorem 5.2.1 we recognise as the density function of a central $\tilde{T}$ distribution, namely $TW(p,k,n+2i,\frac{n}{n+2i}\Sigma)$.

$E(\tilde{T})$ and $E(|\tilde{T}|^{n})$ for the lower non-central case could be found by exploiting the independence of Z and A and using the expressions for the inverted moments of the non-central chi-squared distribution given in Theorem 2.4.3. Alternatively these expected values can be found usig Theorem 5.3.2 as shown in the next two theorems.

Theorem 5.3.3

If $T \sim TW(p,k,n,\Sigma,\lambda)$, then

$$E(\tilde{T}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} \frac{nk}{(n+2i-2)}\Sigma \qquad 5.3.8$$

Proof:

From Theorem 5.3.2

$$E(\tilde{T}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} E(\tilde{T}_i)$$

where $\tilde{T}_i \sim TW(p,k,n+2i,\frac{n}{n+2i}\Sigma)$.

From Theorem 5.2.2

$$E(\tilde{T}_i) = \frac{(n+2i)k}{n+2i-2}\left(\frac{n}{n+2i}\Sigma\right)$$

Hence $$E(\tilde{T}) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} \frac{nk}{(n+2i-2)}\Sigma$$

Theorem 5.3.4

If $\tilde{T} \sim TW(p,k,n,\Sigma,\lambda)$ and $|\tilde{T}|$ is the determinant of $\tilde{T}$ then $E(|\tilde{T}|^h)$ =

$$|n\Sigma|^h \prod_{j=1}^{p} \frac{\Gamma(\frac{1}{2}(K-j)+h)}{\Gamma(\frac{1}{2}(K-j))} \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} \frac{\Gamma(\frac{1}{2}(n+2i-2ph))}{\Gamma(\frac{1}{2}(n+2i))} \qquad 5.3.9$$

Proof: From Theorem 5.3.2

$$E(|\tilde{T}|^h) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} E(|\tilde{T}_i|^h) \qquad 5.3.10$$

where $\quad T_i \sim TW(p,k,n+2i,\frac{n}{n+2i}\Sigma)$

From Theorem 5.2.3

$$E(|\tilde{T}_i|^h) = \left|(n+2i)\left(\frac{n}{n+2i}\Sigma\right)\right|^h \frac{\Gamma(\frac{1}{2}(n+2i-2ph))}{\Gamma(\frac{1}{2}(n+2i))} \prod_{j=1}^{p} \phi(j)$$

where $$\phi(j) = \prod_{j=1}^{p} \frac{\Gamma(\frac{1}{2}(K-j)+h)}{\Gamma(\frac{1}{2}(K-j))} \qquad 5.3.11$$

Substituting in 5.3.10, noting the terms independent of i, gives 5.3.9.

The infinite series in 5.3.9 can be expressed in terms of the hypergeometric function as

$$\frac{e^{-\lambda}\Gamma(\frac{1}{2}(n-2ph))}{\Gamma(\frac{1}{2}n)} {}_1F_1\left(\tfrac{1}{2}(n-2ph);\ \tfrac{1}{2}n;\ \lambda\right) \qquad 5.3.12$$

or using Kummer's first formula (Rainville 1960, page 125)

$$\frac{\Gamma(\frac{1}{2}n-ph)}{\Gamma(\frac{1}{2}n)} {}_1F_1\left(ph, \tfrac{1}{2}n;\ -\lambda\right) . \qquad 5.3.13$$

We now consider an upper non-central $\tilde{T}$ distribution, namely, the distribution of $\tilde{T}$ when Z has a central chi-squared distribution but A has a non-central Wishart distribution. Kshirshagar (1960) gave the distribution for the linear non-central Wishart distribution. In the next theorem we give the distribution for the general case.

Theorem 5.3.5

Let $A \sim W_p(\Sigma,k,\Omega)$ *independently of* $Z \sim \chi^2_n$.

If $\tilde{T} = \left(\frac{Z}{n}\right)^{-1} A$

then

$$f(\tilde{T}) = \frac{etr(-\Omega)|\tilde{T}|^{\frac{1}{2}(k-p-1)}}{\Gamma(\frac{1}{2}n)\Gamma_p(\frac{1}{2}k)|n\Sigma|^{\frac{1}{2}k}} \sum_{i=o}^{\infty} \frac{\Gamma(\frac{1}{2}(n+2i+pk))}{(1+tr(n\Sigma)^{-1}\tilde{T})^{\frac{1}{2}(n+2i+pk)}}$$

$$\times \sum_{\iota} \frac{C_{\iota}[(n\Sigma)^{-1}\Omega\tilde{T}]}{(\frac{1}{2}k)_{\iota} \quad i!} \qquad 5.3.14$$

Proof:

From definition 2.5.5,

$$f(A) = \frac{etr(-\Omega)|A|^{\frac{1}{2}(k-p-1)}etr(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}kp}\Gamma_p(\frac{1}{2}k)|\Sigma|^{\frac{1}{2}k}} {}_0F_1(\tfrac{1}{2}k;\ \tfrac{1}{2}\Sigma^{-1}\Omega A)$$

Expanding the ${}_0F_1$ series using definition 2.3.2,

$$f(A) = \frac{etr(-\Omega)|A|^{\frac{1}{2}(k-p-1)}etr(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}kp}\Gamma_p(\frac{1}{2}k)|\Sigma|^{\frac{1}{2}k}} \sum_{i=o}^{\infty} \sum_{\iota} \frac{C_{\iota}(\frac{1}{2}\Sigma^{-1}\Omega A)}{(\frac{1}{2}k)_{\iota} \quad i!}$$

From definition 2.4.1

$$f(z) = \frac{z^{\frac{1}{2}n-1}e^{-\frac{1}{2}z}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}$$

$$f(A,z) \propto |A|^{\frac{1}{2}(k-p-1)}z^{\frac{1}{2}n-1}\exp(-\tfrac{1}{2}[z+tr\Sigma^{-1}A]) \sum_{i=o}^{\infty} \sum_{\iota} \frac{C_{\iota}(\frac{1}{2}\Sigma^{-1}\Omega A)}{(\frac{1}{2}k)_{\iota} \quad i!} \qquad 5.3.15$$

where the constant of proportionality is

$$\frac{etr(-\Omega)}{\Gamma_p(\frac{1}{2}k)|\Sigma|^{\frac{1}{2}k}2^{\frac{1}{2}(n+kp)}\Gamma(\frac{1}{2}n)} \qquad 5.3.16$$

Transforming in 5.3.15 by

$$A = \left(\frac{Z}{n}\right)\tilde{T}$$

with Jacobian

$$J(A \to \tilde{T}) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p(p+1)}$$

and recalling equations 5.2.20, 5.2.21 and 5.2.24,

$$f(\tilde{T},z) \propto n^{-\frac{1}{2}pk}|\tilde{T}|^{\frac{1}{2}(k-p-1)}z^{\frac{1}{2}(n+pk)-1}\exp(-\tfrac{1}{2}[1+tr(n\Sigma)^{-1}\tilde{T}]z)$$

$$\times \sum_{i=o}^{\infty}\sum_{\iota}\frac{C_\iota((\Sigma^{-1}\Omega\tilde{T}z)/2n)}{(\frac{1}{2}k)_\iota \; i!} \qquad 5.3.17$$

Now by definition 2.3.3(iii) the zonal polynomial in 5.3.17 can be written as

$$C_\iota((\Sigma^{-1}\Omega\tilde{T}z)/2n) = \frac{z^i}{2^i}C_\iota((n\Sigma)^{-1}\Omega\tilde{T}) \qquad 5.3.18$$

since $\left(\frac{z}{2}\right)$ is a scalar.

Substituting 5.3.18 in 5.3.17 and grouping terms in z, gives

$$f(\tilde{T},z) \propto n^{-\frac{1}{2}pk}|\tilde{T}|^{\frac{1}{2}(k-p-1)}\sum_{i=o}^{\infty} z^{\frac{1}{2}(n+2i+pk)-1}\exp(-\tfrac{1}{2}[1+tr(n\Sigma)^{-1}\tilde{T}]z)$$

$$\times \sum_{\iota}\frac{C_\iota((n\Sigma)^{-1}\Omega\tilde{T})}{(\frac{1}{2}k)_\iota 2^i \; i!} \qquad 5.3.19$$

To obtain $f(\tilde{T})$ we integrate termwise over z. Using Theorem 2.2.1, the i^{th} integral has the value

$$\frac{2^{\frac{1}{2}(n+pk+2i)}\Gamma(\frac{1}{2}(n+pk+2i))}{(1+tr(n\Sigma)^{-1}\tilde{T})^{\frac{1}{2}(n+pk+2i)}} \qquad 5.3.20$$

Evaluating the constant terms from 5.3.16, 5.3.19 and 5.3.20 gives

$$f(\tilde{T}) = \frac{\operatorname{etr}(-\Omega)|\tilde{T}|^{\frac{1}{2}(k-p-1)}}{\Gamma_p(\frac{1}{2}k)|n\Sigma|^{\frac{1}{2}k}\Gamma(\frac{1}{2}n)} \sum_{i=o}^{\infty} \frac{\Gamma(\frac{1}{2}(n+pk+2i))}{(1+\operatorname{tr}(n\Sigma)^{-1}\tilde{T})^{\frac{1}{2}(n+pk+2i)}}$$

$$\times \sum_{\iota} \frac{C_{\iota}((n\Sigma)^{-1}\Omega\tilde{T})}{(\frac{1}{2}k)_{\iota}\, i!} \qquad 5.3.21$$

We call this distribution the upper non-central $\tilde{T}$ distribution.

Definition 5.3.2

The upper non-central $\tilde{T}$ distribution will be denoted by

$\tilde{T}W(p,k,n,\Sigma,o,\Omega)$

As remarked in the introduction to this chapter, the case of most practical interest would be the doubly non-central $\tilde{T}$ distribution, when both A and Z have non-central distributions. Such a distribution would arise naturally if we suppose that A and Z are independent statistics constructed from the same sample. Then if A has a non-central Wishart distribution, so will Z, because the true means of the sample vary from observation to observation. In principle the doubly non-central $\tilde{T}$ distribution could be derived as in Theorem 5.3.5 with a non-central chi-squared distribution replacing the central distribution. However, the distribution would be extremely complicated. We shall derive only the "linear case" - i.e. when A has a linear non-central Wishart distribution, with covariance matrix I, k degrees of freedom and non-centrality parameter κ^2.

Suppose that $X_{(\alpha)}$, $\alpha = 1...K$ is a random sample from a normal population, where $X_{(\alpha)} \sim N_p(\mu_{(\alpha)}, \sigma^2 I)$.
Let A^* be the Wishart matrix of the sample and let s^2 be an estimate of the unknown variance σ^2 which is independent of A^* and is such that

$$\frac{ns^2}{\sigma^2} = Z \sim \chi^2_n(\lambda) \qquad 5.3.22$$

where λ is a function of the $\mu_{(\alpha)}$'s and σ^2.

$$\text{Define} \quad \tilde{T} = \frac{1}{s^2} A^*$$

$$= \frac{\sigma^2}{s^2} \left(\frac{1}{\sigma^2} A^*\right)$$

$$= \left(\frac{Z}{n}\right)^{-1} A \qquad 5.3.23$$

Then in general $A \sim W_p(k, I, \Omega)$.

Suppose, however, that the configuration of means is such that A has a linear non-central Wishart distribution, then from definition 2.5.6,

$$f(A) = \frac{e^{-\frac{1}{2}\kappa^2} |A|^{\frac{1}{2}(k-p-1)} etr(-\frac{1}{2}A)}{2^{\frac{1}{2}kp}\Gamma_p(\frac{1}{2}k)} \sum_{j=o}^{\infty} \frac{(\frac{1}{2}k^2)^j a_{11}^j \Gamma(\frac{1}{2}k)}{j!\ 2^j \Gamma(\frac{1}{2}k+j)} \qquad 5.3.24$$

From definition 2.4.3

$$f(z) = \sum_{i=o}^{\infty} \frac{e^{-\lambda}\lambda^i z^{\frac{1}{2}(n+2i)-1} e^{-\frac{1}{2}z}}{i!\ 2^{\frac{1}{2}(n+2i)}\Gamma(\frac{1}{2}(n+2i))} \qquad 5.3.25$$

Then f(A,z) is proportional to

$$\sum_{i,j} \phi(i)\Theta(j)|A|^{\frac{1}{2}(k-p-1)} a_{11}^j z^{\frac{1}{2}(n+2i)-1} \exp(-\tfrac{1}{2}[z+trA]) \qquad 5.3.26$$

where the constant of proportionality is

$$\frac{\exp(-\frac{1}{2}\kappa^2)}{2^{\frac{1}{2}kp}\Gamma_p(\frac{1}{2}k)} \qquad 5.3.27$$

$$\phi(i) = \frac{e^{-\lambda}\lambda^i}{i!\; 2^{\frac{1}{2}(n+2i)}\Gamma(\frac{1}{2}(n+2i))} \qquad 5.3.28$$

$$\Theta(j) = \frac{(\frac{1}{2}\kappa^2)^j\Gamma(\frac{1}{2}k)}{j!\; 2^j\Gamma(\frac{1}{2}k+j)} \qquad 5.3.29$$

Transform from A to $\tilde{T}$ by

$$A = \left(\frac{Z}{n}\right)\tilde{T} \quad \text{with} \quad J(A \to T) = \left(\frac{Z}{n}\right)^{\frac{1}{2}p(p+1)}$$

Then $$a_{11} = \left(\frac{Z}{n}\right)t_{11} \qquad 5.3.30$$

and using 5.2.20, we obtain

$$f(\tilde{T},z) \propto \sum_{i,j} \phi(i)\Theta(j)n^{-\frac{1}{2}(pk+2j)}|\tilde{T}|^{\frac{1}{2}(k-p-1)}t_{11}^j$$

$$\times\; z^{\frac{1}{2}(n+pk+2j+2i)-1}\exp(-\tfrac{1}{2}\left(1+\mathrm{tr}\frac{\tilde{T}}{n}\right)z) \qquad 5.3.31$$

Integrating termwise over z, using Theorem 2.2.1, gives the $(i,j)^{th}$ integral as

$$\frac{2^{\frac{1}{2}(n+pk+2j+2i)}\Gamma(\frac{1}{2}(n+pk+2j+2i))}{\left(1+\mathrm{tr}\frac{\tilde{T}}{n}\right)^{\frac{1}{2}(n+pk+2j+2i)}} \qquad 5.3.32$$

Substituting for $\phi(i)$ and $\Theta(j)$ and evaluating the constant terms gives

$$f(\tilde{T}) = \frac{\exp(-\frac{1}{2}[\kappa^2+2\lambda])\Gamma(\frac{1}{2}k)}{\Gamma_p(\frac{1}{2}k)n^{\frac{1}{2}pk}} \times \sum_{i,j} \psi(i,j)|\tilde{T}|^{\frac{1}{2}(k-p-1)}\left(1+\mathrm{tr}\frac{\tilde{T}}{n}\right)^{-\frac{1}{2}(n+pk+2j+2i)}\left(\frac{t_{11}}{n}\right)^j \qquad 5.3.33$$

where

$$\psi(i,j) = \frac{\Gamma(\frac{1}{2}(n+pk+2j+2i))\lambda^i(\frac{1}{2}\kappa^2)^j}{\Gamma(\frac{1}{2}(n+2i))\Gamma(\frac{1}{2}(k+2j))n^{\frac{1}{2}pk}i!j!}$$

Theorem 5.3.7

If $A \sim W_p(k,I,\kappa^2)$ and $Z \sim \chi^2{}_n(\lambda)$, then $\tilde{T} = \left(\frac{Z}{n}\right)^{-1}A$ has a linear doubly non-central $\tilde{T}$ distribution with density function given by 5.3.3.

5.4 THE DISTRIBUTION OF THE ROOTS OF THE CENTRAL $\tilde{T}$

Returning to the central $\tilde{T}$ distribution, we see that if $\Sigma = I$, then from Theorem 5.2.1

$$f(\tilde{T}) \propto |\tilde{T}|^{\frac{1}{2}(k-p-1)}\left(1+\frac{\mathrm{tr}\tilde{T}}{n}\right)^{-\frac{1}{2}(n+pk)} \qquad 5.4.1$$

If $\lambda_1 > \lambda_2 \ldots > \lambda_p$ are the characteristic roots of $\tilde{T}$, then

$$|\tilde{T}| = \prod_{i=1}^{p} \lambda_i \quad \text{and} \quad \frac{1}{n}\mathrm{tr}\tilde{T} = \frac{1}{n}\sum_{i=1}^{p} \lambda_i \qquad 5.4.2$$

Hence $f(\tilde{T})$ is a function of the characteristic roots of $\tilde{T}$ and can be written as

$$g(\lambda_1,\ldots\lambda_p) \propto \left(1+\sum_{i=1}^{p}\frac{\lambda_i}{n}\right)^{-\frac{1}{2}(n+pk)} \prod_{i=1}^{p}\lambda_i^{\frac{1}{2}(k-p-1)} \qquad 5.4.3$$

where the constant of proportionality is

$$\frac{\Gamma(\frac{1}{2}(n+pk))}{n^{\frac{1}{2}pk}\Gamma(\frac{1}{2}n)\pi^{\frac{1}{4}p(p-1)}\prod_{i=1}^{p}\Gamma(\frac{1}{2}(k+1-i))} \qquad 5.4.4$$

We can now apply the following theorem (Anderson 1958, page 318), which states that:

> If the symmetric matrix B has a density of the form $g(\lambda_1,\ldots\lambda_p)$, where $\lambda_1 >\ldots> \lambda_p$, are the characteristic roots of B, then the joint distribution of the roots is
>
> $$\frac{\pi^{\frac{1}{4}p(p+1)}g(\lambda_1,\ldots\lambda_p)\prod_{i<j}(\lambda_i-\lambda_j)}{\prod_{i=1}^{p}\Gamma(\frac{1}{2}(p-i+1))}$$

Thus it follows that the joint density function of the ordered roots of $\tilde{T}$ is

$f(\lambda_1\ldots\lambda_p)$ =

$$\frac{\pi^{\frac{1}{2}p}\Gamma(\frac{1}{2}(n+pk))\left(1+\sum_{i=1}^{p}\frac{\lambda_i}{n}\right)^{-\frac{1}{2}(n+pk)}\prod_{i=1}^{p}\lambda_i^{\frac{1}{2}(k-p-1)}\prod_{i<j}(\lambda_i-\lambda_j)}{n^{\frac{1}{2}pk}\Gamma(\frac{1}{2}n)\prod_{i=1}^{p}\Gamma(\frac{1}{2}(k+1-i))\Gamma(\frac{1}{2}(p+1-i))} \qquad 5.4.5$$

We summarise this result in

Theorem 5.4.1

If $A \sim W(I,n)$ *and* $Z \sim \chi^2_n$ *independently of* A, *then the density function of the ordered characteristic roots of* $\tilde{T} = \left(\frac{Z}{n}\right)^{-1} A$ *is given by 5.4.5.*

CHAPTER 6

THE MATRIX T DISTRIBUTION

6.1 INTRODUCTION

In previous chapters we defined the central multivariate t as the distribution of the random vector

$$t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} Y \qquad 6.1.1$$

where $Y \sim N(0,\Sigma)$ independently of $Z \sim \chi^2_n$. (Previously we used X instead of Y). The density function of t is

$$\frac{\Gamma(\frac{1}{2}(n+p))}{(n\pi)^{\frac{1}{2}p}\ \Gamma(\frac{1}{2}n)\ |\Sigma|^{\frac{1}{2}}} \left(1 + \frac{t'\Sigma^{-1}t}{n}\right)^{-\frac{1}{2}(n+p)} \qquad 6.1.2$$

With a slight rearrangement 6.1.2 can be written as

$$\frac{\Gamma(\frac{1}{2}(n+p))\quad n^{\frac{1}{2}n}}{\pi^{\frac{1}{2}p}\ \Gamma(\frac{1}{2}n)\ |\Sigma|^{\frac{1}{2}}} \quad (n + t'\Sigma^{-1}t)^{-\frac{1}{2}(n+p)} \qquad 6.1.3$$

Ando and Kaufmann (1965) showed that the density function 6.1.3 also arises from the synthetic representation

$$t = (U^{\frac{1}{2}'})^{-1} X \qquad 6.1.4$$

where $U = U^{\frac{1}{2}}U^{\frac{1}{2}'}$ is a p×p matrix which has a Wishart distribution with covariance matrix Σ^{-1} and (n+p-1) degrees of freedom (i.e. $U \sim W_p(\Sigma^{-1}, n+p-1)$; $U^{\frac{1}{2}}$ is the symmetric square root of U (see definition 2.1.4) and $X \sim N_p(0, nI_p)$ independently of U. (Often in applications Σ is the correlation matrix.)

The equivalence of the two synthetic representations 6.1.1 and 6.1.4 can be demonstrated as follows:

If $U \sim W_p(\Sigma^{-1}, n+p-1)$ then from definition 2.5.4,

$$f(U) = \frac{|U|^{\frac{1}{2}(n-2)} \, etr(-\frac{1}{2}\Sigma U)}{2^{\frac{1}{2}(n+p-1)p} \, \Gamma_p(\frac{1}{2}(n+p-1)) \, |\Sigma^{-1}|^{\frac{1}{2}(n+p-1)}} \qquad 6.1.5$$

If $X \sim N(0,nI_p)$, then from definition 2.5.2

$$f(X) = \frac{\exp(-\frac{1}{2}X'(nI_p)^{-1} X)}{(2\pi)^{\frac{1}{2}p} \, n^{\frac{1}{2}p}} \qquad 6.1.6$$

Now $X'(nI_p)^{-1} X = tr\, X'(nI_p)^{-1} X = n^{-1} tr\, XX'$. 6.1.7

Since X and U are independent

$$f(XU) \propto |U|^{\frac{1}{2}(n-2)} \, etr(-\frac{1}{2}[\Sigma U + n^{-1}XX']) \qquad 6.1.8$$

Transform from X to t in 6.1.8 by

$$X = (U^{\frac{1}{2}'})t \qquad 6.1.9$$

with Jacobian, $J(X \rightarrow t) = |U|^{\frac{1}{2}}$ 6.1.10

Then the joint density of t and U is

$$f(t,U) \propto |U|^{\frac{1}{2}(n-1)} \, etr(-\frac{1}{2}[\Sigma + n^{-1}tt']U) \qquad 6.1.11$$

where the constant of proportionality is

$$\left\{2^{\frac{1}{2}(n+p)p} \, \pi^{\frac{1}{2}p} \, |\Sigma|^{-\frac{1}{2}(n+p-1)} \, n^{\frac{1}{2}p} \, \Gamma_p(\frac{1}{2}(n+p-1))\right\}^{-1} \qquad 6.1.12$$

To obtain f(t), 6.1.11 must be integrated over U.

Now 6.1.11 is the kernel of a $W_p((\Sigma+n^{-1}tt')^{-1},n+p)$ density, hence by Theorem 2.5.2, the required integral has the value

$$2^{\frac{1}{2}(n+p)p}\,\Gamma_p(\tfrac{1}{2}(n+p))\,|\Sigma + n^{-1}tt'|^{-\frac{1}{2}(n+p)} \qquad 6.1.13$$

Consider now the determinant

$$|A| = \begin{vmatrix} \Sigma & -t \\ t' & n \end{vmatrix} \qquad 6.1.14$$

Using Theorem 2.1.1,

$$|A| = |\Sigma|\,|n + t'\Sigma^{-1}t| = n|\Sigma + n^{-1}tt'| \qquad 6.1.15$$

Hence $$|\Sigma^{-1}+ n^{-1}tt'| = n^{-1}|\Sigma|\,|n + t'\Sigma^{-1}t| \qquad 6.1.16$$

Therefore from 6.1.12, 6.1.13 and 6.1.16, noting also by expanding the multivariate gamma functions,

$$\frac{\Gamma_p(\frac{1}{2}(n+p))}{\Gamma_p(\frac{1}{2}(n+p-1))} = \frac{\Gamma(\frac{1}{2}(n+p))}{\Gamma(\frac{1}{2}n)} \qquad 6.1.17$$

the density function of t is

$$f(t) = \frac{\Gamma(\frac{1}{2}(n+p))}{\pi^{\frac{1}{2}p}\,\Gamma(\frac{1}{2}n)\,|\Sigma|^{\frac{1}{2}}}\; n^{\frac{1}{2}n}(n + t'\Sigma^{-1}t)^{-\frac{1}{2}(n+p)} \qquad 6.1.18$$

which is the same as 6.1.3. Hence the representations

(i) $t = \left(\frac{Z}{n}\right)^{-\frac{1}{2}} Y$ where $Y \sim N_p(0,\Sigma)$ independently of $Z \sim \chi^2_n$

(ii) $t = (U^{\frac{1}{2}'})^{-1}X$ where $X \sim N_p(0,nI_p)$ independently of $U \sim W_p(\Sigma^{-1},n+p-1)$

are equivalent.

Notice that in the second representation, the elements or "rows" of X are independent. Dickey (1967) generalised this representation to the matrix case by replacing the p×1 vector X by a p×q matrix X the row vectors of which are independently distributed, each as $N_q(0,Q)$, $Q > 0$. Further Dickey assumed that $U \sim W_p(P, m-q)$. As will be shown later the density function of the p×q matrix

$$T = (U^{\frac{1}{2}'})^{-1} X$$

is proportional to

$$|Q + T'PT|^{-\frac{1}{2}m}$$

Density functions of this form had appeared earlier in the literature, firstly in connection with classical multivariate normal regression (Kshirsagar (1960)) and later in a Bayesian context (Taio and Zellner (1964) and Geisser (1965)). However, it was Dickey who unified these results by showing that the density functions were those of a matrix T distribution.

In the next section, we discuss the distribution of T in detail. In Section 3 we define two non-central matrix T distributions. Firstly, the lower non-central T in which U has a non-central Wishart distribution, and secondly the upper non-central T, in which the rows of X are independent normal vectors with non-zero means. The density function of the upper non-central T cannot be expressed in closed form.

In the final section of the chapter, we discuss the Inverted Matrix T distribution. This distribution was also defined by Dickey (1967) and is a matrix generalisation of the inverted multivariate t distribution discussed in Chapter 3. We consider a non-central form of this distribution, but here again the density cannot be given in closed form.

6.2 THE CENTRAL MATRIX T DISTRIBUTION *(Dickey 1967)*

Theorem 6.2.1. Let T be a p×q random matrix such that

$$T = (U^{\frac{1}{2}\prime})^{-1} X \qquad 6.2.1$$

where $U \sim W_p(P, m-q)$, $P > 0$, $m > p+q+1$, independently of X, the row vectors of which are independently distributed each as $N_q(0,Q)$, $Q > 0$. Let $U^{\frac{1}{2}}$ be the symmetric square root of U. Then the density function of T is

$$f(T) = \frac{\Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq}\,\Gamma_p(\frac{1}{2}(m-q))} |P|^{-\frac{1}{2}(m-q)} \; |Q|^{-\frac{1}{2}p} \; |P^{-1}+TQ^{-1}T'|^{-\frac{1}{2}m} \qquad 6.2.2$$

$$= \frac{\Gamma_q(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq}\,\Gamma_q(\frac{1}{2}(m-p))} |Q|^{\frac{1}{2}(m-p)} \; |P|^{\frac{1}{2}q} \; |Q+T'PT|^{-\frac{1}{2}m} \qquad 6.2.3$$

Proof:

$$f(U) = \frac{|U|^{\frac{1}{2}(m-q-p-1)} \operatorname{etr}(-\frac{1}{2}[P^{-1}U])}{2^{\frac{1}{2}(m-q)p}\;\Gamma_p(\frac{1}{2}(m-q))\;|P|^{\frac{1}{2}(m-q)}} \qquad 6.2.4$$

$$f(X) = \frac{\operatorname{etr}(-\frac{1}{2}(Q^{-1}X'X))}{(2\pi)^{\frac{1}{2}pq}\;|Q|^{\frac{1}{2}p}} \qquad 6.2.5$$

Since X and U are independent

$$f(X,U) \propto |U|^{\frac{1}{2}(m-q-p-1)} \text{etr}(-\tfrac{1}{2}[P^{-1}U + XQ^{-1}X'])$$

6.2.6

where the constant of proportionality is

$$(\pi^{\frac{1}{2}pq} 2^{\frac{1}{2}mp} \Gamma_p(\tfrac{1}{2}(m-q)) |P|^{\frac{1}{2}(m-q)} |Q|^{\frac{1}{2}p})^{-1} \qquad 6.2.7$$

Transform from X to T in 6.2.6 by

$$X = (U^{\frac{1}{2}'})T \text{ with } J(X \rightarrow T) = |U|^{\frac{1}{2}q} \qquad 6.2.8$$

Then the joint density of X and T is

$$f(X,T) \propto |U|^{\frac{1}{2}(m-p-1)} \text{etr}(-\tfrac{1}{2}[P^{-1} + TQ^{-1}T']U) \qquad 6.2.9$$

To obtain f(T), 6.2.9 must be integrated over U. Observing that 6.2.9 is the kernel of a $W_p((P^{-1} + TQ^{-1}T')^{-1},m)$ density, the required integral has the value

$$2^{\frac{1}{2}mp} \Gamma_p(\tfrac{1}{2}m) |P^{-1} + TQ^{-1}T'|^{-\frac{1}{2}m} \qquad 6.2.9$$

Hence from 6.2.7 and 6.2.9, we obtain f(T) as given in 6.2.2.

For the second form of the density function consider

$$|A| = \begin{vmatrix} P^{-1} & -T \\ T' & Q \end{vmatrix} = |P^{-1}||Q+T'PT| = |Q||P^{-1}+TQ^{-1}T'|$$

From which it follows that

$$|P^{-1} + TQ^{-1}T| = |Q^{-1}||P^{-1}||Q+T'PT| \qquad 6.2.10$$

Also

$$\frac{\Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq} \Gamma_p(\frac{1}{2}(m-q))} = \frac{\Gamma_q(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq} \Gamma_q(\frac{1}{2}(m-p))} \qquad 6.2.11$$

which can be easily demonstrated by expanding the multivariate gamma functions. Substituting 6.2.10 and 6.2.11 in 6.2.2 gives the second form of the density function, 6.2.3.

The parameters of the distribution are

P - the covariance matrix of the Wishart distribution.

Q - the covariance matrix of the Matrix Normal distribution.

m - the degrees of freedom.

It is clear that the distribution is centered at the origin. We summarise the above in the next definition.

Definition 6.2.1 If the random p×q matrix T has a density function given by 6.2.2 or 6.2.3, then T has a central matrix T distribution and is denoted by

$$T \sim T(P,Q,0,m). \qquad 6.2.12$$

For an arbitrary centering of the distribution, T can be replaced in 6.2.2 and 6.2.3 by T-C where C is a p×q matrix of constants. Then

$$T + C \sim T(P,Q,C,m) \qquad 6.2.13$$

In this case the density function of T has the form

$$f(T) \propto |P^{-1} + (T-C)Q^{-1}(T-C)'|^{-\frac{1}{2}m} \qquad 6.2.14$$

and if $X_{(i)}$ and $C_{(i)}$ are the i^{th} of X and C respectively, T has the synthetic representation

$$T = (U^{\frac{1}{2}'})^{-1}(X-C)$$

where $U \sim W_p(P,m-q)$ and the rows of X are independently distributed as $N_q(C_{(i)},Q^{-1})$ $i = 1 \cdot\cdot p$.

From the second form of the density 6.2.3, it can be seen that

$$T' = X'(U^{\frac{1}{2}})^{-1} \sim T(Q^{-1},P^{-1},0,m) \qquad 6.2.15$$

and from this a second representation of T is

$$T = Y(V^{\frac{1}{2}})^{-1} \qquad 6.2.16$$

where the q×q matrix $V \sim W(Q^{-1},m-p)$ independently of Y, the column vectors of which are independently distributed each as $N_p(0,P^{-1})$.

A special case of the T distribution occurs when $P = I_p$ and $Q = I_q$. Then

$$T_o = (U^{\frac{1}{2}'})^{-1} X_o \qquad 6.2.17$$

The individual entries of X_o are standard normal variables and so both the rows and columns of X_o are independent. The density function of T_o is

$$f(T_o) = \frac{\Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq}\,\Gamma_p(\frac{1}{2}(m-q))} |I_p + T_oT_o'|^{-\frac{1}{2}m} \qquad 6.2.18$$

$$= \frac{\Gamma_q(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq}\,\Gamma_q(\frac{1}{2}(m-p)}|I_q + T_o'T_o|^{-\frac{1}{2}m} \qquad 6.2.19$$

T_o is closely related to the multivariate Beta distribution.

Theorem 6.2.2. *(Olkin and Rubin (1964))*

If $W = T_oT_o'$ and $q > p$, then

$$f(W) = \frac{\Gamma_p(\frac{1}{2}m)}{\Gamma_p(\frac{1}{2}q)\Gamma_p(\frac{1}{2}(m-q))} |W|^{\frac{1}{2}(q-p-1)} \; |I+W|^{-\frac{1}{2}m} \qquad 6.2.20$$

Proof: Since the density function of T_o is a function of T_oT_o', Hsu's lemma (Anderson (1958) page 319) can be applied and the result follows.

The marginal and conditional distributions of submatrices of T are also T distributions. In particular any row or column of T has a central multivariate t distribution.

Theorem 6.2.3.

If $T = (T_1T_2)$ where T_1 is $p\times q_1, T_2$ is $p\times q_2$ and $q_1 + q_2 = q$, then the marginal distribution of T_2 is $T(P,Q_{22},0,m-q)$. (This is a columnwise partition of T.)

Proof: $T = (U^{\frac{1}{2}'})^{-1} X$ where the rows of X are independently distributed as $N(0,Q)$. If $T = (T_1T_2)$ then $T = (U^{\frac{1}{2}'})^{-1}X_1X_2$ where X_1 is $p\times q_1$ and X_2 is $p\times q_2$. Hence $T_2 = (U^{\frac{1}{2}'})^{-1}X_2$, and from the properties of the normal distribution, the p rows of X_2 are independently distributed each as $N_{q_2}(0,Q_{22})$, where

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}.$$

Applying the same derivation as in Theorem 6.2.1, using U and X_2 gives $f(T_2) \propto |P^{-1} + T_2' Q_{22}^{-1} T_2|^{-\frac{1}{2}(m-q_1)}$.

Hence $T_2 \sim T(P,Q_{22},0,m-q_1)$.

Similarly $T_1 \sim T(P,Q_{11},0,m-q_2)$.

Theorem 6.2.4.

If $T = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where X_1 is $p_1\times q$, X_2 is $p_2\times q$, $p_1+p_2=p$.

then the marginal distribution of X_2 is $T(P_{22.1},Q,0,m-p_1)$. (This corresponds to a row-wise partition of T.)

Proof: A row-wise partition of T is equivalent to a column-wise partition of T'. Thus $T' = X_1'X_2'$ and from 6.2.15, $T' \sim T(Q^{-1},P^{-1},0,m)$.

Partition P^{-1} as

$$P^{-1} = \begin{pmatrix} P^{11} & P^{12} \\ P^{21} & P^{22} \end{pmatrix}$$

then from Theorem 6.3.3, $X_2' \sim T(Q^{-1},P^{22},0,m-p_1)$.
But by Theorem 2.2.1, $P^{22} = P^{-1}_{22.1}$, so

$$X_2' \sim T(Q^{-1}, P^{-1}_{22.1}, 0,m-p_1)$$

and $X_2 \sim T(P_{22.1},Q,0,m-p_1)$.
Similarly, $X_1 \sim T(P_{11.2},Q,0,m-p_2)$

6.3 NON-CENTRAL MATRIX T DISTRIBUTIONS

In this section we consider the distribution of the p×q random matrix

$$T = (U^{\frac{1}{2}'})^{-1} X \qquad 6.3.1$$

when X is distributed as in 6.2, but U has a non-central Wishart distribution.

Suppose that $U \sim W_p(P,m-q,\Omega)$, then U is derived from a sample from a multivariate normal population in which the true mean vectors vary from observation to observation but the covariance matrix P is the same for all observations. If M is the p×(m-q) matrix of true means, then the non-centrality parameter matrix Ω is defined as (Constantine (1963))

$$\Omega = \tfrac{1}{2} MM'P^{-1} . \qquad 6.3.2$$

The form of the non-central Wishart density depends on the rank of Ω . We shall consider the distribution of T for two cases of the non-central Wishart distribution namely

(i) the general case when Ω has full rank

(ii) the canonical form of the linear case, when $P = I_p$, $Q = I_q$ and the rank of Ω is one.

Although (ii) could be obtained from (i) by an appropriate adjustment of the parameters, we shall derive the two distributions separately since the linear case has important applications. These applications are discussed in detail in Chapter 7.

Theorem 6.3.1.

Suppose that the row vectors of the p×q matrix X are each independently distributed as $N_q(0,Q)$, $Q > 0$, and $U \sim W_p(P,m-q,\Omega)$ $P > 0, m > p+q-1$, independently of X. If $U^{\frac{1}{2}}$ is the symmetric square root of U, then the density function of

$$T = (U^{\frac{1}{2}\,\prime})^{-1} X$$

is $f(T) =$

$$\frac{\Gamma_p(\frac{1}{2}m)\ \mathrm{etr}(-\Omega)\ |P^{-1}+TQ^{-1}T'|^{-\frac{1}{2}m}}{\pi^{\frac{1}{2}pq}\ \Gamma_p(\frac{1}{2}(m-q))\ |P|^{\frac{1}{2}(m-q)}\ |Q|^{\frac{1}{2}p}}\ {}_1F_1\left(\tfrac{1}{2}m;\tfrac{1}{2}(m-q);P^{-1}\Omega(P^{-1}+TQ^{-1}T')^{-1}\right) \qquad 6.3.3$$

Proof: The density function of X is

$$f(X) = \frac{etr(-\frac{1}{2}Q^{-1}X'X)}{(2\pi)^{\frac{1}{2}qp}\ |Q|^{\frac{1}{2}p}} \qquad 6.3.4$$

$$f(U) = \frac{etr(-\Omega)\ |U|^{\frac{1}{2}(m-q-p-1)} etr(-\frac{1}{2}P^{-1}U)}{2^{\frac{1}{2}(m-q)p}\ \ \Gamma_p(\frac{1}{2}(m-q-1)\ \ |P|^{\frac{1}{2}(m-q)}}\ {}_0F_1(\tfrac{1}{2}(m-q);\tfrac{1}{2}P^{-1}\Omega U)$$

.6.3.5

Expanding the ${}_0F_1$ series in terms of zonal polynomials using definition 2.3.2, gives

$$f(U) \propto |U|^{\frac{1}{2}(m-q-p-1)}\ etr(-\tfrac{1}{2}P^{-1}U) \sum_{k=o}^{\infty} \sum_{\kappa} \frac{C_\kappa(\frac{1}{2}P^{-1}\Omega U)}{(\frac{1}{2}(m-q))_\kappa\ k!}$$

6.3.6

Since X and U are independent,

$$f(XU) \propto |U|^{\frac{1}{2}(m-q-p-1)}\ etr(-\tfrac{1}{2}(P^{-1}U+XQ^{-1}X'))$$

$$\times \sum_{k=o}^{\infty} \sum_{\kappa} \frac{C_\kappa(\frac{1}{2}P^{-1}\Omega U)}{(\frac{1}{2}(m-q)_\kappa\ \ k!} \qquad .6.3.7$$

where from 6.3.4 and 6.3.5, the constant of proportionality is

$$\left(2^{\frac{1}{2}mp}\ \pi^{\frac{1}{2}pq}\ \Gamma_p(\tfrac{1}{2}(m-q))\ \ |P|^{\frac{1}{2}(m-q)}\ \ |Q|^{\frac{1}{2}p}\right)^{-1} \qquad 6.3.8$$

Transforming in 6.3.7 by

$$X = (U'^{\frac{1}{2}})T \quad \text{with } J(X \to T) = |U|^{\frac{1}{2}q}$$

gives

$$f(T,U) \propto |U|^{\frac{1}{2}(m-p-1)}\ etr(-\tfrac{1}{2}(P^{-1}+TQ^{-1}T')U)$$

$$\times \sum_{k=o}^{\infty} \sum_{\kappa} \frac{C_\kappa(\frac{1}{2}P^{-1}\Omega U)}{(\frac{1}{2}(m-q))_\kappa\ k!} \qquad 6.3.9$$

To find f(T) we must integrate 6.3.9 termwise over U and hence must evaluate integrals of the form

$$\int_{U>0} |U|^{\frac{1}{2}(m-p-1)} \operatorname{etr}(-\tfrac{1}{2}(P^{-1}+TQ^{-1}T')U)\, C_\kappa(\tfrac{1}{2}P^{-1}\Omega U)\,dU \qquad 6.3.10$$

In 6.3.10 let $R = \frac{1}{2}(P^{-1}+TQ^{-1}T')$ and $W = \frac{1}{2}P^{-1}\Omega$. 6.3.11

R is symmetric since both P^{-1} and $TQ^{-1}T'$ are both symmetric. $W = \frac{1}{2}P^{-1}\Omega = \frac{1}{4}P^{-1}MM'P'^{-1}$ is symmetric. Therefore we can write 6.3.10 as

$$\int_{U>0} |U|^{\frac{1}{2}(m-p-1)} \operatorname{etr}(-RU)C_\kappa(WU)\,dU \qquad 6.3.12$$

which by Theorem 2.3.1 has the value

$$\Gamma_p(\tfrac{1}{2}m,\kappa)\ |R|^{-\frac{1}{2}m}\ C_\kappa(WR^{-1}) \qquad 6.3.13$$

$$= \quad 2^{\frac{1}{2}mp}\ |P^{-1}+TQ^{-1}T'|^{-\frac{1}{2}m}\ C_\kappa(P^{-1}\Omega(P^{-1}+TQ^{-1}T')^{-1}) \qquad 6.3.14$$

Hence

$$f(T) \propto 2^{\frac{1}{2}mp}\ |P^{-1}+TQ^{-1}T'|^{-\frac{1}{2}m} \sum_{k=0}^{\infty} \sum_{\kappa} \frac{\Gamma_p(\frac{1}{2}m,\kappa)C_\kappa(P^{-1}\Omega(P^{-1}+TQ^{-1}T')^{-1})}{(\frac{1}{2}(m-q))_\kappa \quad k!}$$

6.3.15

By definition 2.3.2(v)

$$\Gamma_p(\tfrac{1}{2}m,\kappa) = (\tfrac{1}{2}m)_\kappa\Gamma_p(\tfrac{1}{2}m) \qquad 6.3.16$$

by which the infinite series in 6.3.15 becomes

$$\Gamma_p(\tfrac{1}{2}m) \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(\tfrac{1}{2}m)_\kappa}{(\tfrac{1}{2}(m-q))_\kappa} \frac{C_\kappa(P^{-1}\Omega(P^{-1}+TQ^{-1}T')^{-1})}{k!} \qquad 6.3.17$$

By definition 2.3.2, 6.3.17 is

$$\Gamma_p(\tfrac{1}{2}m)\ {}_1F_1\ (\tfrac{1}{2}m;\tfrac{1}{2}(m-q);\ P^{-1}\Omega(P^{-1}+TQ^{-1}T')^{-1}) \qquad 6.3.18$$

Substituting 6.3.18 and the constant term 6.3.8, into 6.3.15, f(T) is as given in 6.3.3.

<u>Corollary 6.3.1.</u> If $P = I_p$ and $Q = I_q$ then

$$f(T) = \frac{\Gamma_p(\tfrac{1}{2}m)\ \mathrm{etr}(-\Omega)\ |I_p+TT'|^{-\frac{1}{2}m}}{\pi^{\frac{1}{2}pq}\ \Gamma_p(\tfrac{1}{2}(m-q))}\ {}_1F_1(\tfrac{1}{2}m;\tfrac{1}{2}(m-q);\Omega(I_p+TT')^{-1})$$

6.3.19

<u>Definition 6.3.1.</u> If the random p×q matrix T has the density function given in 6.3.3, then T has a lower non-central matrix T distribution. This distribution will be denoted by $T \sim T(P,Q,0,m,\Omega)$.

We now derive the distribution of the lower non-central T when the underlying Wishart distribution has rank one and P and Q are both identity matrices.

<u>Theorem 6.3.2.</u> *Let the row vectors of the p×q matrix X each be independently distributed as $N_q(0,I_q)$. Let U have a linear non-central Wishart distribution with covariance matrix I_p, non-centrality parameter λ^2 and m-q degrees of freedom where $m > p+q-1$. If $U^{\frac{1}{2}}$ is the symmetric square root of U, then the p×q matrix $T = (U^{\frac{1}{2}\prime})^{-1}X$ has the density*

$$f(T) = \frac{exp(-\frac{1}{2}\lambda^2)\ \Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq}\ \Gamma(\frac{1}{2}(m-q))} |I_p + TT'|^{-\frac{1}{2}m}\ {}_1F_1(\tfrac{1}{2}m;\tfrac{1}{2}(m-q);\tfrac{1}{2}\lambda^2\tau^{11})$$

6.3.20

where τ^{11} *is the* $(1,1)^{th}$ *element of* $(I_p + TT')^{-1}$.

Proof: The density function of X is

$$f(X) = \frac{etr(-\frac{1}{2}XX')}{(2\pi)^{\frac{1}{2}pq}} \qquad 6.3.21$$

From definition 2.5.6, the density function of U is

$$f(U) = \frac{exp(-\frac{1}{2}\lambda^2)\ |U|^{\frac{1}{2}(m-q-p-1)}\ etr(-\frac{1}{2}U)}{\Gamma_p(\frac{1}{2}(m-q))\ 2^{\frac{1}{2}(m-q)p}}$$

$$\times \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\lambda^2)^k\ \Gamma(\frac{1}{2}(m-q))\ u_{11}^k}{k!\ \ 2^k\Gamma(\frac{1}{2}(m-q)+k)} \qquad 6.3.22$$

where u_{11} is the top left hand element of U.

Since X and U are independent,

$$f(X,U) \propto |U|^{\frac{1}{2}(m-q-p-1)}\ etr(-\tfrac{1}{2}(U+XX')) \sum_{k=0}^{\infty} \phi(k)\ u_{11}^{i} \qquad 6.3.23$$

where the constant of proportionality is

$$\frac{exp(-\frac{1}{2}\ \lambda^2)}{\pi^{\frac{1}{2}pq}\ \Gamma_p(\frac{1}{2}(m-q))2^{\frac{1}{2}mp}} \qquad 6.3.24$$

and $$\phi(k) = \frac{(\frac{1}{2}\lambda^2)^k\ \Gamma(\frac{1}{2}(m-q))}{k!\ \ 2^k\Gamma(\frac{1}{2}(m-q)+k)} \qquad 6.3.25$$

Transforming in 6.3.25 by $X = (U^{\frac{1}{2}'})T$ with $J(X \rightarrow T) = |U|^{\frac{1}{2}q}$ the density function of T and U is

$$f(U,T) \propto |U|^{\frac{1}{2}(m-p-1)} \operatorname{etr}(-\tfrac{1}{2}(I_p + TT')U) \sum_{k=0}^{\infty} \phi(k) u_{11}^k$$

6.3.26

To obtain f(T), 6.3.26 must be integrated termwise over U. Hence we must evaluate integrals of the form

$$\int_{U>0} u_{11}^k |U|^{\frac{1}{2}(m-p-1)} \operatorname{etr}(-\tfrac{1}{2}(I_p + TT')U)\,dU \qquad 6.3.27$$

But 6.3.27 is proportional to $E(u_{11}^k)$, where u_{11} is the $(1,1)^{th}$ element of $U \sim W_p(m,(I_p + TT')^{-1})$. Since $u_{11} \sim \tau^{11} \chi_m^2$ where τ^{11} is the $(1,1)^{th}$ element of $(I_p + TT')^{-1}$, it follows that

$$E(u_{11}^k) = \frac{(\tau^{11})^k\, 2^k\, \Gamma(\frac{1}{2}m+k)}{\Gamma(\frac{1}{2}m)} \qquad 6.3.28$$

and the value of the integral is

$$\frac{2^{\frac{1}{2}mp+k}\, \Gamma_p(\frac{1}{2}m)\, \Gamma(\frac{1}{2}m+k)(\tau^{11})^k}{\Gamma(\frac{1}{2}m)\, |I_p + TT'|^{\frac{1}{2}m}} \qquad 6.3.29$$

Hence

$$f(T) \propto \frac{2^{\frac{1}{2}mp}\, \Gamma_p(\frac{1}{2}m)}{\Gamma(\frac{1}{2}m)\, |I_p + TT'|^{\frac{1}{2}m}} \sum_{k=0}^{\infty} \phi(k)\, 2^k\, \Gamma(\tfrac{1}{2}m+k)(\tau^{11})^k$$

6.3.30

Substituting for the constant term from 6.3.24 and for $\phi(k)$ we find

$$f(T) = \frac{\exp(-\frac{1}{2}\lambda^2)\, \Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq}\, \Gamma(\frac{1}{2}(m-q))} |I_p + TT'|^{-\frac{1}{2}m}$$

6.3.31

$$\times \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\lambda^2)^k\, \Gamma(\frac{1}{2}(m-q))\, \Gamma(\frac{1}{2}m+k)(\tau^{11})^k}{k!\, \Gamma(\frac{1}{2}(m-q)+k)\, \Gamma(\frac{1}{2}m)}$$

But the infinite series in 6.3.31, is $_1F_1(\frac{1}{2}m; \frac{1}{2}(m-q); \frac{1}{2}\lambda^2\tau^{11})$. Hence f(T) is as stated in the theorem.

In the central distribution of the matrix T, an alternative form of the density function could be found by a simple manipulation of the determinant $|P^{-1}+T'Q^{-1}T|$ (see 6.2.10). It does not seem possible to obtain a second form of the density of the lower non-central T in this way, because the argument of the $_1F_1$ series cannot be changed by a manipulation of the matrices. However we know that the two forms of the central case give rise to two synthetic representations of T, namely, $T = (U^{\frac{1}{2}'})^{-1} X$, where $U \sim W_p(P,m-q)$ and the rows of X are $N_q(0,Q)$ or $T = Y(V^{\frac{1}{2}})^{-1}$ where $V \sim W_q(Q^{-1},m-p)$ and the columns of Y are $N_p(0,P^{-1})$(see 6.2.16). If we use the second synthetic form with $V \sim W_q(Q^{-1},m-p,\psi)$ and apply the same derivation as in Theorem 6.3.1, we have

$$f(Y) \propto \text{etr}(-\tfrac{1}{2}PYY') \qquad 6.3.32$$

$$f(V) \propto |V|^{\frac{1}{2}(m-p-q-1)} \text{etr}(-\tfrac{1}{2}QV) \sum_{i=o}^{\infty} \sum_{\imath} \frac{C_{\imath}(\frac{1}{2}Q\psi V)}{(\frac{1}{2}(m-p)_{\imath}\ i!}$$

6.3.33

where $\imath$ now represents all the partitions of i into not more than q parts.

Transforming in the joint density of f(V,Y) by $TV^{\frac{1}{2}} = Y$ with $J(Y \to T) = |V|^{\frac{1}{2}p}$ gives 6.3.34

$$f(V,T) \propto |V|^{\frac{1}{2}(m-q-1)} \text{etr}(-\tfrac{1}{2}(Q+T'PT)V) \times \sum_{i=o}^{\infty} \sum_{\imath} \frac{C_{\imath}(\frac{1}{2}Q\psi V)}{(\frac{1}{2}(m-p))_{\imath}\ i!} \qquad 6.3.35$$

Integration over V gives a second form of the density of the lower non-central T as

$$f(T) = \frac{\Gamma_q(\frac{1}{2}m)\,\mathrm{etr}(-\psi)}{\pi^{\frac{1}{2}pq}\,\Gamma(\frac{1}{2}(m-q))}\ |Q|^{\frac{1}{2}(m-p)}\ |P|^{\frac{1}{2}q}\ |Q+T'PT|^{-\frac{1}{2}m}$$

$$\times\ {}_1F_1\ (\tfrac{1}{2}m;\ \tfrac{1}{2}(m-p);\ Q\psi(Q+T'PT)^{-1}) \qquad 6.3.36$$

We have been unable to convince ourselves of the complete equivalence of the two forms, since the non-centrality parameters and the arguments of the two ${}_1F_1$ series are different. It is clear that this problem needs further investigation. However since we have found that the linear case of the first form has interesting applications and arises in a different context, we shall adopt the synthetic representation and the density function given in Theorem 6.3.1 for the definition of the lower non-central T.

<u>*Theorem 6.3.3.*</u> *If $T \sim T(P,Q,O,m,\Omega)$ and $T = (T_1,T_2)$ where T is $p\times q_1$ and T_2 is $p\times q_2$, $q_1+q_2=q$, then $T_1 \sim T(P,Q_{11},O,m-q_2,\Omega)$ and $T_2 \sim T(P,Q_{22},O,m-q_1,\Omega)$.*

<u>Proof</u>: $T_1 = (T_1T_2) = (U^{\frac{1}{2}'})^{-1}(X_1X_2)$ where $X = (X_1X_2)$ is partitioned columnwise as T. Since the rows of X are independently distributed each as $N_q(0,Q)$ it follows that the rows of X_1 are $N_{q_1}(0,Q_{11})$ and those of X_2 are $N_{q_2}(0,Q_{22})$ where Q_{11} and Q_{22} are the appropriate submatrices of Q. Thus $T_1 = (U^{\frac{1}{2}'})^{-1}\ X_1$ where $U \sim W_p(P,m-q,\Omega)$, and the rows of X_1 are independently $N_{q_1}(0,Q_{11})$ and X_1 and U

are independent. Hence from Theorem 6.3.1, $X_1 \sim T(P,Q_{11},0,m-q_2,\Omega)$. Similarly $X_2 \sim T(P,Q_{22},0,n-q,\Omega)$.

<u>*Theorem 6.3.4.*</u> *If $T \sim T(I_p, I_q, 0, m, \Omega)$; $q \geq p$ and $B = TT'$, then the density function of B is*

$$\frac{etr(-\Omega)\,\Gamma_p(\tfrac{1}{2}m)}{\Gamma_p(\tfrac{1}{2}(m-q))\Gamma_p(\tfrac{1}{2}q)} |B|^{\frac{1}{2}(q-p-1)} |I+B|^{-\frac{1}{2}m} {}_1F_1(\tfrac{1}{2}m;\tfrac{1}{2}(m-q);\Omega(I+TT')^{-1}) \qquad 6.3.38$$

<u>Proof</u>: From 6.3.19,

$$f(T) = \frac{\Gamma_p(\tfrac{1}{2}m)etr(-\Omega)}{\pi^{\frac{1}{2}pq}\,\Gamma_p(\tfrac{1}{2}(m-q))} |I_p+TT'|^{-\frac{1}{2}m} {}_1F_1(\tfrac{1}{2}m;\tfrac{1}{2}(m-q);\Omega(I_p+TT')^{-1}) \qquad 6.3.39$$

The density function of T is a function of TT'. Thus we may write f(T) = f(TT') = f(B) where B = TT'. Applying Hsu's lemma (Anderson (1968) page 319), the density function of B is

$$g(B) = \frac{\pi^{\frac{1}{2}pq} \ |B|^{\frac{1}{2}(q-p-1)} \ f(B)}{\Gamma_p(\tfrac{1}{2}q)} \qquad 6.3.40$$

Substituting f(B) from 6.3.9 we obtain g(B) as stated in 6.3.38.

From Theorem 6.3.4 we see that B has a non-central multivariate Beta Type 2A distribution (De Waal (1968)) with q and m-q degrees of freedom. This result could also be deduced from the synthetic representation as follows:

$$TT' = (U^{\frac{1}{2}'})^{-1}XX'(U^{\frac{1}{2}}) \qquad 6.3.41$$

where $U \sim W_p(I_p, m-q, \Omega)$

and the q rows of X are independently distributed each as $N_q(0,I_q)$. Thus the p columns of X are independently distributed as $N(0,I_p)$, so $XX' \sim W_p(I_p,q)$. The result then follows from the definition of the non-central multivariate Beta distribution. (de Waal (1968)).

To define the upper non-central matrix T distribution we assume that the independent row vectors of X are distributed as $N_q(\mu^{(i)},Q)$ $i = 1 \cdots p$, and $U \sim W_p(P,m-q)$ independently of X. In this case it does not seem possible to express the density of $T = (U^{\frac{1}{2}'})^{-1} X$ in closed form as we shall now show.

We define the p×q matrix μ as

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mu^{(p)} \end{pmatrix} \qquad 6.3.42$$

The density function of X is proportional to

$$f(X) \propto \text{etr}(-\tfrac{1}{2}(X-\mu)Q^{-1}(X-\mu)') \qquad 6.3.43$$

Expanding the exponent, we have

$$f(X) \propto \text{etr}(-\tfrac{1}{2}\mu Q\mu')\text{etr}(-\tfrac{1}{2}XQ^{-1}X')\text{etr}(XQ^{-1}\mu') \qquad 6.3.44$$

The density of U is proportional to

$$f(U) \propto |U|^{\frac{1}{2}(m-q-p-1)}\text{etr}(-\tfrac{1}{2}P^{-1}U) \qquad 6.3.45$$

So $f(X,U) \propto |U|^{\frac{1}{2}(m-q-p-1)}\text{etr}(-\frac{1}{2}(P^{-1}U+XQ^{-1}X'))\text{etr}(XQ^{-1}\mu')$

Transforming from X to T by

$$T = (U^{\frac{1}{2}'})^{-1} X \text{ with } J(X \to T) = |U|^{\frac{1}{2}q}$$

$$f(X,T) \propto U^{\frac{1}{2}(m-p-1)} \operatorname{etr}(-\tfrac{1}{2}(P^{-1}+TQ^{-1}T')U)\operatorname{etr}(U^{\frac{1}{2}'}TQ^{-1}\mu')$$

6.3.46

Now $\operatorname{etr}(U^{\frac{1}{2}'}TQ^{-1}\mu')$ is a scalar and can be expanded in a power series as

$$\operatorname{etr}(U^{\frac{1}{2}'}TQ^{-1}\mu') = \sum_{k=0}^{\infty} \frac{\operatorname{tr}(U^{\frac{1}{2}'}TQ^{-1}\mu')^k}{k!} \qquad 6.3.47$$

Hence to obtain f(T) we must evaluate integrals of the form

$$\int_{U>0} \operatorname{tr}(U^{\frac{1}{2}'}TQ^{-1}\mu')^k \, |U|^{\frac{1}{2}(m-p-1)} \operatorname{etr}(-\tfrac{1}{2}(P^{-1} + TQ^{-1}T')U)dU$$

6.3.48

It does not seem possible to proceed further with the integral. If $U^{\frac{1}{2}'}TQ^{-1}\mu'$ were symmetric, it would be possible to write

$$\operatorname{etr}(U^{\frac{1}{2}'}TQ^{-1}\mu') = {}_0F_0(U^{\frac{1}{2}'}TQ^{-1}\mu') \qquad 6.3.49$$

and obtain an expansion in terms of zonal polynomials, whereby some attack could possibly be made on the integral. However since $U^{\frac{1}{2}'}TQ^{-1}\mu'$ is not symmetric this avenue is closed to us. Hence, we can only say

$$f(T) = \int_{U>0} f(U,T)dU$$

where f(U,T) is given by 6.3.46.

6.4 THE INVERTED MATRIX T DISTRIBUTION

A matrix generalisation of the inverted multivariate t distribution (see Chapter 3) was given by Dickey (1967). If $T \sim T(I_p, I_q, 0, m)$, the inverted matrix T distribution is that of

$$R = \left\{(I_p + TT')^{\frac{1}{2}}\right\}^{-1} T \qquad 6.4.1$$

which, using the synthetic representation of T, can be shown to be equal to

$$R = \left\{(U + XX')^{\frac{1}{2}}\right\}^{-1} X \qquad 6.4.2$$

where $U \sim W_p(I_p, m-q)$ independently of X, the entries of which are standard normal variables (NB. Our T,R,U and X are the same as Dickey's T_o, R_o, U_o and X_o). The density of R is given in the next theorem.

Theorem 6.4.1. *Dickey (1967)*

Let the rows of the p×q matrix X, be independently distributed each as $N_q(0, I_q)$. Let $U \sim W_p(I_p, m-p)$, $m > p+q-1$.

Let $R = \left\{(U + XX')^{\frac{1}{2}}\right\}^{-1} X \qquad 6.4.2$

Then the density function of R is

$$f(R) = \frac{\Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq}\,\Gamma_p(\frac{1}{2}(m-q))} \left|I_p - RR'\right|^{\frac{1}{2}(m-p-q-1)}$$

$$I - RR' > 0.$$

Proof: $$f(X) = \frac{etr(-\frac{1}{2}XX')}{(2\pi)^{\frac{1}{2}pq}} \qquad 6.4.3$$

and

$$f(U) = \frac{U^{\frac{1}{2}(m-q-p-1)}\, etr(-\frac{1}{2}U)}{2^{\frac{1}{2}(m-q)p}\,\Gamma_p(\frac{1}{2}(m-q))} \qquad 6.4.4$$

Then $f(U,X) \propto |U|^{\frac{1}{2}(m-q-p-1)}\, etr(-\frac{1}{2}(U+XX')$ 6.4.5

In 6.4.5, let $G = U-XX'$. Then $J(U,XX' \rightarrow G,XX') = 1$ and we obtain

$$f(X,G) \propto |G-XX'|^{\frac{1}{2}(m-q-p-1)}\, etr(-\frac{1}{2}G) \qquad 6.4.6$$

Let $X = G^{\frac{1}{2}}R$ with $J(X \rightarrow R) = |G|^{\frac{1}{2}q}$ 6.4.7

$$f(R,G) \propto |I-RR'|^{\frac{1}{2}(m-q-p-1)}\, |G|^{\frac{1}{2}(m-p-1)}\, etr(-\frac{1}{2}G)$$

6.4.8

Integrating over G, using Theorem 2.5.2 gives

$$f(R) \propto |I-RR'|^{\frac{1}{2}(m-q-p-1)}\, 2^{\frac{1}{2}m}\, \Gamma_p(\frac{1}{2}m) \qquad 6.4.9$$

Evaluating the constant terms from 6.4.3, 6.4.4 and 6.4.9 gives f(R) as stated in the theorem.

We shall now consider the distribution of R, when U has a non-central Wishart distribution. Unfortunately the density function can only be given in a closed form when the non-centrality parameter is a scalar matrix.

From 6.4.3 and Theorem 2.5.1, the joint density of X and U is

$$f(X,U) \propto |U|^{\frac{1}{2}(m-q-p-1)} etr(-\frac{1}{2}(U+XX')\, {}_0F_1(\frac{1}{2}(m-q);\frac{1}{2}\Omega U)$$

6.4.12

where the constant of proportionality is

$$(\mathrm{etr}(-\Omega)(2^{\frac{1}{2}mp}\,\Gamma_p(\tfrac{1}{2}(m-q))\pi^{\frac{1}{2}pq})^{-1} \qquad 6.4.13$$

Let $U = XX' = G$, then

$$f(X,G) \propto |G-XX'|^{\frac{1}{2}(m-q-p-1)}\mathrm{etr}(-\tfrac{1}{2}G)\,{}_0F_1(\tfrac{1}{2}(m-q);\tfrac{1}{2}\Omega(G-XX'))$$

$$= |G-XX'|^{\frac{1}{2}(m-q-p-1)}\mathrm{etr}(-\tfrac{1}{2}G)\sum_{k=0}^{\infty}\sum_{\kappa}\frac{C_\kappa(\frac{1}{2}\Omega(G-XX'))}{(\frac{1}{2}(m-q))_\kappa\, k!}$$

Let $R = (G^{\frac{1}{2}})^{-1} X$, then $J(X \to R) = |G|^{\frac{1}{2}q}$ and the density function of R and G is

$$f(R,G) \propto |I-RR'|^{\frac{1}{2}(m-p-q-1)}\,|G|^{\frac{1}{2}(m-p-1)}\mathrm{etr}(-\tfrac{1}{2}G)$$

$$\times \sum_{k=0}^{\infty}\sum_{\kappa}\frac{C_\kappa(\frac{1}{2}\Omega G^{\frac{1}{2}}(I-RR')G^{\frac{1}{2}})}{(\frac{1}{2}(m-q))_\kappa\, k!} \qquad 6.4.14$$

To obtain f(R) we must integrate 6.4.16 over G. Hence we need to evaluate integrals of the form

$$\int_{G>0} |G|^{\frac{1}{2}(m-p-1)}\,\mathrm{etr}(-\tfrac{1}{2}G)\,C_\kappa(\tfrac{1}{2}\Omega G^{\frac{1}{2}}(I-RR')G^{\frac{1}{2}})dG \qquad 6.4.15$$

It is not possible to evaluate this integral in the general case since although we can write

$$C_\kappa(\tfrac{1}{2}\Omega G^{\frac{1}{2}}(I-RR')G^{\frac{1}{2}\,\prime}) = C_\kappa(\tfrac{1}{2}G'^{\frac{1}{2}}\Omega G^{\frac{1}{2}}(I-RR')) \qquad 6.4.16$$

we cannot separate G and Ω unless Ω is a scalar matrix.

Suppose that this is the case, and let $\Omega = aI_p$ where a is some scalar. Then 6.4.16 becomes

$$a^k \, C_\kappa(\tfrac{1}{2}G(I-RR')) \qquad 6.4.17$$

and the required integral is

$$a^k \int_{G>0} |G|^{\frac{1}{2}(m-p-1)} \, \mathrm{etr}(-\tfrac{1}{2}G) \, C_\kappa(\tfrac{1}{2}G(I-RR'))dG \qquad 6.4.18$$

which by Theorem 2.5.2 has the value

$$a^k \, 2^{\frac{1}{2}mp} \, \Gamma_p(\tfrac{1}{2}m,\kappa) \, C_\kappa(\tfrac{1}{2}(I-RR')) \qquad 6.4.19$$

$$= \quad 2^{\frac{1}{2}mp} \, \Gamma_p(\tfrac{1}{2}m) \, (\tfrac{1}{2}m)_\kappa \, C_\kappa(\tfrac{1}{2}a(I-RR')) \qquad 6.4.20$$

Thus

$$f(R) \propto |I-RR'|^{\frac{1}{2}(m-p-q-1)} \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(\frac{1}{2}m)_\kappa \, C_\kappa(\frac{1}{2}a(I-RR'))}{(\frac{1}{2}(m-q))_\kappa \quad k!}$$

6.4.21

Evaluating the constant term and noting that if $\Omega = aI_p$ then $\mathrm{etr}(-\Omega) = \exp(-pa)$, we have

$$f(R) = \frac{\exp(-pa) \; \Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq} \; \Gamma_p(\frac{1}{2}(m-q))} \; |I-RR'|^{\frac{1}{2}(m-p-q-1)}$$

$$\times \; {}_1F_1(\tfrac{1}{2}m; \; \tfrac{1}{2}(m-q); \; \tfrac{1}{2}a(I-RR')) \qquad 6.4.22$$

If $\Omega \neq aI_p$, then

$$f(R) = \frac{\mathrm{etr}(-\Omega) \; \Gamma_p(\frac{1}{2}m)}{\pi^{\frac{1}{2}pq} \; \Gamma_p(\frac{1}{2}(m-q))} \; |I-RR'|^{\frac{1}{2}(m-p-q-1)}$$

$$\times \int_{G>0} |G|^{\frac{1}{2}(m-p-1)} \mathrm{etr}(-\tfrac{1}{2}G) \, {}_0F_1(\tfrac{1}{2}(m-q); \; \tfrac{1}{2}G^{\frac{1}{2}}(I-RR')G^{\frac{1}{2}'}))$$

6.4.23

Thus we have

Theorem 6.4.2.

If $R = \left[(U+XX')^{\frac{1}{2}}\right]^{-1} X$, *where the entries of* X *are standard normal variables and* $U \sim W_p(I_p, m-q, \Omega)$, $m > p+q-1$, *independently of* X, *then the density function of* R *is given by 6.4.22 if* $\Omega = aI_p$ *and by 6.4.23 otherwise.*

The distribution of R in both the central and noncentral case is closely related to the multivariate Beta Type I distribution. Since the entries of X are standard normal variables, it follows that the q columns of X are independently distributed each as $N_p(0, I_p)$ and so $XX' \sim W_p(I_p, q)$ if $q > p$. Hence if U has a central Wishart distribution

$$RR' = [U + XX']^{-\frac{1}{2}} \; XX' \; [U + XX']^{-\frac{1}{2}}$$

is distributed as a multivariate Beta Type I with m-q and q degrees of freedom. (Olkin and Rubin(1966)). If $U \sim W_p(I_p, m-q, \Omega)$, then RR' has a non-central multivariate Beta Type IB distribution with m-q and m degrees of freedom (de Waal (1968)).

CHAPTER 7

THE MATRIX T AND MULTIVARIATE REGRESSION

7.1 INTRODUCTION

If a normal vector X is partitioned into two sets of components, the matrix of regression coefficients of $X^{(2)}$ on $X^{(1)}$ is $\beta = \Sigma_{11}^{-1}\Sigma_{12}$ where Σ_{11} and Σ_{12} are the appropriate submatrices of Σ. The sample estimator of β is $B = A_{11}^{-1}A_{12}$ where A_{11} and A_{12} are submatrices of the Wishart matrix A. If $X^{(1)}$ is considered to be a fixed variable then the distribution of B conditional on $X^{(1)} = x^{(1)}$ is matrix normal (Anderson (1958), Chapter 8).

The unconditional distribution of B is matrix T as was shown by Kshirsagar (1960) and Kabe (1968). We give a detailed derivation of these results in the next section and show further that if X is partitioned into three sets of components and one set, $X^{(3)}$ say, is held fixed, then the matrix of partial regression coefficients of $X^{(1)}$ on $X^{(2)}$ conditional on $X^{(3)}$ is also matrix T, with the degrees of freedom reduced by the number of variables held fixed.

In the remaining sections of the chapter we consider non-central forms of the distribution of B. In these sections the statistic B is constructed from a matrix A which has a non-central Wishart distribution. A complete solution to the problem appears to be mathematically

$$(X^{(2)}|x^{(1)}) \sim N_r(\nu^{(2)},\Sigma_{22.1}) \qquad 7.2.2$$

where
$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \qquad 7.2.3$$

and
$$\nu^{(2)} = \eta^{(2)} + \beta' x^{(1)} \qquad 7.2.4$$

with
$$\eta^{(2)} = \mu^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\mu^{(1)} \qquad 7.2.5$$

$$\beta = \Sigma_{11}^{-1}\Sigma_{12} \qquad 7.2.6$$

The $q\times r$ matrix β is the matrix of regression coefficients of $X^{(2)}$ on condition $X^{(1)} = x^{(1)}$.

If $X_{(\alpha)}$, $\alpha = 1,\ldots,N$ is a random sample from X, the maximum likelihood estimators of μ and Σ are

$$\hat{\mu} = \frac{1}{N}\sum_{\alpha=1}^{N} X_{(\alpha)} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{N}A \qquad 7.2.7$$

where
$$A = \sum_{\alpha=1}^{N} (X_{(\alpha)}-\overline{X})(X_{(\alpha)}-\overline{X})' \qquad 7.2.8$$

is Wishart's matrix.

The maximum likelihood estimators of $\eta^{(2)}$ and β are

$$Z^{(2)} = \overline{X}^{(2)} - A_{21}A_{11}^{-1}X^{(1)} \qquad 7.2.9$$

$$B = A_{11}^{-1}A_{12} \qquad 7.2.10$$

where
$$\overline{X} = \begin{pmatrix}\overline{X}^{(1)} \\ \overline{X}^{(2)}\end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix}A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix} \qquad 7.2.11$$

are partitioned as μ and Σ above. These results are standard and can be found in Anderson (1958).

immaterial?

Under the assumption that $\mu = 0$, Kshirsagar (1960) showed that the distribution of $B = A_{11}^{-1}A_{12}$ is matrix T and we give his derivation below.

Since $A \sim W_p(\Sigma,n)$ where $n = N-1$, the density function of A is

$$f(A) = \frac{|A|^{\frac{1}{2}(n-p-1)} etr(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}np}\Gamma_p(\frac{1}{2}n)|\Sigma|^{\frac{1}{2}n}} \qquad 7.2.12$$

Noting that $p = q+r$, we partition A as in 7.2.11 and transform from A_{12} to B by

$$B = A_{11}^{-1}A_{12} \qquad 7.2.13$$

Then $J(A_{21}\to B) = |A_{11}|^r$ 7.2.14

Further

$$|A| = |A_{11}||A_{22}-A_{21}A_{11}^{-1}A_{12}| = |A_{11}||A_{22}-B'A_{11}B| \qquad 7.2.15$$

$$|\Sigma| = |\Sigma_{11}||\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| = |\Sigma_{11}||\Sigma_{22.1}| \qquad 7.2.16$$

and $$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} + \beta\Sigma_{22.1}^{-1}\beta' & -\beta\Sigma_{22.1} \\ -\Sigma_{22.1}\beta' & \Sigma_{22.1} \end{pmatrix} \qquad 7.2.17$$

Hence $$tr\Sigma^{-1}A = tr\Sigma_{22.1}^{-1}A_{22} + trLA_{11} \qquad 7.2.18$$

where $$L = \Sigma_{11}^{-1} + \beta\Sigma_{22.1}^{-1}\beta' - \beta\Sigma_{22.1}^{-1}B'-B\Sigma_{22.1}^{-1}\beta' \qquad 7.2.19$$

The joint distribution of A_{11}, A_{22} and B is

$f(A_{11},A_{22},B) \propto$

$$|A_{11}|^{\frac{1}{2}(n-q+r-1)}|A_{22}-B'A_{11}B|^{\frac{1}{2}(n-q-r-1)} etr(-\tfrac{1}{2}\Sigma_{22.1}^{-1}A_{22}) etr(-\tfrac{1}{2}LA_{11}) \qquad 7.2.20$$

where the constant of proportionality is

$$(2^{\frac{1}{2}np}\Gamma_p(\tfrac{1}{2}n)|\Sigma_{11}|^{\frac{1}{2}n}|\Sigma_{22.1}|^{\frac{1}{2}n})^{-1} \qquad 7.2.21$$

To obtain the density function of B we must integrate 7.2.20 over A_{22} and A_{11}. Integrating with respect to A_{22}, we require

$$\int |A_{22}-B'A_{11}B|^{\frac{1}{2}(n-q-r-1)} \mathrm{etr}(-\tfrac{1}{2}\Sigma_{22.1}^{-1}A_{22})\, dA_{22} \qquad 7.2.22$$

where the integration is over all values of A_{22} such that $W = A_{22}-B'A_{11}B$ is positive definite. Transforming from A_{22} to W, noting that $J(A_{22} \to W) = 1$, the integral becomes

$$\mathrm{etr}(-\tfrac{1}{2}\Sigma_{22.1}^{-1}B'A_{11}B)\int_{W>0} |W|^{\frac{1}{2}(n-q-r-1)} \mathrm{etr}(-\tfrac{1}{2}\Sigma_{22.1}^{-1}W)dW \qquad 7.2.23$$

which by Theorem 2.5.2 has the value

$$2^{\frac{1}{2}(n-q)r}\Gamma_r(\tfrac{1}{2}(n-q))|\Sigma_{22.1}^{-1}|^{\frac{1}{2}(n-q)} \mathrm{etr}(-\tfrac{1}{2}\Sigma_{22.1}^{-1}B'A_{11}B) \qquad 7.2.24$$

In the joint density of B and A_{11}, we then integrate over A_{11} and hence require

$$\int_{A_{11}>0} |A_{11}|^{\frac{1}{2}(n-q+r-1)} \mathrm{etr}(-\tfrac{1}{2}(L+B\Sigma_{22.1}^{-1}B')A_{11})dA_{11} \qquad 7.2.25$$

As above, the integral has the value

$$2^{\frac{1}{2}(n+r)q}\ \Gamma_q(\tfrac{1}{2}(n+r))|L+B\Sigma_{22.1}^{-1}B'|^{-\frac{1}{2}(n+r)} \qquad 7.2.26$$

Substituting for L from 7.2.19, it follows that the density of B is

$$f(B) \propto |\Sigma_{11}^{-1} + (B-\beta)\Sigma_{22.1}^{-1}(B-\beta)'|^{-\frac{1}{2}(n+r)}$$

From 7.2.21, 7.2.24 and 7.2.26, the constant of proportionality is

$$\frac{\Gamma_r(\frac{1}{2}(n-q))\Gamma_q(\frac{1}{2}(n+r))}{\Gamma_p(\frac{1}{2}n)|\Sigma_{11}|^{\frac{1}{2}n}|\Sigma_{22.1}|^{\frac{1}{2}q}} \qquad 7.2.27$$

By Theorem 2.2.2, $\dfrac{\Gamma_r(\frac{1}{2}(n-q))}{\Gamma_p(\frac{1}{2}n)} = \dfrac{1}{\pi^{\frac{1}{2}qr}\Gamma_q(\frac{1}{2}n)}$ 7.2.28

Then

$$f(B) = \frac{\Gamma_q(\frac{1}{2}(n+r))}{\pi^{\frac{1}{2}qr}\Gamma_q(\frac{1}{2}n)} |\Sigma_{11}|^{-\frac{1}{2}n}|\Sigma_{22.1}|^{-\frac{1}{2}q} \times |\Sigma_{11}^{-1}+(B-\beta)\Sigma_{22.1}^{-1}(B-\beta)'|^{-\frac{1}{2}(n+r)} \qquad 7.2.29$$

so B has a matrix T distribution centred at β. Denoting the matrix of sample regression coefficients of $X^{(1)}$ on $X^{(2)}$ by $B_{1/2} = A_{22}^{-1}A_{21}$, we see that its distribution can be obtained from 7.2.29 by interchanging q and r and the subscripts 1 and 2. In an obvious notation we can summarise the above by writing

$$B_{2/1} \sim T(\Sigma_{11},\Sigma_{22.1},\beta_{2/1},n+r) \qquad 7.2.30$$

$$B_{1/2} \sim T(\Sigma_{22},\Sigma_{11.2},\beta_{1/2},n+q) \qquad 7.2.31$$

We shall only use the subscripts on B when it is necessary to distinguish between the two distributions.

Kshirshagar (1960) derived the distribution of B under the assumption that $\mu = 0$. Kabe (1968), obtained the joint distribution of $Z^{(2)}$ (see 7.2.9) and B and showed that the marginal distribution of B is still matrix T as given above. We are of course considering the unconditional distribution of B. If we work with the conditional distribution of $X^{(2)}$ given $X^{(1)} = x^{(1)}$, then conditionally the matrix B has a matrix normal distribution. The distribution theory of B is then the same as in the general linear model and is given in Anderson (1958), Chapter 8. For a comparison between the two models and the distribution of B in the two cases see Troskie (1971). In Chapter 5, we considered the distribution of $\tilde{T} = (\frac{Z}{n})^{-1}A$. Suppose that

$\tilde{T}$ is partitioned as A above, and define $\tilde{B} = T_{11}^{-1}T_{12}$. As Cornish (1954) observed, $(\frac{Z}{n})$ cancels out in $\tilde{T}_{11}^{-1}\tilde{T}_{12}$ and so

$$\tilde{B} = \tilde{T}_{11}^{-1}\tilde{T}_{12} = A_{11}^{-1}A_{12} = B$$

Thus, (Kshirsagar, 1960) the distribution of $\tilde{B}$ is the same as that of B.

Suppose now that $X \sim N(0,\Sigma)$ and is partitioned into three subvectors, $X' = (X^{(1)},X^{(2)},X^{(3)})'$ with q,r and s components respectively. If $X^{(3)}$ is fixed at some value $x^{(3)}$, then the conditional distribution of $(X^{(1)},X^{(2)}|x^{(3)})'$ is again normal with covariance matrix

$$\Sigma_{\cdot 3} = \begin{pmatrix} \Sigma_{11.3} & \Sigma_{12.3} \\ \Sigma_{21.3} & \Sigma_{22.3} \end{pmatrix} \qquad 7.2.32$$

where $\Sigma_{ij.3} = \Sigma_{ij} - \Sigma_{i3}\Sigma_{33}^{-1}\Sigma_{3j}$

If $X_{(\alpha)}$, $\alpha = 1,\ldots,N$ is a random sample from X, and A is the Wishart matrix, then

$$A_{\cdot 3} = \begin{pmatrix} A_{11.3} & A_{12.3} \\ A_{21.3} & A_{22.3} \end{pmatrix} \qquad 7.2.33$$

By Anderson (1958), $A_{\cdot 3} \sim W_{q+r}(\Sigma_{\cdot 3}, n-s)$. The matrix of partial regression coefficients of $(X^{(2)}|x^{(3)})$ on $(X^{(1)}|x^{(3)})$ can be defined as

$$\beta_{\cdot 3} = \Sigma_{11.3}^{-1}\Sigma_{12.3} \qquad 7.2.34$$

and its estimator is

$$B_{\cdot 3} = A_{11.3}^{-1}A_{12.3} \qquad 7.2.35$$

Starting with the density function of $A_{\cdot 3}$, we can apply the same derivation as Kshirsagar (1960) to obtain the result that

$$B_{\cdot 3} \sim T(\Sigma_{11.3}, \Sigma_{22.13}, \beta_{\cdot 3}, n-s+r) \qquad 7.2.36$$

where $$\Sigma_{22.13} = \Sigma_{22.3} - \Sigma_{21.3}\Sigma_{22.3}^{-1}\Sigma_{21.3} \qquad 7.2.37$$

We see that for the partial case, the parameter matrices are replaced by conditional matrices and the degrees of freedom are reduced by the number of variables held fixed in the third set. This is analogous to the results for the distribution of the partial correlation coefficient (Anderson, 1958) and the generalised partial correlation matrix (Troskie (1968), Juritz (1970)).

7.3 THE NON-CENTRAL DISTRIBUTION OF THE MATRIX OF REGRESSION COEFFICIENTS

We now consider the distribution of $B = A_{11}^{-1}A_{12}$ when the matrix A has a non-central Wishart distribution. The formulation of the problem is straightforward but a complete solution to it appears to be mathematically intractable because of the difficulty of evaluating the required integrals. Solutions for certain special cases have appeared in the literature (Basmann (1961, 1963); Kabe (1963, 1964); Richardson (1968); Sawa (1969); Richardson and Wu (1971)). These distributions arose as the result of investigations into the distribution of the two stage least-squares estimators of certain structural parameters in a system of simultaneous linear structural equations. It seems that the connection between these results and the matrix T distribution has not been recognised. In this section we formulate the general problem, to indicate the difficulties that stand in the way of a complete solution.

Let $X_{(\alpha)}, \alpha = 1,\ldots,N$ be independent random variables, with $X_{(\alpha)} \sim N_p(\mu_{(\alpha)},\Sigma)$, i.e. the variables have a common covariance matrix, but some or all of their means differ. Let

$$A = \sum_{\alpha=1}^{N} (X_{(\alpha)}-\overline{X})(X_{(\alpha)}-\overline{X})' \qquad 7.3.1$$

and

$$\tau = \sum_{\alpha=1}^{N} (\mu_{(\alpha)}-\overline{\mu})(\mu_{(\alpha)}-\overline{\mu})' \qquad 7.3.2$$

$$\overline{\mu} = \frac{1}{N} \sum_{\alpha=1}^{N} \mu_{(\alpha)} \qquad 7.3.3$$

Let

$$\Omega = \tfrac{1}{2}\tau\Sigma^{-1} \qquad 7.3.4$$

Then it follows that $A \sim W_p(\Sigma,n,\Omega)$ where $n = N-1$ and the density function of A is

$$f(A) = \frac{\mathrm{etr}(-\Omega)|A|^{\frac{1}{2}(n-p-1)}\mathrm{etr}(-\frac{1}{2}\Sigma^{-1}A)}{2^{\frac{1}{2}np}\,\Gamma_p(\frac{1}{2}n)|\Sigma|^{\frac{1}{2}n}}\ {}_0F_1(\tfrac{1}{2}n;\ \tfrac{1}{2}\Sigma^{-1}\Omega A) \qquad 7.3.5$$

As in 7.2, we partition $X_{(\alpha)}$ into two subvectors with q and r components, and partition A and Σ accordingly. Further let

$$\Omega = \begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{pmatrix} \quad \text{and} \quad \phi = \Sigma^{-1}\Omega = \begin{pmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{pmatrix} \qquad 7.3.6$$

where Ω and ϕ are partitioned as A and Σ. To obtain the distribution of $B = A_{11}^{-1}A_{12}$ from that of A we must transform from A_{12} to B in 7.3.5 by

$$A_{12} = A_{11}B \qquad 7.3.7$$

and using Equations 7.2.13 - 7.2.19, we obtain

$$f(A_{11},A_{22},B) \propto |A_{11}|^{\frac{1}{2}(n-q+r-1)}|A_{22}-B'A_{11}B|^{\frac{1}{2}(n-q-r-1)}$$

$$\times \operatorname{etr}(-\tfrac{1}{2}\Sigma_{22.1}^{-1}A_{22})\operatorname{etr}(-\tfrac{1}{2}LA_{11})\,{}_0F_1(\tfrac{1}{2}n;\ g(\phi,A_{11},A_{22}B)) \qquad 7.3.8$$

where $g(\phi,A_{11},A_{22},B) = \begin{pmatrix} \phi_{11}A_{11}+\phi_{12}B'A_{11} & \phi_{11}A_{11}+\phi_{12}A_{22} \\ \phi_{21}A_{11}+\phi_{22}B'A_{11} & \phi_{21}A_{11}B+\phi_{22}A_{22} \end{pmatrix}$ 7.3.9

To obtain the density of B, 7.3.8 must be integrated over A_{11} and A_{22}. It does not seem possible to evaluate the required integrals, since no expressions are known for integrals involving zonal polynomials of partitioned matrices. However, some attack can be made upon the problem if A has a linear or planar non-central Wishart distribution with $\Sigma = I$, and this will be the subject of the next section.

7.4 THE CANONICAL FORM OF THE NON-CENTRAL DISTRIBUTION OF THE MATRIX OF REGRESSION COEFFICIENTS

In this section we consider the distribution of $B_{2/1} = A_{11}^{-1}A_{12}$ and $B_{1/2} = A_{22}^{-1}A_{21}$. When $A \sim W_p(I_p,n,\kappa^2)$. We shall show that in this case $B_{2/1}$ has a lower non-central matrix T distribution, but $B_{1/2}$ has a different type of distribution. Let $Y_{(\alpha)}$, $\alpha = 1,\ldots,n$ be independent random variables with

(i) $Y_{(1)} \sim N_p(\mu_{(1)}, I_p)$ where $\mu'_{(1)} = |\kappa, 0 \ldots 0|'$

(ii) $Y_{(\alpha)} \sim N_p(0, I_p)$ for $\alpha = 2 \ldots n$ 7.4.1

The distribution of the $Y_{(\alpha)}$'s implies that $A = \sum_{\alpha=1}^{n} Y_{(\alpha)} Y'_{(\alpha)}$ is distributed as the canonical form of the linear non-central Wishart distribution (Anderson (1946)). Suppose that $Y_{(\alpha)}$, $\alpha = 1, \ldots, n$ is partitioned into two subvectors $Y^{(1)}_{(\alpha)}$ and $Y^{(2)}_{(\alpha)}$ with q and $p-q = r$ components respectively and let A be partitioned accordingly (see 7.2.11) and $B = A_{11}^{-1} A_{12}$ be the matrix of sample regression coefficients of $Y^{(2)}_{(\alpha)}$ on $Y^{(1)}_{(\alpha)}$.

The density function of B can be found by transforming the density function of A. Since $A \sim W_p(I_p, n, \kappa^2)$ by Theorem 2.5.6 the density of A is

$$f(A) \propto |A|^{\frac{1}{2}(n-p-1)} \operatorname{etr}(-\tfrac{1}{2}A) \sum_{i=0}^{\infty} \phi(i) a_{11}^{i} \qquad 7.4.2$$

where the constant of proportionality is

$$\frac{\exp(-\frac{1}{2}\kappa^2)}{2^{\frac{1}{2}np}\Gamma_p(\frac{1}{2}n)} \qquad 7.4.3$$

and $$\phi(i) = \frac{(\frac{1}{2}\kappa^2)^i \Gamma(\frac{1}{2}n)}{i!\, 2^i \Gamma(\frac{1}{2}n+i)} \qquad 7.4.4$$

Partitioning A as in 7.2.11, we transform from A_{12} to B by

$$B = A_{11}^{-1} A_{12} \qquad 7.4.5$$

Then $$J(A_{12} \to B) = |A_{11}|^r \qquad 7.4.6$$

Noting that $p = q+r$ and using 7.2.15, the joint density of A_{11}, A_{12} and B is

$$f(A_{11}, A_{12}B) \propto$$

$$|A_{11}|^{\frac{1}{2}(n+r-q-1)} |A_{22}-B'A_{11}B|^{\frac{1}{2}(n-q-r-1)} \operatorname{etr}(-\tfrac{1}{2}A_{11}) \operatorname{etr}(-\tfrac{1}{2}A_{22})$$

$$\times \sum_{i=0}^{\infty} \phi(i) a_{11}^{i} \qquad 7.4.7$$

We first integrate over A_{22}, and require

$$\int |A_{22}-B'A_{11}B|^{\frac{1}{2}(n-q-r-1)} \operatorname{etr}(-\tfrac{1}{2}A_{22})\, dA_{22} \qquad 7.4.8$$

In 7.4.8, let $(A_{22}-B'A_{11}B) = W$, then $J(A_{22} \to W) = 1$ and the integral becomes

$$\operatorname{etr}(-\tfrac{1}{2}B'A_{11}B) \int_{W>0} |W|^{\frac{1}{2}(n-q-r-1)} \operatorname{etr}(-\tfrac{1}{2}W)\, dW \qquad 7.4.9$$

which by Theorem 2.5.2 has the value,

$$\operatorname{etr}(-\tfrac{1}{2}B'A_{11}B)(\Gamma_r(\tfrac{1}{2}(n-q))2^{\frac{1}{2}r(n-q)}) \qquad 7.4.10$$

Hence $f(A_{11}, B) \propto |A_{11}|^{\frac{1}{2}(n-r-q-1)} \operatorname{etr}(-\frac{1}{2}(I+BB')A_{11})$

$$\times \sum_{i=0}^{\infty} \phi(i) a_{11}^{i} \qquad 7.4.11$$

where the constant of proportionality is given by 7.4.3 and the second term of 7.4.10. We must now integrate termwise over A_{11}, and so must evaluate integrals of the form

$$\int_{A_{11}>0} a_{11}^{i} |A_{11}|^{\frac{1}{2}(n+r-q-1)} \operatorname{etr}(-\tfrac{1}{2}(I+BB')A_{11})\, dA_{11} \qquad 7.4.12$$

But 7.4.12 is proportional to $E(a_{11}^{i})$ where a_{11} is the $(1,1)^{th}$ element of $A_{11} \sim W_q(n+r, (I+BB')^{-1})$. Hence the value of the integral is

$$\frac{2^{\frac{1}{2}(n+r)q}\Gamma_q(\frac{1}{2}(n+r))}{|I+BB'|^{\frac{1}{2}(n+r)}} \cdot \frac{2^i\Gamma(\frac{1}{2}(n+r)+i)(\omega^{11})^i}{\Gamma(\frac{1}{2}(n+r))} \qquad 7.4.13$$

where ω^{11} is the $(1,1)^{th}$ element of $(I+BB')^{-1}$. (See Chapter 6, Equations 6.3.27 - 6.3.29 where a similar integral is evaluated.)

Evaluating the constant terms from 7.4.3, 7.4.4, 7.4.10 and 7.4.13, the density function of B is

$$f(B) = \frac{\exp(-\frac{1}{2}\kappa^2)\Gamma_r(\frac{1}{2}(n-q))\Gamma_q(\frac{1}{2}(n+r))}{\Gamma_p(\frac{1}{2}n)|I+BB'|^{\frac{1}{2}(n+r)}}$$

$$\times \sum_{i=o}^{\infty} \frac{(\frac{1}{2}\kappa^2)^i\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}(n+r)+i)(\omega_1^{11})^i}{i!\Gamma(\frac{1}{2}n+i)\Gamma(\frac{1}{2}(n+r))} \qquad 7.4.14$$

The infinite series in 7.4.14 is ${}_1F_1(\frac{1}{2}(n+r);\ \frac{1}{2}n, \frac{1}{2}\ {}^{2}\ {}^{11})$ and by Theorem 2.2.2,

$$\frac{\Gamma_r(\frac{1}{2}(n-q))}{\Gamma_p(\frac{1}{2}n)} = \frac{1}{\pi^{\frac{1}{2}qr}\Gamma_q(\frac{1}{2}n)} \qquad 7.4.15$$

So $$f(B) = \frac{\exp(-\frac{1}{2}\kappa^2)\Gamma_q(\frac{1}{2}(n+r))}{\pi^{\frac{1}{2}qr}\Gamma_q(\frac{1}{2}n)} |I_q+BB'|^{-\frac{1}{2}(n+r)}$$

$$\times\ {}_1F_1(\tfrac{1}{2}(n+r);\ \tfrac{1}{2}n, \tfrac{1}{2}\kappa^2\omega^{11}) \qquad 7.4.16$$

We now compare 7.4.16 with Equation 6.3.20 (Chapter 6, Theorem 6.3.2) and make the following identifications:

7.4.16	6.3.20
κ^2	λ^2
q	p
n+r	m
r	q
n	m-q
I_q+BB'	I_p+TT'
ω^{11}	τ^{11}

Thus we conclude that B has a lower non-central T distribution, centred at the origin with n+r degrees of freedom and non-centrality parameter κ^2, i.e.

$$B \sim T(I_q, I_r, 0, n+r, \kappa^2) \qquad 7.4.17$$

Before we summarise this result in a theorem, let us consider more deeply the conditions under which this result holds. Returning the original sample of the $Y_{(\alpha)}$'s, we have under the partitioning

$$Y_{(\alpha)} = \begin{bmatrix} Y^{(1)}_{(\alpha)} \\ Y^{(2)}_{(\alpha)} \end{bmatrix} \qquad \alpha = 1,\ldots,n$$

For $\alpha = 1$, $Y^{(1)}_{(1)} \sim N_q(\mu^{(1)}, I_q)$ and $Y^{(2)}_{(1)} \sim N_r(0, I_r)$

where the $q\times 1$ vector, $\mu^{(1)} = (\kappa, 0\ldots 0)$.

For $\alpha = 2,\ldots,n$, $Y^{(1)}_{(\alpha)} \sim N_q(0, I_q)$ and $Y^{(2)}_{(\alpha)} \sim N_r(0, I_r)$.

Thus $A_{11} = \sum_{\alpha=1}^{n} Y^{(1)}_{(\alpha)} Y^{(1)\prime}_{(\alpha)}$ has a linear non-central Wishart distribution, and as we have shown above $B_{2/1} = A_{11}^{-1} A_{12}$ has a lower non-central T distribution.

Now under this set-up, $A_{22} = \Sigma\, Y_{(\alpha)}^{(2)} Y_{(\alpha)}^{(2)'}$ has a central Wishart distribution and the distribution of $B_{1/2} = A_{22}^{-1}A_{21}$ is different from that of $B_{2/1}$ as we shall show later.

Note also that if one of the elements of $Y_{(1)}^{(2)}$ had the single non-zero mean, then A_{22} would have a non-central Wishart and so $A_{22}^{-1}A_{21}$ would have a matrix T distribution. The essential point we are making is, that the distribution of B is lower non-central T if and only if A_{ii}^{-1} has a non-central Wishart distribution.

In view of these remarks we can formulate the next theorem precisely.

Theorem 7.4.1(a)

Let $A \sim W_p(I_p, n, \kappa^2)$ *and be partitioned as*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{11} *is* $q\times q$, A_{12} *is* $q\times r$ *and* A_{22} *is* $r\times r$. *Then if the partitioning is such that* $A_{11} \sim W_q(I_r, n, \kappa^2)$ *then* $B_{2/1} = A_{11}^{-1}A_{12} \sim T(I_q, I_r, 0, n+r, \kappa^2)$.

Interchanging q and r and the subscripts 1 and 2 we have

Theorem 7.4.1(b)

Let $A \sim W_p(I_p, n, \kappa^2)$ *and be partitioned as in Theorem 7.4.1(a). Then if the partitioning is such that* $A_{22} \sim W_r(I_r, n, \kappa^2)$ *then*

$$B_{1/2} = A_{22}^{-1}A_{21} \sim T(I_r, I_q, 0, n+q, \kappa^2)$$

The density function of $B_{1/2}$ *is*

$$f(B) = \frac{exp(-\frac{1}{2}\kappa^2)\Gamma_r(\frac{1}{2}(n+q))}{\pi^{\frac{1}{2}qr}\Gamma_r(\frac{1}{2}n)} |I_r+BB'|^{-\frac{1}{2}(n+q)}$$

$$\times {}_1F_1(\tfrac{1}{2}(n+q);\ \tfrac{1}{2}n;\ \tfrac{1}{2}\kappa^2\omega^{11})$$

Of course this result implies that all the properties of the lower non-central T hold for the distribution of B. In particular $V = BB'$ is distributed as multivariate Beta type A (linear case). Much is known about the distribution of V, (Troskie (1966), de Waal (1968), Money (1972)) and it seems that if B arises as the "natural" statistic in any application, it would be better to work with BB'. We shall not develop this aspect of the distribution of B further because we wish to concentrate on the relationship between the distribution of B and the underlying multivariate normal population from which it arose. We shall merely state this as a corollary to Theorem 7.4.1.

__Corollary 7.4.1__

If $B \sim T(I_q, I_r 0, n+r, \kappa^2)$, *and* $r \geq q$ *then* $V = BB'$ *has a linear non-central multivariate Beta Type 2A distribution with* r *and* $n+r$ *degrees of freedom and non-centrality parameter* κ^2.

Referring to the density function of B given in 7.4.16 we see that the argument of the ${}_1F_1$ series depends on ω^{11}, the $(1,1)^{th}$ element of $|I+BB'|$. However this is only because we assumed that the first element of $Y_{(1)}^{(1)}$ had the single non-zero mean value.

It is clear that we would obtain the same form of density function if any other element of $Y_{(1)}^{(1)}$ had the single non-zero mean value, the only change being in the argument of the ${}_1F_1$ series. More precisely, if the i^{th} element of $Y_{(1)}^{(1)}$ has the single non-zero mean value then the argument of the ${}_1F_1$ series is ω^{ii}, where ω^{ii} is the $(i,i)^{th}$ element of $(I+BB')^{-1}$. This is a small point but it must be borne in mind in applications.

We now consider the distribution of $B = A_{22}^{-1}A_{12}$ when the partitioning of A is such that A_{22} has a central Wishart distribution.

Suppose that $Y_{(\alpha)}$, $\alpha = 1,\ldots,n$ are distributed as in 7.4.1, and each $Y_{(\alpha)}$ is partitioned into two sub-vectors $Y_{(\alpha)}^{(1)}$ and $Y_{(\alpha)}^{(2)}$ with dimensions q and r respectively. Then, as before $A = \Sigma Y_{(\alpha)}Y_{(\alpha)}' \sim W_p(I_p,n,\kappa^2)$. The matrix of regression coefficients of $Y^{(1)}$ on $Y^{(2)}$ is

$$B_{1/2} = A_{22}^{-1}A_{21} \tag{7.4.18}$$

The density of B can be found from that of A as follows:

$$f(A) \propto |A|^{\frac{1}{2}(n-p-1)}\mathrm{etr}(-\tfrac{1}{2}A)\sum_{i=o}^{\infty} \phi(i)a_{11}^{i} \tag{7.4.19}$$

where the constant of proportionality is given by 7.4.3 and $\phi(i)$ by 7.4.4. We transform from A_{12} to B in 7.4.19 by $B = A_{22}^{-1}A_{21}$. The Jacobian of the transformation is

$$J(A_{21} \to B) = |A_{22}|^{q} \tag{7.4.20}$$

Noting that

$$|A| = |A_{22}||A_{11}-B'A_{22}B| \qquad 7.4.21$$

$$f(A_{11},A_{22},B) \propto |A_{22}|^{\frac{1}{2}(n+q-r-1)}|A_{11}-B'A_{22}B|^{\frac{1}{2}(n-q-r-1)}$$

$$\times\ \text{etr}(-\tfrac{1}{2}A_{11})\text{etr}(-\tfrac{1}{2}A_{22})\sum_{i=o}^{\infty} \phi(i)a_{11}^{i} \qquad 7.4.22$$

To obtain the density of B, we must integrate over A_{11} and A_{22}. Integrating with respect to A_{11} first we must evaluate integrals of the form

$$\int a_{11}^{i}|A_{11}-B'A_{22}B|^{\frac{1}{2}(n-q-r-1)}\text{etr}(-\tfrac{1}{2}A_{11})dA_{11} \qquad 7.4.23$$

In 7.4.23 let $(A_{11}-B'A_{22}B) = W$

so that $A_{11} = W+B'A_{22}B$ 7.4.24

and $J(A_{11} \to W) = 1$

Now $a_{11} = w_{11} + b'_{(1)}A_{22}b_{(1)}$

where $b_{(1)}$ is the first column of B

and w_{11} is the $(1,1)^{th}$ element of W.

Thus 7.4.23 becomes

$$\text{etr}(-\tfrac{1}{2}B'A_{22}B)\int_{W>0}(w_{11}+b'_{(1)}A_{22}b_{(1)})^{i}|W|^{\frac{1}{2}(n-q-r-1)}\text{etr}(-\tfrac{1}{2}W)dW$$

7.4.25

Expanding $(w_{11}+b'_{(1)}A_{22}b_{(1)})^{i}$ by the binomial theorem 7.4.25 becomes

$$\text{etr}(-\tfrac{1}{2}B'A_{22}B)\sum_{j=o}^{i}\binom{i}{j}(b'_{(1)}A_{22}b_{(1)})^{j}I(W) \qquad 7.4.26$$

where

$$I(W) = \int_{W>0} w_{11}^{i-j}|W|^{\frac{1}{2}(n-q-r-1)}\text{etr}(-\tfrac{1}{2}W)dW \qquad 7.4.27$$

But $I(W)$ is proportional to the $(i-j)^{th}$ moment of a χ^2 variable with $(n-r)$ degrees of freedom and has the value

$$\psi(i,j) = 2^{\frac{1}{2}(n-r)q}\Gamma_q(\tfrac{1}{2}(n-r))2^{i-j}\frac{\Gamma(\frac{1}{2}(n-r)+i-j)}{\Gamma(\frac{1}{2}(n-r))} \qquad 7.4.28$$

Hence 7.4.26 after integration becomes

$$\operatorname{etr}(-\tfrac{1}{2}B'A_{22}B)\sum_{j=o}^{i}\binom{i}{j}(b'_{(1)}A_{22}b_{(1)})^j\psi(i,j) \qquad 7.4.29$$

So $f(A_{22}B) \propto$

$$\sum_{i=o}^{\infty}\sum_{j=o}^{i}\binom{i}{j}\phi(j)\psi(i,j)(b'_{(1)}A_{22}b_{(1)})^j|A_{22}|^{\frac{1}{2}(n+q-r-1)}$$

$$\times \operatorname{etr}(\tfrac{1}{2}(I+BB')A_{22}) \qquad 7.4.30$$

which must be integrated over A_{22} to give $f(B)$. Consider

$$c^{-1}\int(b'_{(1)}A_{22}b_{(1)})^j|A_{22}|^{\frac{1}{2}(n+q-r-1)}\operatorname{etr}(\tfrac{1}{2}(I+BB')A_{22})dA_{22}$$

7.4.31

where $c = 2^{\frac{1}{2}(n+q)r}\Gamma_r(\frac{1}{2}(n+q))|I+BB'|^{-\frac{1}{2}(n+q)}$ 7.4.32

Then 7.4.31 is $E((b'_{(1)}A_{22}b_{(1)})^j)$ where

$A_{22} \sim W_r((I+BB')^{-1},n+q)$. Note also that for the purposes of integration $b_{(1)}$ is a fixed $r\times 1$ vector. We can now use a result given in Rao (1965) page 452, Equation 8 b2 which states:

"if $S \sim W_p(\Sigma,k)$ and L is a fixed vector, then

$$L'SL \sim \sigma_L^2\ \chi_k^2$$

where $\sigma_L^2 = L'\Sigma L$."

Using this result with $S = A_{22}$, $L = b_{(1)}$, $p = r$, $k = n+q$ and $\Sigma = (I+BB')^{-1}$, we see that

$$\sigma_L^2 = b'_{(1)}(I+BB')^{-1}b_{(1)}$$

and $b'_{(1)}A_{22}b_{(1)}$ is distributed as

$$(b'_{(1)}(I+BB')^{-1}b_{(1)})\chi^2_{n+q} \qquad 7.4.33$$

The integral in 7.4.31 is the j^{th} moment of 7.4.33 and so has the value $2^j[b'_{(1)}(I+BB')^{-1}b_{(1)}]^j \cdot \frac{\Gamma(\frac{1}{2}(n+q)+j)}{\Gamma(\frac{1}{2}(n+q))}$ 7.4.34

The integral we require is 7.4.34 multiplied by c. Hence by 7.4.34, 7.4.32, 7.4.28, 7.4.3, 7.4.4 and using Theorem 2.2.2 on the multivariate gamma functions, the density of B is

$$f(B) = \frac{\exp(-\frac{1}{2}\kappa^2)\Gamma_r(\frac{1}{2}(n+q))}{\pi^{\frac{1}{2}qr}\Gamma_r(\frac{1}{2}n)} |I+BB'|^{-\frac{1}{2}(n+q)}$$

$$\times \sum_{i=o}^{\infty} \sum_{j=o}^{i} \binom{i}{j} \frac{(\frac{1}{2}\kappa^2)^i \Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}(n-r)+i-j)\Gamma(\frac{1}{2}(n+q)+j)}{i!\Gamma(\frac{1}{2}n+i))\Gamma(\frac{1}{2}(n-r))\Gamma(\frac{1}{2}(n+q))}$$

$$\times (b'_{(1)}(I+BB')^{-1}b_{(1)})^j \qquad 7.4.35$$

We summarise this result in the next theorem.

Theorem 7.4.2(a)

Let $A \sim W_p(I,n,\kappa^2)$. If A is partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{11} is $q \times q$, A_{21} is $q \times r$ and A_{22} is $r \times r$ and the partitioning of A is such that A_{22} has a central Wishart distribution, then the density function of B

$$B = A_{22}^{-1}A_{21}$$

is given by 7.4.35.

Interchanging q and r and the subscripts 1 and 2 we have

Theorem 7.4.2(b)

If the partitioning of A *is such that* A_{11} *has a central Wishart distribution, then the density function of*

$$B = A_{11}^{-1}A_{12}.$$

can be obtained from 7.4.35 by interchanging q *and* r.

7.5 THE DISTRIBUTION OF TWO STAGE LEAST SQUARES ESTIMATORS

Economic Theories can often be described mathematically by a system of linear equations each of which describes some aspect or sector of the economy at a given time t. Two types of variables appear in the equations, namely

(i) Endogenous variables - those whose values are accounted for by the model.

(ii) Exogenous variables - those whose values are determined outside the economy and are not accounted for by the model.

More generally the classification is into jointly dependent and predetermined variables. Jointly dependent variables are the current endogenous variables, as in (i) above; the predetermined variables are the current exogenous variables and the lagged endogenous variables. We shall assume that no lagged endogenous variables appear in the system so the classification will be (i) and (ii) above. Such systems of equations are called structural equations because they exhibit the explicit relationship between the exogenous and endogenous variables and thus explain the structure of the economy.

For a given time t, a system of G structural equations can be written as

$$By_t + \Gamma x_t = u_t \qquad 7.5.1$$

where

$B = (\beta_{ij})$ is a $G\times G$ matrix of coefficients of the endogenous variables

$\Gamma = (\gamma_{ij})$ is a $G\times K$ matrix of coefficients of exogenous variables

y_t is a $G\times 1$ vector of endogenous variables

x_t is a $K\times 1$ vector of exogenous variables

u_t is a $G\times 1$ vector of unobservable random disturbances

and $u_t \sim N_G(0,\phi)$.

Our aim is to estimate the coefficients of the endogenous variables in the first equation of the system 7.5.1, and to obtain the distribution of the estimators. With suitable scaling of the variables we can assume that each element of the first column of B is -1, so the first equation can be written in full as

$$-y_{1t} + \beta_{12}y_{2t}+\dots+\beta_{1G}y_{Gt} + \gamma_{1t}x_{1t}+\dots+\gamma_{1K}x_{1K} + u_{1t} = 0 \qquad 7.5.2$$

Assuming that B is non-singular, the equations 7.5.1 can be "solved" for the endogenous variables to give the so-called reduced form

$$y_t = \Pi x_t + v_t \qquad 7.5.3$$

where the $G\times K$ matrix $\Pi = B^{-1}\Gamma$ 7.5.4

and $v_t = B^{-1}u_t$ 7.5.5

The $G \times 1$ vector $v_t \sim N_G(0,\Sigma)$ where $\Sigma = B^{-1}\phi B'^{-1}$. Equation 7.5.5 shows that each endogenous variable y_{it} is influenced by each and every disturbance u_{it}, which implies that, in the structural equations 7.5.1, the y's are correlated with the errors and so the ordinary least estimates of B and Γ will be inconsistent. If however it is assumed that x_t and u_t are uncorrelated, the matrix of reduced form coefficients Π can always be estimated by standard methods of multivariate regression (see Koopmans and Hood (1953) page 155). A possible attack on the problem of obtaining estimates of the β's and γ's in the first equation would be to estimate the reduced form and then use equation 7.5.4 to give the required estimates. But, even if Π were known, it would not be possible to do this because we have only $G \times K$ elements of Π from which we must find $(G \times G)$ + $(G \times K)$ elements of B and Γ. This is called the Identification Problem, and is discussed in detail in Koopmans and Hood (1953) or Goldberger (1964). A solution is found by using a priori knowledge of the economic system which tells us that certain variables do not appear in certain equations. In statistical terms this means that some of the coefficients in equations such as 7.5.2 are zero. These restrictions also place restrictions on Π (see Koopmans and Hood (1953)) and if they are correct, they enable us to solve for B and Γ uniquely from Π.

Let us consider in more detail the identification of the single equation 7.5.2. Suppose that a priori information suggests that G_1 endogenous and K_1 exogenous variables appear in the equation. Let $\beta = \left(\beta_1, 0\right)$ be the first row of B where β_1 is the $G_1 \times 1$ coefficient vector of the endogenous variables included in the equation and the remaining $G_2 = G - G_1$ elements of β are zero. Let $\gamma = \left(\gamma_1, 0\right)$ be the first row of Γ where γ_1 is the coefficient vector of the K_1 exogenous variables included in the equation and the remaining $K_2 = K - K_1$ elements of γ are zero. To obtain β and γ from Π it follows from 7.5.4 that we must solve

$$-\beta\Pi = \gamma \qquad 7.5.6$$

Partitioning Π by rows and columns to conform with the partitioning of β and γ, 7.5.6 can be written as

$$-\left(\beta_1, 0\right) \begin{pmatrix} \Pi_1^1 & \Pi_2^1 \\ \Pi_1^2 & \Pi_2^2 \end{pmatrix} = \left(\gamma_1, 0\right) \qquad 7.5.7$$

where the superscript refers to the endogenous coefficients and the subscript to the exogenous coefficients. In both cases the scripts have the value 1 if the variables are included and 2 if they are excluded. Then, (see Koopmans and Hood (1953)) necessary and sufficient conditions for the identification of the first equation are

$$\text{rank}\ (\Pi_2^1) = G_1 - 1 \qquad 7.5.8$$

which is called the "Rank condition for identifiability". But Π_2^1 is a $G_1 \times K_2$ matrix, and since the rank of a matrix cannot exceed the number of columns, 7.5.8 also implies the "Order Condition for identifiability"

$$K_2 \geq G_1 - 1 \qquad 7.5.9$$

It can be seen from 7.5.7, that to obtain the non-zero structural coefficients from the reduced form we need only use the submatrices Π_1^1 and Π_2^1 and in particular for β_1 we need only Π_2^1.

Later on we shall discuss equations in which all G endogenous variables appear. In this case Π is partitioned column-wise as

$$\Pi = (\Pi_1, \Pi_2) \qquad 7.5.10$$

where Π_1 is $G \times K_1$ and Π_2 is $G \times K_2$ and 7.5.7 is now

$$\beta(\Pi_1 \Pi_2) = \left(\gamma_1, 0\right)$$

The rank condition for identifiability is then

$$\text{rank } (\Pi_2) = G-1 \qquad 7.5.11$$

and the order condition is

$$K_2 \geq G-1 \qquad 7.5.12$$

In this case also, to find β we need only consider the submatrix Π_2.

Now in applications Π is unknown and we have only its estimate P which can be partitioned as Π in 7.5.7 or 7.5.10 so that

$$P = \begin{pmatrix} P_1^1 & P_2^1 \\ P_1^2 & P_2^2 \end{pmatrix}$$

if G_1 endogenous variables appear in the first equation, or as

$$P = \left(P_1 \quad P_2\right) \qquad 7.5.13$$

if all G endogenous variables appear in the first equation. However since no restrictions are introduced when estimating P the $G_1 \times K_2$ matrix P_2^1 (or the $G \times K_2$

matrix P_2 in the case of 7.5.3) will only satisfy the rank condition for identifiability if $K_2 = G_1-1$ (or $K_2 = G-1$ in the case of 7.5.13). In this case it is said that the equation is just-identified. If $K_2 > G_1-1$ (or $K_2 > G-1$) then the equation is said to be over-identified. In either case rank $(\Pi_2^1) = G_1-1$ (or rank $(\Pi_2) = G-1$)) - it is only the rank of the estimating matrix P that differs.

If the equation is just-identified the structural parameters can be estimated by indirect least squares (see Goldberger (1964)) but this method is not applicable to an over-identified equation. Methods for estimating the structural parameter of an over-identified equation are given in Goldberger (1964). The various methods yield different estimates for the structural parameters and the sampling distributions of the estimators differ. Of these methods we shall only consider Two Stage Least Squares (abbreviated 2SLS). An alternative name for this method is Generalised Classical Linear/Least Variance Difference - a name given it by Basmann who was one of the originators of the method. This name seems to have fallen out of current statistical usage possibly because of the danger of confusion with the Least Variance Ratio method. Basmann draws a clear distinction between these two methods in Basmann (1960a) and (1960b). Indeed such a distinction is necessary because the two methods give different results.

Let us now consider the estimation of the structural parameters of the endogenous variables (the β's) in the first equation by 2SLS, and for simplicity of exposition assume that all G endogenous variables appear in the equation. We shall arrive at expressions for these estimates in terms of the submatrix P_2 of P. One would not use the expressions we obtain to compute an estimate - this can be accomplished without having to estimate the reduced form (see Goldberger (1964) page 329). In addition following Basmann (1963) who considered the case where $K = G = 3$, we make a number of simplifying assumptions which do not have too marked an effect on the form of the distribution.

Rewriting the first equation as

$$y_{1t} = \beta_{12}y_{2t}+\beta_{13}y_{3t}+\dots+\beta_{1G}y_{Gt}+\gamma_{11}x_{1t}+\dots+\gamma_{1K}x_{K}+u_t \quad 7.5.14$$

we assume that we have observations on the y's and x's for N points of time. From these we wish to estimate the $(G-1)$ coefficients

$$\beta_2' = (\beta_{12},\dots,\beta_{1G})' \quad 7.5.15$$

by 2SLS.

Let $Y = \left(Y_1 \; Y_2\right)$ be the $N\times G$ matrix of observations on the endogenous variables, where Y_1 is the vector of observations on the left hand endogenous variable in 7.5.14, and Y_2 is the $N\times(G-1)$ matrix of observations on the $(G-1)$ right hand endogenous variables.

Let $X = \left(X_1 \; X_2\right)$ be the $N\times K$ matrix of observations on the exogenous variables, where X_1 is the $N\times K_1$ matrix of observations on the exogenous variables included in the first equation and X_2 is the $N\times K_2$ matrix of observations on the exogenous variables excluded from the first equation under the identifiability restrictions. Let us further assume that the observations have been normalised so that

$$\sum_{t=1}^{N} y_{th} = \sum_{t=1}^{N} x_{tk} = 0 \qquad h = 1,\ldots,G,\ k = 1,\ldots,K$$

$$\sum_{t=1}^{N} x_{tk}x_{tm} = \begin{cases} 0 & m \neq k \\ 1 & m = k \end{cases} \qquad 7.5.16$$

These conditions imply that

$$X'X = I_k \quad X_1'X_1 = I_{k_1} \quad X_2'X_2 = I_{k_2} \qquad 7.5.17$$

We first consider the estimation of the reduced form.Following Koopmans and Hood (1953) we form the "moment" matrix

$$M = \begin{pmatrix} Y'Y & Y'X \\ X'Y & X'X \end{pmatrix} \qquad 7.5.18$$

The Least Squares/Maximum Likelihood estimator of Π is

$$P = Y'X(X'X)^{-1}$$

But in view of 7.5.17, $(X'X)^{-1} = I_k$, so

$$P = Y'X \qquad 7.5.19$$

From the partitioning of Y and X into included or excluded variables 7.5.19 can be written

$$P = \begin{pmatrix} Y_1' \\ Y_2' \end{pmatrix} \begin{pmatrix} X_1 & X_2 \end{pmatrix}$$

$$= \begin{pmatrix} Y_1' X_1 & Y_1' X_2 \\ Y_2' X_1 & Y_2' X_2 \end{pmatrix} \qquad 7.5.20$$

$$= \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \qquad 7.5.21$$

For estimation of β_2, we shall show later that we need only consider the two submatrices

$$\begin{pmatrix} P_{12} \\ P_{22} \end{pmatrix} = P_2 \qquad 7.5.22$$

where P_{12} is $1 \times K_2$ and P_{22} is $(G-1) \times K_2$. P_2 is the estimator of Π_2 defined in 7.5.10.

From Basmann (1960a) the 2SLS estimator of β_2 is obtained by minimising

$$Q_1(\beta_2) = G_1(\beta_2) - G_2(\beta_2) \qquad 7.5.23$$

with respect to β_2, where

$$G_1(\beta_2) = (Y_1 - Y_2\beta_2)'[I_N - X_1(X_1'X_1)^{-1}X_1'](Y_1 - Y_2\beta_2)$$

and $G_2(\beta_2) = (Y_1 - Y_2\beta_2)'[I_N - X(X'X)^{-1}X'](Y_1 - Y_2\beta_2)$

It now follows

$$Q_1(\beta_2) = (Y_1 - Y_2\beta_2)'[X(X'X)^{-1}X - X_1(X_1'X_1)^{-1}X_1'](Y - Y_2\beta_2)$$

$$= (Y_1 - Y_2\beta_2)'[XX' - X_1X_1'](Y_1 - Y_2\beta_2) \quad \text{(using 7.5.15)}$$

$$= (Y_1 - Y_2\beta_2)'(X_2X_2')(Y_1 - Y_2\beta_2) \quad \text{(using the partitioning of X)}$$

$$= (Y_1'X_2 - \beta_2'Y_2'X_2)(Y_1'X_2 - \beta_2'Y_2'X_2)' \qquad 7.5.24$$

Referring to the estimate of the reduced form P in Equation 7.5.19 we see that 7.5.24 can be written as

$$Q_1(\beta_2) = (P_{12} - \beta_2'P_{22})(P_{12} - \beta_2'P_{22})' \qquad 7.5.25$$

and we note that $Q_1(\beta_2)$ only involves the elements of P_2 defined in 7.5.22.

Differentiating 7.5.24 with respect to β_2 we see that for a minimum we must solve

$$-P_{22}P_{12}' + P_{22}P_{22}'\beta_2 = 0$$

Hence the minimum value occurs at

$$b = (P_{22}P_{22}')^{-1}P_{22}P_{12}' \quad 7.5.26$$

and b is the 2SLS estimator of β_2.

Now from 7.5.22 $P_2 = \begin{pmatrix} P_{12} \\ P_{22} \end{pmatrix}$ and

$$P_2P_2' = \begin{pmatrix} P_{12} \\ P_{22} \end{pmatrix} \begin{pmatrix} P_{12}' & P_{22}' \end{pmatrix}$$

$$= \begin{pmatrix} P_{12}\, P_{12}' & P_{12}\, P_{22}' \\ P_{22}\, P_{12}' & P_{22}\, P_{22}' \end{pmatrix} \quad 7.5.27$$

Let $P_2P_2' = A$ then 7.5.26 can be written in terms of the submatrices of A as

$$P_2P_2' = A = \begin{pmatrix} a_{11} & a_{(1)}' \\ a_{(1)} & A_{22} \end{pmatrix} \quad 7.5.28$$

where a_{11} is a scalar, $a_{(1)}$ is $(G-1)\times 1$ and A_{22} is $(G-1)\times(G-1)$. Hence the 2SLS estimator of β_2 can be written as

$$b = A_{22}^{-1}\, a_{(1)} \quad 7.5.29$$

We now consider the distribution of b. Consider first the distribution of $A = P_2P_2'$. P_2 is a submatrix of $P = (P_1, P_2)$. Since P is the Least Squares/ Maximum Likelihood estimator of Π, it follows that

$$E(P) = \Pi$$

and with the partitioning; $\Pi = (\Pi_1, \Pi_2)$ we have

$$E(P_2) = \Pi_2.$$

The columns of P_2 are independent normally distributed vectors with the same covariance matrix Σ (defined by 7.5.5) and with mean vectors that vary from column to column.

If $P_{(\alpha)}$ is the α^{th} column of P_2, then

$$E(P_{(\alpha)}) = \Pi_{(\alpha)}$$

where Π_α is the α^{th} column of Π_2. Since P_2 has K_2 such columns, we see that

$$A = P_2P_2' \sim W_G(\Sigma,K_2,\Omega)$$

where the noncentrality parameter is given by

$$\Omega = \tfrac{1}{2}\Pi_2\Pi_2'\Sigma^{-1} \qquad 7.5.30$$

In view of the identifiability conditions it follows that rank $(\Pi_2) = G-1$, so rank $(\Omega) = G-1$. Returning to the columns of P_2: suppose that each column is partitioned as

$$P_\alpha = \begin{bmatrix} P^1_{(\alpha)} \\ P^{(2)}_{(\alpha)} \end{bmatrix} \qquad \alpha = 1,\ldots,K_2$$

where $P^1_{(\alpha)}$ is the α^{th} element of P_{12} and $P^{(2)}_{(\alpha)}$ is the α^{th} column of P_{22} as defined in 7.5.22. Hence we recognise b as the matrix of sample regression coefficients of $P^1_{(\alpha)}$ on $P^{(2)}_\alpha$ and the distribution of b is a special case of the non-central distribution of the matrix of regression coefficients discussed in Section 7.3. We shall now show that under certain circumstances the distribution of b is lower non-central matrix T.

Let us first consider a special case of the distribution of b which was discussed by Basmann (1963) and Kabe (1964). Here it is assumed that $G = 3$, $K_2 = 3$, $\phi = I_3$ and rank $(\Omega) = 1$. We note that in this case only the necessary order condition is satisfied so the equation is not identified in the model. ($\text{rank}(\Omega) = 1 \Rightarrow \text{rank}(\Pi_2) = 1$, but for identification in the model we require $\text{rank}(\Pi_2) = 2$.) However the example does illustrate the connection between the distribution of the 2SLS estimators and the lower non-central matrix T distribution. This example is also discussed in Press (1972) but the summary contains 3 errors which we give below:

(i) The equation is not identified in the model as Press alleges.

(ii) The method of estimation is 2SLS and not Least Generalised Residual Variance.

(iii) The non-central Wishart distribution has 3 degrees of freedom, not N-1.

Proceeding with the example, we suppose that we wish to estimate the first equation in a system of 3 structural equations, each of which contains 3 endogenous variables and K exogenous variables. The first equation is thus

$$y_{t1} = \beta_{2t}y_{2t}+\beta_{3t}y_{3t}+\gamma_{1t}x_{1t}+\dots+\gamma_{1K}x_{Kt}+u_t$$

The identifiability conditions are

$$\gamma_{11} = \gamma_{12} = \gamma_{13} = 0.$$

and we wish to estimate $\beta_2' = (\beta_{2t}\beta_{3t})$ by 2SLS.

Let x_t be the vector of exogenous variables. Partitioning x_t according to the excluded exogenous variables we have $x_t = \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix}$

where
$$x_t^{(1)'} = x_{4t},\dots,x_{Kt}$$
$$x_t^{(2)'} = x_{1t},x_{2t},x_{3t}$$

Then the reduced form can be written

$$y_t = \begin{pmatrix} \Pi_1 & \Pi_2 \end{pmatrix} \begin{pmatrix} x_t^{(1)} \\ x_t^{(2)} \end{pmatrix} + v_t$$

where Π_1 is $3\times(K-3)$ and Π_2 is 3×3. $(K_2 = 3)$.
We need only consider

$$\Pi_2 = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} \\ \Pi_{31} & \Pi_{32} & \Pi_{33} \end{pmatrix}$$

and it is assumed that all the elements of Π_2 are zero except Π_{33}.

The corresponding submatrix of the reduced form is

$$P_2 = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}$$

Assuming that $v_t \sim N_3(0,I_3)$, from the preceding discussion we see that

$$A = P_2 P_2' \sim W(I_3, 3, \kappa^2)$$

and the Wishart distribution has rank 1 with $\kappa^2 = \Pi_{33}^2$.
Partitioning A as in 7.2.28, the 2SLS estimator of $\beta_2' = (\beta_2, \beta_3)$ is given by

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = A_{22}^{-1}\, a(1)$$

Now in this case A_{22} has a non-central Wishart distribution, so by Theorem 7.4.1(b) it follows that b has a lower non-central matrix T distribution.
The parameters of T can be obtained either from Theorem

7.4.1(b) or Theorem 6.3.2 with the following identifications

Theorem 6.3.2	Theorem 7.4.1(b)	Parameters
λ^2	κ^2	Π_{33}^2
p	r	2
m	n+q	4
q	q	1
m-q	n	3
I_p+TT'	I_r+BB'	(I_2+bb')
ω^{11}	ω^{22}	ω^{22}

The density function of b can be obtained either from Theorem 7.4.1(b) or from Equation 6.3.20. In Theorem 7.4.1(b) substituting $n = 3$, $r = 2$, $q = 1$, we obtain

$$|I+BB'| = \left|I+\begin{pmatrix}b_1\\ b_2\end{pmatrix}(b_1\ b_2)\right| = \begin{vmatrix}1+b_1^2 & b_1b_2\\ b_1b_2 & 1+b_2^2\end{vmatrix}$$

$$= 1+b_1^2+b_2^2$$

The argument of the ${}_1F_1$ series is ω^{22}, the lower right hand element of $(I+BB')^{-1}$.

$$\text{Now}\qquad \omega^{22} = \frac{\operatorname{cof}(1+b_2^2)}{|I+BB'|}$$

$$= \frac{1+b_1^2}{1+b_1^2+b_2^2}$$

The constant term becomes

$$\frac{\exp(-\frac{1}{2}\Pi_{33}^2)\Gamma_2(2)}{\pi^{\frac{1}{2}}\Gamma_2(3/2)} = \frac{\exp(-\frac{1}{2}\Pi_{33}^2)\pi^{\frac{1}{2}}\Gamma(2)\Gamma(3/2)}{\pi\Gamma(3/2)\Gamma(\frac{1}{2})} = \frac{\exp(-\frac{1}{2}\Pi_{33}^2)}{\pi}$$

so the density function of b is

$$\frac{\exp(-\frac{1}{2}\Pi_{33}^2)}{\pi(1+b_1^2+b_2^2)} \quad {}_1F_1(2;\ 3/2;\ \frac{\Pi_{33}^2(1+b_1^2)}{2(1+b_1^2+b_2^2)}) \qquad 7.5.31$$

which is the same result as given by Basmann (1963) Equation 4.12 p. 167 and Kabe (1964) Equation 2.8 p. 883. Basmann derived the distribution directly from the joint multivariate normal distribution of the columns of P and his derivation is more complicated than that given above. Kabe (1964) transforms in a 3 dimensional linear non-central Wishart distribution and his derivation follows along the lines of Theorem 7.4.1. The connection between the 2SLS estimators and the lower non-central matrix T is a result of Theorem 7.4.1 and Theorem 6.3.1.

The question of interest is now, are all 2SLS estimators distributed as lower non-central matrix T? Unfortunately we cannot answer this with an unequivocal yes or no. Let us consider the difficulties which stand in our way. For simplicity of exposition we shall assume that all the endogenous variables appear in the equation. The same remarks would apply to an equation which contains only G_1 endogoneous variables, the only changes being that G is replaced by G_1 and Π_2 by Π_2^1 (see equations 7.5.6 and 7.5.7).

This is true because the estimation only involves the observations on the included endogenous variables and the excluded exogenous variables.

To obtain the distribution of the estimators we must transform by $b = A_{22}^{-1}\, a_{(1)}$ say, in a non-central Wishart distribution, the parameters of which are essentially determined by the identifiability restrictions. Firstly, the number of exogenous variables included in the first equation determines the dimension of the distribution, which is G, say. Secondly, the degrees of freedom are determined by the number of excluded exogenous variables, which is K_2, say. The necessary condition for identifiability is $K_2 \geq G-1$. Since we are only considering over-identified equations, $K_2 > G-1$ and so the Wishart distribution is always well-defined. (If $K_2 = G-1$, the Wishart density would not exist and so this method could not be applied. The distribution of the 2SLS estimators in this case has been considered by Basmann et al (1971) for $G = 3$ and the distribution is not lower non-central matrix T.) Thirdly the covariance matrix of the Wishart distribution Σ, say, is determined by the covariance matrix of the disturbances in the structural model (the u_t's) and by the transformation that carries these disturbances over to those of the reduced form (i.e. $v_t = B^{-1}u_t$ see Equation 7.55). Finally the non-centrality parameter of the Wishart distribution is defined as $\Omega = \frac{1}{2}\Pi_2\Pi_2'\Sigma^{-1}$ where Π_2 is defined in Equation 7.5.9. The rank of Ω will be determined by the rank of Π_2. However the identifiability restrictions require that rank $(\Pi_2) = G-1$, and so it follows that rank $(\Omega) = G-1$.

The condition on the rank of the noncentrality parameter is where the difficulty of obtaining the general case of the distribution really lies. The problem is a special case of the general problem outlined in Section 7.3

since we must transform by $b = A_{22}^{-1} a_{(1)}$ in the matrix $A \sim W_G(\Sigma, K_2, \Omega)$ with rank $(\Omega) = G-1$ and then integrate over a_{11} and A_{22}. If the rank of the non-central Wishart distribution is 3 or higher, the integration is extremely difficult to perform and the problem seems to be intractable in the general case. Kabe (1964) considered the case where $G = 3$, $\Sigma = I$ and A has rank 2. The method requires a transformation in a 3 dimensional Wishart distribution with rank 2. At this stage we have been unable to show explicitly that his distribution is lower non-central matrix T and so have not included it here.

Other investigations into the exact distribution of the 2SLS estimators have been confined to the cases where the equation of interest contains 2 endogenous variables and an arbitrary number of excluded exogenous variables. This means that the transformation has to be made in a bivariate Wishart distribution with rank 1, which is of course easier to handle. Also in these cases only one endogenous structural coefficient, β say, has to be estimated, and its sampling distribution is univariate. Again in this case, we can show that its distribution under the hypothesis $\beta = 0$ is, apart from a constant factor, lower non-central univariate t (Marakathavalli (1954)).

Richardson (1968) considered the distribution of the 2SLS estimator of β in the structural equation

$$y_{t1} = \beta y_{t2} + \gamma_{t1} x_{t1} + \dots + \gamma_{tk} x_{tk} + e_t$$

which is a single equation in a system of $G>2$ structural equations. Under the identifiability restrictions he assumes that $K_2 = n$ of the exogenous variables are excluded from this equation. Assuming that the variables have been standardised, the distribution of b, the 2SLS estimator of β, is found by transforming in a $W_2(I_2,n,\Omega)$ distribution where Ω has rank 1. In this case necessary and sufficient conditions for the identification of the equation in the model are satisfied. Under the hypothesis $\beta = 0$, the distribution of b is given by

$$\frac{\exp(-\frac{1}{2}\mu^2)}{B(\frac{1}{2},\frac{1}{2}n)} \quad \frac{1}{(1+b^2)^{\frac{1}{2}(n+1)}} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}(n+1))_j}{(\frac{1}{2}n)_j} \frac{(\frac{1}{2}X^2)^j}{j!} \qquad 7.5.32$$

where $$X^2 = \frac{\mu^2}{(1+b^2)} \qquad 7.5.33$$

Equation 7.5.32 is obtained from Richardson (1968) Equation (3) p. 1218 with $\beta = 0$. Writing $B(\frac{1}{2},\frac{1}{2}n)$ and the reduced factorials in terms of gamma functions and cancelling, we find

$$f(b) = \sum_{j=0}^{\infty} \frac{e^{-\frac{1}{2}\mu^2}}{j!}\left(\frac{\mu^2}{2}\right)^j \frac{1}{B(\frac{1}{2}n+j;\ \frac{1}{2})} \frac{1}{(1+b^2)^{\frac{1}{2}(n+1)+j}} \qquad 7.5.34$$

Now let $v = \sqrt{n}\, b$ then $db = \frac{1}{\sqrt{n}}\, dv$ and 7.5.35

$$f(v) = \sum_{j=0}^{\infty} \frac{e^{-\frac{1}{2}\mu^2}}{j!}\left(\frac{\mu^2}{2}\right)^j \frac{1}{B(\frac{1}{2}n+j;\ \frac{1}{2})} \frac{n^{-\frac{1}{2}}}{(1+\frac{v^2}{n})^{\frac{1}{2}(n+1)+j}} \qquad 7.5.36$$

Comparing 7.5.36 with Marakathavalli (1954) Equation (2) p. 253 we see that $f(v)$ is the density function of a lower non-central univariate t distribution with n degrees of freedom. A similar result given by Kabe (1963) Equation (13) p. 536 for a single equation in a system of 2 equations, with 2 included endogenous and 2 excluded exogenous variables can also be shown to be lower non-central t after a transformation such as is given in 7.3.35. The parameter μ^2 depends on certain elements of the reduced form and is given in Richardson (1968).

In all the examples considered above the covariance matrix of the Wishart distribution was I. This means that the distribution has no regression structure. Sawa (1969) derived the distribution of the 2SLS estimator b under the same set-up as Richardson (1968). However, he assumed that the covariance matrix of the Wishart distribution was

$$\Sigma = \begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix} \qquad 7.5.37$$

Under this general covariance structure, he shows that the distribution of b depends on $(\beta - \frac{\sigma_{12}}{\sigma_{11}})$, i.e. the difference between the value of the structural parameter and that of the regression coefficient of the disturbance terms of the endogenous variables in the reduced form. It is easily shown, using Sawa (1969) Equation 3.25, p. 930, that under the hypothesis $\beta = \frac{\sigma_{12}}{\sigma_{11}}$, the distribution of

$$(K_2)^{\frac{1}{2}}V = \frac{(K_2(\sigma_{11}b-\sigma_{12}))^{\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}} \qquad 7.5.38$$

is univariate lower non-central t with K_2 degrees of freedom and non-centrality parameter

$$\frac{\tau^2}{\sigma^2} = N \sum_{j=K_1+1}^{K_2} \Pi_{ij}^2/\sigma^2 \qquad 7.5.39$$

where Π_{ij} $j = K_1+1,\ldots,K_2$ are the reduced form coefficients corresponding to the excluded exogenous variables in the first equation of the reduced form, and N is the sample size used to estimate b. The transformation in 7.5.38 is essentially the one used by Richardson (1968) to standardise β before deriving the density given in 7.5.32.

Sawa (1969) also demonstrates the explicit relationship between the two stage least squares and ordinary least squares estimator of β. He shows that the density functions of the two estimators have the same functional form, the only change being in the degrees of freedom and the non-centrality parameter. Thus the remarks on the relationship between the distribution of two stage least squares estimators and the lower non-central t distribution also apply to the distribution of ordinary least squares estimators of such a system.

It is now natural to enquire if the lower non-central t distribution could be used to test hypotheses about β. We feel that this application will not usually be possible since we need to know the values of elements of Σ for the standardising transformation and certain of the reduced form coefficients for the non-centrality parameter. These values are usually unknown, but if it should happen that "acceptable" values of Σ and Π_{ij}

are available from previous studies, then it would be possible to test the hypothesis $\beta = \frac{\sigma_{12}}{\sigma_{11}}$ using the lower non-central t distribution.

Finally we give a table of the distributions of the 2SLS estimators that have been given in the literature which are related to either the lower non-central t distribution or the lower non-central matrix T.

THE DISTRIBUTION OF THE 2SLS ESTIMATOR IN AN OVER-IDENTIFIED EQUATION

		UNIVARIATE				BIVARIATE			
	Reference	1	2	3	4	5	6	6	6
	page	631	536	1218	930	167	883	886	891
	Equation	3.47* 3.50	12* 13	3.5*	3.24* 3.25	4.12	2.8	3.17	4.7†
System of Equations	Number of: Equations	2	2	G	G	3	3	3	3
	Endog. variables	2	2	G	G	3	3	3	3
	Exog. variables	4	4	K	K	K	K	K	K
First Equation	Inc. Endog.	2	2	2	2	3	3	3	3
	Exc. Exog.	2	2	$K_2=n$	K_2	3	3	3	3
Parameters of Wishart Distribution	Dimension		2	2	2		3	3	3
	Cov.matrix		I	I	Σ		I	Σ	I
	Rank		1	1	1		1	1	2
	d.o.f		2	K_2	K_2		3	3	3
Dimension of β		1	1	1	1	2	2	2	2

Notes on the Table

The table summarises the cases given in the literature of the exact distribution of the 2SLS estimators of the structural parameters of the endogenous variables in a single over-identified equation.

References

1. Basmann (1961)
2. Kabe (1963)
3. Richardson (1968)
4. Sawa (1969)
5. Basmann (1963)
6. Kabe (1964)

Let b denote the 2SLS estimator.

(i) In row 3 the equations marked with an asterisk give the density function of b when $\beta \neq 0$ or $\beta \neq \sigma_{11}/\sigma_{12}$ in the case of reference 4.

(ii) The remaining equations give the density when $\beta = 0$ or $\beta = \sigma_{12}/\sigma_{11}$ in the case of reference 4. The transformation $v = \sqrt{cb}$ gives a lower non-central t density in the univariable case where c is the number of excluded exogenous variables. In the bivariate case the densities are special cases of the lower non-central matrix T.

(iii) In reference 1 and 5 no parameters are given for the Wishart distribution because the author, Basmann, uses a different method to derive the distributions.

(iv) A further integration is necessary to obtain the density of b from Equation 4.7† of reference 6. The form of the density suggests it may be a special case of the lower non-central T but we have been unable to demonstrate this specifically.

(v) The bivariate density of the 2SLS estimators for a *just* identified equation with 3 endogoneous variables has been given by Basmann et al (1971). The density is not that of a lower non-central matrix T.

CHAPTER 8

A BIVARIATE t DISTRIBUTION AND PARTITIONED HOTELLING'S T^2 TESTS

8.1 INTRODUCTION

In this chapter we consider the distribution of a different type of multivariate t variable that can be constructed from a random sample drawn from a multivariate normal population. We discuss first the bivariate case given by Siddiqui (1967) who derived the central case and Krishnan (1972) who gave the upper non-central distribution.

Let $X_{(\alpha)} = (x_{1\alpha}\ x_{2\alpha})'$ $\alpha = 1 \cdot\cdot N$ be a random sample of size N from a bivariate normal population with

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

i.e. each of the variables has unit variance and the correlation between x_1 and x_2 is ρ. This assumption has no effect on the form of the distribution that follows.

Let $$\bar{x}_1 = \frac{1}{N}\sum_{j=1}^{N} x_{1j} \qquad \bar{x}_2 = \frac{1}{N}\sum_{j=1}^{N} x_{2j} \tag{8.1.1}$$

$$S_1^2 = \frac{1}{N}\sum_{j=1}^{N} (x_{1j}-\bar{x}_1)^2 \qquad S_2^2 = \frac{1}{N}\sum_{j=1}^{N} (x_{2j}-\bar{x}_2)^2$$

Then $$t_1 = \frac{\bar{x}_1 (N-1)^{\frac{1}{2}}}{S_1} \qquad t_2 = \frac{\bar{x}_2 (N-1)^{\frac{1}{2}}}{S_2} \tag{8.1.2}$$

For later purposes, we also note that t_i^2 can be written as

$$t_i^2 = \frac{N(N-1)\bar{x}_i^2}{\sum_j (x_{ij}-\bar{x}_i)^2} \tag{8.1.3}$$

Clearly t_1 and t_2 are both univariate t variables, which are central if $\mu = 0$, and non-central otherwise. nS_1^2 and nS_2^2 follow a multivariate chi-square distribution. The joint density of t_1 and t_2 depends only on ρ, the correlation between the variables. Siddiqui (1967) page 163, Eq.2.2, shows that if $\mu = 0$, the joint density of t_1, t_2 and r, the sample correlation coefficient, is

$$f(t_1,t_2,r) = \frac{\Gamma(n+2)(1-\rho^2)^{\frac{1}{2}(n+1)}}{(2\pi)^{3/2}\,\Gamma\left(n+\frac{3}{2}\right)} \frac{(1-r^2)^{\frac{1}{2}(n-3)}}{\left\{\left(1+\frac{t_1^2}{n}\right)\left(1+\frac{t_2^2}{n}\right)\right\}^{\frac{1}{2}(n+1)}}$$

$$\times \quad (1-b-cr)^{-\frac{1}{2}(n+1)} \quad F\left(\tfrac{1}{2},\tfrac{1}{2};n+3;\tfrac{1}{2}(1+b+cr)\right) \tag{8.1.4}$$

where $n = N-1$

$$b = \frac{\rho t_1 t_2}{n\left\{\left(1+\frac{t_1^2}{n}\right)\left(1+\frac{t_2^2}{n}\right)\right\}^{\frac{1}{2}}} \tag{8.1.5}$$

$$c = \frac{\rho}{\left\{\left(1+\frac{t_1^2}{n}\right)\left(1+\frac{t_2^2}{n}\right)\right\}^{\frac{1}{2}}} \tag{8.1.6}$$

and ${}_2F_1$ is the hypergeometric series (definition 2.3.1).

Siddiqui (1967) obtains the exact distribution of t_1, t_2

for n = 1 and n = 3 and an asymptotic approximation of the density for arbitrary n, using the method of steepest descent. The asymptotic approximation is:

$$f(t_1,t_2) = \frac{\Gamma(n+2)(1-\rho^2)^{\frac{1}{2}(n+1)}}{2\pi n^{\frac{1}{2}} \quad \Gamma\left(n+\frac{3}{2}\right)} \frac{(1-b)^{\frac{1}{2}} \quad {}_2F_1\left(\frac{1}{2},\frac{1}{2};n+\frac{3}{2};\frac{1-b^2+c^2}{1-b}\right)}{\left\{\left(1+\frac{t_1^2}{n}\right)\left(1+\frac{t_2^2}{n}\right)((1-b^2)-c^2)\right\}^{\frac{1}{2}(n+1)}} \qquad 8.1.7$$

When $\rho = 0$, the joint density of $t_1 t_2$ is simply the product of 2 independent student t densities.

As $n \to \infty$ the joint density tends to the bivariate normal.

The joint density of $(t_1 t_2)$ was given in the non-central case ($\mu \neq o$) by Krishnan (1972) in three forms, two of which involve infinite series iterated co-error functions (Krishnan 1972) Eq. 2.14 and Eq. 2.16 page 229) or an infinite series of gamma functions. (ibid Eq. 2.19)

8.2 PARTITIONED T^2 TESTS (JENSEN 1972)

The following generalisation of this bivariate t distribution was considered by Jensen (1972), when developing simultaneous multivariate procedures for testing subsets of mean values using Hotelling's T^2 statistics.

Let $X_{(\alpha)}$ $\alpha = 1,\ldots,N$ be a random sample from a $N_p(\mu,\Sigma)$ population.

Let $\bar{x} = \Sigma X_{(\alpha)}/N$ and $S = (N-1)^{-1}\Sigma_\alpha(X_{(\alpha)}-\bar{x})(X_{(\alpha)}-\bar{x})'$ be the sample mean and variance-covariance matrix.

Let $X_{(\alpha)}$ be partitioned into q subsets of variables, with p_i variables in the i^{th} subset. Partitioning $\mu, \bar{x}$, Σ and S accordingly we have

$$\mu = \begin{vmatrix} \mu^{(1)} \\ \mu^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ \mu^{(q)} \end{vmatrix} \begin{matrix} p_1 \times 1 \\ p_2 \times 1 \\ \\ \\ \\ p_q \times 1 \end{matrix} , \quad \bar{x} = \begin{vmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ \bar{x}^{(q)} \end{vmatrix} \begin{matrix} p_1 \times 1 \\ p_2 \times 1 \\ \\ \\ \\ p_q \times 1 \end{matrix} \qquad 8.2.1$$

$$\Sigma = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1q} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \Sigma_{q1} & & & \Sigma_{qq} \end{vmatrix} \quad S = \begin{vmatrix} S_{11} & S_{12} & \cdots & S_{1q} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ S_{q1} & & & S_{qq} \end{vmatrix}$$

8.2.2

where $\Sigma_1^q p_i = p$ and Σ_{ii} and S_{ii} are positive definite $p_i \times p_i$ matrices of full rank.

It is well known that the likelihood ratio test statistic of the hypothesis (Anderson (1958) Chapter 5)

$$H_0 \; : \mu = \mu_o \qquad 8.2.3$$

is

$$V = N(\bar{x}-\mu_o)'S^{-1}(\bar{x}-\mu_o) \qquad 8.2.4$$

and V is distributed as Hotelling's T^2 with $\nu = N-1$ degrees of freedom and noncentrality parameter $\delta = (\mu-\mu_o)'\Sigma^{-1}(\mu-\mu_o)$.

We denote this distribution by $T^2(p,\nu,\delta)$ and its upper α percentage point by $T^2(p,\nu,\delta)$. Under H_0, $\delta = 0$ and

$$T^2(p,\nu) \sim \frac{(N-1)p}{N-p} F_{p,N-p} \,. \qquad 8.2.5$$

Consider the q hypotheses

$$H_0 \; : \mu^{(i)} = \mu_o^{(i)} \qquad i = 1,\dots,q \qquad 8.2.6$$

A natural test statistic for this hypothesis is

$$V_i = N(\bar{x}^{(i)} - \mu_o^{(i)})' S_{ii}^{-1} (\bar{x}^{(i)} - \mu_o^{(i)}) \quad i = 1,\dots,q \qquad 8.2.7$$

The marginal distribution of V_i is $T^2(p_i,\nu,\delta_i)$ where

$$\delta_i = (\mu^{(i)} - \mu_o^{(i)})' \Sigma_{ii}^{-1} (\mu^{(i)} - \mu_o^{(i)}). \qquad 8.2.8$$

If $q = 2$ and $p_1 = p_2 = 1$ and $\mu_o^{(1)} = \mu_o^{(2)} = 0$, then

$$V_i = \frac{N(N-1)\, \bar{x}_i^2}{\sum (x_{ij} - \bar{x}_i)^2}$$

Comparing with 8.1.3 we see that

$$V_i = t_i^2 \quad i = 1,2$$

So the joint distribution of V_1, V_2 is that given by Siddiqui (1967). The joint distribution of $(V_1 \,..\, V_q)$ is unknown, except for this case and of course the trivial case when the V_i are independent (i.e. $\Sigma_{ij} = 0 \quad i \neq j$) when the joint distribution becomes the product of the marginal distributions.

The joint distribution in the general case appears to be very difficult to obtain and would certainly be so complicated that its practical use would be severely restricted. However tests of sets of hypotheses such as 8.2.6 above are of considerable practical interest.

In many biological and medical experiments where a number of variables are measured on each experimental unit, it often happens that the variables fall naturally into a number of distinct subsets, each of which represents a different aspect of the experiment. The grouping of the variables into subsets is of course determined by biological considerations alone and is not the result of the observations made in the experiment. An investigator often wishes to draw separate conclusions about the behaviour of each subset of measurements and then combine the information obtained at the end to give an overall assessment of the experiment. The statistical formulation of such a problem is given by 8.2.6 and the appropriate statistics are the set V_i given in 8.2.7 . Jensen (1972) has proposed two procedures for the simultaneous testing of the q hypothesis in 8.2.6.

The first procedure rests on the Bonferroni Inequality

Proposition I. (Feller (1967), Jensen (1972)).

Let $\{X\ ,\ldots,X_q\}$ be random variables and $\{R_1,\ldots,R_q\}$ be arbitrary regions such that

$$Pr(X_i \in R_i) = \alpha_i \qquad i = 1,\ldots,q.$$

Then

$$Pr((X_i \in R_1) \cup (X_2 \in R_2) \ldots \cup (X_q \in R_q)) \leq \alpha_1\ \alpha_2 \ldots + \alpha_q$$

whatever the joint distribution of $\{X_1,\ldots,X_q\}$.

Identifying X_i with V_i, $i = 1,\ldots,q$ and R_i as the upper tail region of $T^2(p_i,\nu)$ we conclude

$$Pr(V_1 > T^2_\alpha(p_1,\nu) \cup \ldots \cup V_q > T^2_{\alpha_q}(p_q,\nu)) \leq \alpha_1 + \ldots + \alpha_q \qquad 8.2.9$$

The left hand member of 8.2.9 is defined as the type I probability error rate, i.e. the probability that one or more of the hypotheses are rejected erroneously when all of them are true. The right hand member provides an upper bound

$$\alpha^* = \alpha_1 + \ldots + \alpha_q$$

for this probability.

Using 8.2.9 we can test each of the hypotheses 8.2.6 by comparing V_i with the upper α_i percentage point of the $T^2(p_i,\nu)$ distribution and rejecting H_0 if V_i exceeds this value. The type I error rate for the q hypotheses is at most α^*.

The second procedure depends on the fact that an ellipsoidal confidence region in p dimensional Euclidean space generates ellipsoidal regions simultaneously in all subspaces of fewer dimensions.

<u>Proposition 2</u> <u>(Jensen 1972)</u>

Let $Z \sim N_p(\mu,\Sigma)$ and $\nu S \sim W(\Sigma,\nu)$ where Z and S are independent. Let $\mathcal{C}$ be the class of p×r matrices C with rank r for $r = 1,\ldots,p-1$. Let $T^2_\alpha(p,\nu)$ be a constant such that

$$Pr[(z-\mu)'(\kappa S)^{-1}(z-\mu)) \leq T^2_\alpha(p,\nu)] = 1-\alpha$$

Then

$\Pr[(z-\mu)'C(\kappa C'SC)^{-1}C'(z-\mu) \leq T^2_\alpha(p,\nu)$ for all $C \in \mathcal{C}] = 1-\alpha$ and the ellipsoidal bounds

$$(C'z-C'\mu)'(\kappa C'SC)^{-1}(C'z-C'\mu) \leq T^2_\alpha(p,\nu) \qquad 8.2.10$$

hold simultaneously for all $C \in \mathcal{C}$ with confidence coefficient $1-\alpha$.

Proof: (Given in Jensen 1972)

Reformulating the bounds for hypothesis testing we let

$$H_C; \quad C \in \mathcal{C} \qquad 8.2.11$$

be a family of hypotheses indexed by the choice of C, a typical member of which is

$$H_C; \quad C'\mu = \theta_C \qquad 8.2.12$$

Consider test statistics of the form

$$V_C = (C'z-\theta_C)'(\kappa C'SC)^{-1}(C'z-\theta_C) \qquad 8.2.13$$

such that H_C is rejected if $V_C > T^2_\alpha(p,\nu)$ for all $C \in \mathcal{C}$.

The Type I error rate for the family 8.2.11 is α and the Type I error rate for any proper subset of these is less than α.

If in 8.2.12 and 8.2.13 we identify z with $\bar{x}$, θ_C with $\mu^{(i)}$, κ with N^{-1} and let $C' = (0, I_{p_i}, 0)$ we obtain the test statistic 8.2.4, namely

$$V_i = N(\bar{x}^{(i)}-\mu^{(i)})'S^{-1}_{ii}(\bar{x}^{(i)}-\mu^{(i)})$$

and

$$\Pr[V_i > T^2_\alpha(p,\nu)] < \alpha \quad \text{for } i = 1,\dots,q.$$

Thus we can test the q hypotheses 8.2.6 by referring each V_i to the same upper α percentage of $T^2(p,\nu)$.

The family of hypotheses $\{H_0 : \mu^{(i)} = \mu_o^{(i)}\}$ are a subset of the much larger family, 8.2.11 and so the type I error rate for these is less than α.

Since a single percentage point is used for all q tests we shall call this procedure the Global T^2 test, and the first procedure the Bonferroni T^2 test.

For testing hypotheses of the type considered above, both procedures can be used. The choice of procedure depends only on the parameters of the problem, namely ν , p and q and not on the outcome of random events, so we can choose the procedure which gives the smallest critical value of T^2. A smaller critical value means that we have a test with a higher power and narrower confidence regions.

Can we formulate any rules that will tell us when one procedure is always preferred to the other? We know that

(i) For fixed ν and p, $T^2_\alpha(p,\nu)$ is a decreasing function of α.

(ii) For fixed ν and , $T^2_\alpha(p,\nu)$ is an increasing function of p.

(iii) The number of subsets q into which our p vector is partitioned will influence the critical value of T^2 through its influence on α_i - the more subsets we have, the smaller each α_i must be in order to control the type I error rate. This will tend to inflate the Bonferroni critical values. On the other hand, if q is large the number of variables in each subset tend to be small, which will decrease the critical values. The factors influencing the choice between the two tests can be summarized in the following table.

TABLE I

N = sample size $\nu = N-1$ p = No. of variables

	Bonferroni	Global T
No. of subsets	q	q
No. of variables in ith subset	$p_1, \ldots, p_q$	
Level of significance	$\alpha_1, \ldots, \alpha_q$	α
Critical values	$T^2_{\alpha_1}(p_1, \nu), \quad T^2_{\alpha_q}(p_q, \nu)$	$T^2_\alpha(p, \nu)$

The Global T^2 will be preferred if one or more of the Bonferroni critical constants is greater than $T^2_\alpha(p,\nu)$. To investigate the two procedures, critical values of T^2 were calculated for p = 1,15, with values of α ranging from 0,001 to 0,01 and three sample sizes, N = 20,50 and 100. Later the computations were extended for p = 16,30 and N = 35, 50, 100 using the same values of α.

The tables were inspected by selecting values of p, α and N and noting the Global T critical value. The critical values of the Bonferroni T's were then found for all partitions of p into $q = 2,\ldots,p$ parts with $p_1 \geq p_2 \geq \cdots \geq p_q$ and all tabulated values of α_i such that $\Sigma_1^q \alpha_i = \alpha$. The cases for which the Bonferroni critical value exceeded the Global T were noted.

The following pattern soon revealed itself:

If q is large (i.e. a large number of subsets), the Bonferroni method is best, but has the disadvantage that the individual tests have to be made at a low level of significance in order to control the type I error.

The Global T gave smaller critical values in one situation. This was when q was small and the number of variables in the first group was very much larger than those in the other groups (i.e. $p_1 >> p_i \quad i = 2,\ldots q$). The effect was more marked as the sample size increased and when the first group was tested at a very low level of significance. As the numbers in each group became more balanced, the Bonferroni method gave smaller critical values.

CRITICAL VALUES OF HOTELLINGS T FOR A SAMPE OF SIZE N WITH P VARIABLES

N= 20

VARIABLES

ALPHA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
.10000 I	2.990	5.539	8.173	11.081	14.396	18.261	22.857	28.431	35.343	44.129	55.643	71.306	93.666	127.721	184.567
.05000 I	4.381	7.504	10.719	14.283	18.375	23.189	28.975	36.082	45.023	56.587	72.047	93.592	125.276	175.380	263.269
.04000 I	4.861	8.169	11.576	15.361	19.716	24.855	31.053	38.697	48.357	60.917	77.814	101.539	136.750	193.066	293.459
.03000 I	5.502	9.052	12.712	16.767	21.494	27.068	33.821	42.192	52.831	66.757	85.642	112.410	152.605	217.875	336.478
.02500 I	5.922	9.626	13.449	17.715	22.651	28.511	35.630	44.483	55.776	70.619	90.847	119.691	163.318	234.825	366.380
.02000 I	6.449	10.345	14.372	18.875	24.100	30.322	37.905	47.371	59.501	75.524	97.492	129.041	177.186	256.989	406.034
.01000 I	8.185	12.694	17.385	22.670	28.852	36.283	45.435	56.992	71.999	92.134	120.242	161.502	226.183	337.151	554.167
.00500 I	10.073	15.231	20.639	26.780	34.024	42.810	53.742	67.704	86.072	111.087	146.631	199.925	285.700	437.899	749.341
.00250 I	12.118	17.972	24.159	31.240	39.664	49.973	62.929	79.662	101.954	132.762	177.300	245.485	358.095	564.654	1006.671
.00100 I	15.081	21.934	29.261	37.735	47.927	60.548	76.614	97.665	126.173	166.324	225.682	319.046	478.501	783.927	1476.917

CRITICAL VALUES OF HOTELLINGS T FOR A SAMPE OF SIZE N WITH P VARIABLES

N= 35

VARIABLES

ALPHA	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
.10000 I	53.24	59.85	67.65	76.53	87.24	99.73	114.72	133.02	155.78	184.78	222.81	274.53	348.32	460.55	647.51
.05000 I	63.42	71.69	81.24	92.38	105.51	121.22	140.27	163.82	193.54	232.03	283.51	355.20	460.47	626.57	917.13
.04000 I	66.33	75.67	85.74	97.50	111.65	128.44	148.92	174.31	206.51	248.43	304.83	383.94	501.15	688.23	1020.51
.03000 I	71.30	80.72	91.66	104.48	119.70	138.01	160.41	188.32	223.91	270.53	333.75	423.25	557.36	774.55	1167.87
.02500 I	74.17	84.03	95.48	108.93	124.95	144.24	167.90	197.49	235.34	285.14	352.98	449.57	595.34	833.57	1270.28
.02000 I	77.74	88.14	100.24	114.49	131.48	152.04	177.32	209.05	249.81	303.69	377.51	483.36	644.49	910.74	1406.09
.01000 I	89.23	101.39	115.65	132.55	152.86	177.64	208.43	247.48	298.28	366.44	461.47	600.64	818.13	1189.81	1913.38
.00500 I	101.39	115.49	132.12	151.98	176.01	205.58	242.66	290.19	352.79	437.99	556.61	739.43	1029.00	1540.51	2581.75
.00250 I	114.32	130.56	149.82	172.96	201.17	236.17	280.45	337.31	414.26	519.77	671.90	903.95	1285.48	1981.72	3462.94
.00100 I	132.74	152.13	175.30	203.38	237.93	281.23	336.66	409.44	507.93	646.35	850.25	1169.54	1712.01	2744.95	5073.25

CRITICAL VALUES OF HOTELLINGS T FOR A SAMPE OF SIZE N WITH P VARIABLES

N= 50

VARIABLES

ALPHA		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
.10000	I	2.811	4.934	6.894	8.825	10.778	12.779	14.847	16.997	19.241	21.593	24.065	26.672	29.426	32.346	35.44
.05000	I	4.038	6.514	8.765	10.968	13.187	15.457	17.800	20.236	22.779	25.446	28.253	31.215	34.350	37.678	41.22
.04000	I	4.452	7.033	9.372	11.658	13.960	16.314	18.744	21.269	23.907	26.674	29.587	32.662	35.919	39.378	43.06
.03000	I	4.996	7.709	10.159	12.552	14.959	17.419	19.960	22.600	25.359	28.254	31.303	34.524	37.938	41.565	45.43
.02500	I	5.347	8.141	10.661	13.120	15.593	18.121	20.731	23.444	26.279	29.255	32.390	35.704	39.217	42.952	46.93
.02000	I	5.784	8.675	11.279	13.818	16.371	18.981	21.675	24.477	27.406	30.481	33.722	37.149	40.784	44.651	48.77
.01000	I	7.182	10.365	13.223	16.008	18.807	21.669	24.626	27.703	30.922	34.307	37.878	41.660	45.678	49.960	54.53
.00500	I	8.642	12.104	15.212	18.238	21.282	24.396	27.616	30.969	34.483	38.180	42.088	46.231	50.641	55.347	60.38
.00250	I	10.160	13.895	17.248	20.516	23.806	27.173	30.658	34.292	38.104	42.121	46.373	50.887	55.699	60.844	66.36
.00100	I	12.253	16.343	20.021	23.609	27.227	30.934	34.778	38.792	43.009	47.461	52.182	57.204	62.569	68.317	74.49

CRITICAL VALUES OF HOTELLINGS T FOR A SAMPE OF SIZE N WITH P VARIABLES

N= 50

VARIABLES

ALPHA		16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
.10000	I	38.75	42.28	46.05	50.10	54.47	59.17	64.27	69.82	75.86	82.47	89.74	97.77	106.68	116.61	127.
.05000	I	45.00	49.05	53.39	58.06	63.10	68.56	74.49	80.96	88.04	95.82	104.40	113.92	124.54	136.44	149.
.04000	I	47.00	51.21	55.73	60.60	65.57	71.57	77.78	84.55	91.97	100.13	109.15	119.17	130.35	142.91	157.
.03000	I	49.56	53.99	58.75	63.89	69.44	75.46	82.02	89.19	97.05	105.72	115.31	125.98	137.92	151.34	166.
.02500	I	51.19	55.76	60.67	65.97	71.71	77.94	84.72	92.14	100.30	109.23	119.25	130.34	142.76	156.75	172.
.02000	I	53.19	57.93	63.03	68.53	74.50	80.98	88.04	95.78	104.29	113.68	124.10	135.72	148.74	163.44	180.
.01000	I	59.44	64.72	70.41	76.57	83.26	90.54	98.51	107.26	116.91	127.60	139.50	152.82	167.82	184.82	204.
.00500	I	65.80	71.63	77.94	84.78	92.22	100.35	109.26	119.08	129.94	142.01	155.49	170.63	187.75	207.23	229.
.00250	I	72.30	78.71	85.66	93.21	101.44	110.46	120.37	131.32	143.46	156.99	172.16	189.27	208.67	230.35	256.
.00100	I	81.16	88.38	96.21	104.75	114.10	124.36	135.68	148.23	162.19	177.82	195.42	215.34	236.06	264.15	294.

CRITICAL VALUES OF HOTELLINGS T FOR A SAMPE OF SIZE N WITH P VARIABLES

N= 100

VARIABLES

ALPHA		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
.10000	I	2.757	4.763	6.556	8.268	9.946	11.613	13.282	14.961	16.657	18.375	20.119	21.891	23.695	25.534	27.411
.05000	I	3.937	6.241	8.262	10.174	12.037	13.881	15.720	17.567	19.430	21.313	23.223	25.162	27.136	29.146	31.196
.04000	I	4.331	6.722	8.810	10.781	12.700	14.596	16.487	18.385	20.298	22.232	24.192	26.183	28.208	30.270	32.374
.03000	I	4.848	7.344	9.515	11.561	13.549	15.511	17.467	19.429	21.405	23.402	25.426	27.481	29.571	31.700	33.871
.02500	I	5.180	7.741	9.963	12.054	14.084	16.088	18.084	20.085	22.100	24.136	26.200	28.295	30.425	32.595	34.808
.02000	I	5.591	8.228	10.511	12.656	14.738	16.790	18.834	20.883	22.945	25.029	27.140	29.283	31.463	33.682	35.946
.01000	I	6.898	9.756	12.217	14.524	16.758	18.957	21.145	23.336	25.541	27.768	30.023	32.311	34.639	37.009	39.426
.00500	I	8.244	11.305	13.934	16.393	18.772	21.111	23.437	25.765	28.107	30.471	32.865	35.295	37.766	40.282	42.849
.00250	I	9.626	12.876	15.664	18.269	20.787	23.261	25.720	28.181	30.656	33.155	35.685	38.253	40.864	43.524	46.238
.00100	I	11.503	14.988	17.977	20.766	23.461	26.109	28.740	31.372	34.020	36.691	39.398	42.145	44.940	47.786	50.692

CRITICAL VALUES OF HOTELLINGS T FOR A SAMPE OF SIZE N WITH P VARIABLES

N= 100

VARIABLES

ALPHA		16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
.10000	I	29.33	31.29	33.29	35.34	37.44	39.59	41.80	44.07	46.40	48.79	51.24	53.77	56.37	59.05	61.81
.05000	I	33.29	35.43	37.62	39.86	42.15	44.51	46.92	49.40	51.94	54.56	57.24	60.01	62.86	65.80	68.82
.04000	I	34.52	36.72	38.96	41.26	43.62	46.03	48.51	51.05	53.66	56.35	59.11	61.95	64.87	67.89	71.00
.03000	I	36.09	38.35	40.67	43.04	45.47	47.97	50.52	53.15	55.64	58.62	61.47	64.41	67.43	70.55	73.76
.02500	I	37.07	39.38	41.74	44.16	46.64	49.18	51.78	54.46	57.21	60.04	62.95	65.94	69.03	72.21	75.49
.02000	I	38.26	40.62	43.04	45.51	48.05	50.65	53.31	56.05	58.87	61.76	64.74	67.81	70.97	74.23	77.59
.01000	I	41.39	44.42	47.00	49.64	52.35	55.13	57.99	60.92	63.93	67.03	70.22	73.51	76.90	80.39	84.00
.00500	I	45.47	48.15	50.89	53.70	56.58	59.54	62.58	65.69	68.90	72.20	75.60	79.11	82.72	86.45	90.30
.00250	I	49.01	51.84	54.75	57.72	60.77	63.93	67.11	70.42	73.82	77.32	80.93	84.65	88.49	92.45	96.54
.00100	I	53.66	56.70	59.81	63.00	66.27	69.62	73.08	76.63	80.28	84.05	87.93	91.93	96.07	100.34	104.75

@FIN

BIBLIOGRAPHY

1. Anderson, T.W. (1946): The noncentral Wishart distribution and certain problems of Multivariate statistics. *Ann. Math. Statist.* 17, 409-431.

2. Anderson, T.W. (1958): *An introduction to Multivariate Statistical Analysis.* Wiley, N.Y.

3. Ando, A and Kaufman, G.M. (1965): Bayesian analysis of the independent multinormal process - neither mean nor precision known. *J. Amer. Statist. Assoc.* 60, 347-358.

4. Baldessari, B. (1967): The distribution of a quadratic form of normal variables. *Ann. Math. Statist.* 38, 1700-1704.

5. Basmann, R.L. (1960a): On finite sample distributions of generalised classical linear identifiability test statistics. *J. Amer. Statist. Assoc.* 55, 650-659.

6. Basmann, R.L. (1960b): An expository note on the estimation of simultaneous structural equations. *Biometrics*, 15, 464-480.

7. Basmann, R.L. (1961): A note on the exact finite sample frequency functions of generalised classical linear estimators in two leading over-identified cases. *J. Amer. Statist. Assoc.* 56, 619-636.

8. Basmann, R.L. (1963): A note on the exact finite sample of generalised classical linear estimators in a leading three equation case. *J. Amer. Statist. Assoc.* 58, 161-171.

9. Basmann, R.L. (1963)*: Remarks concerning the application of exact finite sample distribution functions of G.C.L. estimators in econometric statistical inference. *(Not referred to in the text.) *J. Amer. Statist. Assoc.* 58, 943-976.

10. Basmann, R.L., Brown F.L., Dawes, W.S., Schoepfle, G.K. (1971): Exact finite sample density functions of G.C.L. estimators of structural coefficients in a leading identifiable case. *J. Amer. Statist. Assoc.* 66, 122-126.

11. Bechofer, R.E., Dunnett, C.W. and Sobel, M. (1954): A two sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika.* 41, 170-176.

12. Cochran, W.G. (1934): The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Proc. Camb. Phil. Soc.* 30, 178-191.

13. Constantine, A.G. (1963): Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* 34, 1270-1285.

14. Cornish, E.A. (1954a): The multivariate t-distribution associated with a set of normal sample deviates. *Austral. J. Phys.* 7, 531-542.

15. Cornish, E.A. (1954b): The sampling distribution of statistics derived from the multivariate t-distribution. *Austral. J. Phys.* 8, 193-199.

16. De Groot, M.H. (1970): *Optimal Statistical Decisions.* McGraw-Hill Book Company, N.Y.

17. de Waal, D.J. (1968): Non-central multivariate Beta distributions. Unpublished Ph.D. thesis, University of Cape Town.

18. Dickey, J.M. (1967): Matricvariate generalisations of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist.* 38, 511-518.

19. Dunn, O.J. (1965): A property of the multivariate t distribution. *Ann. Math. Statist.* 36, 713-714.

20. Dunn, O.J. and Massey, F.J. (1965): Estimation of multiple contrasts using t-distributions. *J. Am. Statist. Ass.* 50, 573-83.

21. Dunn, O.J. (1968): A note on confidence bounds for a regression line over a finite range. *J. Am. Statist. Ass.* 63, 1028-33.

22. Dunnett, C.W. (1955): A multiple comparison procedure for comparing several treatments with a control. *J. Am. Statist. Assoc.* 50, 1096-1121.

23. Dunnett, C.W. (1964): New tables for multiple comparisons with a control. *Biometrics*, 20 482-491.

24. Dunnett, C.W. and Sobel, M. (1954): A bivariate generalisation of Student t distribution with tables for certain special cases. *Biometrika*, 41, 153-169.

25. Feller, W. (1957): *An Introduction to Probability theory and its applications. Vol. 1.* Wiley, N.Y.

26. Freeman, H. and Kuzmack, A.M. (1972): Tables of multivariate t in six and more dimensions. *Biometrika*, 59, 217-219.

27. Geisser, S. (1965): Bayesian Estimation in Multivariate Analysis. *Ann. Math. Statist.* 36, 150-159.

28. Geisser, S. and Cornfield, J. (1963): Posterior Distributions for Multivariate Normal parameters. *J. Roy. Statist. Soc. Ser. B.* 25, 368-76.

29. Goldberger, A.S. (1964): *Econometric Theory.* Wiley, N.Y.

30. Gosset, W.S. (1908) "Student": On the probable error of a mean. *Biometrika*, 6, 1-25.

31. Graybill, F.A. (1961): *An Introduction to linear statistical models. Volume 1.* McGraw-Hill Book Company, N.Y.

32. Graybill, F.A. (1969): *Introduction to Matrices with applications in Statistics.* Wadsworth Publishing Company Inc. Belmont, California.

33. Graybill, F.A. and Marsaglia, G. (1957): Idempotent matrices and quadratic forms in the general linear hypothesis. *Ann. Math. Statist.* 29, 678-686.

34. Hahn, G.J. (1969): Factors for calculating two sided prediction intervals for samples from a normal distribution. *J. Am. Statist. Ass.* 67, 878-888.

35. Hahn, G.J. (1970): Additional factors for calculating prediction intervals for samples from a distribution. *J. Am. Statist. Ass.* 65, 1668-76.

36. Hahn, G.J. and Hendrickson, R.W. (1971): A table of percentage points of the distribution of the largest absolute value of k student t variates and its applications. *Biometrika*, 58, 323-332.

37. Hewett, J.E. and Bulgren, W.G. (1971): Inequalities for some Multivariate t distributions with applications. *Technometrics*, 13, No. 2, 397-402.

38. James, A.T. (1964): Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* 35, 475-501.

39. Jensen, D.R. (1970): The joint distribution of quadratic forms and related distributions. *Austral. J. Statist.* 12, 13-22.

40. Jensen, D.R. (1972): Some simultaneous multivariate procedures using Hotellings T^2 statistics. *Biometrics*, 28, 39-53.

41. John, S. (1961): On the evaluation of the probability integral of the multivariate t distribution. *Biometrika*, 48, 409-416.

42. Johnson, N.L. and Kotz, S. (1972): *Continuous Multivariate Distributions. Distributions in Statistics: Vol. 4* John Wiley, N.Y. 1972.

43. Juritz, J.M. (1971): The Generalised Partial Correlation Matrix. *S. Afr. Statist. J.* 5, 1-4.

44. Kabe, D.G. (1963): A note on the exact distributions of the G.C.L. estimators in two leading overidentified cases. *J. Amer. Statist.* 58, 535-537.

45. Kabe, D.G. (1964): On the exact distributions of the G.C.L. estimators in a leading three equation case. *J. Amer. Statist. Assoc.* 59, 881-893.

46. Kabe, D.G. (1968): On the distribution of the regression coefficient matrix of a normal distribution. *Austral. J. Statist.* 10. 21.23.

47. Kanofsky, P.B. (1968): Parametric confidence bands on cumulative distribution functions. *Sankhyā Series A.* 30, 369-378.

48. Koopmans, T.C. and Hood, W.C. (1953): *Studies in econometric method, Cowles Commission Monograph 14.* Wiley, N.Y.

49. Krishnan, M. (1967): The Moments of a doubly non-central t distribution. *J. Amer. Statist. Assoc.* 62, 279-287.

50. Krishnan, M. (1972): Series representations of a bivariate singly non-central t distribution. *J. Amer. Statist. Assoc.* 67, 228-231.

51. Krishnan, M. (1970): The bivariate doubly non-central t distribution (abstract). *Ann. Math. Statist.* 41, 1135.

52. Kshirsagar, A.M. (1960): Some extensions of the multivariate t-distribution and the multivariate generalisation of the distribution of the regression coefficient. *Proc. Camb. Phil. Soc.* 56, 80-85.

53. Marakathavalli, N. (1954): The distribution of t, and its applications. *J. Madras Univ. B.* 24, No. 2, 251-272.

54. Miller, K.S. (1968): Some multivariate t-distributions. *Ann. Math. Statist.* 39, 1605-1609.

55. Miller, R.G. (1966): *Simultaneous Statistical Inference.* McGraw-Hill, N.Y.

56. Money, A.H. (1972): Non-central multivariate Beta distributions and their applications. Unpublished Ph.D. thesis, University of Cape Town.

57. Olkin, I. and Rubin, H. (1964): Multivariate beta distributions and independence properties of the Wishart distribution. *Ann. Math. Statist.* 35, 261-269.

58. Owen, D.B. (1968): A Survey of Properties and applications of the non-central t distribution. *Technometrics,* 10, 445-472.

59. Patil, S.A. and Liao, S.H. (1970): The distribution of the ratios of means to the square root of the sum of variances of a bivariate normal sample. *Ann. Math. Statist.* 41, No. 2, 723-728.

60. Patil, S.A. and Kovner, J.L. (1969): On the bivariate doubly noncentral t distributions (abstract). *Ann. Math. Statist.* 40, 1868.

61. Press, S.J. (1972): *Applied Multivariate Analysis.* Holt, Rinehart and Winston, Inc. N.Y.

62. Raiffa, H. and Schlaifer, R. (1961): Applied Statistical Decision Theory. *Studies in Managerial Economics, Harvard University, Boston.*

63. Rainville, E.D. (1960): *Special Functions.* The Macmillan Company, New York.

64. Rao, C.R. (1965): *Linear Statistical Inference and its Applications.* Wiley, N.Y.

65. Rayner, A.A. and Livingstone, D. (1965): On the distribution of quadratic forms in singular normal variates. *S. Afr. J. Agric. Sci.* 8, 357-369.

66. Rayner, A.A. and Niven, B.M. (1970): Conditions for a quadratic form in normal variates to have a chi-squared distribution. *Agrochemophysica,* 2, 93-94.

67. Resnikoff and Lieberman (1957): *Tables of the Non-Central t-distribution.* Stanford Univ. Press.

68. Richardson, D.H. (1968): The exact distribution of a structural coefficient estimator. *J. Amer. Statist. Assoc.* 63, 1214-1226.

69. Richardson, D.H. and Wu, De-Min (1971): A note on the comparison of ordinary and two-stage least squares estimators. *Econometrica,* 39, 973-981.

70. Robbins, H. (1948): The distribution of Student's t when the population means are unequal. *Ann. Math. Statist.* 19, 406-410.

71. Sawa, T. (1969): The exact sampling distribution of ordinary least squares and two stage least squares estimators. *J. Amer. Statist. Assoc.* 64, 923-937.

72. Scheffé, H. (1953): A method for judging all contrasts in the analysis of variance. *Biometrika* 40, 87-104.

73. Searle, S.R. (1971): *Linear Models.* Wiley, N.Y.

74. Siddiqui, M.M. (1967): A bivariate t distribution. *Ann. Math. Statist.* 38, 162-166.

75. Steffens, F.E. (1969): Non-central bivariate t-tables. *CSIR Special Report, Wisk.* 61, 1-15.

76. Stone, M. (1964): Comments on the posterior distribution of Geisser and Cornfield. *J. Roy. Statist.* Series C, 26, 274.

77. Styan, G.P.H. (1970): Notes on the distribution of quadratic forms in singular normal variables. *Biometrika,* 57, No. 3, 567-572.

78. Tiao, G.G. and Guttman I.(1965): The inverted Dirichlet Distribution with applications. *J. Amer. Statist. Assoc.* 60, 793-805.

79. Tiao, G.C. and Zellner, A.(1963): On the Baysian estimation of multivariate regression. *J. Roy. Stat. Soc. B,* 26, 277-285.

80. Troskie, C.G. (1966): Nie-sentrale meerveranderlike beta verdelings. *Tydskr. Natuurwet.* 6, 58-71.

81. Troskie, C.G. (1969): The generalised multiple correlation matrix. *S.A. Statist. Jour.* 3, 109-122.

82. Troskie, C.G. (1971): Regression and Correlation. Proceedings of the Third Symposium on Mathematical Statistics. *C.S.I.R. Special Report,* 89, 21-47.

83. Trout, R and Chow, B. (1972): Table of the percentage points of the trivariate t distribution with application to uniform confidence bands. *Technometrics,* 14, No. 4, 855-879.