SOME ASPECTS OF MULTIVARIATE ANALYSIS

S. N. ROY
Professor of Statistics,
The University of North Carolina

NEW YORK : JOHN WILEY & SONS INC.
CALCUTTA : INDIAN STATISTICAL INSTITUTE
1957

Indian Statistical Series No. 1

Printed in India by Kalipada Mukherjee at Eka Press, Calcutta – 35
and published by the Indian Statistical Institute.

FOREWORD

Professor S. N. Roy's monograph will be a valuable addition to the other important publications on multivariate analysis. The monograph does not attempt to cover the entire field of multivariate analysis but it includes a good deal of new material which would be of interest to advanced students and research workers.

I have much pleasure in writing this foreword. More than fifteen years ago, when Professor Roy was working in the Indian Statistical Institute, he and I and other colleagues had discussed the question of bringing out a series of statistical monographs. Professor Roy had undertaken at that time to prepare one on multivariate analysis and has now completed his voluntary assignment. He has very kindly made over the copyright to the Indian Statistical Institute which is thankfully accepted by the Institute.

We should also like to offer our thanks to Messrs. John Wiley & Sons for the help we have received from them.

P. C. Mahalanobis

27 July 1957.

PREFACE

This monograph does not by any means attempt to cover the entire area of multivariate analysis, or even a major part of it. Aside from certain basic notions and results due to Fisher, Hotelling, Mahalanobis, Karl Pearson, Wilks, Wishart, Yule and some of their predecessors, which have now become current coin, this monograph is primarily concerned with those developments in multivariate analysis in which the author has been specially interested and with which he and some of his collaborators have been associated over several years. Part of the material presented here, as far as the author is aware, has not been published before, while the rest has been collected from papers by various workers in this sector including the author and his collaborators. It will be seen that in this monograph the statistical approach to different problems and the mathematical treatment of all such problems are uniform and perhaps somewhat individual, and that this applies to all specific results, no matter whether they are due to the author and his collaborators, or to other workers in the field or to both groups simultaneously.

What has not been discussed in this monograph has been developed and adequately handled in important papers by Anderson, Bartlett, Bose, Hsu, Kendall, Mahalanobis, Mosteller, Narain, Rao, Votaw, Wald and Brookner, Wilks and several other workers. Three excellent books touching upon but not primarily restricted to this sector, "Advanced Theory of Statistics" by M. G. Kendall, Vol. 2 [35], "Advanced Statistical Methods in Biometric Research" by C. R. Rao [14] and "Mathematical Statistics" by S. S. Wilks [28] have, between them, brought together and competently presented a substantial part of this material. For an adequate, unified and up-to-date presentation of this whole material the author, among others, is looking forward to the forthcoming book by T. W. Anderson, supposed to be dealing perhaps more or less exclusively with multivariate analysis.

The preparation of this monograph within a relatively short period, has been made possible only through the active co-operation of the entire secretarial staff of the department of statistics at Chapel Hill including, in particular, Mrs. Bonnie Baker Fathman, Mrs. Anne Kiley and Mrs. Mary Ann Taylor who did most of the typing and of several students of the author including, in particular, K. V. Ramachandran, A. E. Sarhan, V. N. Murty, R. Bargmann and R. Gnanadeshikan who rendered indispensable mechanical and critical help. This job was supported, in part, by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command. The printing and publication were kindly undertaken by the Indian Statistical Institute and the Eka Press, Calcutta for which mention must be made of J. Roy, S. K. Mitra and R. G. Laha and other members of the staff of the Indian Statistical Institute and the Eka Press for their kind help. To all these individuals and organizations are due the sincerest thanks of the author.

The author would be deeply grateful if errors were brought to his notice and suggestions were made for improvement in form no less than in content.

CONTENTS

APPENDICES

GENERAL INTRODUCTION

Multivariate analysis, till now, has been mostly concerned with point estimation of or testing of hypothesis on parameters or parametric functions occurring in one or more multivariate normal populations, the estimation or testing of hypothesis being of course in terms of random samples drawn from such populations. Except for the last chapter of the main body (i.e., Chapter 15) which deals with categorical data, this monograph also is concerned with "normal variate" data, but here point estimation is not discussed at all; and although testing of hypotheses is discussed a good deal, a careful reader will perceive that the main accent is on obtaining confidence bounds on certain parametric functions, the testing of hypotheses (in so far as it is developed) being largely a means to that end. The parametric functions that figure in this monograph are, in each case, a set of natural measures of departure from the customary null hypothesis, there being, in some simple situations, a single such function (or a single measure), and in some more complicated situations a set of such functions (or a set of measures). Thus, out of the first fourteen chapters of the main body which deal with "normal variate" data, the first twelve chapters constitute a conscious attempt to lead up to confidence bounds on parametric functions (which, in each case, is a measure or a set of measures of deviation from the customary hypothesis), which then are discussed in detail in Chapters 13 and 14.

In each case this measure (or set of measures) of deviation from the customary hypothesis subsumes, as a special case, the prior notion of a distance function between two populations (or dispersion between several populations) which had (i) its beginning in the coefficient of racial likeness of Karl Pearson (who, however, did not conceive of a distance function in this connection), (ii) its second stage of development in the D^2 of Mahalanobis (who may have been motivated, among other things, by a desire to fuse the notion of the coefficient of racial likeness with the notion of the distance function of relativity), (iii) its third stage of development in the distance function between two or more general types of populations, as evolved, among others, by Bhattacharya who in particular, showed a more statistical slant and (iv) a fourth stage of development mostly in the U.S.A. with the same slant which may have been independent of but which, in fact, postdates Bhattacharya's own work.

However, from the general standpoint of this monograph, the reader will notice three large gaps involved in the omission of the important sectors of (a) factor analysis, (b) classification problems and (c) the multivariate generalization of variance components analysis in univariate analysis of variance. The reasons for the omission are the following. Under (a) the author has long been looking for further clarification of the issues and then for some means of bringing (a) into the framework of confidence bounds on suitable parametric functions. Under (c) the author has been looking for some more clarification even within the univariate set-up, and then a suitable multivariate generalization of the univariate set-up and finally a way to bring this into the framework of confidence bounds on proper parametric functions. Under (b) the task for the author was one of merely bringing the problem within the framework of confidence bounds. It has been only long after the manuscript went to the press that all this has been accomplished to the partial satisfaction of the author, and all this with such further developments as may occur meanwhile will be presented in the next edition of the monograph.

In Chapter 15 a small beginning has been made in the direction of a certain type of non-parametric generalization of "normal variate" analysis of variance and multivariate analysis. A lot more has been done in this area since the manuscript went to the press and a great deal more remains to be done. The author hopes to either incorporate all this in a future edition of this monograph or perhaps present it in a separate monograph. Despite all the mathematical elegance and comparative simplicity of "normal variate" analysis of variance and multivariate analysis, one cannot help feeling that the non-parametric approach (whether of this variety or of other varieties) is far more realistic and physically meaningful, and is likely, in the future, to supplant, to a large extent, the existing techniques of "normal variate" analysis of variance and multivariate analysis, including those discussed in the first fourteen chapters of this monograph. Nevertheless, it seems that the customary "normal variate" techniques and concepts (and perhaps also those discussed in this monograph) will long remain a guide and a source of stimulus to non-parametric developments, in both their mathematical and their physical aspects—a point which may be somewhat overlooked by those who are not thoroughly conversant with all the current "normal variate" developments that have occurred over the last forty years.

CHAPTER ONE

Notation, Preliminaries and General Objectives

1.1. *Notation.* As far as possible the following notation and convention will be used, all departures being clearly indicated at the proper places. Greek letters will stand for population parameters and Italic letters over the first half of the alphabet for given (non-stochastic) quantities and over the latter part from, say, r to the end for sample quantities. Matrices and vectors under consideration will consist of real elements (these will be called real matrices or vectors) except occasionally when they might have complex elements (these will be called complex matrices or vectors). Capital letters will stand for matrices, small letters for scalars, bold face small letters for column vectors and for row vectors if they are primed. The transpose of a matrix or a column vector will be denoted by priming such quantities, the conjugate complex transpose of a matrix M by M^*, the set of characteristic roots of M (if it is square) by $c(M)$, its trace by tr M, the modulus of the determinant of such a matrix by $|M|$, the modulus of a scalar m by $|m|$ and the inverse of a matrix M (if it is square and non-singular) by M^{-1}. A real square matrix $M(p\times p)$ will be called $\perp$ if it is orthogonal, i.e., if $MM' = I(p)(= M'M$, necessarily), and if $M(p\times q)$ $(p < q)$ is such that $MM' = I(p)$, then M will be called semi-orthogonal. To indicate the structure, a $p\times q$ matrix, say M, or a $p\times 1$ column vector, say $\mathbf{m}$, will sometimes be written respectively as $M(p\times q)$ or $\mathbf{m}(p\times 1)$. A matrix M whose typical element is m_{ij} will sometimes be denoted by (m_{ij}). The (ij)-th element of a matrix M will be denoted by $(M)_{ij}$ or m_{ij}. A diagonal matrix whose diagonal elements are, say, $a_1, a_2, \ldots, a_p$, will be denoted by D_a. A diagonal matrix with ± 1 for its diagonal elements will be denoted by D_k. $\tilde{A}(p\times p)$ or sometimes simply $\tilde{A}$ will stand for the triangular matrix

$$\begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & 0 \\ \cdot & \cdot & \ldots & \cdot\,\cdot \\ a_{p1} & a_{p2} & \ldots & a_{pp} \end{bmatrix} \qquad \ldots \quad (1.1.1)$$

We have also $|\tilde{A}| = \prod_{i=1}^{p} a_{ii}$, and it is easy to check that if $\tilde{A}$ is non-singular, then $a_{ii} \neq 0$, and $\tilde{A}^{-1}$ will also be a triangular matrix with the same configuration as $\tilde{A}$. The product of two triangular matrices of the same configuration is a triangular matrix of the same configuration. $\tilde{A}'$ is a triangular matrix of the opposite configuration to $\tilde{A}$. If $A(p\times q) = (a_{ij})$, then dA will stand for $\prod_{j=1}^{q}\prod_{i=1}^{p} da_{ij}$ and if $\mathbf{a}'(1\times p) = (a_1, \ldots, a_p)$, then $d\mathbf{a}$ will denote $\prod_{i=1}^{p} da_i$. The Jacobian of the transformation to an independent set of variables, say, $\mathbf{x}$ from any independent set of variables, say, $\mathbf{y}$ (with of course the same number of elements as $\mathbf{x}$) will be denoted by $J(\mathbf{x} : \mathbf{y})$, while a symbol, say, $\frac{\partial(u_1, \ldots, u_k)}{\partial(\theta_1, \ldots, \theta_k)} \left(= \frac{\partial u}{\partial \theta}\right)$ will have the same well known meaning as in the calculus. The terms "positive definite" and "positive semi-definite" will be abbreviated p.d. and p.s.d. respectively. "Almost everywhere", that is, "except for a set of (probability) measure zero" will be referred to as a.e. As usual, p.d.f. and c.d.f. will stand respectively for the probability density function and the cumulative distribution function (of a stochastic variate).

A stochastic variate $x(-\infty < x < \infty)$ will, as usual, be called $N(\xi, \sigma^2)$ if it has the p.d.f.

$$(1/\sigma\sqrt{2\pi}) \exp [-(x-\xi)^2/2\sigma^2], \qquad \ldots \quad (1.1.2)$$

where $-\infty < \xi < \infty$ and $\sigma > 0$. It is well known that $E(x) = \xi$ (to be called the mean) and $E(x-\xi)^2 = \sigma^2$ (to be called the variance). A stochastic vector $\mathbf{x}(p \times 1)$ $(-\infty < x_i < \infty)$ will be called $N(\boldsymbol{\xi}, \Sigma)$ if it has the p.d.f.

$$[1/|\Sigma|^{1/2}(2\pi)^{p/2}] \exp [-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(\mathbf{x}-\boldsymbol{\xi}) (\mathbf{x}'-\boldsymbol{\xi}')], \qquad \ldots \quad (1.1.3)$$

where $-\infty < \xi_i < \infty$ and where Σ is a $p \times p$ symmetric p.d. matrix. It is also well known that

$$E(\mathbf{x}) = \boldsymbol{\xi} \text{ and } E(\mathbf{x}-\boldsymbol{\xi}) (\mathbf{x}'-\boldsymbol{\xi}') = \Sigma. \qquad \ldots \quad (1.1.4)$$

$\boldsymbol{\xi}$ will be called the population mean vector and Σ the population dispersion matrix [see Chapter 3].

The symbols ϵ, $\bigcup$, $\bigcap$, "A statement $\Longleftrightarrow$ another statement", "A statement $\Longrightarrow$ another statement", will all be taken over from the notation and terminology of set theory and measure theory and so also w' for the complement of a set w in a space X. The most powerful critical region of size, say, $\beta_H(< 1)$ (which under fairly general conditions, will exist and which, under slightly less general conditions, will also be unique) of a simple hypothesis H_0 against a simple alternative H (such that $H \epsilon \Omega$ where Ω stands for the domain of possible alternatives) will be denoted by $w(H_0, H, \beta_H)$ and its complement, the acceptance region by $w'(H_0, H, \beta_H)$, to indicate that, in general, both will depend on β_H, H_0 and H. The union of regions $w(H_0, H, \beta_H)$ over different $H \epsilon \Omega$ will be denoted by $\bigcup_{H\epsilon\Omega} w(H_0, H, \beta_H)$ or simply by $\bigcup_H w$, and the intersection of regions $w'(H_0, H, \beta_H)$ over $H \epsilon \Omega$ by $\bigcap_{H\epsilon\Omega} w'(H_0, H, \beta_H)$ or simply by $\bigcap_H w'$. $P(H_0, H, \beta_H)$ will stand for the power of the most powerful critical region of size β_H for H_0 against H. ϕ_H will usually denote the p.d.f. under the hypothesis H.

1.2. *Some preliminaries on testing of hypotheses.* It is well-known that $w(H_0, H, \beta_H)$ and $w'(H_0, H; \beta_H)$ are given respectively by

$$w(H_0, H, \beta_H) : \phi_H \geqslant \lambda\phi_{H_0}, \quad w'(H_0, H, \beta_H) : \phi_H < \lambda\phi_{H_0}, \qquad \ldots \quad (1.2.1)$$

where λ is determined by $P[\mathbf{x}\epsilon w(H_0, H, \beta_H) | H_0] = \beta_H$. It can be shown [43] that

$$P(H_0, H, \beta_H) > \beta_H. \qquad \ldots \quad (1.2.2)$$

Proof: Assume that ϕ is such that w defined by (1.2.1) is unique. Integrating the first inequality of (1.2.1) over $w(H_0, H, \beta_H)$ and the second one over w', we have respectively, $P(H_0, H, \beta_H) \geqslant \lambda\beta_H$ and $1-P(H_0, H, \beta_H) < \lambda(1-\beta_H)$, from which, after a slight reduction, we have (1.2.2).

Note that in general λ will be of the form $\lambda(H_0, H, \beta_H)$ depending on all the elements. Incidentally, *any* critical region of size β for H_0, whose power with respect to an alternative H is greater than or equal to β, will be called an *unbiassed critical region* for H_0 against H.

Along with the more common terminology, namely the most powerful test of H_0 against H, a locally most powerful test of H_0 (in those situations where it is meaningful), a uniformly most powerful test of H_0 (if it exists at all with respect to the whole relevant class of alternatives) we shall also use the less common terminology, namely, an unbiassed test of H_0 against H, a locally unbiassed test and a uniformly unbiassed test. The result (1.2.2) shows that a most powerful test of H_0 against H, a locally most powerful test of H_0 and a uniformly most powerful test of H_0 are also respectively an unbiassed test of H_0 against H, a locally unbiassed test and a uniformly unbiassed test. Of course, in general, unbiassed tests will be a much larger class, of which the most powerful test will be just a member.

The likelihood ratio critical region at a level, say α, of a simple H_0 against the whole class of simple $H \epsilon \Omega$, provided that it exists, will be denoted by $w(\alpha, H_0)$. As is well-known it is given by

$$w(H_0, \alpha) : \phi(\mathbf{x}) \geqslant \mu(H_0, \alpha)\phi_{H_0}(\mathbf{x}), \qquad \ldots \quad (1.2.3)$$

where for a given $\mathbf{x}$, $\phi(\mathbf{x})$ stands for the largest $\phi_H(\mathbf{x})$ (provided that it exists) with respect to variation of H over Ω, and where $\mu(H_0, \alpha)$ is given by

$$P(\mathbf{x} \epsilon w(H_0, \alpha) | H_0) = \alpha. \qquad \ldots \quad (1.2.4)$$

Notice that $\phi(\mathbf{x})$ is a function of $\mathbf{x}$ only, being independent of $H \epsilon \Omega$, but may depend on the *total* domain Ω. The power of this test, aginst any particular alternative $H \epsilon \Omega$, will be denoted by $P(H_0, H, \alpha_H)$.

Assume now that H_0 is a composite hypothesis and $H \epsilon \Omega$ a composite alternative. In earlier papers [40, 41] the author gave a set of sufficient conditions on ϕ_{H_0} for the availability of similar regions for H_0, and a set of (further) restrictions on ϕ_H and ϕ_{H_0} for the availability, among these similar regions, of one which is the most powerful for H_0 against H in the following sense : Suppose that H_0 and H are composite hypotheses, each characterized by some specified and some unspecified elements, so that, if the unspecified elements were specified, both H_0 and H would be simple hypotheses. Now suppose that, among the similar regions for H_0, there is one whose location in the sample space depends on the specified elements of H_0 and possibly on those of H, but not on the unspecified elements of H_0 or H, but which is nevertheless the most powerful critical region for *any* simple hypothesis within H_0 (obtained by specifying the unspecified elements) against *any* simple alternative within H (obtained by specifying the unspecified elements). But this "most powerful" is "most powerful among similar regions". If we drop the restriction of similarity and set up in a straightforward manner the most powerful critical region for the simple hypothesis in question against the simple alternative in question, then we may get a (*non*-similar) region having a larger power than that of the most powerful *similar* critical region just referred to. Such a most powerful critical region may be conveniently called a *bisimilar* region for H_0 against H. The likelihood ratio critical region for composite H_0 against all composite $H \epsilon \Omega$ (which we know how to construct, provided that it exists), can be shown to be a similar region for H_0 under the restrictions just referred to. In this situation the same notation will be used as introduced in the previous paragraph for the case of a simple hypothesis against simple alternatives, and the result (1.2.4) will also hold, it being noted that, while the regions will be independent of the unspecified elements in H_0 and H, $P(H_0, H, \beta_H)$ and $P(H_0, H, \alpha_H)$ however, might depend on the unspecified elements of H though not usually on those of H_0.

1.3. *General objectives.* Throughout this monograph we shall restrict ourselves to very limited objectives, namely solution of certain non-sequential, i.e., fixed sample size two-decision problems, in which, for a preassigned level α or a confidence coefficient $1-\alpha$, we are interested respectively in obtaining (i) a (similar) region test of a composite H_0 which has some kind of reasonably 'good' property against the whole class of relevant (composite) alternatives $H(\epsilon \Omega)$ or (ii) a set of simultaneous confidence bounds on deviations from H_0, naturally occuring in the problems to be considered (all to be explained later), the confidence bounds, again, having some kind of 'good' properties in terms of covering 'wrong' values of the deviations. The scope of the discussion is thus professedly quite narrow and by no means fully adequate for the needs of any possible user of statistics, but that is as far as we can get at the moment. It is hoped that, in the near future, methods and techniques will develop perhaps in extension of those offered here, which can cope with the more recondite problems that are of real interest to the possible users of statistics.

Towards these limited objectives, a heuristic method of test construction will be offered which leads to a certain class of tests including in particular, two members of special importance to be called respectively *type I* and *type II* tests and a generalisation of type I test, to be called an *extended type I test.* The type II test will be identified with the widely known likelihood ratio criterion, but it is the type I and the extended type I test that will be used throughout this report, and, in the specific situations to be considered, it will be possible, in every case, to obtain, by inversion of these tests, suitable confidence bounds on certain deviations or measures of departure from the hypothesis that naturally arise in the case considered. As observed at the outset, the general method is entirely heuristic and, therefore, the test or the set of confidence bounds that emerges as the end product, in any specific problem, has to be justified by its operating characteristics in that situation, no 'good' properties being guaranteed in advance by the general method of test construction itself.

CHAPTER TWO

A Heuristic Class of Tests*

2.1. *Definitions and some remarks.* Consider, for simplicity but without any essential loss of generality (for the definitions could be immediately carried over into the case of composite hypothesis and alternative), a simple hypothesis H_0 against a simple alternative $H \epsilon \Omega$.

(i) Put $\beta_H = \beta(H \epsilon \Omega)$, and set up as the rejection and acceptance regions for H_0, $\bigcup_{H \epsilon \Omega} w(H_0, H, \beta)$ and its complement $\bigcap_{H \epsilon \Omega} w'(H_0, H, \beta)$, to be called respectively $\bigcup_H$ and $\bigcap_H$. *This is defined to be a type I test for H_0 against the whole class $H \epsilon \Omega$, the level of significance α being given by*

$$P(\mathbf{x} \epsilon \bigcup_{H \epsilon \Omega} w(H_0, H, \beta) | H_0) = \alpha(H_0, \beta) \quad (> \beta). \qquad \ldots \quad (2.1.1)$$

Let us for the moment assume non-triviality, that is, that, given $\alpha < 1$, we can find $\beta = \beta(H_0, \alpha) > 0$, for which (2.1.1) will hold.

(ii) Put, in section 1.2, $\lambda(H_0, H, \beta_H) = \mu$ (a preassigned constant) for all $H \epsilon \Omega$ and rewrite $w(H_0, H, \beta_H)$ and $w'(H_0, H, \beta_H)$ as $w^*(H_0, H, \mu)$ and $w^{*\prime}(H_0, H, \mu)$ respectively.

Now set up, as the rejection and acceptance regions for H_0, $\bigcup_H w^*(H_0, H, \mu)$ and its complement $\bigcap_H w^{*\prime}(H_0, H, \mu)$, to be called, respectively, $\bigcup_H^*$ and $\bigcap_H^*$, where the β_H's $(H \epsilon \Omega)$ are subject to $\lambda(H_0, H, \beta_H) = \mu$ (a preassigned constant). *This is defined to be a Type II test for H_0 against the whole class $H \epsilon \Omega$* the level of significance α^* being given by

$$P(\mathbf{x} \epsilon \bigcup_{H \epsilon \Omega} w^*(H_0, H, \mu) | H_0) = \alpha^*(H_0, \mu). \qquad \ldots \quad (2.1.2)$$

Here again let us, for the moment, assume nontriviality, that is, that given $\alpha^*(< 1)$, we can find a μ such that $\beta(H_0, H, \mu) = \beta_H(> 0)$ and that (2.1.2) will hold. This can be easily recognized as the likelihood ratio test by the following consideration. Notice that $w^*(H_0, H, \mu)$ (with a preassigned μ) is given by

$$w^*(H_0, H, \mu) : \phi_H(\mathbf{x}) \geqslant \mu \phi_{H_0}(\mathbf{x}). \qquad \ldots \quad (2.1.3)$$

Any $\mathbf{x}$ would belong to $\bigcup_H w^*(H_0, H, \mu)$ if for that $\mathbf{x}$, there were at least one $H \epsilon \Omega$ for which (2.1.3) holds. It is easy to see that this would be accomplished if for that $\mathbf{x}$ the largest $\phi_H(\mathbf{x})$ (under variation of H over Ω) were $\geqslant \mu \phi_{H_0}(\mathbf{x})$. Hence it is obvious that

$$\bigcup_H w^*(H_0, H, \mu) : \phi(\mathbf{x}) \geqslant \mu \phi_{H_0}(\mathbf{x}), \quad \bigcap_H w^*(H_0, H, \mu) : \phi(\mathbf{x}) < \mu \phi_{H_0}(\mathbf{x}). \quad \ldots \quad (2.1.4)$$

---

* See reference [43] in this connection.

2.2. *An obvious property of the two types of tests.* Notice that $\bigcup_H$ includes all $w(H_0, H, \beta)$ and $\bigcup_H^*$ all $w^*(H_0, H, \mu)$. Now putting

$$P(\mathbf{x}\epsilon\bigcup_H | H) \equiv P(\bigcup_H, H, \alpha) \text{ and } P(\mathbf{x}\epsilon\bigcup_H^* | H) \equiv P(\bigcup_H^*, H, \alpha)$$

we shall have from Section (3.1), for the two types of tests,

$$\beta(H_0, \alpha) \equiv \beta < P(H_0, H, \beta) \leqslant P(\bigcup_H, H, \alpha) \leqslant P(H_0, H, \alpha) \leqslant 1;$$
$$P(H_0, H, \alpha) > \alpha \quad \ldots \quad (2.2.1)$$

$$\beta^*(H_0, H, \alpha) = \beta_H^* < P^*(H_0, H, \mu) \leqslant P(\bigcup_H^*, H, \alpha) \leqslant P(H_0, H, \alpha) \leqslant 1 ;$$
$$P(H_0, H, \alpha) > \alpha. \quad \ldots \quad (2.2.2)$$

(2.2.1) and (2.2.2) give respectively, for all $H\epsilon\Omega$, the lower bounds $P(H_0, H, \beta)$ and $P^*(H_0, H, \mu)$ for $P(\bigcup_H, H, \alpha)$ and $P(\bigcup_H^*, H, \mu)$ which, however, in general, would be far from close except sometimes for large "deviation" from H_0. With more knowledge of the forms of ϕ_{H_0} and ϕ_H, it is often possible to get far closer bounds; even the actual powers are often computable without much difficulty (and turn out to be pretty high) as for example in most of the classical tests on normal populations.

It is easy to see that the results of (2.1) and (2.2) could be easily generalized to cover the case of composite H_0 against composite $H \epsilon \Omega$ provided that we have similar regions for H_0 and a bisimilar region for H_0 against H. This, therefore, need not be separately treated.

2.3. *Display of two classical tests as type I tests.* (i) Almost all classical tests on univariate and multivariate normal populations, (ii) most classical tests on other types of populations and (iii) many tests on multivariate normal populations proposed in recent years are known to be derivable (and indeed many of them have, in fact, been derived) from the "likelihood ratio" principle, so that they belong to type II. The author finds that all the customary tests in category (i), for example the test of significance of (1) a mean; (2) a mean difference, (3) total or partial or multiple correlation and (4) regressions, (5) the F-test in analysis of variance, (6) the test based on Hotelling's T^2, all belong to type I as well. Those classical tests in category (ii) that the author has examined so far also all belong to type I. Coming to those situations that are sought to be handled by tests proposed under category (iii), the author finds that the likelihood ratio tests offered so far, while they automatically belong to type II, do not belong to type I. On the other hand, if, in these situations, one carries out the spirit and method of discriminant analysis, one gets tests which belong to type I in a sense slightly more general than we have indicated so far.

In this section we consider, for illustration, two well known classical tests and show that they belong to type I.

(i) For $N(\xi_1, \sigma^2)$ and $N(\xi_2, \sigma^2)$ the classical test of $H(\xi_1 = \xi_2) = H_0$ against $H(\xi_1 \neq \xi_2) = H$ at a level α is based on a critical region given by

$$t \geqslant t_0 \quad \text{or} \quad \leqslant -t_0 \quad \ldots \quad (2.3.1)$$

where $t \equiv (n_1+n_2-2)^{\frac{1}{2}}\{n_1n_2/(n_1+n_2)\}^{\frac{1}{2}}(\bar{x}_1-\bar{x}_2)/\{(n_1-1)s_1^2+(n_2-1)s_2^2\}^{\frac{1}{2}}$, and t_0 is given by $P(t \geqslant t_0|H_0) = \alpha/2$ and where $(\bar{x}_1, \bar{x}_2)$, (s_1, s_2) stand for the means and standard deviations of two random samples of sizes n_1 and n_2 drawn from $N(\xi_1, \sigma^2)$ and $N(\xi_2, \sigma^2)$, respectively. This is well known as a likelihood ratio test, but it is easily checked as type I as well, in the following way. It is well known that $t \geqslant t_0$ is a one-sided uniformly most powerful (bisimilar) region of size $\alpha/2$ for the composite H_0 against the composite $H(\xi_1 > \xi_2) = H_1$ and so also is $t \leqslant -t_0$ for H_0 against $H(\xi_1 < \xi_2) = H_2$; taking the union we have (2.3.1) of size α

(ii) Consider the testing of a general linear hypothesis in analysis of variance which, as is well known, can be *formally* reduced to the following. Suppose we have random samples of sizes n_i, means $\bar{x}_i$ and standard deviations s_i, drawn respectively from $N(\xi_h, \sigma^2)$ $(h = 1, \ldots, k)$, and suppose we want to test $H(\xi_1 = \xi_2 = \ldots = \xi_k) = H_0$ against the whole class H of $(\xi_1, \ldots, \xi_k)$ violating H_0. Put $n = \sum_{h=1}^{k} n_h$; $\bar{x} = \sum_{h=1}^{k} n_h\bar{x}_h/n$; $\xi = \sum_{h=1}^{k} n_h\xi_h/n$. Now the classical F-test for H_0, which is well known to be a likelihood ratio or type II test has at a level α the critical region given by

$$F \geqslant F_0\,, \qquad \ldots \quad (2.3.2)$$

where
$$F = [\sum_{h=1}^{k} n_h(\bar{x}_h-\bar{x})^2/(k-1)] \div [\sum_{h=1}^{k} (n_h-1)s_h^2/(n-k)]\,,$$

and where F_0 is given by $P(F \geqslant F_0|H_0) = \alpha$.

To recognize this as a type I test as well we proceed as follows. It is observed in earlier papers [40], [21] that among similar regions for H_0 (which exist) there is a most powerful (bisimilar) region for H_0 against any specific $(\xi_1, \ldots, \xi_k) = \boldsymbol{\xi}$ violating H_0, the region of size, say, β being given by

$$t \geqslant t_0\,, \qquad \ldots \quad (2.3.3)$$

where
$$t = \sqrt{n-2}\,\cot\theta\,,$$

and
$$\cos\theta = \sum_{h=1}^{k} n_h(\bar{x}_h-\bar{x})\,(\xi_h-\xi)/[\sum_{h=1}^{k} \{n_h(\bar{x}_h-\bar{x})^2+(n_h-1)s_h^2\} \sum_{h=1}^{k} n_h(\xi_h-\xi)^2]^{\frac{1}{2}}\,,$$

and where t_0 is given by

$$P(t \geqslant t_0|H_0) = \beta.$$

It is also noticed in those papers that this t has exactly the usual t-distribution with $(n-2)$ degrees of freedom. Notice that $t_0 \equiv t_0(n, \beta)$ and $\beta \equiv \beta(n, t_0)$. To obtain now the union of regions: $t \geqslant t_0$ over different sets of $(\xi_1, \ldots, \xi_k)$ we note that a given observation set belongs to the union if for that observation set there is at least one t (obtained by varying over $\xi_1, \ldots, \xi_k$) such that $t \geqslant t_0$. The union is thus easily checked to be given by: the largest t (by varying over $\xi_1, \ldots, \xi_k$) $\geqslant t_0$ (which is fixed). But by

(2.3.3) the largest t would correspond to the largest value of cos θ, and, given $\bar{x}_i$'s and s_i's, the largest value of cos θ (under variation over $\zeta_1, \ldots, \zeta_k$) is easily seen to be given by

$$\cos\,\theta = [\sum_{h=1}^{k} n_h(\bar{x}_h - \bar{x})^2]^{\frac{1}{2}}/[\sum_{h=1}^{k} \{(n_h-1)\ s_h^2 + n_h(\bar{x}_h - \bar{x})^2\}]^{\frac{1}{2}}, \quad \ldots \quad (2.3.4)$$

so that the largest t is given by

$$t = (n-2)^{\frac{1}{2}}[\sum_{h=1}^{k} n_h(\bar{x}_h - \bar{x})^2]^{\frac{1}{2}}/[\sum_{h=1}^{k} (n_h-1)s_h^2]^{\frac{1}{2}}. \quad \ldots \quad (2.3.5)$$

Therefore, the union of regions: $t \geqslant t_0$, is given exactly by (2.3.2), which is the critical region of the F-test. Notice that, given the α of the F-test, F_0 is obtained from (2.3.2) in the form $F_0(k-1,\ n-k;\ \alpha)$; and next by identifying the union of regions $t \geqslant t_0$, with $F \geqslant F_0$ we have

$$t_0 = [(k-1)\,(n-2)\ F_0/(n-k)]^{\frac{1}{2}}$$

and next from (2.3.3) we have

$$\beta \equiv \beta(n,\ t_0) = \beta(k-1,\ n-k;\ \alpha).$$

2.4. *Some further remarks on the two types of tests.* It may be noted (see Section 2.1) that by specializing the β_H's (the sizes of the most powerful critical regions against different alternatives in *two* special ways we get in a heuristic manner the two types of test. By specializing the β_i's in other ways other heuristic principles could be set up, some of which, in special situations, might be "better" than the type I or type II tests. It has already been observed that in many situations type I and type II tests would coincide. This does not mean, however, that in those situations, $\beta(H_0,\ H,\ \alpha)$ of the type II test would be the β of the type I test. Given H_0 and the H's, it would be possible to find a β for type I and a μ for type II such that the same critical region for H_0 against the whole class $H \in \Omega$ could be looked upon as $\bigcup_H w(H_0,\ H,\ \beta)$ in relation to the first type and also as $\bigcup_H w^*(H_0,\ H,\ \mu)$ in relation to the second type.

The following theoretical question or group of questions, now under investigation, is extremely important. Under what general restrictions on the probability law of $\mathbf{x}$ and on H_0 and $H \in \Omega$ would either or both of the tests be nontrivial (in the sense discussed in Section 2.1) and usable (in the sense of having a distribution problem amenable to tabulation), and unbiassed (against all relevant alternatives) and/or admissible and/or reasonably powerful (in the sense of having not too bad a power against all relevant alternatives)? So far as the author is aware, these questions have not yet been adequately discussed in a *general* manner (let alone being answered) even for the likelihood ratio or type II test (which has so long been extensively used in practice), and no attempt will be made in this monograph to discuss these questions.

The advantage, however, of having two such heuristic principles (with the possibility of having two different tests in many situations) is that it gives us more elbow room than we would have with one such principle, in the matter of construction of nontrivial, usable and "pretty good" tests.

One remark on the admissibility of a test (in the Neyman-Pearson set-up) is especially important. In this set-up suppose we have a hypothesis H_0 and a class of alternatives $H \epsilon \Omega$. Assume, for simplicity of discussion, that H_0 and each H are simple hypotheses. Now suppose that there is any critical region of size, say α, for H_0. w_0 will be said to be inadmissible (or admissible) against the whole class $H \epsilon \Omega$ according as we can find (or fail to find) another critical region of size α, say w_1, such that

$$P(\mathbf{x} \epsilon w_1|H) \geqslant P(\mathbf{x} \epsilon w_0|H) \quad \text{for all } H \epsilon \Omega,$$

and

$$P(\mathbf{x} \epsilon w_1|H) > P(\mathbf{x} \epsilon w_0|H) \quad \text{for at least one } H \epsilon \Omega. \quad \ldots \quad (2.4.1)$$

Suppose now that w_0 is an inadmissible critical region in that we can find a w_1 satisfying (2.4.1) and, assume for simplicity of discussion that w_1 itself is admissible. It is easy to satisfy oneself that from any physical point of view w_1 is better than w_0. Suppose now that w_2 is another critical region for H_0 of size α, which is admissible against all $H \epsilon \Omega$. It does not follow from the definition of admissibility that w_2 will necessarily have the property (2.4.1) in relation to w_0. On the contrary it may well be that

$$P(\mathbf{x} \epsilon w_2|H) < P(\mathbf{x} \epsilon w_0|H) \quad \text{for most } H \epsilon \Omega, \quad \ldots \quad (2.4.2)$$

and

$$P(\mathbf{x} \epsilon w_2|H) \geqslant P(\mathbf{x} \epsilon w_0|H) \quad \text{for some } H \epsilon \Omega,$$

and

$$P(\mathbf{x} \epsilon w_2|H) > P(\mathbf{x} \epsilon w_0|H) \quad \text{for some } H \epsilon \Omega.$$

A precise definition of 'most' need not detain us here. In fact, if a most powerful critical region of H_0 against *a specific* $H \epsilon \Omega$ is most powerful in the strict sense of having a power against H, which is $>$ and not just $\geqslant$ that of any other rival, then this critical region will be, by definition, an admissible one against the *whole class* of $H \epsilon \Omega$. But it may have a poor power against most other alternatives. In other words, it is easy to convince ourselves that a particular inadmissible region may, from any physical point of view, be much better than many admissible regions, although there must be at least one admissible test (and usually a whole subclass of such tests) which satisfies (2.4.1) with respect to w_0 and is thus better than w_0 from any physical point of view. This is a point which is apt to be missed by the statistician, especially the theoretical statistician.

2.5. *On the operating characteristics of certain specific tests.* It turns out that in many specific situations (as in the cases to be discussed herein) it is possible to obtain a class of admissible critical regions for H_0 against all $H \epsilon \Omega$, each region having a power which is a function of certain parameters which are naturally interpreted as measures of deviation from H_0. This admissible class may not of course constitute the totality of all admissible critical regions. Now among this class, if there is a subclass which is not only unbiassed against all $H \epsilon \Omega$ but is such that the power of each is a monotonically increasing function of each of the 'deviations', then this subclass is, from any physical point of view, the really valuable subset and will be said to be an admissible, unbiassed subset having the monotonicity property. In situations where this is available and where all that we know about H is that $H \epsilon \Omega$, the rest of the admissible class may, for most purposes, be thrown out. It seems to the author that in such situations, this subset or subclass of critical regions is the best that we can obtain *as a whole* and any further attempt at any choice among this subclass, on the basis of some stronger optimum property or principle, would be open to controversy in that the selection principle would be likely to be artificial and not universally convincing. The author is aware of the asymptotic optimum properties of the likelihood ratio criterion for simple and composite hypotheses, under certain broad restrictions, but there are strong reasons to suppose that these asymptotic optimum properties are not peculiar to the likelihood ratio criterion but must be shared by a large class of criteria or critical regions. Where H_0 is composite there is the further restriction of similarity which, of course, can be relaxed by just requiring that any critical region should have size $\leqslant \alpha \, (< 1)$ under variation of the unspecified elements of H_0, in which case the region will be said to be a *valid* one, a special case of a *valid* region being a similar region. In any actual situation (usually involving a composite H_0), if we can find a similar (or valid) critical region which is (i) unbiassed against all $H \epsilon \Omega$, (ii) has the monotonicity or near monotonicity property to be defined in chapter 10 and is also (iii) admissible, then we shall consider this to be a satisfactory region and any attempt at getting a region with a stronger optimum property would, in most practical situations, be futile for reasons already indicated. However, if, as in most of the situations to be discussed herein, we have a number of rival regions available satisfying (i)—(iii), then it is no doubt an interesting and useful question as to how the powers of the different rivals compare over the whole range of $H \epsilon \Omega$ one rival being better than another over some part of the range with a reversal in another part of the range and so on. In most of the rather complex situations to be discussed in this monograph this would not be possible, because not only are the actual powers not available, but we do not even have, at the moment, methods and techniques of comparing powers (in the sense of greater or less) of two rivals without actually obtaining the powers. It is hoped that such techniques will be available in the near future. It may be noticed here that quite often it is possible to assert properties (i) and (ii) and sometimes also (iii) without explicitly obtaining the power functions. It may also be observed that among similar (or valid) regions satisfying (i)—(iii) an additional consideration for recommendation might be (iv) reasonable simplicity of the null distribution problem, i.e., the distribution problem under H_0. If we are also interested, as we shall be in all the problems hereafter, in simultaneous confidence statements on deviation parameters (or functions thereof), then another additional consideration would be (v) the possibility of inverting the test to obtain (without running into excessively difficult distribution problems) such simultaneous confidence bounds (preferably intervals).

It will be seen that the tests offered in this monograph are similar region tests (in fact, they will be shown to be stronger than that, in a sense to be explained hereafter) having properties (i), (ii), (iv) and (v). There are strong grounds for believing (although we do not yet have a rigorous proof except for the degenerate special cases which will be indicated as we get along) that the tests also satisfy (iii). Furthermore, the tests that are being offered for the different situations are such that it has been possible to obtain for each test a 'pretty good' (easily available) lower bound to the power function (and consequently a lower bound to the *shortness*, i.e., the probability of covering wrong values of the parameters or parametric functions, of the associated set of simultaneous confidence intervals), 'pretty good' in the sense that the lower bound itself is reasonably large and rapidly goes up as the deviations increase. To the tests considered hereafter there are certain rivals (better known but not discussed in this monograph for reasons indicated at the proper places) for which some of the above properties are well known to be true and some of the others are also conjectured by the author to be true, but have not yet been proved.

2.6. *Extended type I test.* Consider a composite hypothesis H_0 against a set of composite alternatives $H_i \epsilon \Omega$, ($i \epsilon$ continuum). It often happens, as for example in the broad situations discussed in Chapter 5, that, while there are similar regions for H_0, there is among these no most powerful (bisimilar) region for H_0 against any $H_i(i \ \epsilon$ continuum), but that we have, instead, the following situation. Suppose we have composite hypotheses H_{0j} (ϵ continuum) such that $\bigcap_j H_{0j} = H_0$ and composite alternatives H_{ij} ($i \ \epsilon$ continuum; $j \ \epsilon$ continuum) such that $\bigcap_j H_{ij} = H_i$. Notice that H_{0j} and H_{ij} have more unspecified elements than H_0 and H_i respectively. It may well be that we have (as in the cases discussed in Chapter 5) not only similar regions for H_{0j} but also, among these, a most powerful (bisimilar) region for H_{0j} against any H_{ij} (one for each i with $j \ \epsilon$ continuum; and then $i \ \epsilon$ continuum). Consider critical regions $w(H_{0j}, H_{ij}, \beta)$ of size β each. Then by our test procedure, over $\bigcap_j \bigcap_i$ of $w'(H_{0j}, H_{ij}, \beta)$ (which we call $\bigcap_{ji}$ for simplicity, we are anyway accepting $\bigcap_j H_{0j}$, that is, H_0 and over its complement $\bigcup_j \bigcup_i w(H_{0j}, H_{ij}, \beta)$ we are rejecting at least one H_{0j} and therefore H_0 itself. Suppose we set this up as a heuristic test for H_0 against the whole class $H_i \epsilon \Omega$. Then the critical region will be $\bigcup_j \bigcup_i w(H_{0j}, H_{ij}, \beta)$ or $\bigcup_{ji}$ of size α, given by

$$P(\mathrm{x} \epsilon \bigcup_{ji} | H_0) = \alpha ,$$

so that $$\alpha = \alpha(H_0, \beta) \text{ and } \beta = \beta(H_0, \alpha). \qquad \ldots \quad (2.6.1)$$

As before, nontriviality will be assumed, and it is easy to check that we shall have for all i and j the following inequality

$$\beta < P(H_{0j}, H_{ij}, \beta) \leqslant P(\bigcup_{ji}, H_i, \alpha) \leqslant 1. \qquad \ldots \quad (2.6.2)$$

It may be noted that while $w(H_{0j}, H_{ij}, \beta)$, a bisimilar region of size β for H_{0j} against H_{ij}, is independent of the unspecified elements of H_{0j} and H_{ij} and while the location of $\bigcup_{ji}$ must be and its *size might* be (as indeed it is for all the cases considered in Chapter 5) independent of the unspecified elements of H_{0j} and H_{ij}, the power $P(H_{0j}, H_{ij}, \beta)$

might involve the unspecified elements of H_{ij}, and $P(H_0, H_i, \alpha)$ involve those of H_i. As observed in Section 2.2, the lower bound to the power of the test, given by (2.6.2), while it is in general easily available, is, at the same time, much too crude. With more knowledge of the probability law a much closer lower bound can often be found as will be exemplified in later sections.

The gist of the heuristic union-intersection principle, in its application to the two-decision problem of the Neyman-Pearson variety, is this. Suppose we have a certain type of hypothesis H_0 against a certain type of alternative H, such that H_0 and H are mutually exclusive sets for which we have an acceptance region $w'(H_0, H)$ and a critical region $w(H_0, H)$ having some optimum properties and also some mathematical simplicity. Suppose, furthermore, that there is an H_0^* formed by the intersection (or union) of H_0's of the previous type and H^* formed by the union (or intersection) of H's of the previous type, such that H_0^* and H^* are also mutually exclusive sets. Then the acceptance region for H^* against H^* is given by $\bigcap_{H_0^* \subset H_0,\, H \subset H^*} w'(H_0, H)$ or $\bigcap_{H_0 \subset H_0^*,\, H^* \subset H} w'(H_0, H)$ and the critical region by $\bigcup_{H_0^* \subset H_0,\, H \subset H^*} w(H_0, H)$ or $\bigcup_{H_0 \subset H_0^*,\, H^* \subset H} w(H_0, H)$. Notice that, in particular, H_0^* may be the same as H_0 and/or H^* may be the same as H. There might be of course other variations on this. It is found in many situations, that if the original test has certain *optimum* properties, then the derived test has some *reasonably good* properties, and no test with a strong optimum property of any physically meaningful kind may be available at all. The same type of heuristic principle can be and has been actually used (though not in this monograph) on more general types of decision problems, too, the general idea being that if a complex decision problem can be built up of less complex decision problems each having a relatively simple decision rule with some optimum properties, then a decision rule for the more complex problem can often be built up from the (relatively simple) decision rules for the less complex problems. In many situations this decision rule will have reasonably good properties, and any rule having strong optimum properties (of any physically meaningful kind) may not be available at all.

CHAPTER THREE

The Multivariate Normal Population

A univariate normal p.d.f. has the form (1.1.2) so that the probability law can be rewritten in the form

$$(1/\sigma\sqrt{2\pi})\ \exp\left[-\tfrac{1}{2}(x-\xi)(\sigma^2)^{-1}(x-\xi)\right]dx, \qquad \ldots \quad (3.1)$$

where $-\infty < x,\ \xi < \infty$, $\sigma > 0$, and $E(x) = \xi$, $E(x-\xi)^2 = V(x) = \sigma^2$. By analogy let us write down for $-\infty < (x_1, \ldots, x_p) = \mathbf{x}'(1\times p) < \infty$, the probability law

$$k\ \exp\left[-\tfrac{1}{2}(\mathbf{x}'-\boldsymbol{\xi}')B^{-1}(\mathbf{x}-\boldsymbol{\xi})\right]d\mathbf{x}, \qquad \ldots \quad (3.2)$$

where $-\infty < \boldsymbol{\xi}' < \infty$, $B(p\times p)$ is symmetric p.d. and k is a positive constant, and B and k have not yet been interpreted in statistical terms.

To obtain k in terms of B, we use (A.3.9) to put $B = \tilde{T}\tilde{T}'$ and have $(\mathbf{x}'-\boldsymbol{\xi}')B^{-1} \times(\mathbf{x}-\boldsymbol{\xi}) = (\mathbf{x}'-\boldsymbol{\xi}')(\tilde{T}^{-1})'\ \tilde{T}^{-1}(\mathbf{x}-\boldsymbol{\xi})$. Now put $\tilde{T}^{-1}(\mathbf{x}-\boldsymbol{\xi}) = \mathbf{y}(p\times 1)$, so that $(\mathbf{x}-\boldsymbol{\xi}) = \tilde{T}\mathbf{y}$ and $J(\mathbf{x} : \mathbf{y}) = |\ \tilde{T}\ | = |B|^{\frac{1}{2}}$. Now $\mathbf{y}$ has the probability law

$$k|B|^{\frac{1}{2}}\ \exp\left[-\tfrac{1}{2}\mathbf{y}'\mathbf{y}\right]d\mathbf{y}, \qquad \ldots \quad (3.3)$$

so that $y_1, \ldots, y_p$ are independent $N(0, 1)$ each varying from $-\infty$ to ∞.

Integrating out over y_i's we have

$$k|B|^{\frac{1}{2}}(\sqrt{2\pi})^p = 1 \text{ or } k = 1/(2\pi)^{p/2}|B|^{\frac{1}{2}}, \qquad \ldots \quad (3.4)$$

which gives k in terms of B.

To interpret B statistically we proceed as follows. We first prove that for any non-null $\mathbf{a}'(1\times p)$

$$\mathbf{a}'\mathbf{x} \text{ is } N(\mathbf{a}'\boldsymbol{\xi}, \mathbf{a}'B\mathbf{a}). \qquad \ldots \quad (3.5)$$

Proof: $\mathbf{a}'(\mathbf{x}-\boldsymbol{\xi}) = \mathbf{a}'\tilde{T}\mathbf{y}$, using the transformation in connection, with (3.3). Now using (A.3.11) put $\mathbf{a}'\tilde{T} = (\mathbf{a}'\tilde{T}\tilde{T}'\mathbf{a})^{\frac{1}{2}}\ \mathbf{l}'(1\times p)$, where $\mathbf{l}'\mathbf{l} = 1$. Next, complete $\mathbf{l}'(1\times p)$ into an $\perp$ matrix $\begin{array}{cl} \left[\begin{array}{c}\mathbf{l}' \\ \mathbf{L}_1\end{array}\right] & \begin{array}{l}1 \\ p-1\end{array} \\ p & \end{array}$ and now make the orthogonal transformation:

$$\mathbf{z}(p\times 1) = \begin{array}{cc} \begin{array}{r}1 \\ p-1\end{array} & \left[\begin{array}{c}\mathbf{l}' \\ L_1\end{array}\right] \\ & p \end{array}\ \mathbf{y}(p\times 1)$$

We have thus $J(\mathbf{y} : \mathbf{z}) = 1$ and $z_1 = \mathbf{l}'\mathbf{y} = \mathbf{a}'\tilde{T}\mathbf{y}/(\mathbf{a}'B\mathbf{a})^{\frac{1}{2}} = \mathbf{a}'(\mathbf{x}-\boldsymbol{\xi})/(\mathbf{a}'B\mathbf{a})^{\frac{1}{2}}$. Going back to (4.3) we now have for $\mathbf{z}$ the probability law

$$\frac{1}{(2\pi)^{p/2}} \exp\left[-\tfrac{1}{2}\mathbf{z}'\mathbf{z}\right] d\mathbf{z}, \qquad \dots \quad (3.5a)$$

so that z_i's are independent $N(0,1)$ and thus z_1 is $N(0,1)$, and, therefore, $\mathbf{a}'(\mathbf{x}-\boldsymbol{\xi})$ is $N(0, \mathbf{a}'\, B\mathbf{a})$ and hence $\mathbf{a}'\mathbf{x}$ is $N(\mathbf{a}'\boldsymbol{\xi}, \mathbf{a}'B\mathbf{a})$, which completes the proof of (4.5).

Putting $\mathbf{a}' = (0, \dots, 0, 1, 0, \dots, 0)$ (0's everywhere else and 1 in the i-th place) we check that x_i is $N(\xi_i, (B)_{ii})$. *This means that any marginal x_i is normally distributed about ξ_i as the mean value with a variance, say $\sigma_{ii} = (B)_{ii} = b_{ii}$*, say $(i = 1, 2, \dots, p)$.

We next prove that

$$\text{if } \underset{p}{\begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix}}\begin{matrix}1\\1\end{matrix} \text{ is of rank 2, then } \begin{matrix}1\\1\end{matrix}\underset{p}{\begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix}} \mathbf{x}(p\times 1) \text{ has the probability law:}$$

$$\frac{1}{(2\pi)|C|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(\mathbf{x}'-\boldsymbol{\xi}')[\mathbf{a}_1 \vdots \mathbf{a}_2]\, C^{-1} \begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix} (\mathbf{x}-\boldsymbol{\xi})\right] d\left[\begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix} \mathbf{x}\right], \dots \quad (3.6)$$

where
$$C(2\times 2) = \begin{matrix}1\\1\end{matrix}\underset{p}{\begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix}} B(p\times p)\, [\underset{1}{\mathbf{a}_1} \vdots \underset{1}{\mathbf{a}_2}]p,$$

and that covariance $(\mathbf{a}'_1\,\mathbf{x}, \mathbf{a}'_2\,\mathbf{x}) = \mathbf{a}'_1\, B\, \mathbf{a}_2$.

Proof: $\begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix}(\mathbf{x}-\boldsymbol{\xi}) = \begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix}\tilde{T}\mathbf{y}$, using the transformation in connection with (3.3).

Now, using (A.3.11), put $\begin{matrix}1\\1\end{matrix}\underset{p}{\begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_2\end{bmatrix}}\ \tilde{T}(p\times p) = \tilde{U}(2\times 2)\ L(2\times p)$, where $LL' = I(2)$.

Next, complete $L(2\times p)$ into an $\perp$ matrix $\underset{p}{\begin{bmatrix}L\\ \cdot\\ L_1\end{bmatrix}}\begin{matrix}2\\ \\ p-2\end{matrix}$ and now make the orthogonal transformation:

$$\mathbf{z}(p\times 1) = \begin{matrix}2\\ \\ p-2\end{matrix}\underset{p}{\begin{bmatrix}L\\ \\ L_1\end{bmatrix}}\ \mathbf{y}(p\times 1). \quad \text{We have thus } J(\mathbf{y} : \mathbf{z}) = 1$$

and
$$\underset{1}{\begin{bmatrix}z_1\\ \\ z_2\end{bmatrix}}\begin{matrix}1\\ \\ 1\end{matrix} = L(2\times p)\mathbf{y}(p\times 1) = \tilde{U}^{-1}\begin{bmatrix}\mathbf{a}'_1\\ \\ \mathbf{a}'_2\end{bmatrix}\ \tilde{T}\mathbf{y} = \tilde{U}^{-1}\begin{bmatrix}\mathbf{a}'_1\\ \\ \mathbf{a}'_2\end{bmatrix}(\mathbf{x}-\boldsymbol{\xi}). \qquad \dots \quad (3.7)$$

Going back to (3.3) we now have for $\mathbf{z}(p\times 1)$ the same probability law as (3.3), so that z_i's are independent $N(0, 1)$ and thus $\begin{bmatrix}z_1\\ z_2\end{bmatrix}$ has the probability law

$$(1/2\pi)\exp\left[-\tfrac{1}{2}[z_1 \vdots z_2]\begin{bmatrix}z_1\\ z_2\end{bmatrix}\right] d\begin{bmatrix}z_1\\ z_2\end{bmatrix} \qquad \dots \quad (3.8)$$

At this point, using (3.7) and noting that

$$J\left[\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} : \begin{bmatrix} \mathbf{a}_1'(\mathbf{x}-\boldsymbol{\xi}) \\ \mathbf{a}_2'(\mathbf{x}-\boldsymbol{\xi}) \end{bmatrix}\right] = |\tilde{U}|^{-1}, \tilde{U}\tilde{U}' = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix} \tilde{T}\,\tilde{T}'(\mathbf{a}_1 : \mathbf{a}_2) = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix} B(\mathbf{a}_1 : \mathbf{a}_2),$$

so that $|\tilde{U}| = \left|\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix} B[\mathbf{a}_1 \;\; \mathbf{a}_2]\right|^{\frac{1}{2}} = |C|^{\frac{1}{2}}$,

we have for $\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix}\mathbf{x}$ the probability law

$$[1/(2\pi)|C|^{\frac{1}{2}}] \exp\left[-\tfrac{1}{2}(\mathbf{x}'-\boldsymbol{\xi}')[\mathbf{a}_1 \;\; \mathbf{a}_2]\, C^{-1} \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix} (\mathbf{x}-\boldsymbol{\xi})\right] d\left[\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix}\mathbf{x}\right], \quad \ldots \quad (3.9)$$

which proves the first part of (3.6). For the second part, we go back to (3.7) and observe that

$$\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix}(\mathbf{x}-\boldsymbol{\xi}) = \tilde{U}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} u_{11} & 0 \\ u_{21} & u_{22} \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} u_{11}z_1 \\ u_{21}z_1+u_{22}z_2 \end{bmatrix}, \quad \ldots \quad (3.10)$$

so that covariance $[\mathbf{a}_1'\mathbf{x}, \mathbf{a}_2'\mathbf{x}] = E(u_{11}z_1 \times \overline{u_{21}z_1+u_{22}z_2}) = [u_{11}\; u_{21}]$ (since z_1 and z_2 are independent $N(0, 1)$) $= (\tilde{U}\tilde{U}')_{12} = \left[\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix} \tilde{T}\tilde{T}'[\mathbf{a}_1 : \mathbf{a}_2]\right]_{12} = \mathbf{a}_1' B \mathbf{a}_2$, from the definition tion of $\tilde{U}$ and $\tilde{T}$ in terms of B. This proves the second part of (3.6).

Now taking $\mathbf{a}_1'$ to be a vector with 1 for its i-th component and 0's for the other components and $\mathbf{a}_2'$ a vector with 1 for its j-th component $(i \neq j)$ and 0's for the other components we have Cov $(x_i, x_j) = \sigma_{ij}$ (say) $= (B)_{ij}$.

Denoting by Σ the variance covariance (or, in other words, the dispersion) matrix of the x_i's and taking into account the statement just before (3.6) we thus see that $\Sigma = B$, which thus provides the statistical interpretation of B. Now (3.2) can be rewritten as

$$\frac{1}{(2\pi)^{p/2}|\Sigma|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(\mathbf{x}'-\boldsymbol{\xi}')\Sigma^{-1}(\mathbf{x}-\boldsymbol{\xi})\right]d\mathbf{x}, \quad \ldots \quad (3.10a)$$

which will be called the p-variate normal distribution. (3.10a) is also otherwise expressed as

$$\mathbf{x} : N(\boldsymbol{\xi}, \Sigma), \quad \ldots \quad (3.11)$$

and denoting by σ_{ij} the elements of Σ we have

$$\mathbf{a}'(1\times p)\mathbf{x}(p\times 1) : N(\mathbf{a}'\boldsymbol{\xi}, \mathbf{a}'\Sigma\mathbf{a}), \quad \ldots \quad (3.12)$$

for any non-null $\mathbf{a}$, of which a special case is

$$x_i : N(\xi_i, \sigma_{ii}), \quad i = 1, 2, \ldots, p, \quad \ldots \quad (3.13)$$

and

$$\begin{matrix} 1 \\ 1 \end{matrix}\underset{p}{\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix}} \mathbf{x}(p\times 1) : N\left[\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix}\boldsymbol{\xi}, \begin{bmatrix} \mathbf{a}_1'\Sigma\mathbf{a}_1 & \mathbf{a}_2'\Sigma\mathbf{a}_1 \\ \mathbf{a}_1'\Sigma\mathbf{a}_2 & \mathbf{a}_2'\Sigma\mathbf{a}_2 \end{bmatrix}\right], \quad \ldots \quad (3.14)$$

for any matrix $\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \end{bmatrix}$ of rank 2, of which a special case is

$$\begin{bmatrix} x_i \\ x_j \end{bmatrix} : N\left[\begin{bmatrix} \xi_i \\ \xi_j \end{bmatrix}, \begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{bmatrix}\right]. \qquad \ldots \quad (3.15)$$

The result (3.15) *means in words, that any marginal* (x_i, x_j) *has the bivariate normal distribution about* ξ_i *and* ξ_j *as means, with variances* σ_{ii} *and* σ_{jj} *and a covariance* σ_{ij}. There is a more general result than (3.14), namely, that for any $A(r\times p)$ (with $r \leqslant p$) of rank r,

$$A(r\times p)\ \mathbf{x}(p\times 1) : N[A(r\times p)\ \boldsymbol{\xi}(p\times 1),\ A(r\times p)\ \Sigma(p\times p)\ A'(p\times r)]. \qquad \ldots \quad (3.16)$$

Denote by σ^{ij} the ij-th element of Σ^{-1} and notice that $\sigma^{ij} = \sigma^{ji}$.

Now, starting from (3.10) and using (3.12)-(3.16), we have the following conditional distributions:

$$x_1 | x_2 : N[E(x_1 | x_2),\ 1/\sigma^{11}_{x_1,x_2}]. \qquad \ldots \quad (3.17)$$

where $\sigma^{11}_{x_1,x_2}$ is the 11-th element of $\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1}$, and $E(x_1|x_2)$ is a linear function of ξ_1, $(x_2-\xi_2)$,

and

$$x_1 | x_2, x_3, \ldots, x_p\ :\ N[E(x_1 | x_2, \ldots, x_p),\ 1/\sigma^{11}], \qquad \ldots \quad (3.18)$$

where σ^{11} is the (1,1)th element of the matrix $\begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1p} \\ . & \ldots & . \\ \sigma_{1p} & \ldots & \sigma_{pp} \end{bmatrix}^{-1}$, and $E(x_1 | x_2, \ldots, x_p)$ is a linear function of ξ_1, $(x_2-\xi_2), \ldots, (x_p-\xi_p)$,

and also

$$x_1, x_2 | x_3, \ldots x_p\ :\ N\left[E\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} x_3, \ldots, x_p\Big),\ \begin{bmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{12} & \sigma^{22} \end{bmatrix}^{-1}\right], \qquad \ldots \quad (3.19)$$

where σ^{11}, σ^{22} and σ^{12} are the 11, 22 and 12 elements of the inverse of the full matrix Σ, and $E\left(\begin{matrix} x_1 \\ x_2 \end{matrix}\middle| x_3, \ldots, x_p\right)$ are two linear functions of $\xi_1, (x_3-\xi_3), \ldots, (x_p-\xi_p)$ and $\xi_2, (x_3-\xi_3), \ldots, (x_p-\xi_p)$.

Taking the customary definition of the correlation coefficient from probability theory as

$$\rho_{x_1, x_2} = \rho_{12} = \text{covariance}\ (x_1, x_2)/[V(x_1)\ \ V(x_2)]^{\frac{1}{2}}, \qquad \ldots \quad (3.20)$$

we define partial correlation coefficient between x_1 and $x_2 | x_3, \ldots, x_p$ as

$$\rho_{x_1, x_2 | x_3, \ldots, x_p} = \rho_{12\cdot 34 \cdots p} = \text{covariance}\ (x_1, x_2 | x_3, \ldots, x_p)/[V(x_1 | x_3, \ldots, x_p)$$

$$\times V(x_2 | x_3, \ldots, x_p)]^{\frac{1}{2}} = -\sigma^{12}/(\sigma^{11}\,\sigma^{22})^{\frac{1}{2}} \text{ using (3.19).} \qquad \ldots \quad (3.21$$

Notice that this is independent of the values of $x_3, \ldots, x_p$.

Going back to (3.13) and (3.17) we can, for the normal population, at any rate, approach the concept of a correlation coefficient another way and reach the same formula as (3.20). This is as follows. We have $V(x_1) = \sigma_{11}$ and, from (3.17), $V(x_1|x_2) = (\sigma_{11}\,\sigma_{22}-\sigma^2_{12})/\sigma_{22}$ which is independent of the value of x_2. Therefore, in this case, it is easy to show that if we define the correlation coefficient, as we intuitively can, as

$$\rho^*_{12} = \left[1-\frac{\text{conditional variance of } x_1|x_2}{\text{total variance of } x_1}\right]^{\frac{1}{2}},$$

we should have

$$\rho^*_{12} = \left[\frac{\sigma_{11}-(\sigma_{11}\,\sigma_{22}-\sigma^2_{12})/\sigma_{22}}{\sigma_{11}}\right]^{\frac{1}{2}} = (\sigma^2_{12}/\sigma_{11}\,\sigma_{22})^{\frac{1}{2}} = \rho_{12}. \qquad \dots \quad (3.22)$$

It is clear that this approach also to the concept of partial correlation would, in the case of a multivariate normal distribution, lead to the same formula as (3.21).

Using this approach to the correlation coefficient between x_1 and $(x_2, \ldots, x_p)$, we define as multiple correlation coefficient between x_1 and $(x_2, x_3, \ldots, x_p)$

$$\rho_{1\cdot 23\cdots p} = \left[1-\frac{\text{conditional variance of } x_1|x_2, \ldots_p}{\text{total variance of } x_1}\right]^{\frac{1}{2}}$$

$$= \left[\frac{\sigma_{11}-1/\sigma^{11}}{\sigma_{11}}\right]^{\frac{1}{2}} = [1-1/\sigma^{11}\,\sigma_{11}]^{\frac{1}{2}}. \qquad \dots \quad (3.23)$$

Notice that this is independent of the values of $x_2, x_3, \ldots, x_p$.

It is easy to check that for a multivariate normal distribution x_1 and x_2 are independent if and only if $\rho_{12} = 0$, and x_1 and $(x_2, \ldots, x_p)$ are independent if and only if $\rho_{1\cdot 23\cdots p} = 0$, and x_1 and x_2 are conditionally independent $|x_3, \ldots, x_p$ if and only if $\rho_{12\cdot 34\cdots p} = 0$.

To tie in with the customary definition (3.20) under which $-1 \leqslant \rho_{12} \leqslant 1$, we allow for both positive and negative square roots in the definition of ρ_{12} and similarly let $\rho_{12\cdot 34\cdots p}$, also take both positive and negative values. But in the definition of $\rho_{1\cdot 23\cdots p}$ we allow only the positive square root, for obvious reasons.

CHAPTER FOUR

Random Samples from *p*-Variate Normal Populations

If $X(p\times m) = (\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_m)p$, where $\mathbf{x}_\lambda$'s $(\lambda = 1, \dots, m)$ are an independent set and each $\mathbf{x}_\lambda$ is $N(\boldsymbol{\xi}, \Sigma)$ and $p < m$, then denoting by $\xi(p\times m)$ the matrix $[\underset{1}{\boldsymbol{\xi}}, \underset{1}{\boldsymbol{\xi}}, \dots \underset{1}{\boldsymbol{\xi}}]p$, we have the following probability law for X:

$$[1/(2\pi)^{\frac{pm}{2}} |\Sigma|^{\frac{m}{2}}] \exp\left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(X-\xi)(X'-\xi')\right] dX. \qquad \dots \quad (4.1)$$

The elements of X, of course, lie between $-\infty$ and ∞. If now we pass over from $X(p\times m)$ to $X_1(p\times m)$ by a transformation: $X(p\times m) = X_1(p\times m)A(m\times m)$, where A is $\perp$ (non-stochastic), then, by (A.5.4), $J(X : X_1) = 1$ and we have also $XX' = X_1X_1{}'$ (since A is $\perp$).

Putting now $\overline{\mathbf{x}}'(1\times p) = (\bar{x}_1, \dots, \bar{x}_p)$, where $\bar{x}_i = \sum\limits_{\lambda=1}^{m} x_{i\lambda}/m$ $(i = 1, \dots, p)$, it is easy to see that

$$X(p\times m)\xi'(m\times p) = m\overline{\mathbf{x}}(p\times 1)\boldsymbol{\xi}'\,(1\times p), \xi(p\times m)X'(m\times p) = m\boldsymbol{\xi}(p\times 1)\overline{\mathbf{x}}'(1\times p)$$
$$\text{and } \xi(p\times m)\xi'\,(m\times p) = m\boldsymbol{\xi}(p\times 1)\boldsymbol{\xi}'(1\times p). \qquad \dots \quad (4.2)$$

Hence

$$(X-\xi)(X'-\xi') = XX'-m\overline{\mathbf{x}}\boldsymbol{\xi}'-m\boldsymbol{\xi}\overline{\mathbf{x}}'+m\boldsymbol{\xi}\boldsymbol{\xi}'. \qquad \dots \quad (4.3)$$

If we now choose the $\perp$ transformation matrix (from X to X_1) such that

$$A = \begin{bmatrix} \sqrt{1/m} & \sqrt{1/m} & \dots & \sqrt{1/m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \cdot & \cdot & \dots & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{bmatrix} = \begin{bmatrix} \sqrt{1/m} \quad \sqrt{1/m} \quad \dots \quad \sqrt{1/m} \\ B(\overline{m-1}\times m) \end{bmatrix} \text{(say)}, \qquad \dots \quad (4.4)$$

we have

$$X_1 = XA' = \left[\begin{array}{c:c} \sqrt{m}\,\bar{x}_1 & \\ \sqrt{m}\,\bar{x}_2 & \\ & X(p\times m)B'(m\times \overline{m-1}) \\ \sqrt{m}\,\bar{x}_p & \end{array}\right] = [\underset{1}{\sqrt{m}\,\overline{\mathbf{x}}} \; \vdots \; \underset{m-1}{Y}]p, \qquad \dots \quad (4.5)$$

(say). Thus

$$XX' = X_1X_1' = m\,\overline{\mathbf{x}}\,\overline{\mathbf{x}}'+YY', \qquad \dots \quad (4.6)$$

and hence substituting for XX', the right hand side of (4.3) becomes

$$YY'+m\,\overline{\mathbf{x}}\overline{\mathbf{x}}'-m\,\overline{\mathbf{x}}\boldsymbol{\xi}'-m\boldsymbol{\xi}\overline{\mathbf{x}}'+m\boldsymbol{\xi}\boldsymbol{\xi}' \text{ or } YY'+m(\overline{\mathbf{x}}-\boldsymbol{\xi})(\overline{\mathbf{x}}'-\boldsymbol{\xi}'). \qquad \dots \quad (4.7)$$

Remembering now that $J(X : X_1) = 1$ and transforming from X to X_1 we have for X_1 the probability law:

$$(1/(2\pi)^{\frac{pm}{2}}|\Sigma|^{\frac{m}{2}}) \exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\{YY'+m(\bar{\mathbf{x}}-\boldsymbol{\xi})(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\}\right] d(\sqrt{m}\,\bar{\mathbf{x}})dY. \quad \ldots \quad (4.8)$$

Let us put

$$m-1 = n \text{ and } \bar{X}(p\times m) = [\underset{1}{\bar{\mathbf{x}}} \ldots \underset{1}{\bar{\mathbf{x}}}]p = \begin{bmatrix} \bar{x}_1 & \bar{x}_1 & \ldots & \bar{x}_1 \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \bar{x}_p & \bar{x}_p & \ldots & \bar{x}_p \end{bmatrix} \quad \ldots \quad (4.9)$$

and recall that for a sample $X(p\times m)$, the sample dispersion matrix S or (s_{ij}) is fined by

$$nS = n(s_{ij}) = (X-\bar{X})(X'-\bar{X}'). \quad \ldots \quad (\ .10)$$

It is easy to check that

$$X\bar{X}' = \bar{X}X' = m\bar{\mathbf{x}}\bar{\mathbf{x}}', \text{ or } (X-\bar{X})(X'-\bar{X}') = XX'-m\bar{\mathbf{x}}\bar{\mathbf{x}}', \quad \ldots \quad (4.11)$$

so that, using (4.6), (4.10), and (4.11) we have

$$YY' = XX'-m\bar{\mathbf{x}}\bar{\mathbf{x}}' = (X-\bar{X})(X'-\bar{X}'). \quad \ldots \quad (4.12)$$

We note that if, as in this case, the elements of X vary from $-\infty$ to ∞, so do those of $\bar{\mathbf{x}}$ and Y, to make the transformation one to one. Now integrating out (4.8) over $\bar{\mathbf{x}}(-\infty$ to $\infty)$ we have for Y the probability law:

$$\left(1/(2\pi)^{\frac{pn}{2}}|\Sigma|^{\frac{n}{2}}\right)\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}YY'\right]dY, \quad \ldots \quad (4.13)$$

and integrating out over Y we have for $\bar{\mathbf{x}}$ the probability law:

$$(1/(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}) \exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}m(\bar{\mathbf{x}}-\boldsymbol{\xi})(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\right]d(\sqrt{m}\,\bar{\mathbf{x}})$$

or

$$(1/(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}) \exp\left[-\frac{m}{2}(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\Sigma^{-1}(\bar{\mathbf{x}}-\boldsymbol{\xi})\right] d(\sqrt{m}\,\bar{\mathbf{x}}), \quad \ldots \quad (4.14)$$

which shows that $\bar{\mathbf{x}}$ is $N(\boldsymbol{\xi}, \frac{1}{m}\Sigma)$.

For the purpose of any study of the sample or population dispersion matrices, we could, without any loss of generality, start right off (as we will quite often do) from (4.13) replacing Y by X, but with the understanding that now X is not the original matrix of $p\times m$ observations, but is a part of the transformed matrix (considered under (4.5)), being $p\times n$ in structure. *We shall customarily call this the reduced matrix.*

For an $X = \begin{array}{cc} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} & \begin{matrix} p \\ q \end{matrix} \\ m & \end{array}$ $(p+q < m)$ consisting of m independent $((p+q)\times 1)$ column vectors, the reduced matrix $Y = \begin{array}{cc} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} & \begin{matrix} p \\ q \end{matrix} \\ n & \end{array}$ $(p+q \leqslant n)$ will have the probability law:

$$(1/(2\pi)^{\frac{(p+q)n}{2}}|\Sigma|^{\frac{n}{2}}) \exp\left[-\tfrac{1}{2}\operatorname{tr}\begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22}\end{bmatrix}^{-1}\begin{bmatrix}Y_1 \\ Y_2\end{bmatrix}[Y'_1 : Y'_2]\right] dY_1\, dY_2, \quad \dots \quad (4.15)$$

where $\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22}\end{bmatrix}$ is the partitioned population dispersion matrix (symmetric p.d.) for the $(p+q)$ normal variates.

For k random samples of sizes m_h from k $N(\boldsymbol{\xi}_h, \Sigma)$, we have for Y_h's and $\bar{\mathbf{x}}'_h\ (1\times p) = (\bar{x}_{1h}, \dots, \bar{x}_{ph})$ $(h = 1, 2, \dots, k)$ the joint probability law:

$$(1/(2\pi)^{\frac{pm}{2}}|\Sigma|^{\frac{m}{2}}) \exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\left\{\sum_{h=1}^{k} Y_hY'_h + \sum_{h=1}^{k} m_h(\bar{\mathbf{x}}_h-\boldsymbol{\xi}_h)(\bar{\mathbf{x}}'_h-\boldsymbol{\xi}'_h)\right\}\right]$$

$$\times \prod_{h=1}^{k} dY_h \prod_{h=1}^{k} d(\sqrt{m_h}\bar{\mathbf{x}}_h), \quad \dots \quad (4.16)$$

where $m = \sum_{h=1}^{k} m_h$. We next put $n_h = m_h - 1, n = \sum_{h=1}^{k} n_h = m-k, Y(p\times n) = [\underset{n_1}{Y_1}, \underset{n_2}{Y_2}, \dots \underset{n_k}{Y_k}]p$; $\bar{\mathbf{x}} = \sum_{h=1}^{k} m_h\bar{\mathbf{x}}_h/m$, $\boldsymbol{\xi} = \sum_{h=1}^{k} m_h\boldsymbol{\xi}_h/m$, with components $\bar{x}_i = \sum_{h=1}^{k} m_h\bar{x}_{ih}/m$, $\xi_i = \sum_{h=1}^{k} m_h\xi_{ih}/m$ $(i = 1, \dots, p)$, and finally set

$$\left.\begin{aligned} X(p\times k) &= \begin{bmatrix} \sqrt{m_1}\,\bar{x}_{11} & \dots & \sqrt{m_k}\,\bar{x}_{1k} \\ \dots & \dots & \dots \\ \sqrt{m_1}\,\bar{x}_{p1} & \dots & \sqrt{m_k}\,\bar{x}_{pk} \end{bmatrix} = [\sqrt{m_1}\,\bar{\mathbf{x}}_1\ \sqrt{m_2}\,\bar{\mathbf{x}}_2 \dots \sqrt{m_k}\,\bar{\mathbf{x}}_k] \\ \text{and}\quad \xi(p\times k) &= \begin{bmatrix} \sqrt{m_1}\,\xi_{11} & \dots & \sqrt{m_k}\,\xi_{1k} \\ \dots & \dots & \dots \\ \sqrt{m_1}\,\xi_{p1} & \dots & \sqrt{m_k}\,\xi_{pk} \end{bmatrix} = [\sqrt{m_1}\,\boldsymbol{\xi}_1 \dots \sqrt{m_k}\,\boldsymbol{\xi}_k]. \end{aligned}\right\} \dots \quad (4.17)$$

Using now an $\perp$ $A(k \times k)$ of the structure

$$\begin{bmatrix} \sqrt{m_1/m} & \sqrt{m_2/m} & \dots & \sqrt{m_k/m} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix} \text{ or } \begin{bmatrix} \sqrt{m_1/m} \quad \dots \quad \sqrt{m_k/m} \\ B(k-\overline{1} \times k) \end{bmatrix} \text{(say)}, \quad \dots \quad (4.18)$$

and transforming from X and ξ to X_1 and ξ_1 such that

$$X_1 = XA = [\sqrt{m}\, \underset{1}{\bar{\mathbf{x}}} \,\vdots\, \underset{k-1}{Z}]p \quad \text{and } \xi_1 = \xi A' = [\sqrt{m}\, \underset{1}{\boldsymbol{\xi}} \,\vdots\, \underset{k-1}{\zeta}]p, \qquad \dots \quad (4.19)$$

remembering that $\sum\limits_{h=1}^{k} Y_h Y_h' = YY'$, and substituting in (4.16), it is easy to check for $\bar{\mathbf{x}}(p \times 1)$, $Z(p \times \overline{k-1})$ and $Y(p \times n)$ the following probability law :

$$[1/(2\pi)^{\frac{pm}{2}} |\Sigma|^{\frac{m}{2}}] \exp \left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} \{YY' + (Z-\zeta)(Z'-\zeta') + m(\bar{\mathbf{x}}-\boldsymbol{\xi})(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\} \right]$$

$$\times\, dYdZd(\sqrt{m}\, \bar{\mathbf{x}}). \qquad \dots \quad (4.20)$$

As before, all elements of $(Y, Z, \bar{\mathbf{x}})$ vary from $-\infty$ to ∞, and now integrating out over $\bar{\mathbf{x}}$, we have for (Y, Z) the joint probability law :

$$[1/(2\pi)^{\frac{p(m-1)}{2}} |\Sigma|^{\frac{m-1}{2}}] \exp \left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} \{YY' + (Z-\zeta)(Z'-\zeta')\} \right] dYdZ. \qquad \dots \quad (4.21)$$

Denoting by $(s_{ij})_h$, $\bar{x}_{ih}$, ξ_{ih} the dispersion matrix of the h^{th} sample, the mean of the h^{th} sample for the i^{th} variate and the mean of the h^{th} population for the i^{th} variate $(i, j = 1, 2, \dots, p;\ \ h = 1, 2, \dots, k)$, we note that

$$\left.\begin{aligned} YY' = \left[\sum_{h=1}^{k} n_h(s_{ij}). \right] \quad ZZ' &= \left[\sum_{h=1}^{k} m_h(\bar{x}_{ih}-\bar{x}_i)(\bar{x}_{jh}-\bar{x}_j) \right], \\ Z\zeta' &= \left[\sum_{h=1}^{k} m_h(\bar{x}_{ih}-\bar{x}_i)(\xi_{jh}-\xi_j) \right], \\ \zeta Z' = (Z\zeta')' \text{ and } \zeta\zeta' &= \left[\sum_{h=1}^{k} m_h(\xi_{ih}-\xi_i)(\xi_{jh}-\xi_j) \right], \end{aligned}\right\} \qquad \dots \quad (4.22)$$

where all the elements of the right hand side are either defined explicitly in terms of the original set of observations or parameters or are directly calculable in terms of that set. We shall denote $\left[\sum\limits_{h=1}^{k} n_h(s_{ij})_h \right]/n$ by S (to be called the *sample "within" dispersion matrix*), $\left[\sum\limits_{h=1}^{k} m_h(\bar{x}_{ih}-\bar{x}_i)(\bar{x}_{jh}-\bar{x}_j) \right]/(k-1)$ by S^* (to be called the *sample* *"between" dispersion matrix*), $\left[\sum_{h=1}^{k} m_h(\xi_{ih}-\xi_i)(\xi_{jh}-\xi_j)\right]/(k-1)$ by Σ^* (to be called the *population "between" dispersion matrix*) and the vectors $\bar{\mathbf{x}}$ and $\boldsymbol{\xi}$ defined by (4.16) will be called respectively the sample and the population grand mean vector.

For $k = 2$ it is easy to check that $Z(p\times\overline{k-1})$ and $\zeta(p\times\overline{k-1})$ become respectively the column vectors, say, $\mathbf{z}(p\times 1)$ and $\boldsymbol{\zeta}(p\times 1)$ given by

$$\mathbf{z}(p\times 1) = m_{12}^{1/2}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2),\ \boldsymbol{\zeta}(p\times 1) = m_{12}^{1/2}(\boldsymbol{\xi}_1-\boldsymbol{\xi}_2), \qquad \dots (4.23)$$

where
$$m_{12} = \frac{m_1 m_2}{m_1+m_2},$$

and we have for $Y(p\times(n_1+n_2))$ and $(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)$ the probability law :

$$[1/(2\pi)^{\frac{p(m-1)}{2}}|\Sigma|^{\frac{m-1}{2}}] \exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\left\{YY'+m_{12}^{1/2}(\bar{\mathbf{x}}_{1}-\boldsymbol{\xi}_1-\bar{\mathbf{x}}_2+\boldsymbol{\xi}_2)\right.\right.$$
$$\left.\left.\times\ (\bar{\mathbf{x}}_1'-\boldsymbol{\xi}_1'-\bar{\mathbf{x}}_2'+\boldsymbol{\xi}_2')\right\}\right]\times dY\ d[m_{12}^{1/2}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)]. \qquad \dots (4.24)$$

For the simple regression set-up corresponding to a one-way classification we have an $X(p\times n)$ (with $p < n$) with a probability law

$$[1/(2\pi)^{pn/2}|\Sigma|^{n/2}] \exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\{(X-E(X))(X'-E(X'))\}\right] dX, \qquad \dots (4.25)$$

where Σ is symmetric p.d. and

where
$$E(X)(p\times n) = \xi(p\times n) + \mu(p\times q)\ U(q\times n). \qquad \dots (4.26)$$

Here $q < n$ but might be $\geqslant p$ or $< p$. Without any loss of generality it is assumed that $\xi(p\times n) = (\boldsymbol{\xi} \dots \boldsymbol{\xi})p$, i.e., ξ is a matrix of unknown parameters consisting of the same n columns of unknown parameter vector $\boldsymbol{\xi}(p\times 1)$. $\mu(p\times q)$ is also an unknown parameter matrix, while U is a matrix of rank q of the so-called "concomitant" variates, i.e., a set of observations which are supposed to stay constant with the probabilistic set-up of the experiment and the analysis. Again, without any loss of generality, but for simplicity of discussion, we can assume that the row sums of U are zero (for each row), i.e., that $n\bar{\mathbf{u}} = \sum_{j=1}^{n} \mathbf{u}_j = 0$ (where $\mathbf{u}_j$ denotes the j-th column vector of the U matrix. As in (4.1)-(4.13) denote the j-th column vector of the X matrix by $\mathbf{x}_j$ and set $\bar{\mathbf{x}}(p\times 1) = \sum_{j=1}^{n} \mathbf{x}_j(p\times 1)/n$ and use the orthogonal transformation $X_1 = XA$

$= [\underset{1}{\sqrt{n}\,\bar{\mathbf{x}}} : \underset{n-1}{Y}]\,p$, $U_1 = UA = [\underset{1}{\sqrt{n}\,\bar{\mathbf{u}}} : \underset{n-1}{V}]\,q = [\underset{1}{0} : \underset{n-1}{V}]\,q$, $\zeta_1 = \zeta A = [\underset{1}{\sqrt{n}\,\boldsymbol{\xi}} : \underset{n-1}{0}]\,p$,

where
$$A = \begin{bmatrix} 1/\sqrt{n} & a_{12} & \dots & a_{1n} \\ 1/\sqrt{n} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ 1/\sqrt{n} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

is an $\perp$ matrix. Since $J(X : X_1) = J(X : \sqrt{n}\,\bar{\mathbf{x}}, Y) = 1$, we have, as before, the joint distribution of $\bar{\mathbf{x}}$ and Y given by

$$[1/(2\pi)^{pn/2}|\Sigma|^{n/2}] \exp\left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} \left\{n(\bar{\mathbf{x}}-\boldsymbol{\xi})(\bar{\mathbf{x}}'-\boldsymbol{\xi}') + (Y-\mu V)(Y'-V'\mu')\right\}\right]$$
$$d(\sqrt{n}\,\bar{\mathbf{x}})dY. \qquad \dots \quad (4\ 27)$$

Integrating out over $\bar{\mathbf{x}}$ from $-\infty$ to ∞, we have for Y the probability law

$$[1/(2\pi)^{p(n-1)/2}|\Sigma|^{\frac{n-1}{2}}] \exp\left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(Y-\mu V)(Y'-V'\mu')\right] dY. \quad \dots \quad (4.28)$$

Using the results that (i) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ and (ii) for any square matrix A, $\operatorname{tr} A = \operatorname{tr} A'$, and recalling that Σ^{-1} is symmetric, we have

$$\operatorname{tr} \Sigma^{-1} \mu V\, Y' = \operatorname{tr} Y\, V'\mu'\, \Sigma^{-1} = \operatorname{tr} \Sigma^{-1}\, Y\, V'\, \mu'. \qquad \dots \quad (4.29)$$

Substituting in (4.28) we rewrite (4.28) as

$$[1/(2\pi)^{p(n-1)/2}|\Sigma|^{\frac{n-1}{2}}] \exp\left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(YY'-2Y\, V'\mu'+\mu VV'\mu')\right] dY. \quad \dots \quad (4.30)$$

Now notice that $YY' = XX'-n\bar{\mathbf{x}}\,\bar{\mathbf{x}}'$, $VV' = UU'-n\,\bar{\mathbf{u}}\,\bar{\mathbf{u}}' = UU'$ and $YV' = XU'-n\,\bar{\mathbf{x}}\,\bar{\mathbf{u}}' = XU'$(since $\bar{\mathbf{u}} = 0$). At this point let us use (A.3.11) to set

$$V(q\times\overline{n-1}) = \tilde{T}(q\times q)\, L_1(q\times\overline{n-1}), \text{ where } L_1 L_1' = I(q), \qquad \dots \quad (4.31)$$

and use (A.1.7.) to complete L_1 into an $\perp$ martix $\underset{n-1}{\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}} \begin{array}{c} q \\ n-1-q \end{array}$. Next use on (4.30) the transformation

$$Y(p\times\overline{n-1}) = Y_1(p\times\overline{n-1}) \underset{n-1}{\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}} \begin{array}{c} q \\ n-1-q \end{array} \qquad \dots \quad (4.32)$$

or

$$Y_1(p\times\overline{n-1}) = Y(p\times\overline{n-1}) \quad [\underset{q}{L_1'} : \underset{n-1-q}{L_2'}]\ n-1 = [\underset{q}{Z_1} : \underset{n-1-q}{Z_2}]p, \text{ say.}$$

Check that $Y_1 Y_1' = YY' = Z_1 Z_1' + Z_2 Z_2'$, $YV' = YL_1'\tilde{T}' = Z_1\tilde{T}'$, $\tilde{T}\tilde{T}' = VV'$, recall that $J(Y : Z_1, Z_2) = 1$, and obtain for Z_1 and Z_2 the probability law

$$[1/(2\pi)^{p(n-1)/2}|\Sigma|^{\frac{n-1}{2}}] \exp[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(Z_1Z_1' + Z_2Z_2' - 2Z_1\tilde{T}'\mu' + \mu\tilde{T}\tilde{T}'\mu')] \, dZ_1 \, dZ_2. \quad \ldots \quad (4.33)$$

This shows that the joint distribution of $Z_1(p\times q)$ and $Z_2(p\times \overline{n-1-q})$ is exactly of the same form as of Z and Y in (4.21), the m being replaced by n here and the $k-1$ there being replaced by q here. It may be interesting to check again what is implicit in the above, namely that the expression under the exponential is expressible very simply in terms of the original set-up. Verify that $Z_1\tilde{T}' = YV' = XU'$, $\tilde{T}\tilde{T}' = VV' = UU'$, $Z_1Z'_1 = YL_1'L_1Y' = YV\tilde{T}'^{-1}\tilde{T}^{-1}VY'$ (using (4.31)) $= YV'(VV')^{-1}VY' = XU'(UU')^{-1}UX'$ and $Z_2Z'_2 = YY' - YL'_1 L_1Y' = XX' - n\bar{\mathbf{x}}\bar{\mathbf{x}}' - XU'(UU')^{-1}UX'$.

The way to handle more general regression problems which arise from other types of designs will be indicated in Chapter 12.

CHAPTER FIVE

Statement of the Specific Problems to be discussed

The problems will be formulated in terms of testing of hypotheses, and, in each case, the associated problem in terms of simultaneous confidence interval estimation will also be indicated, although the latter will be discussed in full in sections 14.1-14.11. For each hypothesis to be considered here, the associated (set of) simultaneous confidence bounds will be referred to as A.S.C.B. It will be seen later that corresponding to each hypotheses and its class of alternatives (to be presently stated) there is a 'natural, and 'physically meaningful' set of parameters (or rather functions of the primitive population parameters) which can be easily interpreted as measures of deviations from the hypotheses. It will be also seen that the tests of hypotheses going to be offered here are such that, for each test, it is possible to obtain by inversion (and without running into any very difficult distribution problems) a set of simultaneous confidence bounds on these 'deviations' from the hypotheses. In this section, for most (though not for all) of the hypotheses stated, the structure of the corresponding 'deviations, are also stated without any attempt to show just why and how they are 'appropriate' or 'natural' ; this is done later. The following are the problems: (i) For $N(\boldsymbol{\xi}(p \times 1), \Sigma(p \times p))$ (where Σ is symmetric p.d.), to test $H_0 : \Sigma = \Sigma_0$ against $H : \Sigma \neq \Sigma_0$; the associated simultaneous confidence bounds, as will be seen later, will be bounds on characteristic roots of Σ, i.e., on all $c(\Sigma)$, or by using (A.2.5), bounds on $\mathbf{a}'(1 \times p)\, \Sigma(p \times p)\, \mathbf{a}(p \times 1)$ (for all arbitrary vectors $\mathbf{a}'(1 \times p)$ of unit length each); (ii) for $N(\boldsymbol{\xi}_h(p \times 1), \Sigma_h(p \times p))$ ($h = 1, 2,$ and Σ_1 and Σ_2 are both symmetric p.d.), to test $H_0 : \Sigma_1 = \Sigma_2$ against H: $\Sigma_1 \neq \Sigma_2$; the A.S.C.B. will be bounds on all $c(\Sigma_1 \Sigma_2^{-1})$, or using (A.2.6), on $\mathbf{a}'(1 \times p)\Sigma_1(p \times p)\, \mathbf{a}(p \times 1)/\mathbf{a}'(1 \times p)\, \Sigma_2(p \times p)\, \mathbf{a}(p \times 1)$ (for all arbitrary non-null $\mathbf{a}'(1 \times p)$); (iii) for $N(\boldsymbol{\xi}_r(p \times 1), \Sigma(p \times p))$ ($r = 1, \ldots, k$; Σ is symmetric p.d.), to test $H_0 : \boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = \ldots = \boldsymbol{\xi}_k$ against H: not H_0, i.e., violation of at least one equality ; then the A.S.C.B. will be on $\mathbf{a}'(1 \times p)\, \eta(p \times k)\mathbf{b}(k \times 1)$ (for all arbitrary non-null $\mathbf{a}'(1 \times p)$ and arbitrary $\mathbf{b}(k \times 1)$ of unit length), where η stands for the $(p \times k)$ population matrix with k column vectors (each $p \times 1$) $\sqrt{n_1}(\boldsymbol{\xi}_1 - \boldsymbol{\xi})$, $\sqrt{n_2}(\boldsymbol{\xi}_2 - \boldsymbol{\xi})$, ..., $\sqrt{n_k}(\boldsymbol{\xi}_k - \boldsymbol{\xi})$ and $\boldsymbol{\xi} = \sum_{h=1}^{k} n_h \boldsymbol{\xi}_h / \sum_{h=1}^{k} n_h$. Notice that η will be of rank $\leqslant \min(p, k-1)$; (iv) for $N[\boldsymbol{\xi}((p+q) \times 1), \Sigma((p+q) \times (p+q))]$, where Σ is symmetric p.d. of the form $\begin{array}{cc} \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array}$ $(p \leqslant q)$, to test $H_0 : \Sigma_{12}(p \times q) = 0$ against $\Sigma_{12} \neq 0$; the A.S.C.B. will be on $\mathbf{a}'(1 \times p)\Sigma_{12}(p \times q)\Sigma_{22}^{-1}(q \times q)\mathbf{b}(q \times 1)$ (for all arbitrary unit-length vectors $\mathbf{a}'(1 \times p)$ and $\mathbf{b}(q \times 1)$) [6].

A number of useful problems can be formally tied up with problem (iii), of which the more important are the following: (iii*a*) For $N(\boldsymbol{\xi}_h, \Sigma)$ ($h = 1, 2$), to test $H_0 : \boldsymbol{\xi}_1 = \boldsymbol{\xi}_2$ against $H : \boldsymbol{\xi}_1 \neq \boldsymbol{\xi}_2$, the A.S.C.B. being now on $\mathbf{a}'(1 \times p)(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)(p \times 1)$ (for all arbitrary non-null $\mathbf{a}'(1 \times p)$); (iii*b*) for $N(\boldsymbol{\xi}, \Sigma)$, to test $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi}_0$ against H : $\boldsymbol{\xi} \neq \boldsymbol{\xi}_0$, the A.S.C.B. being on $\mathbf{a}'(1 \times p)\boldsymbol{\xi}(p \times 1)$ (for all arbitrary non-null $\mathbf{a}'(1 \times p)$); (iii*c*) given an observation matrix $X(p \times n)$ $(p < n)$ of stochastic variates with independent $p\times 1$ column vectors $\mathbf{x}_h$ $(h = 1, 2, \ldots, n)$ having p.d.f.'s $N(E(\mathbf{x}_h), \Sigma)$, where $E(X')(n\times p) = A(n\times m)\boldsymbol{\xi}(m\times p)$ $(m < n$; $\boldsymbol{\xi}$ is a matrix of unknown population parameters and A is a non-stochastic matrix of rank $r \leqslant m < n$, whose elements are supposed to be given by the particular experimental situation), to test H_0 : $C(q\times m)$ $\boldsymbol{\xi}$ $(m\times p) = 0$, where C is such that H_0 is testable (see (12.7.5)) against H : not H_0; the A.S.C.B. will be given in section 14.6; this H_0 is called the general multivariate linear hypothesis which includes the usual problems of multivariate analysis of variance and covariance as particular cases and also of course the problem (iii) as a very special case; next (iii*d*) for $N(\boldsymbol{\xi}_h, \Sigma)$ $(h = 1, 2, \ldots, k)$, where $\boldsymbol{\xi}_h = \begin{array}{cl} \left[\begin{array}{c} \boldsymbol{\xi}_{1h} \\ \boldsymbol{\xi}_{2h} \end{array}\right] & \begin{array}{l} p \\ q \end{array} \\ 1 & \end{array}$ (say) and Σ is symmetric p.d. of the structure $\begin{array}{cl} \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{array}\right] & \begin{array}{l} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array}$, to test $H_0 : \boldsymbol{\xi}_{1h}(p\times 1) = \Sigma_{12}(p\times q)\ \Sigma_{22}^{-1}(q\times q)\ \boldsymbol{\xi}_{2h}(q\times 1)$ $(h = 1, 2, \ldots, k)$ against H: not H_0; the A.S.C.B. will be given in section (14.7); this H_0 is called the hypothesis of (a particular kind of) multicollinearity of the means; and finally (iii*e*) for the linear regression model of (4.26) to test H_0: $\mu(p\times q) = 0$ (or, say $= \mu_0$) against $H : \mu \neq 0$ (or say $\neq \mu_0$); the A.S.C.B. will be given in section 14.8.

Formally tied up with (iv) is the following : (iv) for $N(\boldsymbol{\xi}, \Sigma)$ where Σ is symmetric p.d. of the form $\Sigma = \begin{array}{cl} \begin{array}{ccc} p & q & r \end{array} & \\ \left[\begin{array}{ccc} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma'_{12} & \Sigma_{22} & \Sigma_{23} \\ \Sigma'_{13} & \Sigma'_{23} & \Sigma_{33} \end{array}\right] & \begin{array}{l} p \\ q \\ r \end{array} \end{array}$ $(p \leqslant q)$, to test $H_0 : \Sigma_{12\cdot 3} = 0$, where $\Sigma_{12.3}(p\times q) = \Sigma_{12}(p\times q) - \Sigma_{13}(p\times r)\ \Sigma_{33}^{-1}(r\times r)\ \Sigma'_{23}(r\times q)$ and where $\Sigma_{22.3}(q\times q) = \Sigma_{22}(q\times q) - \Sigma_{23}(q\times r)\ \ \Sigma_{33}^{-1}(r \times r)\Sigma'_{23}(r\times q)$; the A.S.C.B. will be on $\mathbf{a}'(1\times p) \times\Sigma_{12.3}(p\times q)\ \Sigma_{22.3}^{-1}(q\times q)\ \mathbf{b}(q\times 1)$ (for all arbitrary unit length vectors $\mathbf{a}'(1\times p)$ and $\mathbf{b}(q\times 1)$).

In addition to those considered in the two previous paragraphs there are several other problems whose solutions can be formally thrown back upon those of (i)-(v) and these need not be discussed or even stated separately here. But even within the very restricted set-up (considered in this monograph) of non-sequential, one stage, fixed sample-size, two-decision problems of the classical type there are several problems of great practical and theoretical interest which have had to be excluded, because of the fact that (so far as the author is aware) no suitable and reasonably easy techniques are known at the moment. Among such problems (unfortunately to be omitted) a particularly important one is the following : for $N(\boldsymbol{\xi}_h, \Sigma_h)(h = 1, 2, \ldots, k > 2)$, to test $H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$ against H : not H_0; and of course the A.S.C.B. on 'appropriate deviations' from H_0.

In what follows chapter 6 will give the derivation of the proposed tests for H_0 in the situations (i)-(v) and make the formal identification of (iiia)-(iiie) with (iii) and of (iva) with (iv), chapters 9—11 will give the operating characteristics of the proposed tests, 14 will deal with all the set of simultaneous confidence bounds associated with each test of chapter 6, the operating characteristics of the proposed set of simultaneous confidence bounds in each case being easily available from chapters 9—11.

CHAPTER SIX

Tests for the Null Hypothesis*

6.1. *Direct type I construction not possible.* It is well known [40, 41] that for each composite H_0 above there are infinitely many similar regions but no most powerful (bisimilar) region against any specific composite alternative, i.e., any composite alternative in which the specifiable elements are given special values. Thus direct type I construction will not work here.

6.2. *Reduction to pseudo-univariate and pseudo-bivariate problems.* At this point suppose that, starting from an $\mathbf{x}(p \times 1)$ which is $N(\boldsymbol{\xi}, \Sigma)$ we consider a linear compound $\mathbf{a}'\mathbf{x}$ (with an arbitrary constant, i.e., non-stochastic $\mathbf{a}'(1 \times p)$ of nonzero modulus). This $\mathbf{a}'\mathbf{x}$ is a scalar well known to be $N(\mathbf{a}'\boldsymbol{\xi}, \mathbf{a}'\Sigma\mathbf{a})$. Notice that $\mathbf{a}'\boldsymbol{\xi}$ and $\mathbf{a}'\Sigma\mathbf{a}$ are also scalars. Suppose also that given

$$\mathbf{x} = \begin{array}{c} \left[\begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \end{array}\right] \end{array} \begin{array}{c} p \\ q \end{array} : N \left[\begin{array}{c} \left[\begin{array}{c} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{array}\right] \\ 1 \end{array} \begin{array}{c} p \\ q \end{array}, \begin{array}{c} \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_2 & \Sigma_{22} \end{array}\right] \\ \begin{array}{cc} p & q \end{array} \end{array} \begin{array}{c} p \\ q \end{array} \right] (p \leqslant q),$$

we consider linear compounds $\mathbf{a}'\mathbf{x}_1$, $\mathbf{b}'\mathbf{x}_2$(where $\mathbf{a}$ $(p \times 1)$ and $\mathbf{b}(q \times 1)$ are each non-null and non-stochastic); then these two scalars $\mathbf{a}'\mathbf{x}_1$ and $\mathbf{b}'\mathbf{x}_2$ are well known to be distributed as a bivariate normal with a correlation coefficient

$$\rho(\mathbf{a},\ \mathbf{b}) = \rho_{12} = \mathbf{a}'\Sigma_{12}\mathbf{b}/[(\mathbf{a}'\Sigma_{11}\mathbf{a})^{\frac{1}{2}}(\mathbf{b}'\Sigma_{22}\mathbf{b})^{\frac{1}{2}}]. \quad \ldots \quad (6.2.1)$$

Now suppose that, in place of H_0 of (*i*)-(iv) of chapter 5 we consider respectively (v) $H(\mathbf{a}'\Sigma\mathbf{a} = \mathbf{a}'\Sigma_0\mathbf{a})$ $(= H_{0\mathbf{a}})$ against *all* $H(\mathbf{a}'\Sigma\mathbf{a} \neq \mathbf{a}'\Sigma_0\mathbf{a})$ $(= H_{\mathbf{a}})$, ($\mathbf{a}$ fixed), (vi) $H(\mathbf{a}'\Sigma_1\mathbf{a} = \mathbf{a}'\Sigma_2\mathbf{a})$ $(= H_{0\mathbf{a}})$ against *all* $H(\mathbf{a}'\Sigma_1\mathbf{a} \neq \mathbf{a}'\Sigma_2\mathbf{a})$ $(= H_{\mathbf{a}})$, ($\mathbf{a}$ fixed), (vii) $H(\mathbf{a}'\boldsymbol{\xi}_1 = \mathbf{a}'\boldsymbol{\xi}_2 = \ldots = \mathbf{a}'\boldsymbol{\xi}_k)$ $(= H_{0\mathbf{a}})$ against *all* $H_{\mathbf{a}}$ $(\neq H_{0\mathbf{a}})$, ($\mathbf{a}$ fixed), (viii) $H(\mathbf{a}'\ \Sigma_{12}\mathbf{b} = 0)$ $(= H_{0\mathbf{ab}})$ against all $H(\mathbf{a}'\ \Sigma_{12}\mathbf{b} \neq 0)$ $(= H_{\mathbf{ab}})$, ($\mathbf{a}$, $\mathbf{b}$ fixed).

We now consider the *totality of all* non-null $\mathbf{a}$ for (v)-(vii) and *all* non-null $\mathbf{a}$ and $\mathbf{b}$ for (viii). Notice that (*a*) $\bigcap_{\mathbf{a}} H(\mathbf{a}'\Sigma\mathbf{a} = \mathbf{a}'\Sigma_0\mathbf{a}) = H(\Sigma = \Sigma_0)$, (b)$\bigcap_{\mathbf{a}} H(\mathbf{a}'\Sigma_1\mathbf{a} = \mathbf{a}'\Sigma_2\ \mathbf{a}) = H(\Sigma_1 = \Sigma_2)$, (c) $\bigcap_{\mathbf{a}} H\ (\mathbf{a}'\boldsymbol{\xi}_1 = \mathbf{a}'\boldsymbol{\xi}_2 = \ldots = \mathbf{a}'\boldsymbol{\xi}_k) = H(\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = \ldots = \boldsymbol{\xi}_k)$ and (d)$\bigcap_{\mathbf{a},\mathbf{b}} H(\mathbf{a}'\Sigma_{12}\mathbf{b} = 0) = H(\Sigma_{12} = 0)$. We could have worked in terms of any subset of $\mathbf{a}$'s which led by intersection to the same H_0, but this we do not do here. It may be noted that, by the procedure to be used here, apart from set-theoretic difficulties which, however, do not arise in these applications, the total set of $\mathbf{a}$'s or any subset of it (of the kind considered) will *uniquely* define an *extended type I test associated with the total set or with that particular subset.* Next suppose that, in the alternative, under (v)-(viii), we substitute "*specific*" for "*all*" and thus have four new situations (ix)-(xii). It is well known that for each of the situations (ix)-(xii, we have one most powerful (bisimilar) region, so that from these we can construct respective (*modified in a sense to be explained in section* 6.3) *type I tests* for the pseudo-univariate

---

* See reference [43] in this connection.

situations (v) and (vi), *straight type I tests* for the pseudo-univariate situation (vii) and the pseudo-bivariate situation (viii). From these *modified type I* and straight *type I tests* we can try to construct the respective *extended type I tests* for the situations (i)-(iv). This ties up (see section 4) the p-variate problems (i)-(iii) with the pseudo-univariate problems (v)-(vii), the $(p+q)$-variate problem (iv) with the pseudo-bivariate problem (viii).

6.3. *Modified type I tests.* We now take over the notation and symbols from section 4.

(v) Starting from (4.13), put $\chi^2_{\mathrm{a}} = n\mathrm{a}'S\mathrm{a}/\mathrm{a}'\Sigma_0\mathrm{a}$ and notice that, at a level β_2, for $H(\mathrm{a}'\Sigma\mathrm{a} = \mathrm{a}'\Sigma_0\mathrm{a})(= H_{0\mathrm{a}})$ against all $H(\mathrm{a}'\Sigma\mathrm{a} > \mathrm{a}'\Sigma_0\mathrm{a})$ we have the one-sided uniformly most powerful (bisimilar) region: $\chi^2_{\mathrm{a}} \geqslant \chi^2_{\beta_2}(n)$, and, at a level β_1, for $H_{0\mathrm{a}}$ against all $H(\mathrm{a}'\Sigma\mathrm{a}<\mathrm{a}'\Sigma_0\mathrm{a})$ we have the one-sided uniformly most powerful (bisimilar) region : $\chi^2_{\mathrm{a}} \leqslant \chi^{2\prime}_{\beta}(n)$, where $\chi^{2\prime}_{\beta_1}(n)$ and $\chi^2_{\beta_2}(n)$ are the upper β_2 and lower β_1 points of the χ^2-distribution with d.f.n. Notice that χ^2_{a} has the central χ^2-distribution with d.f.n. Now consider the union $[\chi^2_{\mathrm{a}} \geqslant \chi^2_{\beta_2}(n)] \bigcup [\chi^2_{\mathrm{a}} \leqslant \chi^{2\prime}_{\beta_1}(n)] = \bigcup(\mathrm{a})$, say, which, if we decide to call it a new critical region, will be one of size $\beta_1+\beta_2 = \beta$ (say). Notice that given β, we can regard β_1 and β_2 now as flexible, subject to $\beta_1+\beta_2 = \beta$. At this point, we can so choose β_1 and β_2, i.e., the tail ends $\chi^{2\prime}_{\beta_1}$ and $\chi^{2\prime}_{\beta_2}$ as to make $\bigcup(\mathrm{a})$ a locally unbiassed (here it will turn out to be also locally most powerful) critical region (in the neighbourhood of $H_{0\mathrm{a}'}$). It will be seen that the condition of unbiassedness imposes a relation between $\chi^2_{\beta_1}(n)$ and $\chi^2_{\beta_2}(n)$ which involves only n but is independent of the total size of the region β. We now call these tail ends $\chi^2_{1\beta}(p, n)$ and $\chi^2_{2\beta}(p,n)$. We now recall from (1.2.2) that this $\bigcup(\mathrm{a})$ is also a uniformly unbiassed region (having, in fact, the stronger property of monotonicity) and is also admissible. With this choice of $\chi^2_{1\beta}(n)$ and $\chi^2_{2\beta}(n)$ we have now for $H(\mathrm{a}'\Sigma\mathrm{a} = \mathrm{a}'\Sigma_0\mathrm{a})$ against all $H(\mathrm{a}'\Sigma\mathrm{a} \neq \mathrm{a}'\Sigma_0\mathrm{a})$ a modified type I critical region of size β.

$$\chi^2_{\mathrm{a}} = n\mathrm{a}'S\mathrm{a}/\mathrm{a}'\Sigma_0\,\mathrm{a} \geqslant \chi^2_{2\beta}(n) \text{ or } \leqslant \chi^2_{1\beta}(n), \qquad \ldots \quad (6.3.1)$$

which is uniformly unbiassed, monotonic and is also admissible.

(vi) Starting from the product of two distributions like (4.13), put $F_{\mathrm{a}} = \mathrm{a}'S_1\mathrm{a}/\mathrm{a}'S_2\mathrm{a}$ and notice. as in the previous case, that, at a level β_2, for $H(\mathrm{a}'\Sigma_1\mathrm{a}=\mathrm{a}'\Sigma_2\mathrm{a})$ $(= H_{0\mathrm{a}})$ against all $H(\mathrm{a}'\,\Sigma_1\mathrm{a} > \mathrm{a}'\,\Sigma_2\,\mathrm{a})$ we have the one-sided uniformly most powerful (bisimilar) region : $F_{\mathrm{a}} \geqslant F_{\beta_2}(n_1, n_2)$, and, for $H_{0\mathrm{a}}$ against all $H(\mathrm{a}'\Sigma_1\mathrm{a} < \mathrm{a}'\,\Sigma_2\mathrm{a})$, the one-sided uniformly most powerful region : $F_{\mathrm{a}} \leqslant F'_{\beta_1}(n_1, n_2)$, where $F'_{\beta_1}(n_1, n_2)$ and $F_{\beta_2}(n_1, n_2)$ are the lower β_1 and upper β_2 points of the F-distribution with d.f. n_1 and n_2. Notice that F_{a} has the ordinary F-distribution with d.f. n_1 and n_2. Take the union of the two regions and as in the previous case call it a new critical region, say $\bigcup(\mathrm{a})$ of size $\beta_1+\beta_2 = \beta$(say), and given β, pick out the tails $F'_{\beta_1}(n_1, n_2)$ and $F_{\beta_2}(n_1, n_2)$ so as to make $\bigcup(\mathrm{a})$ a locally unbiassed region (in the neighbourhood of $H_{0\alpha}$), notice that this imposes an extra relation between $F'_{\beta_1}(n_1, n_2)$ and $F_{\beta_2}\,(n_1,\, n_2)$ which involves only n_1, n_2 and not the total size of the region β. Recall also from (1.2.2) that this is a uniformly unbiassed region (also having the monotonicity property) and also admissible. As before, with this choice of F'_{β_1} and F_{β_2} to be called $F_{1\beta}(n_1, n_2)$ and $F_{2\beta}(n_1, n_2)$ we have now for $H(\mathbf{a}'\Sigma_1\mathbf{a} = \mathbf{a}'\Sigma_2\,\mathbf{a})$ against all $H(\mathbf{a}'\,\Sigma_1\mathbf{a} \neq \mathbf{a}'\,\Sigma_2\mathbf{a})$ a modified type I critical region of size β (uniformly unbiassed, monotonic and admissible)

$$F_{\mathbf{a}} = \mathbf{a}'S_1\mathbf{a}/\mathbf{a}'S_2\mathbf{a} \geqslant F_{2\beta}(n_1, n_2) \text{ or } \leqslant F_{1\beta}(n_1, n_2). \qquad \dots \quad (6.3.2)$$

(vii) Start from (4.16)-(4.22) and recall from (ii) of section 2.3 that for $H(\mathbf{a}'\boldsymbol{\xi}_1 = \mathbf{a}'\boldsymbol{\xi}_2 = \dots = \mathbf{a}'\boldsymbol{\xi}_k)$ $(= H_{0\mathbf{a}})$ against any *specific* $H_{\mathbf{a}}(\neq H_{0\mathbf{a}})$, there is the most powerful (bisimilar) critical region (of size, say γ) which is a one-sided t—region, and by taking the union of these regions (for fixed $\mathbf{a}$ but by variation over $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_k$), we have the straight type I region of size, say β, given by (notice that $F_{\mathbf{a}}$ has the ordinary F-distribution with d.f. n_1 and n_2)

$$F_{\mathbf{a}} = \mathbf{a}'S^*\mathbf{a}/\mathbf{a}'S\mathbf{a} \geqslant F_\beta(n_1, n_2), \qquad \dots \quad (6.3.3)$$

where $F_\beta(n_1, n_2) = F_\beta$ (say) is obtained from $P(F_{\mathbf{a}} \geqslant F_\beta \,|\, H_{0\mathbf{a}}) = \beta$.

This is well known to be a type II or likelihood ratio region as well and is also well known to have a number of desirable properties (including uniform unbiassedness, the stronger property of monotonicity and also admissibility).

(viii) Start from (4.15) and put

$$\begin{array}{c} \left[\begin{array}{cc} S_{11} & S_{12} \\ S'_{12} & S_{22} \end{array}\right]\begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} \end{array} = \frac{1}{n}\left[\begin{array}{c} Y_1 \\ Y_2 \end{array}\right] \begin{array}{c}[Y'_1 : Y'_2]n. \\ \begin{array}{cc} p & q \end{array}\end{array} \qquad \dots \quad (6.3.4)$$

Next put

$$r_{\mathbf{ab}} = \mathbf{a}'S_{12}\mathbf{b}/(\mathbf{a}'S_{11}\mathbf{a})^{\frac{1}{2}}(\mathbf{b}'S_{22}\mathbf{b})^{\frac{1}{2}}, \qquad \dots \quad (6.3.5)$$

and notice that, at a level β, for $H(\mathbf{a}'\,\Sigma_{12}\mathbf{b} = 0)$ $(=H_{0\mathbf{ab}})$ against *all* $H(\mathbf{a}\Sigma_{12}\mathbf{b} > 0)$ we have the one-sided uniformly most powerful (bisimilar) region : $r_{\mathbf{ab}} \geqslant r_\beta(n-1)$ and for $H_{0\mathbf{ab}}$ against all $H(\mathbf{a}\Sigma_{12}\mathbf{b} < 0)$ the one-sided uniformly most powerful (bisimilar) region : $r_{\mathbf{ab}} \leqslant -r_\beta(n-1)$, where $r_\beta(n-1)$ $(= r_\beta$, say) is given by

$$P(r_{\mathbf{ab}} \geqslant r_\beta \,|\, H_{0\mathbf{ab}}) = \beta. \qquad \dots \quad (6.3.6)$$

Notice also that this $r_{\mathbf{ab}}$ has the distribution of the central correlation coefficient with d.f. $(n-1)$. Taking the union of the two regions we shall have a straight type I critical region of size 2β given by

$$[r_{\mathbf{ab}} \geqslant r_\beta(n-1)] \cup [r_{\mathbf{ab}} \leqslant -r_\beta(n-1)]. \qquad \dots \quad (6.3.7)$$

This is well known to be a type II or likelihood ratio region as well and it is also well known that this has a number of desirable properties (including uniform unbiassedness, the stronger property of monotonicity and also admissibility).

6.4. *Actual construction of extended type I regions.*

(i) By the test procedure (7.3.1) over $\chi^2_{1\beta}(n) \leqslant \chi^2_{\mathbf{a}} \leqslant \chi^2_{2\beta}(n)$ we accept $H(\mathbf{a}'\Sigma\mathbf{a} = \mathbf{a}'\Sigma_0\mathbf{a})$, so that, by using the heuristic principle of section 2.6 over $\bigcap_{\mathbf{a}}[\chi^2_{1\beta}(n) \leqslant \chi^2_{\mathbf{a}} \leqslant \chi^2_{2\beta}(n)]$ we accept $\bigcap_{\mathbf{a}} H(\mathbf{a}'\Sigma\mathbf{a} = \mathbf{a}'\Sigma_0\mathbf{a}) = H(\Sigma = \Sigma_0) = H_0$, and thus over its complement $\bigcup_{\mathbf{a}}[\chi^2_{\mathbf{a}} > \chi^2_{2\beta}(n)$ or $< \chi^2_{1\beta}(n)]$ we reject H_0. This may be set up as the extended type I test. To obtain $\bigcap_{\mathbf{a}}[\chi^2_{1\beta} \leqslant \chi^2_{\mathbf{a}} \leqslant \chi^2_{2\beta}(n)]$ we note that a particular S would belong to the intersection if , for that S, $\chi^2_{1\beta} \leqslant \mathbf{a}'S\mathbf{a}/\mathbf{a}'\Sigma_0\mathbf{a} \leqslant \chi^2_{2\beta}$ for all non-null $\mathbf{a}$. This statement $\Longleftrightarrow \chi^2_{1\beta} \leqslant$ smallest $\mathbf{a}'S\mathbf{a}/\mathbf{a}'\Sigma_0\mathbf{a} \leqslant$ largest $\mathbf{a}'S\mathbf{a}/\mathbf{a}'\Sigma_0\mathbf{a} \leqslant \chi^2_{2\beta}$, the "largest" and "smallest" being under variation of $\mathbf{a}$ (for given S). Now, given S, and of course Σ_0, the largest and smallest values of $\mathbf{a}'S\mathbf{a}/\mathbf{a}'\Sigma_0\mathbf{a}$ are easily seen from (A.2.5) to be the largest and smallest roots, say c_1 and c_p of the p-th degree equation in c:

$$|S-c\Sigma_0| = 0, \qquad \dots \quad (6.4.1)$$

all the p roots $c_1, c_2, \dots, c_p$ being in this situation, a.e. positive, since Σ_0 is given to be symmetric p.d. and S is, by definition and the assumptions, a.e., p.d. Starting out from the (modified) type I test (6.3.1) for H_0 we have for H_0, i.e., $H(\Sigma = \Sigma_0)$ the extended type I critical region

$$c_p \geqslant \chi^2_{2\beta}(n) \text{ and/or } c_1 \leqslant \chi^2_{1\beta}(n). \qquad \dots \quad (6.4.2)$$

To find $\chi^2_{2\beta}$ and $\chi^2_{1\beta}$ we make use of the condition of local unbiassedness (which involves only n) (see (v) of 6.3) and also 11.5) and write down the further condition (which now completely determines $\chi^2_{2\beta}$ and $\chi^2_{1\beta}$)

$$P(\chi^2_{1\beta} \leqslant c_1 \leqslant c_p \leqslant \chi^2_{2\beta}\,|\,H_0) = 1-\alpha. \qquad \dots \quad (6.4.3)$$

Notice from (A.7.1.1) that under H_0 the distribution of $c_1, \dots, c_p$ and thus also of c_1 and c_p turn out to be independent of Σ, depending only on p and n thus the c.d.f. (6.4.3) depends only on α, p and n, so that it will now be proper to write the tail ends as $c_{1\alpha}(p, n)$ and $c_{2\alpha}(p, n)$.

(ii) The general nature of the arguments will be exactly the same as in the previous case. Starting from (6.3.2), over $F_{1\beta} \leqslant F_{\mathbf{a}} \leqslant F_{2\beta}$ we accept $H(\mathbf{a}'\Sigma_1\mathbf{a} = \mathbf{a}'\Sigma_2\mathbf{a})$, so that, by using the principle of section 2.6 over $\bigcap_{\mathbf{a}}\left[F_{1\beta} \leqslant \dfrac{\mathbf{a}'S_1\mathbf{a}}{\mathbf{a}'S_2\mathbf{a}} \leqslant F_{2\beta}\right]$ we accept $\bigcap_{\mathbf{a}} H(\mathbf{a}'\Sigma_1\mathbf{a} = \mathbf{a}'\Sigma_2\mathbf{a}) = H(\Sigma_1 = \Sigma_2) = H_0$, and thus over its complement $\bigcup_{\mathbf{a}}\left[\dfrac{\mathbf{a}'S_1\mathbf{a}}{\mathbf{a}'S_2\mathbf{a}} > F_{2\beta} < \text{or } F_{1\beta}\right]$ we reject H_0. As before we set it up as the extended type I test and, using (A.2.6), notice that the statement $F_{1\beta} \leqslant \mathbf{a}'S_1\mathbf{a}/\mathbf{a}'S_2\mathbf{a} \leqslant F_{2\beta}$ (where S_1, S_2 and $F_{1\beta}$ and $F_{2\beta}$ are held fixed and $\mathbf{a}$ alone is varied) $\Longleftrightarrow F_{1\beta} \leqslant c_1 \leqslant c_p \leqslant F_{2\beta}$, where c_1 and c_p are the smallest and largest roots of the p-th degree equation in c:

$$|S_1-cS_2| = 0, \qquad \dots \quad (6.4.4)$$

all the p roots being here, a.e., positive, since S_1 and S_2 are by the conditions of the problem, a.e., p.d. Starting out from the (modified) type I test (6.3.2) for H_0 we have thus for $H(\Sigma_1 = \Sigma_2)$ the extended type I region

$$c_p > F_{2\beta}(n_1, n_2) \text{ and /or } c_1 < F_{1\beta}(n_1, n_2). \qquad \dots \quad (6.4.5)$$

As in the previous case, given α, to determine $F_{2\beta}$ and $F_{1\beta}$ we first take over (see (vi) of 6.3 and also 11.1) the relation (involving only n_1 and n_2) between $F_{2\beta}$ and $F_{1\beta}$ imposed by the condition of local unbiasedness and write down the further condition (which now completely determines $F_{2\beta}$ and $F_{1\beta}$)

$$P(F_{1\beta} \leqslant c_1 \leqslant c_p \leqslant F_{2\beta} | H_0) = 1-\alpha. \qquad \dots \quad (6.4.6)$$

Notice from (A.7.2) that under H_0 the distribution of $c_1, \dots, c_p$ and thus also of c_1 and c_p happen to be independent of the common value of Σ_1 and Σ_2 and also of $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, depending only on p, n_1 and n_2 and thus the c.d.f. (6.4.6) depends only on α, p, n_1 and n_2, so that the tail ends $F_{2\beta}$ and $F_{1\beta}$ can be more appropriately written as $c_{1\alpha}(p, n_1, n_2)$ and $c_{2\alpha}$ (p, n_1, n_2). The actual distribution problem on which depends the evaluation of the left side of (6.4.6) is solved in section (A.9.7).

(iii) By the test procedure (6.3.3), over $F_{\mathbf{a}} = \mathbf{a}'S^*\mathbf{a}/\mathbf{a}'S\mathbf{a} \leqslant F_\beta(n_1, n_2)$ we accept $H(\mathbf{a}'\boldsymbol{\xi}_1 = \dots = \mathbf{a}'\boldsymbol{\xi}_k)$ $(= H_{0\mathbf{a}})$, so that using the principle of section 2,6 over $\bigcap_{\mathbf{a}} \left[F_{\mathbf{a}} = \dfrac{\mathbf{a}'S^*\mathbf{a}}{\mathbf{a}'S\mathbf{a}} \leqslant F_\beta\right]$ we accept $\bigcap_{\mathbf{a}} H(\mathbf{a}'\boldsymbol{\xi}_1 = \dots = \mathbf{a}'\boldsymbol{\xi}_k) = H(\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = \dots = \boldsymbol{\xi}_k)$ $= H_0$ and over its complement $\bigcup_{\mathbf{a}}[F_{\mathbf{a}} > F_\beta]$ we reject H_0. We set it up as the extended type I test and, using (A.2.6), notice that the statement $\mathbf{a}'S^*\mathbf{a}/\mathbf{a}'S\mathbf{a} \leqslant F_\beta$ (where S^* and S and F_β are held fixed and $\mathbf{a}$ alone varied) $\Longleftrightarrow c_r \leqslant F_\beta$ where c_r is the largest root of the p-th degree equation in c

$$|S^* - cS| = 0. \qquad \dots \quad (6.4.7)$$

From the definitions and assumptions of section 6.2 and chapter 4 it is easy to check that S is, a.e., p.d. while S^* is, a.e., at least p.s.d. of rank $r = \min(p, k-1)$. It will of course be, a.e., p.d. if $p \leqslant k-1$. In any case, we can say that, out of the p roots of (6.4.7), $p-r$ will be always zero, while r roots, to be called $c_1 \leqslant c_2 \leqslant \dots \leqslant c_r$ will be, a.e., positive where $r = \min(p, k-1)$. Starting out from the straight type I test (6.3.3) for $H_{0\mathbf{a}}$ we have thus for $H(\Sigma^* = 0)$ the extended type I region:

$$c_r \geqslant F_\beta(n_1, n_2), \qquad \dots \quad (6.4.8)$$

where, given the size α of (7.4.8), F is to be determined by

$$P(c_r \geqslant F_\beta | H_0) = \alpha. \qquad \dots \quad (6.4.9)$$

Notice from (A.7.5) that under H_0 the distribution of $c_1, \dots, c_r$ and thus also of c_r happen to depend only on p, n_1 and n_2, i.e., on p, $k-1$, $n-k$ (where n is the total number of observations and k the total number of samples or populations), being independent of all other nuisance parameters. Also the c.d.f. (6.4.9) depends only on α, p, $k-1$, $n-k$. Thus the tail end F_β can now be more appropriately written as

$c_\alpha(p\ , k-1, n-k)$. The actual distribution problem on which the evaluation of the left side of (6.4.9) depends is solved in section 7.6 and chapter 8.

(iv) By the test procedure (6.3.7), over $r^2_{\mathbf{ab}} = \dfrac{(\mathbf{a}'S_{12}\mathbf{b})^2}{(\mathbf{a}'S_{11}\mathbf{a})(\mathbf{b}'S_{22}\mathbf{b})} \leqslant r^2_\beta\,(n-1)$ we accept $H(\mathbf{a}'\Sigma_{12}\mathbf{b} = 0)$, so that, using the principle of section 2.6 over $\bigcap_{\mathbf{ab}}[r^2_{\mathbf{ab}} \leqslant r^2_\beta(n-1)]$ we accept $\bigcap_{\mathbf{ab}}(\mathbf{a}'\ \Sigma_{12}\mathbf{b} = 0) = H(\Sigma_{12} = 0) = H_0$, and over its complement $U_{\mathbf{ab}}\ [r^2_{\mathbf{ab}} > r^2_\beta\ (n-1)]$ we reject H_0. As before, we set this up as an extended type I test and, using (A.2.3), notice tha tthe statement $(\mathbf{a}'S_{12}\mathbf{b})^2/(\mathbf{a}'S_{11}\mathbf{a})(\mathbf{b}'S_{22}\mathbf{b}) \leqslant r^2_\beta(n-1)$ (where S_{11}, S_{12}, S_{22} and r_β are held fixed and $\mathbf{a}$, $\mathbf{b}$ alone varied) $\Longleftrightarrow c_p \leqslant r^2_\beta\ (n-1)$, where c_p is the largest root of the p-th degree equation in c:

$$|cS_{11}-S_{12}S_{22}^{-1}S'_{12}| = 0. \qquad \dots \quad (6.4.10)$$

From the definitions and assumptions of chapter 4 and section 6.2 it is easy to see that, a.e., S_{11} is p.d. and so also $S_{12}S_{22}^{-1}S'_{12}$, so that, a.e., all roots will be positive. Under these conditions it is known (from (A.1.16)) that the p roots will all, a.e., lie between 0 and 1, satisfying the condition, $0 < c_1 < c_2 < \dots < c_p < 1$. Starting out from the straight type I test (6.3.7) for $H_{0\mathbf{ab}}$ we have thus for $H(\Sigma_{12} = 0)$ the extended type I region:

$$c_p \geqslant r^2_\beta(n-1), \qquad \dots \quad (6.4.11)$$

where, given that α is the size of (6.4.11), r_β is to be determined by

$$P(c_p \geqslant r^2_\beta\,|\,H_0) = \alpha. \qquad \dots \quad (6.4.12)$$

Notice from (A.7.3) that under H_0 the distribution of $c_1, \dots, c_p$ and thus also of c_p happen to depend only on p, q and n, being independent of all other nuisance parameters. Thus the c.d.f. (6.4.12) also depends only on α, p, q, n and hence the tail end r^2_β can be now more appropriately written as $c_\alpha(p, q, n)$. The actual distribution on which the evaluation of the left side of (6.4.12) depends is solved in section 7.4 and chapter 8.

CHAPTER SEVEN

Reduction of Some Distribution Problems and Some Actual Distributions*

7.1. *Distribution of rectangular co-ordinates.* As in (A.8.6), put $X(p\times n) = \tilde{T}(p\times p)L(p\times n)$, subject to $LL' = I(p)$, observe that $\tilde{T}$ and L_I have the distribution

$$2^p[1/(2\pi)^{\frac{pn}{2}}|\Sigma|^{\frac{n}{2}}]\exp[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\tilde{T}\tilde{T}']\prod_{i=1}^{p}t_{ii}^{n-i}\,d\tilde{T}\;\;dL_I\Big/\left|\frac{\partial(LL')}{\partial(L_D)}\right|_{L_I}. \quad \ldots \quad (7.1.1)$$

Now, using (A.8.6.3) to integrate out over L_I, we have the following distribution for $\tilde{T}$ [31]:

$$\left[1/(2)^{\frac{pn}{2}-p}\pi^{\frac{p(p-1)}{4}}|\Sigma|^{\frac{n}{2}}\prod_{i=1}^{p}\Gamma\left(\frac{n-i+1}{2}\right)\right]\exp[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\tilde{T}\tilde{T}']\prod_{i=1}^{p}t_{ii}^{n-i}d\tilde{T}. \ldots \quad (7.1.2)$$

From (7.1.2), by using (A.6.1.12) and the fact that $|nS| = |\tilde{T}|^2 = \prod_{i=1}^{p} t_{ii}^2$ we have the following distribution for S (usually known as the Wishart distribution):

$$\left[\left(\frac{n}{2}\right)^{\frac{np}{2}}/\pi^{\frac{p(p-1)}{4}}|\Sigma|^{\frac{n}{2}}\prod_{i=1}^{p}\Gamma\left(\frac{n-i+1}{2}\right)\right]\exp[-\tfrac{1}{2}\operatorname{tr}n\Sigma^{-1}S]\,|S|^{\frac{n-p-1}{2}}dS. \quad \ldots \quad (7.1.3)$$

7.2: *Distribution of characteristic roots of the sample dispersion matrix S.* Using the results of (A.7.1) we start, without any loss of generality, from the canonical form (A.7.1.1), use (A.3.6) to set $X(p\times n) = M(p\times p)\times D_{\sqrt{c}}(p\times p)\;L(p\times n)$ where $LL' = I(p)$ and M is $\perp$, with a positive first row, take over from (A.6.3.1) the Jacobian $J(X : M_I, c\text{'s}, L_I)$ and have for M_I, c's and L_I the distribution:

$$2^p\left[1/(2\pi)^{\frac{pn}{2}}\prod_{i=1}^{p}\gamma_i^{n/2}\right]\exp[-\tfrac{1}{2}\operatorname{tr}D_{1/\gamma}MD_cM']\prod_{i=1}^{p}c_i^{\frac{n-p-1}{2}}\;dc_i$$

$$\times\bmod\left[\prod_{i<j=1}^{p-1}(c_i-c_j)\right]\frac{dM_I}{\left|\dfrac{\partial(MM')}{\partial(M_D)}\right|_{M_I}}\frac{dL_I}{\left|\dfrac{\partial(LL')}{\partial(L_D)}\right|_{L_I}} \quad \ldots \quad (7.2.1)$$

Using (A.8.6.3) to integrate out over L_I, we have for c and M_I the distribution

$$2^pF(p,n)\left[1/(2\pi)^{\frac{pn}{2}}\prod_{i=1}^{p}\gamma_i^{n/2}\right]\exp[-\tfrac{1}{2}\operatorname{tr}D_{1/\gamma}MD_cM']$$

$$\times\prod_{i=1}^{p}c_i^{\frac{n-p-1}{2}}dc_i\;\bmod\left[\prod_{i<j=1}^{p-1}(c_i-c_j)\right]dM_I\Big/\left|\frac{\partial(MM')}{\partial(M_D)}\right|_{M_I}. \quad \ldots \quad (7.3.2)$$

---

* See references [24, 31, 32, 54] in this connection.

This is the point to which the reduction of the distribution problem for the general case can be conveniently carried out. If, however all $\gamma_i =$ a constant $= 1$ (without any loss of generality), then $\operatorname{tr} D_{1/\gamma}MD_cM' = \operatorname{tr} MD_cM' = \operatorname{tr} D_c = \sum_{i=1}^{p} c_i$, and now remembering that the first row of M is positive and using (A.8.6.4)—(A.8.6.8) to integrate over M_I, we have for c's the distribution on the null hypothesis all γ_i's $=$ const. $= 1$:

$$\left[1/(2)^{\frac{pn}{2}}\pi^{\frac{pn}{2}-p}\right]F(p,n)\ F(p,p)\exp\left[-\tfrac{1}{2}\sum_{i=1}^{p}c_i\right]\prod_{i=1}^{p}c_i^{\frac{n-p-1}{2}}dc_i \bmod \left[\prod_{i<j=1}^{p-1}(c_i-c_j)\right] \quad \ldots \quad (7.2.3)$$

where $F(p, n)$ and $F(p, p)$ are given by (A.8.6.3) [31].

7.3. *Distribution of characteristic roots of* $S_1S_2^{-1}$. As in (7.2.2), using the results of (A.7.2), we start without any loss of generality, from the canonical form (A.7.2.1), use (A.3.8) to set $X_1(p\times n_1) = A(p\times p)D_{\sqrt{c}}(p\times p)L_1(p\times n)$ and $X_2(p\times n_2) = A(p\times p)\ L_2(p\times n_2)$, where A is non-singular and $L_1L_1' = L_2L_2' = I(p)$, take over from (A.6.2.11) the Jacobian $J(X_1, X_2 : A, c\text{'s}, L_{1I}, L_{2I})$ and obtain for A, c's, L_{1I} and L_{2I} the distribution

$$2^p\left[1/(2\pi)^{\frac{p(n_1+n_2)}{2}}\prod_{i=1}^{p}\gamma_i^{n_1/2}\right]\exp\left[-\tfrac{1}{2}\operatorname{tr}(D_{1/\gamma}AD_cA'+AA')\right]|A|^{n_1+n_2-p}\,dA$$

$$\times\prod_{i=1}^{p}c_i^{(n_1-p-1)/2}dc_i \bmod \left[\prod_{i<j=1}^{p-1}(c_i-c_j)\right]\left[dL_{1I}\Big/\left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}}\right]$$

$$\times\left[dL_{2I}\Big/\left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}\right] \quad \ldots \quad (7.3.1)$$

Using (A.8.6.3) to integrate out over L_{1I} and L_{2I} we have for A and c's the distribution

$$2^p\left[1/(2\pi)^{\frac{p(n_1+n_2)}{2}}\prod_{i=1}^{p}\gamma_i^{n_1/2}\right]F(p,n_1)\ F(p,n_2)\exp\left[-\tfrac{1}{2}\operatorname{tr}\{D_{1/\gamma}AD_cA'+AA'\}\right]$$

$$\times|A|^{n_1+n_2-p}dA\prod_{i=1}^{p}c_i^{(n_1-p-1)/2}dc_i \bmod \left[\prod_{i<j=1}^{p-1}(c_i-c_j)\right] \quad \ldots \quad (7.3.2)$$

As before, this is the point to which, for the general case, the reduction of the distribution problem can be conveniently carried out. If, however, all γ's, i.e., all

$c(\Sigma_1 \Sigma_2^{-1}) = 1$ (which by (A.1.13), happens if and only if $\Sigma_1 = \Sigma_2$), then further reduction is possible and (7.3.2) reduces to

$$2^p(1/2\pi)^{\frac{p(n_1+n_2)}{2}} F(p, n_1)F(p, n_2) \exp\left[-\tfrac{1}{2}\operatorname{tr}(AD_{1+c}A')\right]|A|^{n_1+n_2-p}dA \times \text{(factors}$$

involving the c's taken over from (7.3.2)). ... (7.3.3)

Now putting $AD_{\sqrt{1+c}} = B$ and using (A.5.2) we have for B and c's the distribution

$$2^p[1/2\pi]^{\frac{p(n_1+n_2)}{2}} F(p, n_1)\, F(p, n_2) \exp\left[-\tfrac{1}{2}\operatorname{tr} BB'\right]|B|^{n_1+n_2-p}\, dB$$

$$\times \prod_{i=1}^{p} c_i^{(n_1-p-1)/2}\, dc_i/(1+c_i)^{\frac{n_1+n_2}{2}} \bmod \left[\prod_{i<j=1}^{p-1} (c_i-c_j)\right]. \quad \text{... (7.3.4)}$$

Now using (A.8.7.1) and remembering that

$$\underset{\textit{over } B \textit{ with a positive first row}}{\int} \exp\left[-\tfrac{1}{2}\operatorname{tr} BB'\right]|B|^q\, dB$$

$$= \frac{1}{2^p} \underset{\textit{over } B \textit{ (unrestricted)}}{\int} \exp\left[-\tfrac{1}{2}\operatorname{tr} BB'\right]|B|^q\, dB, \quad \text{... (7.3.5)}$$

we have for c's the distribution (on the null hypothesis $\Sigma_1 = \Sigma_2$):

$$\left[\pi^{p/2} \prod_{i=1}^{p} \Gamma\left(\frac{n_1+n_2-i+1}{2}\right) \Big/ \prod_{i=1}^{p} \Gamma\left(\frac{n_1-i+1}{2}\right) \Gamma\left(\frac{n_2-i+1}{2}\right) \Gamma\left(\frac{p-i+1}{2}\right)\right]$$

$$\times \prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}}\, dc_i/(1+c_i)^{\frac{n_1+n_2}{2}} \bmod \left[\prod_{i<j=1}^{p-1} (c_i-c_j)\right]. \quad \text{... (7.3.6)}$$

7.4. *Distribution of canonical correlations.* As in the two previous cases, start, without any loss of generality, from the canonical distribution form (A.7.3.5) and let $c_i(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma'_{12}) = \gamma_i$ (say) $(i = 1, 2, \ldots, p)$. Next, use (A.3.17) to set $X_2\ (q \times n) = \tilde{T}\,(q \times q)\ L_2(q \times n)$ and $X_1(p \times n) = U(p \times p)$ $\times\ [M_1(p \times \overline{n-q}) \vdots D_{\sqrt{c}}\,(p \times p) M_2(p \times q)] \underset{n}{\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}}\begin{matrix} n-q \\ q \end{matrix}$, where $\tilde{T}, U, M_1, M_2, L_1$ and L_2 are characterized in (A.3.17) and $e_i = (1-c_i)/c_i$. Now take over from (A.6.5.1) the Jacobian $J(X_1, X_2 : \tilde{T}, U, c\text{'s}, M_{1I}, M_{2I}, L_{2I})$ and obtain for $\tilde{T}$, U, c's M_{1I}, M_{2I} and L_{2I} the distribution

$$2^{p+q}[1/(2\pi)^{\frac{n(p+q)}{2}} \prod_{i=1}^{p} (1-\gamma_i)^{\frac{n}{2}}] \exp\left\{ -\tfrac{1}{2} \operatorname{tr} \left[\begin{array}{c|c} D_{1/1-\gamma} & -[D_{\sqrt{\gamma}/(1-\gamma)} \mid 0] \\ -\begin{bmatrix} D_{\sqrt{\gamma}/(1-\gamma)} \\ 0 \end{bmatrix} & \left[\begin{array}{c|c} D_{1/1-\gamma} & 0 \\ 0 & I(q-p) \end{array}\right] \end{array} \right] \right.$$

$$\left. \times \begin{bmatrix} UD_{1+e}U' & UM_2\tilde{T} \\ \tilde{T}'M'_2U' & \tilde{T}\tilde{T}' \end{bmatrix} \right\} |U|^{n-p}dU \prod_{i=1}^{q} t_{ii}^{n-i} \; d\tilde{T} \prod_{i=1}^{p} e_i^{\frac{n-p-q-1}{2}} de_i \bmod \left[\prod_{i<j=1}^{p-1} (e_i - e_j) \right]$$

$$\times \left[dM_{1I} \Big/ \left| \frac{\partial(M_1M'_1)}{\partial(M_{1D})} \right|_{M_{1I}} \right] \left[dM_{2I} \Big/ \left| \frac{\partial(M_2M'_2)}{\partial(M_{2D})} \right|_{M_{2I}} \right] \left[dL_{2I} \Big/ \left| \frac{\partial(L_2L'_2)}{\partial(L_{2D})} \right|_{L_{2I}} \right] . \quad (7.4.1)$$

Using (A.8.6.3) to integrate out over M_{1I} and L_{2I}, we have for $\tilde{T}$, U, M_{2I} and e's the distribution

$$2^{p+q} \; [1/(2\pi)^{\frac{n(p+q)}{2}} \prod_{i=1}^{p} (1-\gamma_i)^{\frac{n}{2}}] \; F(p, n-q) \; F(q, n)$$

$$\times \exp\left\{ -\tfrac{1}{2} \operatorname{tr} \left[\begin{array}{c|c} D_{1/1-\gamma} & -[D_{\sqrt{\gamma}/1-\gamma} \mid 0] \\ -\begin{bmatrix} D_{\sqrt{\gamma}/1-\gamma} \\ 0 \end{bmatrix} & \left[\begin{array}{c|c} D_{1/1-\gamma} & 0 \\ 0 & I(q-p) \end{array}\right] \end{array} \right] \begin{bmatrix} UD_{1+e}U' & UM_2\tilde{T} \\ \tilde{T}'M'_2U' & \tilde{T}\tilde{T}' \end{bmatrix} \right\}$$

$$\times |U|^{n-p}dU \prod_{i=1}^{q} t_{ii}^{n-i} \; d\tilde{T} \prod_{i=1}^{p} e_i^{\frac{n-p-q-1}{2}} de_i \bmod [\prod_{i<j=1}^{p-1} (e_i - e_j)] \, dM_{2I} \Big/ \left| \frac{\partial(M_2M'_2)}{\partial(M_{2D})} \right|_{M_{2I}} . \quad (7.4.2)$$

This is the point to which, in the general case, the reduction of the distribution problem can be conveniently carried out. However, if all $c(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma'_{12}) = 0$, i.e., all γ_i's $= 0$, (which according to (A.1.17) happens if and only if $\Sigma_{12} = 0$), then further reduction is possible and (7.4.2) reduces to

$$2^{p+q}[1/2\pi]^{\frac{n(p+q)}{2}} \; F(p, n-q) \; F(q, n) \; \exp[-\tfrac{1}{2}(\operatorname{tr} UD_{1+e}U' + \operatorname{tr} \tilde{T}\tilde{T}')] \, |U|^{n-p}dU$$

$$\times \prod_{i=1}^{q} t_{ii}^{n-i} \; d\tilde{T} \prod_{i=1}^{p} e_i^{\frac{n-p-q-1}{2}} de_i \bmod \left[\prod_{i<j=1}^{p-1} (e_i - e_j) \right] dM_{2I} \Big/ \left| \frac{\partial(M_2M'_2)}{\partial(M_{2D})} \right|_{M_{2I}} \quad \ldots \quad (7.4.3)$$

Note that $$\operatorname{tr} \tilde{T}\tilde{T}' = \sum_{i \geqslant j=1}^{q} t_{ij}^2$$

and $$\int_{\tilde{T}} \exp\left[-\tfrac{1}{2}\operatorname{tr} \tilde{T}\tilde{T}'\right] \prod_{i=1}^{q} t_{ii}^{n-i} \prod_{i \geqslant j=1}^{q} dt_{ij} = 2^{-q+qn/2}\, \pi^{q(q-1)/4} \prod_{i=1}^{q} \Gamma\left(\frac{n-i+1}{2}\right)$$

and hence integrating out over $\tilde{T}$ and M_{2I}, obtain for U and e's the distribution

$$2^p\left[1/2^{\frac{np}{2}} \pi^{\frac{n(p+q)}{2} - \frac{q(q-1)}{4}}\right] \prod_{i=1}^{q} \Gamma\left(\frac{n-i+1}{2}\right) F(p, n-q)\ F(p, q)\ F(q, n)$$

$$\times \exp\left[-\tfrac{1}{2}\operatorname{tr} UD_{1+e}U'\right]\ |U|^{n-p} dU \prod_{i=1}^{p} e_i^{\frac{n-p-q-1}{2}} de_i \bmod\left[\prod_{i<j=1}^{p-1}(e_i - e_j)\right]. \qquad \ldots \quad (7.4.4)$$

Now put $UD_{\sqrt{1+e}} = V$, use (A.5.2) and (A.8.7) to integrate out over V and obtain for e's the distribution

$$\text{Const}\left[\prod_{i=1}^{p} e_i^{(n-p-q-1)/2} de_i/(1+e_i)^{n/2}\right] \bmod\left[\prod_{i<j=1}^{p-1}(e_i - e_j)\right]. \qquad \ldots \quad (7.4.5)$$

Putting $e_i = (1-c_i)/c_i$ we have for c_i's, on the null hypothesis $\Sigma_{12} = 0$, i.e., $c(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}') = 0$, i.e., $\gamma_i = 0$, the distribution,

$$\text{Const} \prod_{i=1}^{p} (1-c_i)^{\frac{n-p-q-1}{2}} c_i^{\frac{q-p-1}{2}} dc_i \bmod\left[\prod_{i<j=1}^{p-1}(c_i - c_j)\right],$$

where the

$$\text{Const} = \pi^{p/2} \prod_{i=1}^{p} \Gamma\left(\frac{n-i+1}{2}\right) \Big/ \prod_{i=1}^{p} \Gamma\left(\frac{n-q-i+1}{2}\right) \Gamma\left(\frac{q-i+1}{2}\right) \Gamma\left(\frac{p-i+1}{2}\right) \qquad \ldots \quad (7.4.6)$$

An important special case is when $p = 1$ and this we shall consider both on the null and on the non-null hypothesis. In this case there is only one non-zero (and here positive) e or c and only one posible non-zero (and here positive) γ. Thus $\bmod\left[\prod_{i<\gamma=1}^{p-1}(e_i - e_j)\right]$ or $\bmod\left[\prod_{i<\gamma=1}^{p-1}(c_i - c_j)\right]$ will drop out. On the null hypothesis $\gamma = 0$, the distribution of e and c, as special cases of (7.4.5) and (7.4.6), will respectively be

$$\frac{\Gamma(n/2)}{\Gamma(q/2)\Gamma(\overline{n-q}/2)}\left[e^{(n-q-2)/2} de/(1+e)^{n/2}\right], \qquad \ldots \quad (7.4.7)$$

$$\frac{\Gamma(n/2)}{\Gamma(q/2)\Gamma(\overline{n-q}/2)}\left[(1-c)^{(n-q-2)/2} c^{(q-2)/2} dc\right]. \qquad \ldots \quad (7.4.8)$$

For the distribution on the non-null hypothesis $\gamma \neq 0$, we start from (7.4.2), put $U(1\times 1)=u$ (a scalar), so that $UD_{1+e}U'=(1+e)u^2, M_2(1\times q)=\mathbf{m}_2'(1\times q)=(m_1, m_2, \ldots, m_q)$ (say), and take $m_2, \ldots, m_q$ to be the so-called independent elements of $\mathbf{m}_2'$, so that

$$dM_{2I}\Big/\left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right|_{M_{2I}} = \prod_{i=2}^{q} dm_i\Big/2\Big(1-\sum_{i=2}^{q} m_i^2\Big)^{\frac{1}{2}},$$

and obtain for e the distribution

$$2^q\left[1/(2\pi)^{\frac{n(q+1)}{2}}(1-\gamma)^{\frac{n}{2}}\right]F(1, n-q)F(q, n)\,e^{\frac{n-q-2}{2}}\,de$$

$$\times \int_{u,\,\tilde{T},\,\mathbf{m}_{2I}'} \exp\left[-\frac{1}{2(1-\gamma)}\{(1+e)u^2-2\sqrt{\gamma}ut_{11}m_1+t_{11}^2\}-\tfrac{1}{2}\sum_{i=2}^{q}\sum_{j=1}^{i} t_{ij}^2\right]$$

$$\times\, u^{n-1}du \prod_{i=1}^{q} t_{ii}^{n-i}d\tilde{T} \prod_{i=2}^{q} dm_i\Big/\Big(1-\sum_{i=2}^{q} m_i^2\Big)^{\frac{1}{2}}. \qquad \ldots \quad (7.4.9)$$

Now putting $m_1 = \cos\theta$ so that $\sum_{i=2}^{q} m_i^2 = \sin^2\theta$, we note from (A.8.4) that

$$\int_{\sin\theta \leqslant \left(\sum_{i=2}^{q} m_i^2\right)^{\frac{1}{2}} \leqslant \sin\theta+d(\sin\theta)} \exp \pm\left[\frac{\sqrt{\gamma}}{1-\gamma}\,ut_{11}\Big(1-\sum_{i=2}^{q} m_i^2\Big)^{\frac{1}{2}}\right] \prod_{i=2}^{q} dm_i\Big/\Big(1-\sum_{i=2}^{q} m_i^2\Big)^{\frac{1}{2}}$$

$$= \left[(q-1)\pi^{\frac{q-1}{2}}\Big/\Gamma\Big(\frac{q+1}{2}\Big)\right]\exp\left[\pm\frac{\sqrt{\gamma}}{1-\gamma}\,ut_{11}\cos\theta\right](\sin\theta)^{q-2}d\theta. \qquad \ldots \quad (7.4.10)$$

Using (7.4.10) and also integrating out over t_{ij}'s ($j = 1, 2, \ldots, i$ and $i = 2, 3, \ldots, p$), and setting $v^2 = (1+e)u^2$, we have (7.4.9) reducing to

$$\text{Const (which is easily obtained) } e^{\frac{n-q-2}{2}}\,de/(1+e)^{\frac{n}{2}}$$

$$\times \int_{v,\,t_{11},\,\theta} \exp\left[-\frac{1}{2(1-\gamma)}(v^2+t_{11}^2)\right]\cosh\left[\frac{\sqrt{\gamma}}{1-\gamma}\,vt_{11}\cos\theta/\sqrt{1+e}\right](vt_{11})^{n-1}dv\,dt_{11}(\sin\theta)^{q-2}d\theta,$$

$$\ldots \quad (7.4.11)$$

the limits of v and t_{11} being from 0 to ∞ and of θ from 0 to π. To evaluate the integral

$$\int_{\theta=0}^{\pi}\int_{x,\,y=0}^{\infty} \exp\left[-a(x^2+y^2-2bxy\cos\theta)\right](xy)^n dx\,dy\,(\sin\theta)^m d\theta$$

we proceed as follows

$$\int_0^{\pi} \exp\,[2abxy\cos\theta]\,(\sin\theta)^m d\theta = 2\int_0^{\pi/2} \cosh\,(2abxy\cos\theta)\sin^m\theta\,d\theta$$

$$= \Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{1}{2}\right)(abxy)^{-\frac{m}{2}} I_{m/2}(2abxy), \qquad \ldots \quad (7.4.12)$$

where I stands for the Bessel I function in the usual notation [52] (Watson's Bessel functions, p. 79, formula (9)). Thus we have

$$\text{Integral} = \Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{1}{2}\right)(ab)^{-\frac{m}{2}} \int_{x,\,y=0}^{\infty} \exp[-a(x^2+y^2)](xy)^{n-\frac{m}{2}} I_{m/2}(2abxy)dxdy. \qquad \ldots \quad (7.4.13)$$

To evaluate this put $xy = z$ and $x/y = e^v$, so that $J(x, y : z, v) = 1/2$ and the range of z and v could be taken as : $0 \leqslant z < \infty$ and $-\infty < v < \infty$. Thus we have, from the symmetry of the integrand,

$$\text{Integral} = \Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{m+1}{2}\right)(ab)^{-\frac{m}{2}} \int_{z=0}^{\infty}\int_{v=0}^{\infty} \exp\,[-2az\cosh v]\,z^{n-\frac{m}{2}} I_{m/2}(2abz)dz\,dv. \qquad \ldots \quad (7.4.14)$$

But putting $\nu = 0$ in formula (9), p. 181, Watson's Bessel functions, and noting that K stands for the Bessel K-function in the usual notation we have

$$\int_{v=0}^{\infty} \exp\,[-2az\cosh v]dv = K_0(2az). \qquad \ldots \quad (7.4.15)$$

$$\text{Hence}\quad \text{Integral} = \Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{1}{2}\right)(ab)^{-\frac{m}{2}} \int_0^{\infty} I_{m/2}(2abz)K_0(2az)\,z^{n-\frac{m}{2}}\,dz. \qquad \ldots \quad (7.4.16)$$

Now putting $\mu = 0$, $\nu = m/2$ and $\lambda = n-m/2$ in (1) of (13.45), p. 410, Watson's Bessel functions and checking up on the validity conditions indicated there, we have

$$\int_0^{\infty} K_0(2az)J_{m/2}(2acz)z^{n-\frac{m}{2}}dz = \left[(2ac)^{\frac{m}{2}}\Gamma^2\left(\frac{n+1}{2}\right)2^{n-2-\frac{m}{2}}\Big/(2a)^{n+1}\,\Gamma\left(\frac{m+2}{2}\right)\right]$$

$$\times\ {}_2F_1\left(\frac{n+1}{2}, \frac{n+1}{2}; \frac{m+2}{2}; -c^2\right) \qquad \ldots \quad (7.4.17)$$

If now we put $c = ib$, we shall have $J_{m/2}(2aibz) = (i)^{\frac{n}{2}} I_{m/2}(2abz)$, so that substituting in (7.4.16) we should have

$$\text{Integral} = \left[2^{-2}\Gamma^2\left(\frac{n+1}{2}\right)\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{1}{2}\right)\Big/ a^{n+1}\Gamma\left(\frac{m+2}{2}\right)\right]$$

$$\times\, {}_2F_1\left(\frac{n+1}{2}, \frac{n+1}{2}; \frac{m+2}{2}; b^2\right). \qquad \ldots \quad (7.4.18)$$

Substituting in (7.4.11) we have for e the distribution proportional to

$${}_2F_1\left(\frac{n}{2}, \frac{n}{2}; \frac{q}{2}; \frac{\gamma}{1+e}\right) e^{\frac{n-q-2}{2}} de/(1+e)^{\frac{n}{2}}. \qquad \ldots \quad (7.4.19)$$

Now putting $c = 1/(1+e)$, we have for c the distribution,

$$\frac{\Gamma\left(\frac{n}{2}\right)(1-\gamma)^{n/2}}{\Gamma\left(\frac{n-q}{2}\right)\Gamma\left(\frac{q}{2}\right)}\, {}_2F_1\left(\frac{n}{2}, \frac{n}{2}; \frac{q}{2}, \gamma c\right)(1-c)^{\frac{n-q-2}{2}} c^{\frac{q-2}{2}} dc. \qquad \ldots \quad (7.4.20)$$

7.5. *Distribution of partial canonical correlations.* Notice that (A.7.4.5) can be rewritten as

$$\left[1/(2\pi)^{\frac{n(p+q+r)}{2}} \prod_{i=1}^{p}(1-\gamma_i)^{\frac{n}{2}}\right] \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\begin{array}{c:c} I(p) & 0 \\ \begin{bmatrix} -D_{\sqrt{\gamma/1-\gamma}} \\ 0 \end{bmatrix} & \left[\begin{array}{c:c} D_{\sqrt{1/1-\gamma}} & 0 \\ 0 & I(q-p) \end{array}\right] \end{array}\right]\right.$$

$$\left.\times \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} (X_1' \quad X_2') \left[\begin{array}{c:c} I(p) & -[D_{\sqrt{\gamma/1-\gamma}} \;\vdots\; 0] \\ 0 & \left[\begin{array}{c:c} D_{\sqrt{1/1-\gamma}} & 0 \\ 0 & I(q-p) \end{array}\right] \end{array}\right] - \tfrac{1}{2}\operatorname{tr} X_3X_3' \right\} dX_1dX_2dX_3.$$

$$\ldots \quad (7.5.1)$$

Now using (A.3.19) make the transformation $X_3(r\times n) = \tilde{T}(r \times r) L_3(r \times n)$ subject to $L_3 L_3' = I(r)$ and $\begin{array}{cc} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} & \begin{array}{c} p \\ q \end{array} \\ n & \end{array} = \begin{array}{cc} \begin{array}{c} p \\ q \end{array} & \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \\ & n-r \quad r \end{array} \begin{array}{cc} \begin{bmatrix} L \\ L_3 \end{bmatrix} & \begin{array}{c} n-r \\ r \end{array} \\ n & \end{array}$ where L is a completion of L_3 to make $\begin{bmatrix} L \\ L_3 \end{bmatrix} \perp$. We have, by (A.6.6),

$$J(X : Z, \tilde{T}, L_{3I}) = 2^r \prod_{i=1}^{r} t_{ii}^{n-i} \Big/ \left|\frac{\partial(L_3L_3')}{\partial(L_{3D})}\right|_{L_{3I}}$$

We have also $X_3X_3' = \tilde{T}\tilde{T}'$. Hence from (7.5.1) we have for Z, $\tilde{T}$, L_{3I}, the distribution

$$2^r\left[1/(2\pi)^{\frac{n(p+q+r)}{2}} \prod_{i=1}^{p} (1-\gamma_i)^{\frac{n}{2}}\right] \exp\left\{-\tfrac{1}{2}\,\mathrm{tr}\left[\begin{array}{ccc} D_{1/1-\gamma} & \vdots & -[D_{\sqrt{\gamma}/1-\gamma} \;\vdots\; 0] \\ -\begin{bmatrix} D_{\sqrt{\gamma}/1-\gamma} \\ 0 \end{bmatrix} & \vdots & \begin{bmatrix} D_{1/1-\gamma} & \vdots & 0 \\ 0 & \vdots & I \end{bmatrix} \end{array}\right]\right.$$

$$\times \begin{bmatrix} Z_{11} \\ Z_{21} \end{bmatrix} [Z_{11}' \;\vdots\; Z_{21}'] - \tfrac{1}{2}\,\mathrm{tr}\left[\begin{array}{ccc} D_{1/1-\gamma} & \vdots & -[D_{\sqrt{\gamma}/1-\gamma} \;\vdots\; 0] \\ -\begin{bmatrix} D_{\sqrt{\gamma}/1-\gamma} \\ 0 \end{bmatrix} & \vdots & \begin{bmatrix} D_{1/1-\gamma} & \vdots & 0 \\ 0 & \vdots & I \end{bmatrix} \end{array}\right] \begin{bmatrix} Z_{12} \\ Z_{22} \end{bmatrix} [Z_{12}' \;\vdots\; Z_{22}']$$

$$\left. - \tfrac{1}{2}\,\mathrm{tr}\, \tilde{T}\tilde{T}' \right\} dZ_{11} \quad dZ_{21} \quad dZ_{12} \quad dZ_{22} \prod_{i=1}^{r} t_{ii}^{n-i} d\tilde{T} dL_{3I} \Big/ \left|\frac{\partial(L_3L_3')}{\partial(L_{3D})}\right|_{L_{3I}}. \qquad \ldots \quad (7.5.2)$$

It is thus seen that (Z_{11}, Z_{21}), (Z_{12}, Z_{22}) and $(\tilde{T}, L_{3I})$ are distributed as three independent sets. Therefore integrating out over $(\tilde{T}, L_{3I})$ and (Z_{12}, Z_{22}) and noting from (A.3.19) that the c's of this section are exactly the same as $c[(Z_{11}\ Z_{11}')^{-1}(Z_{11}Z_{21}') \times (Z_{21}\ Z_{21}')^{-1}(Z_{21}\ Z_{11}')]$, it becomes evident that both on the null and on the non-null hypothesis the distribution of these c's are exactly the same as the c's in section (7.4) with n of that section being replaced here by $n-r$.

7.6. *Distribution of characteristic roots connected with the multivariate analysis of variance.* Without any loss of generality we start from the canonical form (A.7.5.6) and consider two cases separately, namely where (i) $p \leqslant n_1$ and (ii) $p > n_1$ involving respectively, a.e., p non-zero and n_1 non-zero c's.

(i) For case (i) use (A.3.8) to set $X_1(p\times n_1) = A(p\times p)\, D_{\sqrt{c}}(p\times p)\, L_1(p\times n_1)$ and $X_2(p\times n_2) = A(p\times p) L_2(p\times n_2)$, where A, L_1, L_2 satisfy the conditions of (A.3.8), take over from (A.6.2.11) the Jacobian $J(X_1, X_2 : A, c\text{'s}, L_{1I}, L_{2I})$ and obtain for A, c's and L_{1I} and L_{2I} the distribution

$$2^p[1/2\pi]^{\frac{p(n_1+n_2)}{2}} \exp\left[-\tfrac{1}{2}\left\{\mathrm{tr}\, AD_{1+c}A' + \sum_{i=1}^{s} \gamma_i - 2\sum_{i=1}^{s} (AD_{\sqrt{c}}L_1)_{ii}\gamma_i^{\frac{1}{2}}\right\}\right]$$

$$\times\ |A|^{n_1+n_2-p} dA \prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}} dc_i \bmod \left[\prod_{i<j=1}^{p-1} (c_i-c_j)\right] dL_{1I} \Big/ \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}} dL_{2I} \Big/ \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}$$

$$\ldots \quad (7.6.1)$$

Use (A.8.6.3) to integrate out over L_{2I} and obtain for A, c's and L_{1I} the distribution

$$2^p[1/2\pi]^{\frac{p(n_1+n_2)}{2}} F(p, n_2) \exp\left[-\tfrac{1}{2}\left\{\text{tr}\, AD_{1+c}A'+\sum_{i=1}^{s}\gamma_i-2\sum_{i=1}^{s}(AD_{\sqrt{c}}L_1)_{ii}\gamma_i^{\frac{1}{2}}\right\}\right]$$

$$\times |A|^{n_1+n_2-p}dA\prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}} dc_i\left[\text{mod}\prod_{i<j=1}^{p-1}(c_i-c_j)\right]\frac{dL_{1I}}{\left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}}}. \quad \ldots \quad (7.6.2)$$

This is the point to which, for the general case, the distribution problem can be conveniently reduced. If γ_i's $= 0$, i.e., all $c(\Sigma_1\Sigma_2^{-1})$'s $= 0$ (which by (A.1.13), happens if and only if $\Sigma_1 = 0$, i.e., $\xi = 0$), then further reduction is possible and using (A.8.6.3) to integrate out over L_{1I} we have for A and c's the distribution

$$2^p[1/2\pi]^{\frac{p(n_1+n_2)}{2}} F(p, n_1)\ F(p, n_2) \exp\left[-\tfrac{1}{2}\,\text{tr}\, AD_{1+c}A'\right]|A|^{n_1+n_2-p}dA$$

$$\times \prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}} dc_i \;\text{mod}\left[\prod_{i<j=1}^{p-1}(c_i-c_j)\right]. \quad \ldots \quad (7.6.3)$$

Now as in (7.3), integrating out over A we have for c_i's, i.e. for $c(X_1\, X_1'\, (X_2\, X_2')^{-1})$'s the following distribution on the null hypothesis $\xi = 0$:

$$\pi^{p/2}\prod_{i=1}^{p}\Gamma\left(\frac{n_1+n_2-i+1}{2}\right)\Big/\prod_{i=1}^{p}\Gamma\left(\frac{n_1-i+1}{2}\right)\Gamma\left(\frac{n_2-i+1}{2}\right)\Gamma\left(\frac{p-i+1}{2}\right)$$

$$\times\left[\prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}} dc_i/(1+c_i)^{\frac{n_1+n_2}{2}}\right] \text{mod}\left[\prod_{i<j=1}^{p-1}(c_i-c_j)\right], \quad \ldots \quad (7.6.4)$$

which is exactly the same form as (7.3.6).

(ii) For case (ii) use (A.3.14) to set $X_1(p\times n_1) = \begin{array}{cc} p-n_1 & \left[\begin{array}{c} U_1 \\ U_2 \end{array}\right] \\ n_1 & \\ & n_1 \end{array} D_{\sqrt{c}}\,(n_1\times n_1)$ $\times L_1(n_1\times n_1)$ and $X_2(p\times n_2) = U(p\times p)L_2(p\times n_2)$, where U, L_1, L_2 and c satisfy the conditions of (A.3.14), take over from (A.6.7.8) the Jacobian $J(X_1, X_2{:}c\text{'s}, U, L_{1I}, L_{2I})$, and obtain for C, U, L_{1I}, L_{2I} the distribution

$$2^p[1/2\pi]^{\frac{p(n_1+n_2)}{2}} \exp\left[-\tfrac{1}{2}\left\{\operatorname{tr} UD_{1+c}(p)U' + \sum_{i=1}^{s}\gamma_i - 2\sum_{i=1}^{s}\left[\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} D_{\sqrt{c}} L_1\right]_{ii} \gamma_i^{\frac{1}{2}}\right\}\right]$$

$$\times |U|^{n_1+n_2-p} \prod_{i=1}^{p-n_1} (\tilde{U}_3)_{ii}^{p-n_1-i} dU \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} dc_i \bmod \left[\prod_{i<j=1}^{n_1-1} (c_i-c_j)\right]$$

$$\times \left[dL_{1I} \Big/ \left|\frac{\partial(L_1 L_1')}{\partial(L_{1D})}\right|_{L_{1I}}\right] \left[dL_{2I} \Big/ \left|\frac{\partial(L_2 L_2')}{\partial(L_{2D})}\right|_{L_{2I}}\right], \quad \ldots \quad (7.6.5)$$

where $D_{1+c}(p) = \begin{array}{c} \begin{bmatrix} D_c & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} n_1 \\ p-n_1 \end{array} \\ \begin{array}{cc} n_1 & p-n_1 \end{array} \end{array} + I(p).$

Using (A.8.6.3) to integrate out over L_{2I}, we have for c's, U, L_{1I} the distribution

$$2^p[1/2\pi]^{\frac{p(n_1+n_2)}{2}} F(p, n_2) \exp\left[-\tfrac{1}{2}\left\{\operatorname{tr} UD_{1+c}(p)U' + \sum_{i=1}^{s}\gamma_i - 2\sum_{i=1}^{s}\left[\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} D_{\sqrt{c}} L_1\right]_{ii} \gamma_i^{\frac{1}{2}}\right\}\right]$$

$$\times |U|^{n_1+n_2-p} dU \prod_{i=1}^{p-n_1} (\tilde{U}_3)_{ii}^{p-n_1-i} \prod_{i=1}^{n_1} c_i^{p-n_1-1} dc_i \bmod \left[\prod_{i<j=1}^{n_1-1} (c_i-c_j)\right] dL_{1I} \div \left|\frac{\partial(L_1 L_1')}{\partial(L_{1D})}\right|_{L_{1I}}. \quad \ldots \quad (7.6.6)$$

As in case (i) we shall stop here so far as the non-central distribution is concerned. For the central case, i.e., for the null hypothesis that γ_i's$=0$, i.e., all $c(\Sigma_1\Sigma_2^{-1})$'s $= 0$ (which happens if and only if $\xi = 0$) further reduction is possible as in case (*i*) and, using (A.8.6.3) to integrate out over L_{1I}, we have for U and c's the distribution

$$2^p[1/2\pi]^{\frac{p(n_1+n_2)}{2}} F(p, n_1)\, F(p, n_2) \exp(-\tfrac{1}{2}\operatorname{tr} UD_{1+c}U')\, |U|^{n_1+n_2-p} dU$$

$$\times \prod_{i=1}^{p-n_1} (\tilde{U}_3)_{ii}^{p-n_1-i} \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} dc_i \bmod \left[\prod_{i<i=1}^{n_1-1} (c_i-c_j)\right]. \quad \ldots \quad (7.6.7)$$

Now putting $UD_{\sqrt{1+c}} = V$ and using (A.8.8.2) we integrate out over V and obtain for c's the distribution,

$$\pi^{n_1/2}\Big[\prod_{i=1}^{p} \Gamma\Big(\frac{n_1+n_2+1-i}{2}\Big)\prod_{i=1}^{p-n_1} \Gamma\Big(\frac{p-n_1+1-i}{2}\Big)\Big/ \prod_{i=1}^{p} \Gamma\Big(\frac{n_2+1-i}{2}\Big)\, \Gamma\Big(\frac{p-i+1}{2}\Big)$$

$$\times \prod_{i=1}^{n_1} \Gamma\Big(\frac{n_1+1-i}{2}\Big)\Big]\Big[\prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} dc_i/(1+c_i)^{\frac{n_1+n_2}{2}}\Big] \bmod \Big[\prod_{i<j=1}^{n_1-1} (c_i-c_j)\Big]. \quad \ldots \quad (7.6.8)$$

Another way to handle the distribution problem in this case is not to use the transformation (A.3.14) and its Jacobian, but to use the transformation (A.3.15) and its Jacobian given by (A.6.4). This gives us for $\tilde{T}$, c's, L_I, L_{1I} and L_{2I} the distribution

$$\text{Const}\,\exp\Big[-\tfrac{1}{2}\Big\{\operatorname{tr}(\tilde{T}\tilde{T}'+\tilde{T}L_1'D_cL_1\tilde{T}')-2\sum_{i=1}^{n_1} \gamma_i^{\frac{1}{2}} \sum_{j=1}^{i} t_{ij}\, c_i^{\frac{1}{2}}\, l_{1,ij}\Big\}\Big]$$

$$\times \prod_{i=1}^{p} t_{ii}^{n_1+n_2-i}\, d\tilde{T} \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j)$$

$$\times \frac{dL_I}{\left|\frac{\partial(LL')}{\partial(L_D)}\right|_{L_I}} \quad \frac{dL_{1I}}{\left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}}} \quad \frac{dL_{2I}}{\left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}} \quad \ldots \quad (7.6.8.1)$$

It should be remembered that $L(n_1 \times n_1)$ is $\perp$ while $L_1(n_1 \times p)$ and $L_2(p \times n_2)$ (with $n_1 < p \leqslant n_2$) are semi-orthogonal. $\tilde{T}$ and D_c are of course $p \times p$ and $n_1 \times n_1$ respectively. Integrating over L_I and L_{2I} by using (A.8.6), and absorbing in the constant, we have for c's the distribution,

$$\text{Const} \quad \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}}\, dc_i \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j)$$

$$\times \int_{\tilde{T}} \int_{L_{1I}} \exp\Big[-\tfrac{1}{2}\Big\{\operatorname{tr}(\tilde{T}\tilde{T}'+\tilde{T}L_1'D_cL_1\tilde{T}')-2\sum_{i=1}^{n_1} \gamma_i^{\frac{1}{2}} \sum_{j=1}^{i} t_{ij}\, c_i^{\frac{1}{2}}\, l_{1,ij}\Big\}\Big]$$

$$\times \prod_{i=1}^{p} t_{ii}^{n_1+n_2-i} \quad d\tilde{T} \quad \frac{dL_{1I}}{\left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}}} \quad \ldots \quad (7.6.8.2)$$

Since $L_1(n_1 \times p)$ is semiorthogonal, let us adjoin to it $M_1(\overline{p-n_1} \times p)$ so as to make

$$\begin{array}{c} \left[\begin{array}{c} L_1 \\ \\ M_1 \end{array}\right]\begin{array}{l} n_1 \\ \\ p-n_1 \end{array} \\ p \end{array}, \text{ i.e., } \begin{array}{cc} [L_1' & M_1']p \\ n_1 & p-n_1 \end{array}$$

orthogonal.

Now taking into account (A.8.6.11) it is easy to check by putting $[L_1' \;\; M_1'] = M(p\times p)$ that the integral occurring in (7.6.8.2) can by replaced by

$$\prod_{i=r+1}^{p} \left(\frac{n-i+1}{2}\right) \Big| \; \pi^{\frac{n(p-r)}{2}-\frac{(p-r)(p+r-1)}{4}}$$

$$\times \int_{\tilde{T}} \int_{M_I} \exp\left[-\tfrac{1}{2}\left(\operatorname{tr} \tilde{T}M\, D_{1+c}M'\tilde{T}'-2\sum_{i=1}^{n_1} \gamma_i^{\frac{1}{2}} \sum_{j=1}^{i} t_{ij}c_i^{\frac{1}{2}} m_{ji}\right)\right]$$

$$\times \prod_{i=1}^{p} t_{ii}^{n_1+n_2-i}\, d\,\tilde{T} \frac{dM_I}{\left|\dfrac{\partial(MM')}{\partial(M_D)}\right|_{M_I}} \qquad \ldots \quad (7.6.8.3)$$

where D_{1+c} stands for a $p\times p$ diagonal matrix with diagonal elements $1+c_1$, $1+c_2, \ldots,$ $1+c_{n_1}, 1, \ldots, 1$, and m_{ij} are the elements of M. Using first the transformation $\tilde{T}M=U$, then $U\,D_{\sqrt{1+c}} = V$, we observe that (7.6.8.3) now reduces to

$$\text{Const} \int_V \exp\left[-\tfrac{1}{2}\left(\operatorname{tr} \; VV'-2\sum_{i=1}^{n_1} \gamma_i^{\frac{1}{2}}(VD_{c^{\frac{1}{2}}/(1+c)^{\frac{1}{2}}})_{ii}\right)\right] |V|^{n_1+n_2-p}\, dV.$$

$$\ldots \quad (7.6.8.4)$$

Thus for the distribution of c's we have

$$\text{Const} \;\; \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} dc_i \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j) \qquad \ldots \quad (7.6.8.5)$$

$$\times \int_V \exp\left[-\tfrac{1}{2}\left(\operatorname{tr} \; VV'-2\sum_{i=1}^{n_1} \gamma_i^{\frac{1}{2}}(VD_{c^{\frac{1}{2}}/(1+c)^{\frac{1}{2}}})_{ii}\right)\right] |V|^{n_1+n_2-p}\, dV,$$

where the elements of V vary from $-\infty$ to ∞. For the general, i.e., the non-null hypothesis we would leave the distribution problem at this. On the null hypothesis, i.e., when γ_i's $= 0$, the integral becomes

$$\int_V \exp\left[-\tfrac{1}{2} \operatorname{tr} VV'\right] |V|^{n_1+n_2-p} dV,$$

which we can evaluate by using (A.8.7) and which we can then absorb into the constant, thus reducing (7.6.8.5) to

$$\text{Const} \left[\prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} dc_i/(1+c_i)^{\frac{n_1+n_2}{2}}\right] . \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j). \qquad \ldots \quad (7.6.8.6)$$

The constant is easily seen to simplify into

$$\text{Const} = \pi^{n_1^2/2} \prod_{i=1}^{n_1} \Gamma\left(\frac{n_1+n_2-i+1}{2}\right) \Big/ \prod_{i=1}^{n_1} \Gamma\left(\frac{n_2-p+n_1-i+1}{2}\right)$$

$$\times\ \Gamma\left(\frac{n_1-i+1}{2}\right)\Gamma\left(\frac{p-i+1}{2}\right).$$

It may be noticed that (7.6.4) which is the distribution of the roots on the null hypothesis in the case $p \leqslant n_1, n_2$, goes over into (7.6.8) which is the distribution of the roots on the null hypothesis in the case $n_1 < p \leqslant n_2$, if we make the substitution $(p, n_1, n_2) \rightarrow (n_1, p, n_2-p+n_1)$. It can be proved by a general reasoning, without working through the distributions, that this tie-up will be true both for the two null hypothesis distributions as well as for the two non-null hypothesis distributions.

A special case of (ii), namely where $n_1 = 1$, is of considerable importance and in this case not only the central but also the non-central distribution is easily available. Notice that in this case $n_1 = 1$, $X_1(p\times n_1) = \mathbf{x}(p\times 1)$ (say). $\xi(p\times n_1) = \boldsymbol{\xi}(p\times 1)$ (say), $L_1(1\times 1)$, subject to $L_1\, L_1' = I(1)$, is equal to $\pm\ 1$ so that L_{1I} drops out and $1 \Big/ \left| \frac{\partial(L_1 L_1')}{\partial(L_{1D})} \right|_{L_{1I}} = 1/2$ which we absorb in the constant. There is only one non-zero (and positive) c which is equal to $\mathbf{x}'(X_2\, X_2')^{-1}\mathbf{x}$ [since, in general, $\operatorname{tr} D_c = \sum_{i=1}^{n_1} c_i = \operatorname{tr}(X_1\, X_1'(X_2\, X_2')^{-1}) = \operatorname{tr} X_1'(X_2\, X_2')^{-1}X_1$ and in this case $\sum_{i=1}^{n_1} c_i = c$ and $\operatorname{tr} X_1'(X_2\, X_2')^{-1}X_1 = \mathbf{x}'(X_2\, X_2')^{-1}\mathbf{x}$] and also one possible non-zero (and positive) γ which is equal to $\boldsymbol{\xi}'\, \Sigma^{-1}\, \boldsymbol{\xi}$. Also both in (7.6.6) and (7.6.8) the factor mod $\left[\prod_{i<j=1}^{n_1-1} (c_i-c_j)\right]$ will drop out.

Substituting in (7.6.8) we have thus in this case, for c, the central distribution, [7, 17],

$$\frac{\Gamma\left(\frac{n_2+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n_2-p+1}{2}\right)} c^{\frac{p-2}{2}}\, dc/(1+c)^{\frac{n_2+1}{2}}. \qquad \ldots \quad (7.6.9)$$

For the non-central distribution, i.e., when $\gamma \neq 0$, we have, in this case, $\sum_{i=1}^{n_1} \gamma_i = \gamma$, $\sum_{i=1}^{s} \left[\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} D_{\sqrt{c}} L_1\right]_{ii} \sqrt{\gamma_i} = \pm\, u_{11}\sqrt{c\gamma}$ and thus it is easily checked that by substituting in (7.6.6) and remembering that L_1 could take just two values ± 1, we shall have, for U and c, the joint distribution

$$\text{Const}\ \exp\left[-\tfrac{1}{2}\left(\text{tr}\ UD_{1+c}(p)U'+\gamma\mp 2u_{11}\sqrt{c\gamma}\right)\right]|U|^{n_2+1-p}\prod_{i=1}^{p-1}(\tilde{U}_3)_{ii}^{p-1-i}\,dU\,c^{p-2}dc \qquad \ldots \quad (7.6.10)$$

or

$$\text{Const}\ \exp\left[-\tfrac{1}{2}\ \text{tr}\ UD_{1+c}(p)U'\right]\cosh\left(u_{11}\sqrt{c\gamma}\right)|U|^{n_2+1-p}$$

$$\times\prod_{i=1}^{p-1}(\tilde{U}_3)_{ii}^{p-1-i}dU\,c^{p-2}\,dc. \qquad \ldots \quad (7.6.11)$$

Now putting
$$UD_{\sqrt{1+c}}=\begin{array}{c|cc|} p-1 & U_1 & \tilde{U}_3 \\ 1 & U_2 & U_4 \\ & 1 & p-1 \end{array}\ \begin{array}{|cc|c} \sqrt{1+c} & \vdots\ 0 & 1 \\ 0 & \vdots\ I & p-1 \\ 1 & p-1 & \end{array} = \begin{array}{|cc|c} V_1 & \tilde{V}_3 & p-1 \\ V_2 & V_4 & 1 \\ 1 & p-1 & \end{array}$$

$= V$(say), we have for V and c the joint distribution

$$\text{Const}\ \exp\left(-\tfrac{1}{2}\ \text{tr}\ VV'\right)\cosh\left(v_{11}\sqrt{\gamma c/1+c}\right)|V|^{n_2+1-p}\prod_{i=1}^{p}(\tilde{V}_3)_{ii}^{p-1-i}dV$$

$$\times\ c^{p-2}dc/(1+c)^{\frac{(n_2+1)}{2}} \qquad \ldots \quad (7.6.12)$$

To obtain the distribution of c by integrating over V, we use the same artifice as in (A.8.8), change over to a solid matrix W and obtain for c the distribution

$$\frac{\text{Const}}{F(p-1,p-1)}\left[c^{\frac{p-2}{2}}dc/(1+c)^{\frac{n_2+1}{2}}\right]\int_W \exp\left(-\tfrac{1}{2}\ \text{tr}\ WW'\right)$$

$$\times\ \cosh\left(w_{11}\sqrt{\gamma c/1+c}\right)|W|^{n_2+1-p}dW, \qquad \ldots \quad (7.6.13)$$

where $F(p-1,\ p-1)$ is given by (A.8.6.3).

To effect the integration over W, denote the row vectors of W by $\mathbf{w}_i'$ $(i=1,2,\ldots,p)$ and make the transformation

$$\begin{array}{|c|c} \mathbf{w}_2' & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathbf{w}_p' & 1 \\ p & \end{array} = \begin{array}{c|c|} 1 & \mathbf{g}_2' \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \mathbf{g}_p' \\ & p \end{array}\ L(p\times p) \text{ (where } L \text{ is arbitrary } \perp \text{ such that the first row vector}$$

of L is along $\mathbf{w}_1'$). Thus, though L will depend on $\mathbf{w}_1'$, the Jacobian of this transformation is easily checked to be unity. In this case the new matrix is

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ g_{21} & g_{22} & \cdots & g_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ g_{p1} & g_{p2} & \cdots & g_{pp} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \cdots w_{1p} \\ g_{21} & \\ \cdot & G \\ g_{p1} & \end{bmatrix} \text{ (say).}$$

$$= \begin{bmatrix} & \mathbf{w}_1' & \\ [\mathbf{g}^* & & \mathbf{G}] \end{bmatrix} \text{ (say).}$$

Thus

$$WW' = \left[\begin{array}{c|c} \mathbf{w}_1'\mathbf{w}_1 & \mathbf{w}_1'\alpha'\begin{bmatrix}\mathbf{g}^{*\prime} \\ \mathbf{G}'\end{bmatrix} \\ \hline [\mathbf{g}^* \quad \mathbf{G}]\alpha\mathbf{w}_1 & \mathbf{g}^*\mathbf{g}^{*\prime}+\mathbf{G}\mathbf{G}' \end{array}\right]$$

$$= \left[\begin{array}{c|c} \mathbf{w}_1'\mathbf{w}_1 & [\mathbf{w}_1'\mathbf{w}_1)^{\frac{1}{2}} \quad 0]\begin{bmatrix}\mathbf{g}^{*\prime} \\ \mathbf{G}'\end{bmatrix} \\ \hline [\mathbf{g}^* \quad \mathbf{G}]\begin{bmatrix}(\mathbf{w}_1'\mathbf{w}_1)^{\frac{1}{2}} \\ 0\end{bmatrix} & \mathbf{g}^*\mathbf{g}^{*\prime}+\mathbf{G}\mathbf{G}' \end{array}\right]$$

$$= \left[\begin{array}{c|c} \mathbf{w}_1'\mathbf{w}_1 & (\mathbf{w}_1'\mathbf{w}_1)^{\frac{1}{2}}\mathbf{g}^{*\prime} \\ \hline \mathbf{g}^*(\mathbf{w}_1'\mathbf{w}_1)^{\frac{1}{2}} & \mathbf{g}^*\mathbf{g}^{*\prime}+\mathbf{G}\mathbf{G}' \end{array}\right]$$

It is easy to see that $|W| = |WW'|^{\frac{1}{2}} = (\mathbf{w}_1'\mathbf{w}_1)^{\frac{1}{2}}|GG'|^{\frac{1}{2}}$. Also tr WW' goes over into tr $\mathbf{w}_1'\mathbf{w}_1 + \sum_{i=2}^{p} g_{i1}^2 + \text{tr } GG'$.

Now by using (A.8.7.1) it is easy to integrate out

$$\int \exp\left[-\tfrac{1}{2}\left(\sum_{i=2}^{p} g_{i1}^2 + \text{tr } GG'\right)\right] \prod_{i=2}^{p} dg_{i1}|G|^{n_2-p+1}\, dG,$$

and obtain a constant which we absorb into the constant factor and obtain for c the distribution

$$\text{Const}\left[c^{\frac{p-2}{2}} dc/(1+c)^{\frac{n_2+1}{2}}\right]$$

$$\times \int_{w_{1j}(j=1,2,..,p)} \exp\left[-\tfrac{1}{2}\sum_{j=1}^{p} w_{1j}^2\right] \cosh\left(w_{11}\sqrt{\gamma c/1+c}\right)\left(\sum_{j=1}^{p} w_{1j}^2\right)^{\frac{n_2-p+1}{2}} \prod_{j=1}^{p} dw_{1j}. \quad \ldots \quad (7.6.14)$$

To evaluate this integral we put $\sum_{j=1}^{p} w_{ij}^2 = r^2$ and $w_{11} = r \cos \theta$, so that, using (A.8.5.1), we have the integral = constant (which is easily obtained)

$$\times \int_{r=0}^{\infty} \int_{\theta=0}^{\pi/2} \exp\left[-\tfrac{1}{2} r^2\right] \cosh\left(r \cos \theta \sqrt{\gamma c/(1+c)}\right) r^{n_2} \, dr \, (\sin \theta)^{p-2} \, d\theta.$$

We note that the integral over θ could be taken from 0 to $\pi/2$ and the result multiplied by 2.

Using now the formula (9) p. 79 and formula (2) p.393 of Watson's Bessel Functions and remembering the relation between Bessel I and Bessel J functions, this integral reduces to const ${}_1F_1\left(\frac{n_2+1}{2}\,;\frac{p}{2}\,;\tfrac{1}{2}\gamma\frac{c}{1+c}\right)$, so that the distribution of c comes out in the form

$$\frac{\Gamma\left(\frac{n_2+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n_2-p+1}{2}\right)} \exp\left(-\frac{\gamma}{2}\right)\left[c^{\frac{p-2}{2}}/(1+c)^{\frac{n_2+1}{2}}\right] {}_1F_1\left(\frac{n_2+1}{2}\,;\frac{p}{2}\,;\tfrac{1}{2}\,\frac{\gamma c}{(1+c)}\right), [1]. \quad \dots \quad (7.6.15)$$

It is of interest to note that on the null hypothesis $\gamma = 0$, the confluent hyper-geometric function reduces to a constant and (7.6.15) goes over, as it should, into (7.6.9)

7.7. *The distribution of charateristic rooots connected with the multivariate regression model of* (4.25)-(4.33). Going back to (4.33) and noting the identity of this distribution form with that of (4.21), it is easy to check from section (A.7.5) that the distribution of the $c[Z_1 Z_1' (Z_2 Z_2')^{-1}]$'s (= c_i's say) could not involve as parameters obtaining anything except $c[(\mu\tilde{T}\tilde{T}'\mu')\,\Sigma^{-1}]$'s = $c[(\mu U U'\mu')\Sigma^{-1}]$'s (= γ_i's, say). The problem of the distribution of $c[Z_1 Z_1'(Z_2 Z_2')^{-1}]$'s can thus be thrown back, where $p \leqslant q$, on the case (i) of (7.6) and, when $p > q$, the case (ii) of (7.6), in both cases putting $n_1 = q$ and $n_2 = n-1-q$. The complete reduction of the distribution problem, i.e., the derivation of the joint distribution of the c_i's on the null hypothesis $\gamma_i = 0$ $(i = 1, \dots, p)$ (which, in this case, can happen if and only if $\mu = 0$, since U is not presumably 0), can be effected in exactly the same manner as for the distribution on the null hypothesis in the cases (i) and (ii) of section 7.6. Turning now to $(Z_1 Z_1')\,(Z_2 Z_2')^{-1}$ and checking with (4.25)-(4.33), we see that $(Z_1 Z_1')\,(Z_2 Z_2')^{-1}$ $= [XU'(UU')^{-1}UX'] \times [XX'-n\bar{\mathbf{x}}\bar{\mathbf{x}}'-XU'(UU')^{-1}UX']^{-1}$.

An important special case is that of $p = 1$ which we can handle by putting, in case (*i*) of (7.6), $p = 1$, $n_1 = q$ and $n_2 = n-q$, $L_1(p\times n_1) = \mathbf{1}'(1\times q)$ and $A(1\times 1)$ = a (say), a scalar.

Substituting in (7.6.2) we have for c (note that there is only one non-zero c here) the distribution

$$2[1/2\pi]^{\frac{n}{2}} \; F(1, n-q) c^{\frac{q-2}{2}} \; dc$$

$$\times \int_{a=0}^{\infty} \int_{l_2, \dots, l_q} \exp\left[-\tfrac{1}{2}\left(a^2(1+c)+\gamma\pm 2a\sqrt{c\gamma}\,l_1\right)\right] a^{n-1} da \prod_{i=2}^{q} dl_i/(1-\sum_{i=2}^{q} l_i^2)^{\frac{1}{2}}. \quad \dots \quad (7.7.1)$$

To evaluate the integral proceed as follows

$$\int\limits_{l_2,\,\dots,\,l_q} \exp\,[\pm a\sqrt{c\gamma}\,l_1] \prod_{i=2}^{q} dl_i/(1-\sum_{i=2}^{q} l_i^2)^{\frac{1}{2}} = \int_{\theta=0}^{\pi/2} \cosh\,(a\sqrt{c\gamma}\cos\theta)(\sin\theta)^{q-2}\,d\theta,$$

so that substituting in (7.7.1) we have the integral under (7.7.1) reducing to

$$\text{Const} \int_{a=0}^{\infty}\int_{\theta=0}^{\pi/2} \exp\,[-\tfrac{1}{2}\,a^2(1+c)]\cosh\,(a\sqrt{c\gamma}\cos\theta)a^{n-1}da(\sin\,\theta)^{q-2}d\theta.$$

Taking account of the discussion after (7.6.14) this integral reduces to const ${}_1F_1\left(\frac{n}{2}\,;\,\frac{q}{2}\,;\,\frac{1}{2}\,\frac{\gamma c}{1+c}\right)\Big/\,(1+c)^{n/2}$, and hence we have for c the distribution

$$\frac{\Gamma(n/2)}{\Gamma(q/2)\,\Gamma(\overline{n-q}/2)}\exp\left(-\frac{\gamma}{2}\right)[c^{\frac{q-2}{2}}\,dc/(1+c)^{\frac{n}{2}}]{}_1F_1\left(\frac{n}{2}\,;\frac{q}{2}\,;\frac{1}{2}\,\frac{\gamma c}{1+c}\right), \quad \dots \quad (7.7.2)$$

the const factor being easily evaluated (since the constant factor at each stage is known and carried over to the next stage).

If $\gamma = 0$, this reduces to

$$\frac{\Gamma(n/2)}{\Gamma(q/2)\;\Gamma(\overline{n-q}/2)}\,c^{\frac{q-2}{2}}\,dc/(1+c)^{\frac{n}{2}}\;. \quad \dots \quad (7.7.3)$$

For e given by $c = e/(1-e)$ we have on the non-null and the null hypothesis the respective distributions

$$\frac{\Gamma(n/2)}{\Gamma(q/2)\;\Gamma(\overline{n-q}/2)}\exp\left(-\frac{\gamma}{2}\right)(1-e)^{\frac{n-q-2}{2}}\,e^{\frac{q-2}{2}}\,de\;{}_1F_1\left(\frac{n}{2}\,;\,\frac{q}{2}\,;\,\tfrac{1}{2}\gamma e\right) \dots \quad (7.7.4)$$

and

$$\frac{\Gamma(n/2)}{\Gamma(q/2)\;\Gamma(\overline{n-q}/2)}\,(1-e)^{\frac{n-q-2}{2}}\,e^{\frac{q-2}{2}}\,de\;. \quad \dots \quad (7.7.5)$$

7.8. *Reduction of the various joint distributions of the characteristic roots to a common standard form.* On the respective null hypotheses consider the distributions

(7.3.5), (7.4.5), (7.5.4) and (7.5.8) and check that they can all be reduced to the following common standard form (expressed in terms of (7.4.5)).

$$\frac{1}{s!}\left[\pi^{s/2}\prod_{i=1}^{s}\Gamma\left(\frac{2m_1+2m_2+s+i+2}{2}\right)\Big/\prod_{i=1}^{s}\Gamma\left(\frac{2m_1+i+1}{2}\right)\Gamma\left(\frac{2m_2+i+1}{2}\right)\Gamma\left(\frac{i}{2}\right)\right]$$

$$\times\prod_{i=1}^{s}x_i^{m_1}(1-x_i)^{m_2}\times\operatorname{mod}\left[\prod_{i<j=1}^{s-1}(x_i-x_j)\right]\prod_{i=1}^{s}dx_i, \qquad\ldots\quad(7.8.1)$$

where $0 \leqslant x_1, \ldots, x_s \leqslant 1$ and, a.e., $0 < x_1, \ldots, x_s < 1$. If, however, we order the x's as $0 \leqslant x_1 \leqslant x_2 \leqslant \ldots x_s \leqslant 1$, or, a.e., $0 < x_1 < x_2 < \ldots < x_s < 1$, then the above distribution can be rewritten as

$$\left[\pi^{s/2}\prod_{i=1}^{s}\Gamma\left(\frac{2m_1+2m_2+s+i+2}{2}\right)\Big/\prod_{i=1}^{s}\Gamma\left(\frac{2m_1+i+1}{2}\right)\Gamma\left(\frac{2m_2+i+1}{2}\right)\Gamma\left(\frac{i}{2}\right)\right]\prod_{i=1}^{s}x_i^{m_1}(1-x_i)^{m_2}$$

$$\times\prod_{i>j=2}^{s}(x_i-x_j)\prod_{i=1}^{s}dx_i\,. \qquad\ldots\quad(7.8.2)$$

It is well known that $\prod_{i<j=1}^{s-1}(x_i-x_j)$ can be written in another form, namely, as a Vandermonde determinant

$$\begin{vmatrix} x_1^{s-1} & x_2^{s-1} & \ldots & x_s^{s-1} \\ x_1^{s-2} & x_2^{s-2} & \ldots & x_s^{s-2} \\ . & . & \ldots & . \\ 1 & 1 & \ldots & 1 \end{vmatrix} \qquad\ldots\quad(7.8.3)$$

Denoting now the constant factor (within the square brackets) of (7.8.1) or (7.8.2) by $k(s, m_1, m_2)$, we can rewrite (7.8.1) and (7.8.2) respectively as

$$\frac{1}{s!}\,k(s, m_1, m_2)\prod_{i=1}^{s}x_i^{m_1}(1-x_i)^{m_2}dx_i \operatorname{mod}\begin{vmatrix} x_1^{s-1} & \ldots & x_s^{s-1} \\ x_1^{s-2} & \ldots & x_s^{s-2} \\ . & \ldots & . \\ 1 & \ldots & 1 \end{vmatrix} \qquad\ldots\quad(7.8.4)$$

and

$$k(s, m_1, m_2)\prod_{i=1}^{s}x_i^{m_1}(1-x_i)^{m_2}dx_i\begin{vmatrix} x_s^{s-1} & \ldots & x_1^{s-1} \\ . & \ldots & . \\ 1 & \ldots & 1 \end{vmatrix} \qquad\ldots\quad(7.8.5)$$

or

$$k(s,\ m_1,\ m_2)\begin{vmatrix} x_s^{m_1+s-1}(1-x_s)^{m_2} & \ldots & x_1^{m_1+s-1}(1-x_1)^{m_2} \\ x^{m_1+s-2}(1-x_s)^{m_2} & \ldots & x_1^{m_1+s-2}(1-x_1)^{m_2} \\ \vdots & \ldots & \\ x_s^{m_1}(1-x_s)^{m_2} & \ldots & x_1^{m_1}(1-x_1)^{m_2} \end{vmatrix} . \prod_{i=1}^{s} dx_i\ .\ \ldots \qquad (7.8.6)$$

The following substitutions are to be made in (7.8.1) or (7.8.4) in order to obtain (7.3.6), (7.4.6), (7.5.4) and (7.5.8). For (7.3.6) put $x_i = c_i/(1+c_i)$, $s = p$, $m_1 = \frac{n_1-p-1}{2}$, $m_2 = \frac{n_2-p-1}{2}$, for (7.4.6) put $x_i = c_i$, $s = p$, $m_1 = \frac{q-p-1}{2}$, $m_2 = \frac{n-p-q-1}{2}$, for (7.6.4) put $x_i = c_i/(1+c_i)$, $s = p$, $m_1 = \frac{n_1-p-1}{2}$, $m_2 = \frac{n_2-p-1}{2}$ and for (7.6.8) put $x_i = c_i/(1+c_i)$, $s = n_1$, $m_1 = \frac{p-n_1-1}{2}$, $m_2 = \frac{n_2-p-1}{2}$. It is of interest to note that ordering the x_i's is exactly equivalent to ordering the c_i's of (7.3.6), (7.6.4) and (7.6.8); in other words, $0 \leqslant x_1 \leqslant \ldots \leqslant x_s \leqslant 1 \Longleftrightarrow 0 \leqslant c_1 \leqslant \ldots \leqslant c_s < \infty$ and $0 < x_1 < \ldots < x_s < 1 \Longleftrightarrow 0 < c_1 < \ldots < c_s < \infty$.

CHAPTER EIGHT

On the C.D.F. of the Largest and/or the Smallest Root

In this chapter starting from (7.8.2) or (7.8.5) we shall obtain the c.d.f. of x_s, i.e., $P(x_s \leqslant x_0)$, where x_0 is a given constant $\leqslant 1$ (from which it is easy to check that one can obtain the c.d.f. of x_1 by merely interchanging m_1 and m_2), and also obtain $P(x_0' \leqslant x_1 \leqslant x_s \leqslant x_0)$; where x_0' and x_0 are also given constants subject to $0 \leqslant x_0' \leqslant x_0 \leqslant 1$. Starting from (7.8.5) and putting in (A.9.6.13) $m_i = m_1+i-1$ and $n = m_2$ ($i = 1, 2, \ldots, s$) we have

$$P(0 \leqslant x_1 \leqslant \ldots \leqslant x_s \leqslant x) = P(x_s \leqslant x) = k(s, m_1, m_2)$$

$$\times \beta\left[x; \begin{pmatrix} m_1+s-1, m_2 & m_1+s-2, m_2 & \ldots & m_1, m_2 \\ & & \ldots & . \\ m_1+s-1, m_2 & m_1+s-2, m_2 & \ldots & m_1, m_2 \end{pmatrix}, \right]. \quad \ldots \quad (8.1)$$

where β is to be successively and completely reduced with the help of the fundamental formula (A.9.6.13).

For the c.d.f of the smallest root x_1 we note that $P(x_1 \leqslant x) = 1-P(x_1 \geqslant x) = 1-P(x \leqslant x_1 \leqslant \ldots \leqslant x_s \leqslant 1)$. Going back to the c.d.f. of $(x_1, \ldots, x_s)$ and using the transformation $x_i = 1-z_i$ ($i = 1, 2, \ldots, s$) we have

$$k(s, m_1, m_2) \int_x^1 \int_{x_1}^1 \cdots \int_{x_{s-1}}^1 \prod_{i=1}^s x_i^{m_1}(1-x_i)^{m_2} \prod_{i>j=2}^s (x_i-x_j) \prod_{i=1}^s dx_i$$

$$= k(s, m_1, m_2) \int_0^{1-x} \int_0^{z_1} \cdots \int_0^{z_{s-1}} \prod_{i=1}^s z_i^{m_2}(1-z_i)^{m_1} \prod_{i>j=2}^s (z_j-z_i) \prod_{i=1}^s dz_i \, . \quad \ldots \quad (8.2)$$

It is now easy to see that the integral on the right hand side of (8.2) is exactly the same as that on the right hand side of (8.1) with just the interchange of m_1 and m_2, which shows that the c.d.f. of the smallest root can be thrown back on that of the largest root and vice versa. The final reduction of the exact c.d.f. of the largest or the smallest root is necessarily lengthy and need not be given here. When m_1+m_2 is large, which is the case in most practical applications,there is a good approximation in relatively far simpler terms (especially when percentage points, up to, say, 10% are needed) which will be given in a later monograph.

In this chapter the final reduction for the exact c.d.f. of the largest root in the case of $s = 2,3, 4$ will be given. This is as follows:

For $s = 2$,

$$P(x_2 \leqslant x) = \frac{k(2, m_1, m_2)}{m_1+m_2+2}[-\beta_0(x; m_1+1, m_2+1)\ \beta(x; m_1, m_2)+2\beta(x; 2m_1+1, 2m_2+1)]. \quad \dots \ (8.3)$$

For $s = 3$,

$$P(x_3 \leqslant x) = \frac{k(3, m_1, m_2)}{m_1+m_2+3}[2\beta(x; 2m_1+3, 2m_2+1)\ \beta(x; m_1, m_2)-2\beta(x; 2m_1+2, 2m_2+1)$$

$$\times \beta(x; m_1+1, m_2) - \frac{1}{k(2, m_1, m_2)}\beta_0(x; m_1+2, m_2+1)P(x_2 \leqslant x]. \quad \dots \ (8.4)$$

For $s = 4$,

$$P(x_4 \leqslant x) = \frac{k(4, m_1, m_2)}{m_1+m_2+4}\Big[-\beta_0(x; m_1+3, m_2+1)\frac{1}{k(3, m_1, m_2)}P(x_3 \leqslant x)$$

$$+2\beta(x; 2m_1+5, 2m_2+1)\frac{1}{k(2, m_1, m_2)}P(x_2 \leqslant x)+\frac{2}{m_1+m_2+3}\beta(x; 2m_1+3, 2m_2+1)$$

$$\times\Big\{-\beta_0(x; m_1+2, m_2+1)\ \beta(x; m_1+1, m_2)+2\beta(x; 2m_1+3, 2m_2+1)\Big\}$$

$$-\frac{2}{m_1+m_2+3}\beta(x; 2m_1+4, 2m_2+1)\Big\{-\beta_0(x; m_1+2, m_2+1)\ \beta(x; m_1, m_2)$$

$$+\frac{m_1+2}{k(2, m_1, m_2)}P(x_2 \leqslant x)+2\beta(x; 2m_1+2, 2m_2+1)\Big\}\Big] \quad \dots \ (8.5)$$

Again starting from (7.8.5) and putting in (A.9.7.2) $m_i = m_1+i-1$ and $n = m_2$ $(i = 1, 2, \dots, s)$ we have

$$P(x_0 \leqslant x_1 \leqslant \dots \leqslant x_s \leqslant x) = P(x_0 \leqslant x_1 \leqslant x_s \leqslant x)$$

$$= k(s, m_1, m_2)\beta\left[x, x_0; \begin{pmatrix} m_1+s-1, m_2 & m_1+s-2, m_2 & \dots & m_1, m_2 \\ m_1+s-1, m_2 & m_1+s-2, m_2 & \dots & m_1, m_2 \\ \cdot & \cdot & \dots & \cdot \\ m_1+s-1, m_2 & m_1+s-2, m_2 & \dots & m_1, m_2 \end{pmatrix}\right] \quad \dots \ (8.6)$$

where β is to be successively and completely reduced with the help of the fundamental formula (A.9.7.2). Below is given the complete reduction of the left side of (8.6) for $s = 2$, 3, and 4, the left side of (8.6) being conveniently denoted by P_2, P_3, P_4, etc.

$$P_2 = \frac{k(2, m_1, m_2)}{(m_1+m_2+2)}[2\beta(x, x_0;\, 2m_1+1,\, 2m_2+1) - \beta(x, x_0;\, m_1, m_2)\{\beta_0(x;\, m_1+1,\, m_2+1)$$

$$+\beta_0(x_0;\, m_1+1,\, m_2+1)\}], \qquad \ldots \quad (8.7)$$

$$P_3 = \frac{k(3, m_1, m_2)}{(m_1+m_2+3)}[2\beta(x, x_0;\, m_1, m_2)\, \beta(x, x_0;\, 2m_1+3,\, 2m_2+1)$$

$$-\, 2\beta(x, x_0;\, m_1+1,\, m_2)\, \beta(x, x_0;\, 2m_1+2,\, 2m_2+1)$$

$$-\, \frac{P_2}{k(2, m_1, m_2)}\{\beta_0(x;\, m_1+2,\, m_2+1) - \beta_0(x_0;\, m_1+2,\, m_2+1)\}], \qquad \ldots \quad (8.8)$$

$$P_4 = \frac{k(4, m_1, m_2)}{(m_1+m_2+4)}\Big[2\beta(x, x_0;\, 2m_1+5,\, 2m_2+1)\frac{P_2}{k(2, m_1, m_2)}$$

$$-\, \frac{P_3}{k(3, m_1, m_2)}\{\beta_0(x;\, m_1+3,\, m_2+1) + \beta_0(x_0;\, m_1+3,\, m_2+1)\}$$

$$+\, \frac{2\beta(x, x_0;\, 2m_1+3,\, 2m_2+1)}{(m_1+m_2+3)}\{-\beta_0(x;\, m_1+2,\, m_2+1)\, \beta(x, x_0;\, m_1+1,\, m_2)$$

$$-\, \beta_0(x_0;\, m_1+2,\, m_2+1)\, \beta(x, x_0;\, m_1+1,\, m_2) + 2\beta(x, x_0;\, 2m_1+3,\, 2m_2+1)\}$$

$$-\, \frac{2\beta(x, x_0;\, 2m_1+4,\, 2m_2+1)}{(m_1+m_2+3)}\Big\{-\beta_0(x;\, m_1+2,\, m_2+1)\beta(x, x_0;\, m_1, m_2)$$

$$-\, \beta_0(x_0;\, m_1+2,\, m_2+1)\, \beta(x, x_0;\, m_1, m_2) + 2\beta(x, x_0;\, 2m_1+2,\, 2m_2+1) + \frac{m_1+2}{k(2, m_1, m_2)}P_2\Big\}\Big].$$

$$\ldots \quad (8.9)$$

For larger values of s the reduction of P_s to exact expressions like those given by the right side of (8.7)-(8.9) will no doubt be increasingly lengthy but the remarks made after (8.2) will apply to this situation as well, so that if m_1+m_2 is reasonably large, as would be the case in most practical situations, we can use much shorter expressions as good approximations.

CHAPTER NINE

Operating Characteristics and Lower Bounds on the Power Functions of the Test Regions*

9.1. *The operating characteristics of the test regions.* As of the moment the exact (small sample) power functions of the regions (6.4.2), (6.4.5), (6.4.8) and (6.4.11) seem to be, in the general cases, quite intractable. At any rate, so far as the author is aware, no method is known at the moment by which the requisite distribution problems could be solved and the final c.d.f.'s be given, except in very symbolic (and, for practical purposes, quite useless) forms. However, it is possible even without exact expressions for c.d.f.'s, to obtain a number of useful semi-qualitative and semi-quantitative properties of the power functions, which, as will be presently seen, are about all that would really matter for most practical purposes.

We observe from (A.7.1), (A.7.2), (A.7.5) and (A.7.3) respectively that the powers of the critical regions (6.4.2), (6.4.5), (6.4.8) and (6.4.11) depend only on the corresponding sets of populations roots, namely $c(\Sigma\,\Sigma_0^{-1})$'s (to be called γ's) for the first case, $c(\Sigma_1\Sigma_2^{-1})$'s (to be called γ's) for the second case, $c(\Sigma^*\Sigma^{-1})$'s (to be called γ's) for the third case and $c(\Sigma_{11}^{-1}\Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{12}')$'s (to be called γ's) for the fourth case. For convenience we write down the respective powers for the four cases as

$$P[c_p \geqslant c_{2\alpha}(p, n) \text{ and/or } c_1 \leqslant c_{1\alpha}(p, n)\,|\,H] = P(\alpha, p, n;\, \gamma_1, \gamma_2, \ldots, \gamma_p) \qquad \ldots \quad (9.1.1)$$

$$P[c_p \geqslant c_{2\alpha}(p, n_1, n_2) \text{ and/or } c_1 \leqslant c_{1\alpha}(p, n_1, n_2)\,|\,H] = P(\alpha, p, n_1, n_2;\, \gamma_1, \gamma_2, \ldots, \gamma_p), \quad (9.1.2)$$

$$P[c_r \geqslant c_\alpha(p, k-1, n-k)\,|\,H] = P(\alpha, p, k-1, n-k;\, \gamma_1, \gamma_2, \ldots, \gamma_r) \text{ and} \qquad \ldots \quad (9.1.3)$$

$$P[c_p \geqslant c_\alpha(p, q, n)\,|\,H] = P(\alpha, p, q, n;\, \gamma_1, \gamma_2, \ldots, \gamma_p). \qquad \ldots \quad (9.1.4)$$

Notice that, depending on the rank of Σ^* and Σ_{12}, some of the γ's of (9.1.3) and (9.1.4) might be zero but the most general case will be one in which as many as are set down will be positive. Notice also that in (9.1.3), $r = \min\,(p, k-1)$. Recall now from (A.3.3) that for (9.1.1), $0 \leqslant$ all γ's $<\infty$, from (A.1.9) that for (9.1.2) and (9.1.3), $0 \leqslant$ all γ's $< \infty$ and from (A.1.14) that for (9.1.4) $0 \leqslant$ all γ's < 1.

With this introduction we shall consider the power functions (9.1.1) and (9.1.2) for the problems of one dispersion matrix and two dispersion matrices and the power function (9.1.3) for the problem of multivariate analysis of variance and the power function (9.1.4) for the problem of independence between two sets of variates. In section 9.2, to each power function a lower bound will be obtained which will be called a lower bound function and which will be seen to involve (aside from the degrees of freedom) just those deviation parameters that occur in the power function itself. In Chapter 10 it will be shown that for each power function the lower bound function monotonically increases as each deviation parameter separately tends away from its

---

* See reference [43] in this connection.

value under the null hypothesis. Although, under the null hypothesis, the lower bound function does not assume the value α which is the significance level of the test, this value is attained soon enough under deviations from the hypothesis. Thus the power function stays greater than a monotonically increasing function of each deviation parameter and is also shown to be unbiassed against all deviations from the hypothesis for which the lower bound function is greater than or equal to the size α of the test. In chapter 11, for each of the power functions (9.1.3) and (9.1.4) another such monotonic lower bound function is obtained which is believed to be closer than the lower bound functions of section 9.2; also for each of the power functions (9.1.1) and (9.1.2) some near monotonic properties are proved.

9.2. *Lower bounds on the power functions.* The lower bounds are obtained in three different stages to be called (9.2*a*), (9.2b) and (9.2c).

9.2*a*. *Reduction to canonical forms.* Without any loss of generality we can, for the case of (6.4.2), start right from the canonical form (A.7.1.1); for the case of (6.4.5) from the canonical form (A.7.2.1); for the case of (6.4.8) from the canonical form (A.7.5.6); and for the case of (6.4.11) from the canonical form (A.7:3.5). for the case of (6.4.2) there is an additional point to be noted. Putting together (A.7.1) and (6.4.1) it is clear that it will be appropriate, instead of using as we did in (A.7.1) the transformation

$$X(p\times n) = \mu(p\times p)\ Y(p\times n), \quad \text{where } \Sigma = \mu D_\gamma \mu'$$

(γ's being the roots of Σ), to use the transformation $X(p\times n)$

$$= \mu(p\times p)\Delta_0(p\times p)\ Y(p\times n), \text{ where } \Sigma_0 = \Delta_0\, \Delta_0' \text{ and } \Delta_0^{-1}\Sigma(\Delta_0')^{-1} = \mu D_\gamma \mu',$$

γ's being the roots of $\Delta_0^{-1}\Sigma(\Delta_0')^{-1}$, i.e., of $\Sigma(\Delta_0\, \Delta_0')^{-1}$, i.e., of $\Sigma\Sigma_0^{-1}$. Under this transformation the γ's of the canonical form (A.7.1.1) will really be the roots of $\Sigma\ \Sigma_0^{-1}$ and the roots of $(YY')/n$ will really be the roots of the equation (6.4.1) and thus we have an exact tie-up with the problem involving the power function of (6.4.2).

9.2b. *The inclusion within the test regions* (6.4.2), (6.4.5), (6.4.8) *and* (6.4.11) *of regions having simpler probability measures* (*under the respective non-null hypotheses*).

(i) We recall from (6.4) and the canonical form (A.7.1.1) that the test region (6.4.2) is really $U_{\mathbf{a}}[\mathbf{a}'YY'\mathbf{a}/n\mathbf{a}'\mathbf{a} \geqslant c_{2\alpha}(p, n)$ or $\leqslant c_{1\alpha}(p, n)]$, where Y has the distribution (A.7.1.1). We also notice from the canonical form (A.7.1.1) that the p functions $\mathbf{a}_i'YY'\mathbf{a}_i/\gamma_i n\mathbf{a}_i'\mathbf{a}_i$ $(i = 1, 2, \ldots, p)$ (with $\mathbf{a}_i'$ being a $1\times p$ row vector having 1 for the i-th element and 0 for all other elements) are distributed as p independent χ^2's with d.f. n each. Putting these two facts together we have that the test region (6.4.2) includes the union of p regions, each composed of the tail ends of a central χ^2-region, all the p χ^2's being independent.

(ii) We recall from (6.4) and the canonical form (A.7.2.1) that the test region (6.4.5) is really $U_{\mathbf{a}}[n_2\mathbf{a}'Y_1Y_1'\mathbf{a}/n_1\mathbf{a}'Y_2Y_2'\mathbf{a} \geqslant c_{2\alpha}(p, n_1, n_2)$ or $\leqslant c_{1\alpha}(p, n_1, n_2)]$, where Y_1Y_2 have the distribution (A.7.2.1). We also notice from the canonical form (A.7.3·1) that the p functions $n_2\mathbf{a}_i'Y_1Y_1'\mathbf{a}_i/\gamma_i n_2\mathbf{a}_i'Y_2Y_2'\mathbf{a}_i$ $(i = 1, 2, \ldots, p)$ (with $\mathbf{a}_i'$ being a $1\times p$

row vector having 1 for the i-th element and 0 for all other elements) are distributed as p independent F's with d.f. n_1 and n_2 each. Putting these facts together it is easy to see that the test region (6.4.5) includes the union of p regions, each composed of the tail ends of a central F-region, all the p F's being independent.

(iii) From the canonical form (A.7.5.6) and from (6.4) we notice that the test region (6.4.8) will really be $U_{\mathbf{a}}[(n-k)\mathbf{a}'Y^*Y^{*\prime}\mathbf{a}/(k-1)\mathbf{a}'YY'\mathbf{a} \geqslant c_\alpha(p, k-1, n-k)]$ where Y^* and Y have the distribution (A.7.5.6). We also notice from the canonical form (A.7.4.5) that the p-functions $(n-k)\mathbf{a}_i'Y^*Y^{*\prime}\mathbf{a}_i/(k-1)\mathbf{a}_i'YY'\mathbf{a}_i$ $(i = 1, 2, \ldots, p)$ (with $\mathbf{a}_i'$ being a $1 \times p$ row vector having 1 for the i-th element and 0 for all other elements) are distributed as p independent F's out of which at least $p-r$ are central and at the most r are non-central with non-centrality parameters $(\gamma_1, \gamma_2, \ldots, \gamma_r)$ (notice that if $s \leqslant r = \min(p, k-1)$, then, out of these γ's, s will be positive and the rest, i.e., $r-s$ will be zero). Putting these together we observe that the test region (6.4.8) includes the union of p regions, out of which at least $p-r$ are central F-regions and at the most r are non-central F-regions with non-central parameters γ_i's $(i = 1, 2, \ldots, r)$, all F's being independent and each being based on d.f. n_1 and n_2.

(iv) We notice from the canonical form (A.7.3.5) and from (6.4) that the test region (6.4.11) will really be $U_{\mathbf{a},\mathbf{b}}[(\mathbf{a}'Y_1Y_2'\mathbf{b})^2/(\mathbf{a}'Y_1Y_1'\mathbf{a})(\mathbf{b}'Y_2Y_2'\mathbf{b}) \geqslant c_\alpha(p, q, n)]$, where Y_1 and Y_2 have the distribution (A.7.3.5). We notice further that there are p functions $(\mathbf{a}_i'Y_1Y_2'\mathbf{b}_i)^2/(\mathbf{a}_i'Y_1Y_1'\mathbf{a}_i)(\mathbf{b}_i'Y_2Y_2'\mathbf{b}_i)$ $(i = 1, 2, \ldots, p)$ (with $\mathbf{a}_i'(1 \times p)$ being a row vector with 1 for the i-th element and 0 for all other elements and $\mathbf{b}_i'(1 \times q)$ being a row vector with 1 for the i-th element and 0 for all other elements) which are distributed as the squares of p independent correlation coefficients (some of them central and some non-central, the respective non-centrality or deviation parameters being γ_i (notice that out of these p γ_i's, t are positive and the rest, i.e., $p-t$ are zero, where $t \leqslant p \leqslant q$ is the rank of $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$, i.e., of Σ_{12} and all lie between 0 and 1; the positive γ's can be conveniently arranged as $0<\gamma_1<\ldots<\gamma_t<1$).

Putting these together it is easy to check that the test region (6.4.11) includes the union of p regions out of which $(p-t)$ are central correlation (square) regions and t are non-central ones, all being independent and each based on d.f. $(n-1)$. When $q>p$ it is possible to improve on this in the following manner. Pick out linked $\mathbf{a}_i'$ and $\mathbf{b}_i'$ $(i = 1, 2, \ldots, p-1)$ and at the last stage an $\mathbf{a}_p'$ with a set of $\mathbf{b}_i$'s $(i = p, p+1, \ldots, q)$ such that there are p independently distributed correlation squares, of which $(p-1)$ are total correlation (squares) and the last one is a multiple correlation (square) between the p-th variate of the Y_1-set and the $(p, p+1, \ldots, q)$ variates of the Y_2-set. The deviation parameters being γ_i's $(0<\gamma_1<\ldots<\gamma_t<1)$, we could so arrange that the first $p-t$ sample (total) correlation (squares) had zero deviation parameters to go with, the next $t-1$ sample (total) correlation (squares) had respective (one each) deviation parameters $(\gamma_1, \gamma_2, \ldots, \gamma_{t-1})$ to go with, and the last sample (multiple) correlation (square) had γ_t to go with. Notice that the distributions of the square of a correlation (central and non-central) are easily available from those of the central and non-central multiple correlation coefficient (see (7.4.9) and (7.4.20)) by putting $p = 1$.

9.2c. *Actual construction of the lower bounds.* From (9.2b) it is now easy to write down the lower bounds of the power functions (9.1.1)-(9.1.4) of (6.4.2), (6.4.5), (6.4.8) and (6.4.11) as follows:

$$P(\alpha, p, n;\ \gamma_1, \gamma_2, \ldots, \gamma_p) > 1-\prod_{i=1}^{p}[1-P(\chi^2 \geqslant c_{2\alpha}(p, n)/\gamma_i \text{ or } \leqslant c_{1\alpha}(p, n)/\gamma_i)] \quad \ldots \quad (9.2.1)$$

(each χ^2 being based on d.f. n),

$$P(\alpha, p, n_1, n_2;\ \gamma_1, \gamma_2, \ldots, \gamma_p) > 1-\prod_{i=1}^{p}[1-P(F \geqslant c_{2\alpha}(p, n_1, n_2)/\gamma_i$$

or $\leqslant c_{1\alpha}(p, n_1, n_2)/\gamma_i))]$, (each F being based on d.f. n_1 and n_2), ... (9.2.2)

$$P(\alpha, p, k-1, n-k;\ \gamma_1, \gamma_2, \ldots, \gamma_s) > 1-[1-P(\text{central } F \geqslant c_\alpha(p, k-1, n-k)]^{p-s}$$

$$\times \prod_{i=1}^{s}[1-P\ (\text{non-central } F \geqslant c_\alpha\ (p, k-1,\ n-k)\,|\,\gamma_i)], \ \ldots \quad (9.2.3)$$

(each F being based on d.f. $k-1$ and $n-k$), and finally

$$P(\alpha, p, q, n;\ \gamma_1, \gamma_2, \ldots, \gamma_t) > 1-[1-P(r^2 \geqslant c_\alpha(p, q, n)\,|\,\text{null hypothesis}]^{p-t}$$

$$\times \prod_{i=1}^{t}[1-P(r^2 \geqslant c_\alpha(p, q, n)\,|\,\rho_i^2 = \gamma_i)], \quad \ldots \quad (9.2.4)$$

(each r^2 being based on d.f. $(n-1)$.

If $p<q$, it is easy to check from (9.2b) that this lower bound could be improved by the following

$$P(\alpha, p, q, n;\ \gamma_1, \ldots, \gamma_t) > 1-[1-P(r^2 \geqslant c_\alpha(p, q, n)\,|\,\text{null hypothesis})]^{p-t}$$

$$\times \prod_{i=1}^{t-1}[1-P(r^2 \geqslant c_\alpha(p, q, n)\,|\,\rho_i^2 = \gamma_i)][1-P(R^2 \geqslant c_\alpha(p, q, n)\,|\,\rho_t^2 = \gamma_t)], \ \ldots \quad (9.2.5)$$

where all the factors except the last are on squares of (total) correlations distributed with d.f. $n-1$, while the last is on the square of a multiple correlation, distributed with $n-1$ and $q-p$ d.f. and also where, out of the $p-1$ total correlations, $p-t$ are central, $t-1$ are non-central with non-centrality parameters $\gamma_1, \ldots, \gamma_{t-1}$ and the multiple correlation is non-central with the non-centrality parameter γ_t.

To compute in any situation the right sides of (9.2.1)-(9.2.5) we observe that aside from the central, i.e., ordinary, χ^2 and F and the total correlation (squared)-distributions (the last one being transformable to an F-distribution), which are all well known and have their percentage points tabulated, we need, in addition, tables for the c.d.f. of non-central F and non-central multiple correlation (connected with the multivariate normal population). These tables are available in part [11, 22, 50] and could be easily extended with modern computing facilities.

It may be noted that if in (9.2.3) we put $k = 2$, i.e., $s = 0$ or 1, then each side of (9.2.3) is computationally accessible, the left side being the power function of Hotelling's T^2, while the right side is also easily available (in this as in all other cases).

It is of considerable importance at this stage to ask how "good" the lower bounds indicated in (9.2.1), (9.2.2) and (9.2.3) or (9.2.4) are. A lower bound to the power could be said to be "good" if it were (i) close to the actual power, and/or (ii) if it were itself pretty large, being greater than the level of significance α for reasonably large values of the deviation parameters and possibly getting larger as those parameters increase. For all the three tests condition (ii) has been numerically checked to be true over a fairly wide range of test values of the several parameters involved. With regard to condition (i), *in general*, that is, *for small samples*, not only do we not know the actual power (in which case the search for a lower bound would have been redundant) but at the moment we do not even know an upper bound on the expression: (actual power—given lower bound to it)$\div$actual power. In large samples, however, the situation improves and it turns out that the relative error is "small", so that the given lower bounds are "good" also in the sense (i).

CHAPTER TEN

The Monotonic Character of the Lower Bounds on the Power Functions

10.1. *The problem of one dispersion matrix.* For convenience we rewrite (9.2.1) as

$$P(\alpha, p, n; \gamma_1, \gamma_2, \ldots, \gamma_p) > 1 - \prod_{i=1}^{p} P\Big(\frac{c_{1\alpha}(p, n)}{\gamma_i} \leqslant \chi^2 \leqslant \frac{c_{2\alpha}(p, n)}{\gamma_i}\Big), \ldots \quad (10.1.1)$$

each χ^2 being based on d.f. n.

Now denoting, for shortness, the factors in the product on the right side of (10.1.1) by $P_1, \ldots, P_p$, we shall show that $\partial P_i/\partial\left(\frac{1}{\gamma_i}\right)$ is positive or negative according as γ_i is $>$ or $<$ 1, or in other words, P_i decreases as γ_i tends away from 1, provided that c_{1_α} and c_{2_α} are so chosen that

$$\Big[\partial P_i/\partial\left(\frac{1}{\gamma_i}\right)\Big]_{\gamma_i=1} = 0.$$

Proof: Aside from a constant and positive factor of proportionality which is free from γ_i, we have

$$P_i = \int_{c_{1\alpha}/\gamma_i}^{c_{2\alpha}/\gamma_i} e^{-\frac{1}{2}\chi^2}(\chi^2)^{\frac{n-2}{2}} d(\chi^2), \quad \ldots \quad (10.1.2)$$

and thus

$$\frac{\partial P_i}{\partial\left(\frac{1}{\gamma_i}\right)} = e^{-c_{2\alpha}/2\gamma_i}\left(\frac{c_{2\alpha}}{\gamma_i}\right)^{\frac{n-2}{2}} c_{2_\alpha} - e^{-c_{1\alpha}/2\gamma_i}\left(\frac{c_{1\alpha}}{\gamma_i}\right)^{\frac{n-2}{2}} c_{1_\alpha}, \quad \ldots \quad (10.1.3)$$

where
$$e^{-c_{2\alpha}/2}(c_{2_\alpha})^{\frac{n}{2}} - e^{-c_{1\alpha}/2}(c_{1_\alpha})^{\frac{n}{2}} = 0.$$

It is easy to check from (10.1.3) that $\partial P_i/\partial\left(\frac{1}{\gamma_i}\right)$ is positive if $\gamma_i > 1$ and negative if $\gamma_i < 1$ and also that $P_i \to 0$ as $\gamma_i \to \infty$ or $\to 0$.

Thus the right side of (10.1.1) monotonically increases as each γ_i, separately, tends away from unity and the left side, which is the power function of the test, always stays greater than this monotonic function. Furthermore although at H_0, i.e., when all γ_i's $= 1$, this monotonic function is $<\alpha$, it becomes greater than or equal to α for all γ_i's statisfying

$$\prod_{i=1}^{p} P\left(\frac{c_{1\alpha}}{\gamma_i} \leqslant \chi^2 \leqslant \frac{c_{2\alpha}}{\gamma_i}\right) \leqslant 1-\alpha. \qquad \ldots \quad (10.1.4)$$

This means that the test itself is unbiased at least against all alternatives γ_i's satisfying (10.1.4).

10.2. *The problem of two dispersion matrices.* As in the previous case we rewrite (9.2) as

$$P(\alpha, p, n_1, n_2; \gamma_1, \gamma_2, \ldots, \gamma_p) > 1-\prod_{i=1}^{p} P\left(\frac{c_{1\alpha}(p, n_1, n_2)}{\gamma_i} \leqslant F \leqslant \frac{c_{2\alpha}(p, n_1, n_2)}{\gamma_i}\right), \qquad \ldots \quad (10.2.1)$$

each F being based on d.f. n_1 and n_2. Now denoting, for shortness, the factors in the product on the right side of (10.2.1) by $P_1, P_2, \ldots, P_p$ we can show exactly as in the previous case that $\partial P_i/\partial\left(\frac{1}{\gamma_i}\right)$ is positive or negative according as $\gamma_i >$ or < 1, provided that $c_{1\alpha}$ and $c_{2\alpha}$ are so chosen that $\left[\partial P_i/\partial\left(\frac{1}{\gamma_i}\right)\right]_{\gamma_i=1} = 0$. It is also easy to check that $P_i \rightarrow 0$ as $\gamma_i \rightarrow \infty$ or $\rightarrow 0$. Thus, as before, the right side of (10.2.1) monotonically increases to 1 as each γ_i, separately, tends away from unity and the expression to the left side of (10.2.1) which is the power function of the test always stays greater than this monotonic function. As in the previous case it follows here also that the test itself is unbiassed at least against all alternative γ_i's satisfying

$$\prod_{i=1}^{p} P\left\{\frac{c_{1\alpha}}{\gamma_i} \leqslant F \leqslant \frac{c_{2\alpha}}{\gamma_i}\right\} \leqslant 1-\alpha. \qquad \ldots \quad (10.2.2)$$

10.3. *The problem of multivariate analysis of variance.* We rewrite (9.2.3) as

$$P(\alpha, p, k-1, n-k; \gamma_1, \gamma_2, \ldots, \gamma_s) > 1-[P(\text{central } F < c_\alpha(p, k-1, n-k))]^{p-s}$$

$$\times \prod_{i=1}^{s} P(\text{non-central } F < c_\alpha(p, k-1, n-k) \mid \gamma_i), \qquad \ldots \quad (10.3.1)$$

each F being based on d.f. $k-1$ and $n-k$ and $s = \min(p, k-1)$. It is well known that $P(\text{non-central } F < c_\alpha \mid \gamma)$ is a monotonically decreasing function of $|\sqrt{\gamma}|$, which has been and can be proved in various ways, perhaps the simplest proof being the following.

It is well known that with a canonical p.d.f.. we can, except for a constant and positive factor of proportionality not involving γ, write

$$P\,(\text{non-central } F < c_\alpha|\gamma) = \int\limits_D \exp\left\{-\tfrac{1}{2}\left(\sum_{i=1}^{k-1} x_i^2 + \sum_{i=1}^{n-k} y_i^2\right)\right\} \prod_i^{k-1} dx_i \prod_i^{n-k} dy_i, \tag{10.3.2}$$

where the domain of integration D is

$$(x_1+\sqrt{\gamma})^2 + \sum_{i=2}^{k-1} x_i^2 \leqslant c_\alpha \sum_{i=1}^{n-k} y_i^2,$$

and x_i's and y_i's are otherwise capable of going from $-\infty$ to ∞. We can thus rewrite the right side of (10.3.2) as

$$\int\limits_{\substack{x_i(i\,=\,2,\,..,\,k-1)\\ y_i(i\,=\,1,\,..,\,n-k)}} \Pi\, dx_i\, \Pi\, dy_i \exp\left[-\tfrac{1}{2}\left(\sum_{i=1}^{k-1} x_i^2 + \sum_{i=1}^{n-k} y_i^2\right)\right]\left\{\int_{-\sqrt{\gamma}-f}^{-\sqrt{\gamma}+f} \exp(-\tfrac{1}{2}\, x_1^2)dx_1\right\}, \quad \text{...(10.3.3)}$$

where $f^2 = c_a \sum\limits_{i=1}^{n-k} y_i^2 - \sum\limits_{i=2}^{k-1} x_i^2$ and only the positive square root of f^2 is supposed to be taken and the x_i's and y_i's vary from $-\infty$ to ∞ subject to f^2 always staying non-negative. Thus we have $\dfrac{\partial P}{\partial(\sqrt{\gamma})} = \int \Pi dx_i\ \Pi dy_i$
$\exp\left[-\tfrac{1}{2}\left(\sum\limits_{i=2}^{k-1} x_i^2 + \sum\limits_{i=1}^{n-k} y_i^2\right)\right]\left\{\exp\left(-\tfrac{1}{2}(\sqrt{\gamma}+f)^2\right) - \exp\left(-\tfrac{1}{2}(-\sqrt{\gamma}+f)^2\right)\right\}$
Remembering that f is, a.e., positive it is easily seen that according as $\sqrt{\gamma}$ is positive or negative $(\sqrt{\gamma}+f)^2$ is, a.e., $>$ or $<$ $(-\sqrt{\gamma}+f)^2$, so that $\exp\left(-\tfrac{1}{2}(\sqrt{\gamma}+f)^2\right)$ is, a.e., $<$ or $>$ $\exp\left(-\tfrac{1}{2}(-\sqrt{\gamma}+f)^2\right)$, so that $\partial P/\partial(\sqrt{\gamma})$ is negative or positive. This means that P is a monotonically decreasing function of $|\sqrt{\gamma}|$ and it is easy to check that in this case it$\rightarrow 0$ as $|\sqrt{\gamma}| \rightarrow \infty$.

Thus the right side of (10.3.1) is a monotonically increasing function of each $\sqrt{\gamma_i}$ separately, tending to unity as each $|\sqrt{\gamma_i}| \rightarrow \infty$, and the left side of (10.3.1), which is the power function of the test, stays greater than this monotonic function. As in the previous cases the test is unbiassed at least against all alternatives satisfying

$$[P\,(\text{central } F < c_\alpha)]^{p-s} \prod_{i=1}^{s} P\,(\text{non-central } F < c_\alpha|\gamma_i) \leqslant 1-\alpha. \quad \text{...,} \quad (10.3.4)$$

10.4. *The problem of test of independence between two sets of variates.* We rewrite (9.2.5) as

$$P(\alpha, p, q, n;\, \gamma_1, \gamma_2, \ldots, \gamma_p) > 1 - \prod_{i=1}^{p-1} P(\text{non-central } r^2 < c_\alpha(p, q, n)|\gamma_i)$$

$$\times\ P(\text{non-central } R^2 < c_\alpha(p, q, n)|\gamma_p), \quad \ldots \quad (10.4.1)$$

where all r^2's are based on d.f. $n-1$ and R^2 is the square of a multiple correlation based on d.f. $n-1$ and $q-p$ and where γ_p is the largest population canonical correlation coefficient. Notice that for a particular alternative some of the γ's might be zero and in any case the γ's vary from 0 to 1. As in the previous case it is well known and can be proved in various ways that both $P(\text{non-central } r^2 < c_\alpha | \gamma)$ and $P(\text{non-central } R^2 < c_\alpha | \gamma)$ are each a monotonically decreasing function of $|\sqrt{\gamma}|$, which$\rightarrow 0$ as $|\sqrt{\gamma}| \rightarrow 1$. The simplest proof of this theorem can be developed exactly on the same lines as in the previous case. But this need not be spelled out here.

Thus the right side of (10.4.1) is a monotonically increasing function of each $|\sqrt{\gamma_i}|$ separately, tending to unity as each $|\sqrt{\gamma_i}| \rightarrow 1$, and the left side of (10.4.1), which is the power function of the test, stays greater than this monotonic function. As in the previous case, this test is unbiassed at least against all alternatives satisfying

$$\prod_{i=1}^{p-1} P(\text{non-central } r^2 < c_\alpha | \gamma_i) \times P(\text{non-central } R^2 < c_\alpha | \gamma_p) \leqslant 1-\alpha. \quad \ldots \quad (10.4.2)$$

CHAPTER ELEVEN

Other Monotonic Lower Bounds on the Power Functions*

11.1. *Multivariate analysis of variance test.* We start from the canonical form (A.7.5.5) and denote by c_r the largest characteristic root of $(Y_1\, Y_1')(Y_2\, Y_2')^{-1}$, by H_0 the $H(\gamma_i = 0)$ $(i = 1, 2, \ldots, s)$ and by H its complement, and observe that for a given positive c_0, $P(c_r \leqslant c_0 | H) =$ a function of $\gamma_1, \gamma_2, \ldots, \gamma_s = \psi_1(\gamma_1, \gamma_2, \ldots, \gamma_s)$, say. We shall prove that (11.1.1) $P(c_s \leqslant c_0 | H)$, i.e., $\psi_1(\gamma, \gamma_2, \ldots, \gamma_s)$ *stays less than a monotonically decreasing function of each* $|\sqrt{\gamma_i}|$ *separately* (*notice that each* $\gamma_i \geqslant 0$), *which is different from the decreasing function on the right side of* (10.3.1).

Proof: We recall from (A.2.2) that the largest characteristic root c_r of $(Y_1\, Y_1')\,(Y_2\, Y_2')^{-1}$ can be written as $\text{Sup}_{\mathbf{a}}(\mathbf{a}'Y_1\, Y_1'\mathbf{a})/(\mathbf{a}'Y_2 Y_2'\mathbf{a})$ and the domain $c_r \leqslant c_0$ can be rewritten as

$$\text{Sup}_{\mathbf{a}}\ (\mathbf{a}'Y_1\, Y_1'\mathbf{a})/(\mathbf{a}'Y_2Y_2'\mathbf{a}) \leqslant c_0, \text{ or } \textstyle\bigcap_{\mathbf{a}}[(\mathbf{a}'Y_1Y_1'\mathbf{a}/\mathbf{a}'Y_2Y_2'\mathbf{a}) \leqslant c_0]. \quad \ldots \quad (11.1.2)$$

It is easy to see now that the canonical p.d.f. based on (A.7.5.5) can be rewritten as

$$\text{Const}\ \exp\left[-\tfrac{1}{2}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_1} x_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{n_2} y_{ij}^2\right)\right] \quad \ldots \quad (11.1.3)$$

and region (11.1.2) can be rewritten as

$$\text{Sup}_{\mathbf{a}}[\mathbf{a}'(X+\delta)(X'+\delta')\mathbf{a}/\mathbf{a}'YY'\mathbf{a}] \leqslant c_0, \text{ or}$$

$$\textstyle\bigcap_{\mathbf{a}}[\mathbf{a}'(X+\delta)(X'+\delta')\mathbf{a}/\mathbf{a}'YY'\mathbf{a} \leqslant c_0], \quad \ldots \quad (11.1.4)$$

where $\delta(p \times n_1)$ is such that $\delta_{ij} = \sqrt{\gamma_i}$(if $i=j=1, 2, \ldots, s$) and $= 0$ otherwise, and where

$$Y_1(p \times n_1) = X(p \times n_1) + \delta(p \times n_1) \text{ and } Y_2(p \times n_2) = Y(p \times n_2). \quad \ldots \quad (11.1.5)$$

Notice that all the components of X and Y will vary from $-\infty$ to ∞. Notice also that $r = \min(p, n_1)$, and s, i.e., the number of non-zero population roots might go up to r. Observe further that the constant factor in (11.1.3) does not, in this case, involve the γ_i's. The problem is now one of integrating (11.1.3) over the domain (11.1.4) (which let us call $\bigcap_{\mathbf{a}}$, for shortness) and showing that the integral stays greater than a monotonically decreasing function of each γ_i or $|\sqrt{\gamma_i}|$, separately. It will

---

* See reference [35] in this connection.

suffice to show the monotonic character of this integral with respect to variation of, say, $|\sqrt{\gamma_1}|$. To this end, remembering that $\mathbf{a}'$ is a non-null row vector $(a_1, a_2, \ldots, a_p)$ we might, without any loss of generality, put $a_1 = 1$ and rewrite (11.1.4) as

$$\bigcap_{\mathbf{a}}[\{(x_{11}+\sqrt{\gamma_1})+a_2x_{21}+\ldots+a_px_{p1}\}^2+\sum_{j=2}^{n_1-1}\{\sum_{i=1}^{p}a_i(x_{ij}+\delta_{ij})\}^2$$

$$\leqslant c_0\sum_{j=1}^{n_2}(\sum_{i=1}^{p}a_iy_{ij})^2], \qquad \ldots \quad (11.1.6)$$

where $\delta_{ij} = \sqrt{\gamma_i}$ (if $i = j = 1, 2, \ldots, s$) and $= 0$ otherwise, and where $a_1 = 1$. To carry out the integration of (11.1.3) over (11.1.6), we first integrate out over x_{11} and then check the total integral, which we call I_1, is proportional to

$$I_1 = \int\Big[\int_{\mathrm{Sup}_{\mathbf{a}}\, l_{2\mathbf{a}}}^{\mathrm{Inf}_{\mathbf{a}}\, l_{1\mathbf{a}}} \exp(-\tfrac{1}{2}x_{11}^2)dx_{11}\Big]\exp[-\tfrac{1}{2}(\Sigma x_{i1}^2+\Sigma z_{ij}^2+\Sigma y_{ij}^2)]\, dx_{i1}dz_{ij}dy_{ij}, \qquad \ldots \quad (11.1.7)$$

the symbols being defined in the following way. For y_{ij}'s, as in (10.1.3), $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, n_2$, but for x_{i1}'s $i = 2, 3, \ldots, p$. Also $z_{ij} = x_{ij}$ with $i = 1, 2, \ldots, p$ and $j = 2, 3, \ldots, k-1$, and

$$l_{1\mathbf{a}} = [-\sqrt{\gamma_1}-f_{1\mathbf{a}}+f_{2\mathbf{a}}] \text{ and } l_{2\mathbf{a}} = [-\sqrt{\gamma_1}-f_{1\mathbf{a}}-f_{2\mathbf{a}}], \qquad \ldots \quad (11.1.8)$$

where $$f_{1\mathbf{a}} = \sum_i a_ix_{i1} \text{ and } f_{2\mathbf{a}} = [c_0\sum_j(\sum_i a_iy_{ij})^2-\sum_j(\sum_i a_i(z_{ij}+\delta_{ij})^2]^{\frac{1}{2}}. \qquad \ldots \quad (11.1.9)$$

Furthermore, (i) the constant of proportionality in (11.1.7) is free from γ_i's, (ii) $x_{21}, \ldots, x_{p1}$ vary from $-\infty$ to ∞ while y_{ij}'s and z_{ij}'s from $-\infty$ to ∞ subject to $f_{2\mathbf{a}}$ always staying real, (iii) for $f_{2\mathbf{a}}$ only the positive square root is to be taken, (iv) $f_{1\mathbf{a}}$ and $f_{2\mathbf{a}}$ are free from γ_1. Now with $a_1 = 1$, let $\mathbf{a}^*$ denote the value of $\mathbf{a}$ for which $f_{2\mathbf{a}}$ is a minimum. Then it is clear that this $\mathbf{a}^*$ is free from γ_1 and x_{i1}'s and is a function of z_{ij}'s, y_{ij}'s, c_0 and possibly also of δ_{ij}'s. Notice that $a_1^* = 1$. Also let $l_{1\mathbf{a}^*}$ and $l_{2\mathbf{a}^*}$ stand for the values of $l_{1\mathbf{a}}$ and $l_{2\mathbf{a}}$ on substitution of $\mathbf{a}^*$ for $\mathbf{a}$. It is now clear that $Inf_{\mathbf{a}}l_{1\mathbf{a}} < l_{1\mathbf{a}^*}$, $\mathrm{Sup}\,_{\mathbf{a}}l_{2\mathbf{a}} > l_{2\mathbf{a}^*}$, so that Interval $[\mathrm{Sup}\,_{\mathbf{a}}l_{2\mathbf{a}}, \mathrm{Inf}\,_{\mathbf{a}}l_{1\mathbf{a}}] <$ Interval $[l_{2\mathbf{a}^*}, l_{1\mathbf{a}^*}]$.

Let us now introduce an I_1^* such that, aside from a constant and positive factor of proportionality (the same as for I_1) it is defined by

$$I_1^* = \int\Big[\int_{l_{2\mathbf{a}^*}}^{l_{1\mathbf{a}^*}} \exp(-\tfrac{1}{2}x_{11}^2)dx_{11}\Big]\exp\{-\tfrac{1}{2}(\Sigma x_{i1}^2+\sum_{i,j} z_{ij}^2+\sum_{i,j} y_{ij}^2)\}$$

$$\prod_i\prod_j dx_{i1}\,dz_{ij}\,dy_{ij}. \qquad \ldots \quad (11.1.10)$$

It will be seen that, while I_1 is the integral of the, a.e., positive function (11.1.3) over the domain (11.1.6) which is the intersection of a class of domains, I_1^* is the integral of the same, a.e., positive function (11.1.3) over the intersection of a subclass of the previous class. In fact, the subclass is formed by excluding from (11.1.6) all a's for which $\text{Inf}_{\mathbf{a}} l_{1\mathbf{a}} \leqslant l_{1\mathbf{a}} < l_{1\mathbf{a}*}$ and/or $l_{2\mathbf{a}*} < l_2 \leqslant \text{Sup}_{\mathbf{a}} l_{2\mathbf{a}}$. This shows that $I_1 < I_1^*$.

It is now easy to check that, aside from a constant and positive factor of proportionality, we have

$$\frac{\partial I_1^*}{\partial(\sqrt{\gamma_1})} = \int [\exp(-\tfrac{1}{2}l_{2\mathbf{a}*}^2) - \exp(-\tfrac{1}{2}l_{1\mathbf{a}*}^2)][\exp\{-\tfrac{1}{2}(\sum_i x_{i1}^2 + \sum_{i,j} z_{ij}^2 + \sum_{i,j} y_{ij}^2)\}] \Pi\, dx_{i1} dz_{ij} dy_{ij}$$

$$= \int [\exp\{-\tfrac{1}{2}(f_{1\mathbf{a}*} + \sqrt{\gamma_1} + f_{2\mathbf{a}*})\}^2 - \exp\{-\tfrac{1}{2}(f_{1\mathbf{a}*} + \sqrt{\gamma_1} - f_{2\mathbf{a}*})\}^2]$$

$$\times [\exp\{-\tfrac{1}{2}(\sum_i x_{i1}^2 + \sum_{i,j} z_{ij}^2 + \sum_{i,j} y_{ij}^2)\}] \prod_i \prod_j dx_{i1}\, dz_{ij}\, dy_{ij}, \qquad \ldots \quad (11.1.11)$$

by using (11.1.9). The domain of variation of x_{i1}'s, z_{ij}'s and y_{ij}'s has been already defined immediately after (11.1.9). It will be proved that the expression on the right side of (11.1.11) is negative for positive values of $\sqrt{\gamma_1}$ and positive for negative values of $\sqrt{\gamma_1}$, or, in other words, I_1^* is a monotonically decreasing function of $|\sqrt{\gamma_1}|$. To prove this we proceed as follows.

We recall from the remarks preceding (11.1.9) that $f_{2\mathbf{a}*}$ is a function of z_{ij}'s, y_{ij}'s, c_0 and possibly also of the other δ_{ij}'s, while $f_{1\mathbf{a}*}$ is just a linear function of x_{i1}'s with a coefficient vector $\mathbf{a}^*$ which is a function of the same quantities that occur in $f_{2\mathbf{a}*}$. Thus, since x_{i1}'s are each a $N(0, 1)$, therefore, the conditional distribution of $f_{1\mathbf{a}*}$, given $\mathbf{a}^*$, that is, given z_{ij}'s and y_{ij}'s, is normal with zero mean and variance $\sigma_{\mathbf{a}*}^2 = \sum_{i=1}^{p} a_i^{*2}$.

Therefore, aside from a constant and positive factor of proportionality, we can rewrite (11.1.11) as

$$\frac{\partial I_1^*}{\partial(\sqrt{\gamma_1})} = \int \Big[\exp\Big\{-\tfrac{1}{2}(f_{1\mathbf{a}*} + \sqrt{\gamma_1} + f_{2\mathbf{a}*})^2\Big\} - \exp\Big\{-\tfrac{1}{2}(f_{1\mathbf{a}*} + \sqrt{\gamma_1} - f_{2\mathbf{a}*})^2\Big\}\Big]$$

$$\times \exp\Big(-\frac{1}{2\sigma_{\mathbf{a}*}^2} f_{1\mathbf{a}*}^2\Big) df_{1\mathbf{a}*} \exp\Big\{-\tfrac{1}{2}(\Sigma z_{ij}^2 + \Sigma y_{ij}^2)\Big\} \prod_i \prod_j dz_{ij} dy_{ij}. \qquad \ldots \quad (11.1.12)$$

Integrating out over $f_{1\mathbf{a}*}$ it is easy to check that the right side reduces to

$$\int^{\mathbf{a}} \Big[\exp\Big\{-\frac{1}{2(1+\sigma_{\mathbf{a}*}^2)}(\sqrt{\gamma_1} + f_{2\mathbf{a}*})^2\Big\} - \exp\Big\{-\frac{1}{2(1+\sigma_{\mathbf{a}*}^2)}(\sqrt{\gamma_1} - f_{2\mathbf{a}*})^2\Big\}\Big]$$

$$\times \exp\Big\{-\tfrac{1}{2}(\Sigma z_{ij}^2 + \Sigma y_{ij}^2)\Big\} \prod_i \prod_j dz_{ij} dy_{ij}. \qquad \ldots \quad (11.1.13)$$

Remembering that $f_{2\mathbf{a}*}$ is, a.e., positive, it is now easy to check that, according as $\sqrt{\gamma_1}$ is positive or negative, we have a.e., $(\sqrt{\gamma_1} + f_{2\mathbf{a}*})^2 >$ or $< (\sqrt{\gamma_1} - f_{2\mathbf{a}*})^2$, that is, a.e.,

$$\exp\Big\{-\frac{1}{2(1+\sigma_{\mathbf{a}*}^2)}(\sqrt{\gamma_1} + f_{2\mathbf{a}*})^2\Big\} < \text{ or } > \exp\Big\{-\frac{1}{2(1+\sigma_{\mathbf{a}*}^2)}(\sqrt{\gamma_1} - f_{2\mathbf{a}*})^2\Big\} \ldots (11.1.14)$$

Thus, the integral (11.1.13) is negative or positive according as $\sqrt{\gamma_1}$ is positive or negative, which proves that I_1^* is a monotonically decreasing function of each $|\sqrt{\gamma_i}|$ separately, so that the power of the test stays greater than a monotonically increasing function of each $|\sqrt{\gamma_i}|$ separately and is unbiassed at least against all alternatives γ_i's for which $I_1^* \leqslant 1-\alpha$.

11.2. *Test of independence between two sets of variates.* We start from the canonical form (A.7.3.5), denote by c_p the largest characteristic root of $(Y_1Y_1')^{-1}(Y_1Y_2')(Y_2Y_2')^{-1}(Y_2Y_1')$, by H_0 the $H(\gamma_i = 0)$ $(i = 1, 2, \ldots, p)$ and by H its complement and then observe that, for a given $c_0(<1)$, $P(c_p \leqslant c_0|H) =$ a function of $\gamma_1, \ldots, \gamma_p$ $= \psi_2(\gamma_1, \ldots, \gamma_p)$, say. We shall prove that

$$P(c_p \leqslant c_0|H), \text{ i.e., } \psi_2(\gamma_1, \ldots, \gamma_p), \qquad \ldots \quad (11.2.1)$$

stays less than a monotonically decreasing function of each $|\sqrt{\gamma_i}|$ separately (notice that each $\gamma_i \geqslant 0$ and <1), which is different from the decreasing function on the right side of (11.4.1).

Proof: It will suffice to prove this monotonicity with respect to any one parameter, say γ_1. Toward this end we proceed as follows. We first rewrite the canonical p.d.f. based on (A.7.3.5) in the expanded form

$$\text{Const exp}\left[-\tfrac{1}{2}\left\{\sum_{i=1}^{p}\frac{1}{1-\gamma_i}\sum_{k=1}^{n}(x_{ik}^2+y_{ik}^2-2\gamma_i^{\frac{1}{2}}x_{ik}y_{ik})+\sum_{i=p+1}^{q}\sum_{k=1}^{n}y_{ik}^2\right\}\right] \qquad \ldots \quad (11.2.2)$$

by putting $Y_1 = X$ and $Y_2 = Y$ (the elements of the latter matrices being x_{ik} and y_{ik}) and letting all new variates also vary from $-\infty$ to ∞. We next use (A.3.9) to find a triangular $\tilde{U}(q\times q)$ such that

$$YY' = \tilde{U}\tilde{U}' \text{ and } u_{ij} = 0 \text{ if } j > i\,(j = 2, 3, \ldots, q). \qquad \ldots \quad (11.2.3)$$

We recall from (A.3.9) that given Y, the elements of $\tilde{U}$ can be uniquely determined by adopting a convention, say, that $u_{ii} > 0$ $(i = 1, 2, \ldots, q)$, provided that Y is of rank q, as it will, almost everywhere, be. Now (see (A.3.15)) it is possible to choose an orthogonal transformation: $X(p\times n)$ $\Gamma(n\times n) = X^*$ and $Y = Y$ (notice that although Γ might involve Y, yet the Jacobian is 1), such that

$$XX' = X^*X^{*\prime},\ YY' = YY' \text{ and } XY' = X^*\left[\begin{array}{c} I(q) \\ \hline 0(n-q\times q)\end{array}\right]\tilde{U}' \qquad \ldots \quad (11.2.4)$$

This is easily seen if we put $Y(q\times n) = \tilde{U}(q\times q)L(q\times n)$ (where $LL' = I(q)$, complete $L(q\times n)$ into an orthogonal matrix $\underset{n}{\begin{bmatrix} L \\ M\end{bmatrix}}\begin{matrix} q \\ n-q\end{matrix} = \Gamma'$, say and then put $X = X^*\Gamma'$, so that $XX' = X^*X^{*\prime}$ and $XY' = X^*\begin{bmatrix} L \\ M\end{bmatrix}L'\tilde{U}' = X^*\left[\begin{array}{c} I(q) \\ \hline 0(n-q\times q)\end{array}\right]\tilde{U}'$. The p.d.f. of X^*, Y can now be conveniently written as

$$\text{Const exp}\left[-\tfrac{1}{2}\sum_{k=1}^{n}\sum_{i=1}^{p}(x_{ik}^*-\rho_{ik}u_{ik})^2/(1-\rho_{ik}^2)-\tfrac{1}{2}\sum_{i\geqslant j=1}^{q}u_{ij}^2\right], \qquad \ldots \quad (11.2.5)$$

where $\rho_{ik} = 0$ if $k > i$ and $= \sqrt{\gamma_i}$ if $k \leqslant i$. We now put

$$(x^*_{ik}-\rho_{ik}u_{ik})/(1-\rho^2_{ik})^{\frac{1}{2}} = z_{ik} \qquad \dots \quad (11.2.6)$$

(being the elements of a matrix $Z(p\times n)$) and $\beta_{ik} = \rho_{ik}/(1-\rho^2_{ik})^{\frac{1}{2}}$, and obtain the p.d.f. of z_{ik} and y_{ik} (which vary from $-\infty$ to ∞) in the form

$$\text{Const } \exp\left[-\tfrac{1}{2}\left(\sum_{k=1}^{n}\sum_{i=1}^{p} z^2_{ik}+\sum_{i\geqslant j=1}^{q} u^2_{ij}\right)\right], \qquad \dots \quad (11.2.7)$$

where u_{ij}'s are given in terms of y_{ik}'s by (11.2.3). Notice from (11.2.2), (11.2.4) and (11.2.6) that finally

$$(Y_1Y_1')_{ii'} = \sum_{k=1}^{n}(z_{ik}+\beta_{ik}u_{ik})(z_{i'k}+\beta_{i'k}u_{i'k})(1-\rho^2_{ik})^{\frac{1}{2}}(1-\rho^2_{i'k})^{\frac{1}{2}}, \qquad \dots \quad (11.2.8)$$

$$(Y_1Y_2')_{ij} = \sum_{k=1}^{n}(z_{ik}+\beta_{ik}u_{ik})(1-\rho^2_{ik})^{\frac{1}{2}}u_{jk} \text{ and } (Y_2\,Y_2')_{jj'} = \sum_{k=1}^{min(j,j')} u_{jk}u_{j'k}, \qquad \dots \quad (11.2.8)$$

$(i, i' = 1, 2, \dots, p;\ j, j' = 1, 2, \dots, q)$. Next we recall from (A.2.3) that the largest characteristic root c_p of $(Y_1\,Y_1')^{-1}(Y_1Y_2')(Y_2\,Y_2')^{-1}(Y_2\,Y_1')$ can be written as

$\text{Sup}_{\,\mathbf{a},\mathbf{b}}(\mathbf{a}'Y_1\,Y_2'\mathbf{b})^2/(\mathbf{a}'Y_1Y_1'\mathbf{a})(\mathbf{b}'Y_2Y_2'\mathbf{b})$ and the domain $c_p \leqslant c_0$ as

$$\text{Sup}_{\,\mathbf{a},\mathbf{b}}(\mathbf{a}'Y_1Y_2'\mathbf{b})^2/(\mathbf{a}Y_1\,Y_1'\mathbf{a})(\mathbf{b}'Y_2\,Y_2'\mathbf{b}) \leqslant c_0 \qquad \dots \quad (11.2.9)$$

or $\bigcap_{\mathbf{a},\mathbf{b}}[(\mathbf{a}'Y_1\,Y_2'\mathbf{b})^2/(\mathbf{a}'Y_1\,Y_1'\mathbf{a})(\mathbf{b}'Y_2\,Y_2'\mathbf{b}) \leqslant c_0]$

or alternatively as

$$\bigcap_{\mathbf{a},\mathbf{b}}[(\mathbf{a}'Y_1Y_2'\mathbf{b})^2/\{(\mathbf{a}'Y_1\,Y_1'\mathbf{a})(\mathbf{b}'Y_2\,Y_2'\mathbf{b})-(\mathbf{a}'Y_1\,Y_2'\mathbf{b})^2\} \leqslant c_0' \text{ (say)}] \qquad \dots \quad (11.2.10)$$

or, using (11.2.8), as

$$\bigcap_{\mathbf{a},\mathbf{b}}[\{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{j} a_i(z_{ik}+\beta_{ik}u_{ik})b_ju_{jk}\}^2 \leqslant c_0'\ \{(\sum_{k=1}^{p}(\sum_{j=k}^{q} b_ju_{jk})^2)\}$$

$$\times\ [\sum_{i=1}^{n}(\sum_{i=1}^{p} a_i(z_{ik}+\beta_{ik}u_{ik}))^2]-\text{expression on the left of} \leqslant \text{in (11.2.11)}\}. \qquad \dots \quad (11.2.11)$$

Now, by using certain standard inequalities, and taking the intersection over **b**, it is easy to check that (11.2.11) reduces to

$$\bigcap_{\mathbf{a}}\left[\sum_{k=1}^{q}\left(\sum_{i=1}^{p} a_i\,(z_{ik}+\beta_{ik}u_{ik})\right)^2 \leqslant c_0' \sum_{k=q+1}^{n}\left(\sum_{i=1}^{p} a_iz_{ik}\right)^2\right], \qquad \dots \quad (11.2.12)$$

in which, without any loss of generality, we can take $a_1 = 1$. The problem now is one of integrating out (11.2.7) over (11.2.12), the u_{ij}'s being given by (11.2.3). To carry out the integration of (11.2.7) over (11.2.12) we proceed exactly as in the previous case, integrate out over z_{11} and then check that, aside from a constant and positive factor of proportionality, the total integral which we call I_2 is given by [see (11.1.7)]

$$I_2 = \int \Big[\int_{\mathrm{Sup}\, l_{2a}}^{\mathrm{Inf}_a l_{1a}} \exp\,(-\tfrac{1}{2}z_{11}^2)dz_{11}\Big] \exp\Big\{-\tfrac{1}{2}(\Sigma z_{ik}^2+\Sigma u_{ik}^2)\Big\}\,\Pi dz_{ik}du_{ik}. \qquad \ldots \quad (11.2.13)$$

In (11.2.13), (i) z_{11} is omitted from z_{ik}'s, (ii) u_{ik}'s (for $i \neq k$) and z_{ik}'s vary from $-\infty$ to ∞ and u_{ii}'s from 0 to ∞, all subject to l_{1a} and l_{2a} staying real and (iii) l_{1a} and l_{2a} are given by

$$l_{1a} = [-\sqrt{\gamma_1/(1-\gamma_1)}\,u_{11}-f_{1a}+f_{2a}] \text{ and } l_{2a}=[-\sqrt{\gamma_1/(1-\gamma_1)}u_{11}-f_{1a}-f_{2a}], \quad \ldots \quad (11.2.14)$$

in which f_{1a} and f_{2a} are defined by

$$f_{1a} = \sum_{i=2}^{p} a_i(z_{i1}+\beta_{i1}u_{i1})$$

and

$$f_2 = [c' \sum_{k=q+1}^{n} (\sum_{i=1}^{p} a_i z_{ik})^2-\sum_{k=2}^{q} (\sum_{i=1}^{p} (z_{ik}+\beta_{ik}u_{ik}))^2]^{\frac{1}{2}}. \qquad \ldots \quad (11.2.15)$$

Arguing now exactly in the same manner as in subsection (11.1) we can establish that I_2 stays less than a monotonically decreasing function I_2^* and thus the power of the test stays greater than a monotonically increasing function of $|\sqrt{\gamma_1/(1-\gamma_1)}|$, that is, of $|\gamma_1^{\frac{1}{2}}|$, that is, of any $|\gamma_i^{\frac{1}{2}}|$, from considerations of symmetry. Also the test is unbiassed against at least all alternatives γ_i's for which $I_2^* \leqslant 1-\alpha$.

There are reasons to believe that the lower bounds indicated in sections 11.1 and 11.2 are closer than the lower bounds for the corresponding problems, indicated in sections 10.3 and 10.4.

11.3. *Test of independence between two sets of variates under the regression model of* (4.25)-(4.33). It will be observed from section (7.7) that the distribution of the roots (and therefore that of the largest root) in this case can be identified with that of case (i) of (7.6) when $p \leqslant q$ and with that of case (ii) of (7.6) when $p > q$, in both cases, by putting $n_1 = q$ and $n_2 = n$. It will be also observed that this identification holds for the distributions on both the null and the non-null hypothesis. We have, therefore, exactly the same kind of monotonicity property in this situation as in the case (11.2) and no separate proof need, therefore, be given for this case.

11.4. *Modified test for the equality of two dispersion matrices against a special class of alternatives.* We take over from (6.4.5) the acceptance region for the hypothesis $H_0 : \Sigma_1 = \Sigma_2$ and rewrite it as

$$c_{1\alpha}(p, n_1, n_2) \leqslant \text{all } c_i\text{'s} \leqslant c_{2\alpha}(p, n_1, n_2), \qquad \ldots \quad (11.4.1)$$

where $c_{1\alpha}$ and $c_{2\alpha}$ are so chosen as to satisfy

$$P(c_{1\alpha} \leqslant \text{all } c_i\text{'s} \leqslant c_{2\alpha} | \Sigma_1 = \Sigma_2) = 1-\alpha \qquad \ldots \quad (11.4.2)$$

and

$$\left[\frac{\partial P(c_{1\alpha} \leqslant \text{all } c_i\text{'s} \leqslant c_{2\alpha} \mid \Sigma_1 \neq \Sigma_2)}{\partial \gamma_i}\right]_{\gamma_1 = \gamma_2 = \ldots = \gamma_p = 1} = 0 \; (i = 1, 2, \ldots, p)$$

or

$$\left[\frac{\partial P(c_{1\alpha}, c_{2\alpha}, \gamma_1, \ldots, \gamma_p)}{\partial \gamma_i}\right]_{\gamma_1 = \ldots = \gamma_p = 1} = 0 \quad (i = 1, 2, \ldots, p), \qquad \ldots \quad (11.4.3)$$

remembering that if $\Sigma_1 \neq \Sigma_2$, the probability P is, aside from the degrees of freedom n_1 and n_2 and the limits $c_{1\alpha}$ and $c_{2\alpha}$, purely a function of γ_i's, the characteristic roots of $\Sigma_1 \Sigma_2^{-1}$. It will be shown here that if $\gamma_1 = \ldots = \gamma_p = \gamma$ (say) (which means that $\Sigma_1 \Sigma_2^{-1}$ itself is equal to $\gamma I(p)$, i.e., $\Sigma_1 = \gamma\Sigma_2$), then P monotonically decreases, i.e., the power of the test monotonically increases as this common γ tends away from 1 which is the value of γ on the null hypothesis $\Sigma_1 = \Sigma_2$.

Proof: We start from the canonical probability

$$\left[\text{Const} \Big/ \prod_{i=1}^{p} \gamma_i^{\frac{n_1}{2}}\right] \exp\left[-\tfrac{1}{2} \operatorname{tr}(D_{1/\gamma_i} X_1 X_1' + X_2 X_2')\right] dX_1\, dX_2,$$

where the constant factor is a pure constant not involving the parameters, D_{1/γ_i} stands for a diagonal matrix whose diagonal elements are $1/\gamma_1, \ldots, 1/\gamma_p$, X_1 and X_2 are $p \times n_1$ and $p \times n_2$ $(p < n_1, n_2)$ and where the c_i's of (11.4.1) are the roots of $X_1 X_1'(X_2 X_2')^{-1}$. We first show that the p equations under (11.4.3) are really equivalent to one equation. To prove this we note that aside from the constant factor,

$$P = \int\limits_{c_{1\alpha} \leqslant \text{ all } c[X_1 X_1'(X_2 X_2')^{-1}] \leqslant c_{2\alpha}} \prod_{i=1}^{p} (1/\gamma_i)^{\frac{n_1}{2}} \exp\left[-\tfrac{1}{2} \operatorname{tr}(D_{1/\gamma_i} X_1 X_1' + X_2 X_2')\right] \times dX_1\, dX_2. \qquad \ldots \quad (11.4.4)$$

Hence

$$\frac{\partial P}{\partial\left(\frac{1}{\gamma_i}\right)} = \int\limits_{c_{1\alpha} \leqslant \text{ all } c[X_1 X_1'(X_2 X_2')^{-1}] \leqslant c_{2\alpha}} \prod_{j=1}^{p} (1/\gamma_j)^{\frac{n_1}{2}} \left[\frac{n_1}{2}\gamma_i - \tfrac{1}{2}(D_{1/\gamma_j} X_1 X_1')_{ii}\right]$$

$$\times \quad \exp\left[-\tfrac{1}{2} \operatorname{tr}(D_{1/\gamma_j} X_1 X_1' + X_2 X_2')\right] dX_1\, dX_2. \qquad \ldots \quad (11.4.5)$$

Now using the transformation

$$\begin{aligned} X_1(p \times n_1) &= U(p \times p) D_{\sqrt{c}}(p \times p) L_1(p \times n_1) \\ \text{and} \quad X_2(p \times n_2) &= U(p \times p) L_2(p \times n_2), \end{aligned} \qquad (11.4.6)$$

where U is non-singular and $L_1L_1' = L_2 L_2' = I(p)$, and integrating out over L_{1I} and L_{2I}, we observe that, aside from a positive and constant factor of proportionality, (11.4.5) reduces to

$$\frac{\partial P}{\partial(1/\gamma_i)} = \int_D \prod_{j=1}^{p} (1/\gamma_j)^{\frac{n_1}{2}} \left[\frac{n_1}{2}\gamma_i - \tfrac{1}{2}(D_{1/\gamma_j} U D_{c_k} U')_{ii}\right] \qquad \dots \quad (11.4.7)$$

$$\times \quad \exp\left[-\tfrac{1}{2}\operatorname{tr}(D_{1/\gamma_j} U D_{c_k} U' + UU')\right] |U|^{n_1+n_2-p}\, dU \prod_{k=1}^{p} c_k^{\frac{n_1-2}{2}}\, dc_k \prod_{k>k'} (c_k - c_{k'}),$$

where D_{c_k} stands for a diagonal matrix with diagonal elements $c_1, \dots, c_p$ and the domain D is

$$c_{1\alpha} \leqslant \text{all } c_i\text{'s} \leqslant c_{2\alpha} \text{ and } -\infty < \text{all } u_{ij}\text{'s} < \infty. \qquad \dots \quad (11.4.8)$$

We have thus

$$\left\{\frac{\partial P}{\partial(1/\gamma_i)}\right\}_{\gamma_1 = \dots = \gamma_p = 1} = \int_D \left[\frac{n_1}{2} - \tfrac{1}{2}(UD_{c_j}U')_{ii}\right] \exp\left[-\tfrac{1}{2}\operatorname{tr}(UD_{c_j}U' + UU')\right]$$

$$\times \quad |U|^{n_1+n_2-p}\, dU \prod_{j=1}^{p} c_j^{\frac{n_1-p-1}{2}}\, dc_j \prod_{j>j'} (c_j - c_{j'}). \qquad \dots \quad (11.4.9)$$

Having regard to the definition of the domain given by (11.4.8) and the structure of the integral on the right side of (11.4.9) it is easy to check that this integral is invariant under a change of the subscript i, so that the expression on the left side of (11.4.9) is the same for $i = 1, 2, \dots, p$ and hence the p equations (11.4.3) are equivalent to really one equation. Now adding p formally different looking integrals like the right side of (11.4.9) over $i = 1, 2, \dots, p$ and cancelling a factor we see that (11.4.3) is really equivalent to

$$\int_D \left[1 - \frac{1}{p}\operatorname{tr} UD_{c_i}U'\right] \exp\left[-\tfrac{1}{2}\operatorname{tr}(UD_{c_i}U' + UU')\right] |U|^{n_1+n_2-p} dU$$

$$\times \prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}}\, dc_i \prod_{i>j} (c_i - c_j) = 0. \qquad \dots \quad (11.4.10)$$

It is easy to check that the left side of (11.4.10) is the same as if we had put all γ_i's $= \gamma$ in (11.4.4) and then differentiated the integral with respect to γ and then put $\gamma = 1$. As will be presently seen this will enable us to rewrite (11.4.10) in a simpler form (which can also be derived in a straightforward though rather lengthier manner). At this point, remembering the definition of D by (11.4.8), we merely observe that (11.4.10) gives one relation between $c_{1\alpha}$ and $c_{2\alpha}$ which we call the *condition of local unbiassedness* and then (11.4.2) added to this determines $c_{1\alpha}$ and $c_{2\alpha}$ completely.

Going back now to the problem of proving the monotonicity of the integral on the right of (11.4.4) under the special assumption that $\gamma_1 = \ldots = \gamma_p = \gamma$ (say), we proceed as follows.

Putting $D_{1/\sqrt{\gamma_i}} X_1 = Y_1$ and $X_2 = Y_2$ we note that (11.4.4) reduces to

$$P = \int\limits_{c_{1\alpha} \leqslant \text{ all } c\,[D_{\gamma_i} Y_1 Y_1'(Y_2 Y_2')^{-1}] \leqslant c_{2\alpha}} \exp\left[-\tfrac{1}{2}\operatorname{tr}(Y_1 Y_1' + Y_2 Y_2')\right] dY_1\, dY_2. \quad \ldots \quad (11.4.11)$$

Now putting all $\gamma_i = \gamma$ it is easy to check that this reduces to

$$P = \int\limits_{\frac{c_{1\alpha}}{\gamma} \leqslant \text{ all } c[Y_1 Y_1'(Y_2 Y_2')^{-1}] \leqslant \frac{c_{2\alpha}}{\gamma}} \exp\left[-\tfrac{1}{2}\operatorname{tr}(Y_1 Y_1' + Y_2 Y_2')\right] dY_1\, dY_2. \quad \ldots \quad (11.4.12)$$

We are thus back on the problem of the distribution of the characteristic roots on the null hypothesis and we have, therefore, aside from a constant and positive factor of proportionality not involving the γ,

$$P = \int\limits_{c_{1\alpha}/\gamma \leqslant c_1 \leqslant \ldots \leqslant c_p \leqslant c_{2\alpha}/\gamma} \left[\prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}} dc_i/(1+c_i)^{\frac{n_1+n_2}{2}}\right] \prod_{i>j} (c_i - c_j)$$

$$= \int\limits_{c_p=\frac{c_{1\alpha}}{\gamma}}^{c_{2\alpha}/\gamma} \ldots \int\limits_{c_2=\frac{c_{1\alpha}}{\gamma}}^{c_3} \int\limits_{c_1=\frac{c_{1\alpha}}{\gamma}}^{c_2} f(c_1, \ldots, c_p) \prod_{i=1}^{p} dc_i, \quad \ldots \quad (11.4.13)$$

where

$$f(c_1, \ldots, c_p) = \prod_{i=1}^{p}\left[c_i^{\frac{n_1-p-1}{2}}/(1+c_i)^{\frac{n_1+n_2}{2}}\right] \prod_{i>j} (c_i - c_j) = \prod_{i=1}^{p} c^{im}(1+c_i)^{-n} \prod_{i>j} (c_i - c_j)\ (\text{say}).$$

$$\ldots \quad (11.4.14)$$

It is now easy to check that

$$\frac{\partial P}{\partial(1/\gamma)} = [c_{2\alpha} \int\limits_{c_{1\alpha}/\gamma}^{c_{2\alpha}/\gamma} \int\limits_{c_{1\alpha}/\gamma}^{c_{p-1}} \ldots \int\limits_{c_{1\alpha}/\gamma}^{c_2} f\left(c_1, \ldots, c_{p-1}, \frac{c_{2\alpha}}{\gamma}\right) \prod_{i=1}^{p-1} dc_i$$

$$-c_{1\alpha} \int\limits_{c_{1\alpha}/\gamma}^{c_{2\alpha}/\gamma} \int\limits_{c_{1\alpha}/\gamma}^{c_p} \ldots \int\limits_{c_{1\alpha}/\gamma}^{c_3} f\left(\frac{c_{1\alpha}}{\gamma},\ c_2, \ldots, c_p\right) \prod_{i=2}^{p} dc_i \quad \ldots \quad (11.4.15)$$

$$= \frac{c_{2\alpha}(c_{2\alpha}/\gamma)^m}{\left\{1 + \frac{c_{2\alpha}}{\gamma}\right\}} \int\limits_{c_{1\alpha}/\gamma}^{c_{2\alpha}/\gamma} \ldots \int\limits_{c_{1\alpha}/\gamma}^{c_2} \prod_{i=1}^{p-1} c_i^m (1+c_i)^{-n} dc_i \prod_{i>j=1}^{p-2} (c_i - c_j) \prod_{i=1}^{p-1} \left\{\frac{c_{2\alpha}}{\gamma} - c_i\right\}$$

$$- \frac{c_{1\alpha}(c_{1\alpha}/\gamma)^m}{\left\{1 + \frac{c_{2\alpha}}{\gamma}\right\}^n} \int\limits_{c_{1\alpha}/\gamma}^{c_{2\alpha}/\gamma} \ldots \int\limits_{c_{1\alpha}/\gamma}^{c_3} \prod_{i=2}^{p} c_i^m (1+c_i)^{-n}\, dc_i \prod_{i>j=2}^{p-1} (c_i - c_j) \prod_{i=2}^{p} \left\{c_i - \frac{c_{1\alpha}}{\gamma}\right\}$$

$$= k_1(\gamma) I_1(\gamma) - k_2(\gamma) I_2(\gamma), \text{ say.}$$

The condition of local unbiassedness is that

$$k_1(1)I_1(1) = k_2(1)I_2(1). \qquad \ldots \quad (11.4.16)$$

We shall now show that subject to (11.4.16) the last expression on the right of (11.4.15) > 0 if $\gamma > 1$ and < 0 if $\gamma < 1$. The proof will thus be complete if we can show that according as $\gamma > 1$ or < 1 we have

$$\frac{\left(1+\frac{c_{1\alpha}}{\gamma}\right)^n}{\left(1+\frac{c_{2\alpha}}{\gamma}\right)^n}\,\frac{I_1(\gamma)}{I_2(\gamma)} > \left(\frac{c_{1\alpha}}{c_{2\alpha}}\right)^{m+1}, \text{ i.e., } > \left(\frac{1+c_{1\alpha}}{1+c_{2\alpha}}\right)^n \frac{I_1(1)}{I_2(1)} \text{ or } < \left(\frac{1+c_{1\alpha}}{1+c_{2\alpha}}\right)^n \frac{I_1(1)}{I_2(1)} \qquad \ldots \quad (11.4.17)$$

Now according as $\gamma > 1$ or < 1 we have

$$\left(1+\frac{c_{1\alpha}}{\gamma}\right)\Big/\left(1+\frac{c_{2\alpha}}{\gamma}\right) > \text{ or } < (1+c_{1\alpha})/(1+c_{2\alpha}). \qquad \ldots \quad (11.4.18)$$

Thus (11.4.17) will be proved if we show that $I_1(\gamma)$ is an increasing function of γ and $I_2(\gamma)$ a decreasing function of γ.

Now

$$\frac{\partial I_1(\gamma)}{\partial(1/\gamma)} = -\frac{c_{1\alpha}(c_{1\alpha}/\gamma)^m}{\left(1+\frac{c_{1\alpha}}{\gamma}\right)^n} I \text{ and } \frac{\partial I_2(\gamma)}{\partial(1/\gamma)} = \frac{c_{2\alpha}(c_{2\alpha}/\gamma)^m}{\left(1+\frac{c_{2\alpha}}{\gamma}\right)^n} I, \qquad \ldots \quad (11.4.19)$$

where I stands for the positive quantity

$$\int_{c_{1\alpha}/\gamma}^{c_{2\alpha}/\gamma}\int_{c_{1\alpha}/\gamma}^{c_{p-1}}\cdots\int_{c_{1\alpha}/\gamma}^{c_3}\prod_{i=2}^{p-1} c_i^m(1+c_i)^{-n}\,dc_i \prod_{i>j=2}^{p-2}(c_i-c_j)\prod_{i=2}^{p-1}\left(c_i-\frac{c_{1\alpha}}{\gamma}\right)$$

$$\times \prod_{i=2}^{p-1}\left(\frac{c_{2\alpha}}{\gamma}-c_i\right)\left(\frac{c_{2\alpha}}{\gamma}-\frac{c_{1\alpha}}{\gamma}\right). \qquad \ldots \quad (11.4.20)$$

It is thus easy to see that $\frac{\partial I_1(\gamma)}{\partial(1/\gamma)} < 0$ and $\frac{\partial I_2(\gamma)}{\partial(1/\gamma)} < 0$, so that P of (11.4.13) monotonically decreases or the power of the test (6.4.5) monotonically increases as γ tends away from 1.

11.5. *Modified test of the hypothesis that a population dispersion matrix has a given (matrix) value against a special class of alternatives.* We take over from (6.4.4) the acceptance region for the hypothesis $H_0 : \Sigma = \Sigma_0$ and rewrite it as

$$c_{1\alpha}(p, n) \leqslant \text{ all } c_i\text{'s} \leqslant c_{2\alpha}(p, n), \qquad \ldots \quad (11.5.1)$$

where $c_{1\alpha}$ and $c_{2\alpha}$ are chosen so as to satisfy

$$P(c_{1\alpha} \leqslant \text{all } c_i\text{'s} \leqslant c_{2\alpha} | \Sigma = \Sigma_0) = 1-\alpha \qquad \dots \quad (11.5.2)$$

and

$$\left[\frac{\partial P(c_{1\alpha} \leqslant \text{all } c_i\text{'s} \leqslant c_{2\alpha} | \Sigma \neq \Sigma_0)}{\partial \gamma_i}\right]_{\gamma_1=\dots=\gamma_p=1} = 0 \; (i = 1, 2, \dots, p), \qquad \dots \quad (11.5.3)$$

or

$$\left[\frac{\partial P(c_{1\alpha}, c_{2\alpha}, \gamma_1, \dots, \gamma_p)}{\partial \gamma_i}\right]_{\gamma_1=\dots=\gamma_p=1} = 0 \; (i = 1, 2, \dots, p).$$

Here the c_i's are the characteristic roots of $\frac{1}{n}(XX')\Sigma_0^{-1}$, γ_i's are the characteristic roots of $\Sigma\Sigma_0^{-1}$ and $X(p \times n)(p \leqslant n)$ is the reduced observation matrix. Exactly along the same lines as in the previous case it can be proved that (i) the p equations (11.5.3) are really equivalent to one equation and that (ii) if $\gamma_1 = \gamma_2 = \dots = \gamma_p = \gamma$ (say), in other words if $\Sigma\Sigma_0^{-1} = \gamma$, i.e., $\Sigma = \gamma\Sigma_0$, then the P of (11.5.3) monotonically decreases, i.e., the power of the test monotonically increases as γ tends away from 1, which is the value on the null hypothesis.

11.6. It can be shown by very lengthy and tedious calculations that the two tests considered in 11.1 and 11.2 for multivariate analysis of variance and for independence between two sets of variates as also the modified tests considered in 11.4 and 11.5 for one and two dispersion matrices have each of them the monotonicity property, and not just the near monotonicity property which has been proved in chapters 10 and 11. But this lengthy proof is not being offered, in the hope that a much simpler and more elegant proof may be forthcoming in the near future.

CHAPTER TWELVE

Least Squares and Univariate Analysis of Variance and Covariance with Multivariate Extensions

12.1. *Statement of the problems.* Let $\mathbf{x}(n\times 1)$ denote a set of n uncorrelated stochastic variates with the same (unknown) variance σ^2 and let $E(\mathbf{x})$ be subject to the constraint:

$$E(\mathbf{x}) = A(n\times m)\boldsymbol{\xi}(m\times 1), \qquad \dots \quad (12.1.1)$$

where $m \leqslant n$ and $\boldsymbol{\xi}\,(m\times 1)$ is a set of unknown parameters (to be estimated) and A is a matrix of rank $r \leqslant m \leqslant n$, whose elements are given by the particular experimental design.

Problem I: Given a non-null $\mathbf{c}'(1\times m)$ (subject to certain restrictions to be brought out in (12.2)) and given $\mathbf{x}$, it is required to obtain for $\mathbf{c}'\,\boldsymbol{\xi}$ a linear estimate $\mathbf{b}'(1\times n)\mathbf{x}(n\times 1)$ such that (i) $E(\mathbf{b}'\mathbf{x})=\mathbf{c}'\boldsymbol{\xi}$ (for all $\boldsymbol{\xi}$) and (ii) *variance* $(\mathbf{b}'\mathbf{x})$ is to be a minimum. $\mathbf{c}'\boldsymbol{\xi}$ will be said to be linearly estimable (or sometimes just "estimable") if and only if (i) is satisfied.

Problem II: Given $\mathbf{c}'$ and $\mathbf{x}$ as above, it is required to obtain $\hat{\boldsymbol{\xi}}$ so that $(\mathbf{x}'-\hat{\boldsymbol{\xi}}'A')\,(\mathbf{x}-A\hat{\boldsymbol{\xi}})$ is a minimum. It will then be incidentally verified that $\mathbf{b}'\mathbf{x}$ of Problem I $= \mathbf{c}'\hat{\boldsymbol{\xi}}$ of Problem II.

Problem III: To the model of Problem I add the further condition that the x_i's are independent $N(E(x_i),\sigma^2)$ $(i = 1, 2, \dots, n)$. Let us now try to obtain (in terms of given elements) the customary F-test for the hypothesis

$$C(q\times m)\boldsymbol{\xi}(m\times 1) = \mathbf{0}(q\times 1), \qquad \dots \quad (12.1.2)$$

where $r \leqslant m$ (r being the rank of the A-matrix of (12.1.1)) and C is a given matrix of rank $s \leqslant \min\,(r, q)$.

The model indicated, under which the hypothesis (12.1.2) is tested, is usually called the linear hypothesis model, or in more recent years, the model I of analysis of variance. The hypothesis (12.1.2) is called the linear hypothesis. Now going back to (12.1.1) we observe that for the usual types of experiments when they do not involve regression, A is a matrix whose elements are ordinarily 0 or 1. For experiments which also involve regression on the so-called "concomitant variates' which are really certain observations supposed to stay constant within the probabilistic set-up of the experiment and the analysis, A is a matrix some of whose elements involve these "non-stochastic" observations, the rest of the elements being pure constants, mostly 0 or 1. That will be called the general regression set-up under the so-called model I of analysis of variance and covariance.

12.2. *Solution of Problem I.* Assume that $A'(m\times n)$ is such that $A_1(r\times n)$ can be taken as a basis and let $A'(m\times n)$ of (12.1.1) be factorized into:

$$\begin{array}{r}r\\m-r\end{array}\underset{n}{\begin{bmatrix}A_1'\\A_2'\end{bmatrix}} = \begin{array}{r}r\\m-r\end{array}\underset{r}{\begin{bmatrix}\tilde{T}_1\\T_2\end{bmatrix}} L(r\times n), \qquad \dots \quad (12.2.1)$$

such that $LL' = I(r)$, and let $L_1((n-r)\times n)$ be an arbitrary completion of L in the sense of (A.1.15), so that

$$L_2(n\times n) = \underset{n}{\begin{bmatrix}L\\L_1\end{bmatrix}}\begin{array}{l}r\\n-r\end{array} \text{ is } \perp . \qquad \dots \quad (12.2.2)$$

Notice that L_1 is not unique. Also observe that

$$I(n) = \begin{bmatrix}L\\L_1\end{bmatrix}[L' : L_1'] = \begin{bmatrix}LL' & LL_1'\\L_1L' & L_1L_1'\end{bmatrix} = [L' : L_1']\begin{bmatrix}L\\L_1\end{bmatrix} = L'L+L_1'L_1. \dots \quad (12.2.3)$$

Furthermore, with an A having the structure (12.2.1), let (12.1.1) be rewritten as

$$E(\mathbf{x}) = n\,\underset{r\quad m-r}{[A_1 : A_2]}\underset{1}{\begin{bmatrix}\xi_1\\\xi_2\end{bmatrix}}\begin{array}{l}r\\m-r\end{array}, \qquad \dots \quad (12.2.4)$$

and let $\mathbf{c}'\xi$ be rewritten as

$$1\,\underset{r\quad m-r}{[\mathbf{c}_1' : \mathbf{c}_2']}\underset{1}{\begin{bmatrix}\xi_1\\\xi_2\end{bmatrix}}\begin{array}{l}r\\m-r\end{array} \qquad \dots \quad (12.2.5)$$

Now condition (i) (of unbiassedness) of Problem I of (12.1) becomes

$$[\mathbf{c}_1' : \mathbf{c}_2']\begin{bmatrix}\xi_1\\\xi_2\end{bmatrix} = E(\mathbf{b}'\mathbf{x}) = \mathbf{b}'E(\mathbf{x}) \quad = \mathbf{b}'[A_1 : A_2]\begin{bmatrix}\xi_1\\\xi_2\end{bmatrix} \qquad \dots \quad (12.2.6)$$

$$= \mathbf{b}'(A_1\,\xi_1+A_2\,\xi_2),$$

and, since this is to be true of all ξ_1 and ξ_2, we should have

$$\mathbf{b}'A_1 = \mathbf{c}_1' \text{ and } \mathbf{b}'A_2 = \mathbf{c}_2', \qquad \dots \quad (12.2.7)$$

which imposes a number of restrictions ($\leqslant m$) on $\mathbf{b}'(1\times n)$ but by no means fully determines $\mathbf{b}'$ (which has to be determined).

Substituting in (12.2.7) for A_1 and A_2 from (12.2.1) we have

$$\mathbf{b}'L'\tilde{T}_1' = \mathbf{c}_1' \text{ or } \mathbf{b}'L' = \mathbf{c}_1'(\tilde{T}_1')^{-1}, \text{ and } \mathbf{b}'L'T_2' = \mathbf{c}_2'. \qquad \dots \quad (12.2.8)$$

Now to minimize $V(\mathbf{b}'\mathbf{x})$ subject to (12.2.8) we proceed as follows:

$$V(\mathbf{b}'\mathbf{x}) = \sigma^2\mathbf{b}'\mathbf{b} \text{ (since } \mathbf{x} \text{ is an uncorrelated set with a common variance } \sigma^2) \quad \ldots \quad (12.2.9)$$

$$= \sigma^2\mathbf{b}'[L' : L_1']\begin{bmatrix} L \\ L_1 \end{bmatrix}\mathbf{b} \text{ (using (12.2.2))} = \sigma^2[\mathbf{b}'L'L\mathbf{b}+\mathbf{b}'L_1'L_1\mathbf{b}]$$

$$= \sigma^2[\mathbf{c}_1'(\tilde{T}_1')^{-1}(\tilde{T}_1)^{-1}\mathbf{c}_1+\mathbf{b}'L_1'L_1\mathbf{b}] \text{ (using (12.2.8))}.$$

The minimum $V(\mathbf{b}'\mathbf{x})$ is thus reached when

$$\mathbf{b}'L_1' = 0, \quad \ldots \quad (12.2.10)$$

so that, combining (12.2.2), (12.2.8) and (12.2.10), we have

$$\mathbf{b}' = \mathbf{c}_1'(\tilde{T}_1')^{-1}L, \quad \ldots \quad (12.2.11)$$

and hence

$$\mathbf{b}'\mathbf{x} = \mathbf{c}_1'(\tilde{T}_1')^{-1}L\mathbf{x} = \mathbf{c}_1'(\tilde{T}_1')^{-1}(\tilde{T}_1)^{-1}A_1'\mathbf{x} \text{ (using (12.2.1))}$$

$$= \mathbf{c}_1'(\tilde{T}_1\tilde{T}_1')^{-1}A_1'\mathbf{x} = \mathbf{c}_1'(A_1'A_1)^{-1}A_1'\mathbf{x}. \quad \ldots \quad (12.2.12)$$

This gives the "unbiassed minimum variance" estimate of $\mathbf{c}'\boldsymbol{\xi}$.

Restriction on c'. Now, going back to (12.2.8) we have

$$\mathbf{c}_2' = \mathbf{b}'A_2 = \mathbf{c}_1'(\tilde{T}_1')^{-1}LA_2 = \mathbf{c}_1'(\tilde{T}_1')^{-1}(\tilde{T}_1)^{-1}A_1'A_2 \text{ (using (12.2.1))}$$

$$= \mathbf{c}_1'(A_1'A_1)^{-1}A_1'A_2. \quad \ldots \quad (12.2.13)$$

We have thus that, in order that $\mathbf{c}'\boldsymbol{\xi}$, i.e., $[\mathbf{c}_1' : \mathbf{c}_2']\begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix}$ may be "estimable" (in the sense indicated), $\mathbf{c}_2'$ must be related to $\mathbf{c}_1'$ by (12.2.13), which can be expressed in another form that is more suggestive. From (12.2.1) we have

$$A_2 = L'T_2' = A_1(\tilde{T}_1')^{-1}T_2' \text{ or } A_2' = T_2(\tilde{T}_1)^{-1}A_1', \quad \ldots \quad (12.2.14)$$

which on substitution into (12.2.13) yields

$$\mathbf{c}_2' = \mathbf{c}_1'(A_1'A_1)^{-1}A_1'A_1(\tilde{T}_1')^{-1}T_2' = \mathbf{c}_1'(\tilde{T}_1')^{-1}T_2'. \quad \ldots \quad (12.2.15)$$

Thus $\mathbf{c}_2'$ is related to $\mathbf{c}_1'$ by the same post factor by which A_2 is related to A_1.

Invariance of the linear estimate (12.2.13) *under choice of* A_1'. If, instead of A_1' and $\mathbf{c}_1'$, we choose another set of independent row vectors, say A_3' and $\mathbf{c}_3'$ to match it, then in place of the right hand side of (12.2.12) we should have the linear estimate given by replacing the subscript 1 by 3. But remembering that

$$A_3'(r\times n) = T_3(r\times r)\ L(r\times n), \quad \ldots \quad (12.2.16)$$

where T_3 is obtained by picking out from the right hand side of (12.2.1) the rows corresponding to A'_3 and is necessarily non-singular (since A'_3 is of rank r), and using (12.2.15) and (12.2.16), we have

$$\mathbf{c}'_3(A'_3A_3)^{-1}A'_3\mathbf{x} = \mathbf{c}'_1(\tilde{T}'_1)^{-1}T'_3\{T_3(\tilde{T}_1)^{-1}A'_1A_1(\tilde{T}'_1)^{-1}T'_3\}^{-1}T_3(\tilde{T}_1)^{-1}A'_1\mathbf{x}$$

$$= \mathbf{c}'_1(\tilde{T}'_1)^{-1}T'_3(T'_3)^{-1}\tilde{T}'_1(A'_1A_1)^{-1}\tilde{T}_1(T_3)^{-1}T_3(\tilde{T}_1)^{-1}A'_1\mathbf{x} = \mathbf{c}'_1(A'_1A_1)^{-1}A'_1\mathbf{x}, \quad \dots \quad (12.2.17)$$

which proves the invariance.

Variance of the "unbiassed minimum variance" estimate. From (12.2.9), (12.2..11) this variance is given by

$$V(\mathbf{b}'\mathbf{x}) = \sigma^2\mathbf{b}'\mathbf{b} = \sigma^2\mathbf{c}'_1(\tilde{T}'_1)^{-1}LL'(\tilde{T}_1)^{-1}\mathbf{c}_1 = \sigma^2\mathbf{c}'_1(\tilde{T}_1\tilde{T}'_1)^{-1}\mathbf{c}_1 = \sigma^2\mathbf{c}'_1(A'_1A_1)^{-1}\mathbf{c}_1, \quad \dots \quad (12.2.18)$$

which again by the method of the previous paragraph, can be shown to be invariant under choice of A'_1.

12.3. *Solution of problem II or the "Least squares solution".*

$$(\mathbf{x}'-\boldsymbol{\xi}'A')(\mathbf{x}-A\boldsymbol{\xi}) = (\mathbf{x}'-\boldsymbol{\xi}'A')[L' : L'_1]\begin{bmatrix}L\\L_1\end{bmatrix}(\mathbf{x}-A\boldsymbol{\xi})$$

$$= \left[\mathbf{x}' - \boldsymbol{\xi}'\begin{pmatrix}\tilde{T}_1\\T_2\end{pmatrix}L\right] [L' : L'_1]\begin{bmatrix}L\\L_1\end{bmatrix}\left[\mathbf{x}-L'(\tilde{T}'_1 : T'_2)\boldsymbol{\xi}\right] \quad \dots \quad (12.3.1)$$

$$= \left[\mathbf{x}'L'-\boldsymbol{\xi}'\begin{pmatrix}\tilde{T}_1\\T_2\end{pmatrix}\right][L\mathbf{x}-(\tilde{T}'_1 : T'_2)\boldsymbol{\xi}]+\mathbf{x}'L'_1L_1\mathbf{x}$$

(using (12.2.3)). It is now quite easy to see that given $\mathbf{x}$ and A the minimum value $(\mathbf{x}'-\boldsymbol{\xi}A')(\mathbf{x}-A\boldsymbol{\xi})$, under variation of $\boldsymbol{\xi}$, will be attained if

$$L\mathbf{x} = [\tilde{T}'_1 : T'_2]\boldsymbol{\xi}. \quad \dots \quad (12.3.2)$$

If we now want the "least squares estimate" $\mathbf{c}'\hat{\boldsymbol{\xi}}$ of an "estimable linear function" $\mathbf{c}'\boldsymbol{\xi}$, we have from the above: $\mathbf{c}'\hat{\boldsymbol{\xi}} = \mathbf{c}'_1\hat{\boldsymbol{\xi}}_1+\mathbf{c}'_2\hat{\boldsymbol{\xi}}_2 = \mathbf{c}'_1\hat{\boldsymbol{\xi}}_1+\mathbf{c}'_1(\tilde{T}'_1)^{-1}T'_2\hat{\boldsymbol{\xi}}_2$ (from (12.2.16) $= \mathbf{c}'_1(T'_1)^{-1}(\tilde{T}'_2\hat{\boldsymbol{\xi}}_1+T'_1\hat{\boldsymbol{\xi}}_2)^1 = \mathbf{c}'_1(\tilde{T}'_1)^{-1}L\mathbf{x}$ (from (12.3.2)) $= \mathbf{c}'_1(\tilde{T}'_1)^{-1}(T_1)^{-1}A'\mathbf{x}$ (from (12.2.1)) $= \mathbf{c}'_1(A'_1A_1)^{-1}A_1\mathbf{x}$, which proves the identity of the "least squares solution" of an "estimable linear function" with the "unbiassed minimum variance solution".

12.4. *Solution of problem III.* It is well known that

if $\mathbf{x}(n\times 1)$ is a set of n uncorrelated $N(E(\mathbf{x}), \sigma^2I(n))$ (and thus also independent variates) and if $L(p\times n)$ $(p \leqslant n)$ is subject to $LL' = I(p)$, then $L(p\times n)\mathbf{x}(n\times 1)$ is a set of p uncorrelated $N(LE(\mathbf{x}), \sigma^2I(p))$ variates. ... (12.4.1)

It is also well known that

if $\mathbf{u}(p\times 1)$ is an independent $N(0, \sigma^2)$ set and so is $\mathbf{v}(q\times 1)$ and if $\mathbf{u}$ and $\mathbf{v}$ are mutually independent, then $\mathbf{u}'\mathbf{u}/\sigma^2$ is a χ^2 with p degrees of freedom, $\mathbf{v}'\mathbf{v}/\sigma^2$ is a χ^2 with q degrees of freedom and $q\mathbf{u}'\mathbf{u}/p\mathbf{v}'\mathbf{v}$ is an F with degrees of freedom p and q. ... (12.4.2)

Going back to the model of Problem III in (12.1) and to (12.2.1)-(12.2.3) we observe that

if $\mathbf{x}(n\times 1)$ is an uncorrelated $N(E(\mathbf{x}),\sigma^2 I(n))$ set, then $L(r\times n)\mathbf{x}(n\times 1)$ is an uncorrelated $N(LE(\mathbf{x}), \sigma^2 I(p))$ set and $L_1((n-r)\times n)\mathbf{x}(n\times 1)$ is an uncorrelated $N(L_1 E(\mathbf{x}), \sigma^2 I(n-r)$ set which is also independent of the $L\mathbf{x}$ set, since $LL_1' = 0$. ... (12.4.3)

Now from (12.1.1) and (12.2.1)-(12.2.3) we have $E(\mathbf{x}) = L'[\tilde{T}_1' : T_2']\boldsymbol{\xi}$, so that $L_1\, E(\mathbf{x}) = L_1\, L'[\tilde{T}_1' :\, T_2']\boldsymbol{\xi} = 0$. Thus $L_1\mathbf{x}$ is an independent $N(0, \sigma^2)$ set, whence it follows that we have a χ^2 (with $n-r$ degrees of freedom) given by:

$$\mathbf{x}'L_1'L_1\mathbf{x}/\sigma^2 \text{ or } \mathbf{x}'(I(n)-L'L)\mathbf{x}/\sigma^2 \text{ or } [\mathbf{x}'\mathbf{x}-\mathbf{x}'A_1(\tilde{T}_1\, \tilde{T}_1')^{-1}A_1'\mathbf{x}]/\sigma^2 \quad \ldots \quad (12.4.4)$$

or

$$[\mathbf{x}'\mathbf{x}-\mathbf{x}'A_1(A_1'\, A_1)^{-1}A_1'\mathbf{x}]\sigma^2.$$

Consider now the hypothesis $C(q\times m)\boldsymbol{\xi}(m\times 1) = \mathbf{0}$, where C is of rank $s \leqslant \min\,(q, r)$, r being the rank of the A-matrix and thus being $\leqslant m<n$. Let us rewrite the hypothesis as

$$\begin{array}{cc} \begin{array}{c} s \\ q-s \end{array} & \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \\ & \begin{array}{cc} r & m-r \end{array} \end{array} \begin{array}{c} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} \\ 1 \end{array} \begin{array}{c} r \\ m-r \\ \end{array} = \mathbf{0}, \qquad \ldots \quad (12.4.5)$$

where $[C_{11} \quad C_{12}]$ are a set of s *independent* row vectors and $\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ a matrix, each row of which is of the nature of $\mathbf{c}_1'$ of (12.2). In this case the hypothesis (12.4.5) will be said to be 'testable'.

It is now easy to see that the hypothesis $C\boldsymbol{\xi} = \mathbf{0}$ is equivalent to $C_{11}\boldsymbol{\xi}_1+C_{12}\boldsymbol{\xi}_2 = \mathbf{0}$, so that we shall work in terms of this latter. Going back to (12.2.12), (12.2.3) and (12.2.15) we note that

$$C_{12} = C_{11}(\tilde{T}_1')^{-1}T_2' \qquad \ldots \quad (12.4.6)$$

and $$\mathbf{0} = C_{11}\boldsymbol{\xi}_1+C_{12}\boldsymbol{\xi}_2 = E[C_{11}(A_1'A_1)^{-1}A_1'\mathbf{x}] = E[C_{11}(\tilde{T}_1\, \tilde{T}_1')^{-1}\tilde{T}_1\, L\mathbf{x}]. \quad \ldots \quad (12.4.7)$$

Now $(\tilde{T}_1\tilde{T}_1')^{-1}\tilde{T}_1 L$ is a $r\times n$ matrix of rank r and C_{11} is a $s\times r$ matrix $(s \leqslant r)$ of rank s. Then using (A.1.6) we note that $C_{11}(\tilde{T}_1\tilde{T}_1')^{-1}\tilde{T}_1 L$, which is a $s\times n$ matrix, must be of rank $s \leqslant \min\,(q, r)$ (note that $r \leqslant m < n$). Let

$$C_{11}(\tilde{T}_1\tilde{T}_1')^{-1}\tilde{T}_1 L = \tilde{V}(s\times s)M(s\times n), \qquad \ldots \quad (12.4.8)$$

where $MM' = I(s)$, and $\tilde{V}$ of course is non-singular. Then we have

$$E(M\mathbf{x}) = (\tilde{V})^{-1}E[C_{11}(\tilde{T}_1\tilde{T}_1')^{-1}\tilde{T}_1 L\mathbf{x}] = \mathbf{0} \qquad \ldots \quad (12.4.9)$$

(from (12.4.7)) and furthermore

$$ML_1' = (\tilde{V})^{-1}C_{11}(\tilde{T}_1\tilde{T}_1')^{-1}\tilde{T}_1 LL_1' = 0, \qquad \ldots \quad (12.4.10)$$

so that

$M\mathbf{x}$ is a s-set of independent $N(0, \sigma^2)$, $L_1\mathbf{x}$(of (12.4.9)) is a $(n-r)$—set of independent $N(0, \sigma^2)$, $M\mathbf{x}$ and $L_1\mathbf{x}$ are mutually independent, ... (12.4.11)

and hence

$(n-r)\mathbf{x}'M'M\mathbf{x}/s\mathbf{x}'L_1'L_1\mathbf{x}$ is an F with degrees of freedom s and $n-r$. ... (12.4.12)

Using (12.4.4), (12.4.8) and of course (12.2.1) and (A.3.11), we can reduce (12.4.12) to

$$\frac{(n-r)\mathbf{x}'A_1(A_1'A_1)^{-1}C_{11}'[C_{11}(A_1'A_1)^{-1}C_{11}']^{-1}C_{11}(A_1'A_1)^{-1}A_1'\mathbf{x}}{s[\mathbf{x}'\mathbf{x}-\mathbf{x}'A_1(A_1'A_1)^{-1}A_1'\mathbf{x}]}, \quad \dots \quad (12.4.13)$$

which is an F(with degrees of freedom s and $n-r$) for testing the hypothesis $C\boldsymbol{\xi} = 0$ and which is expressed in terms of quantities directly observed or given by the experimental design and the hypothesis to be tested. The form (12.4.13) can be shown to be invariant under the kind of choice indicated in (12.2), i.e. under the choice of a basis of A, in much the same way as there.

12.5. *Conditions that k different linear hypotheses may be testable in a quasi-independent manner.* Suppose instead of the hypothesis (12.1.2) we have the following hypotheses:

$$C^{(i)}(q_i\times m)\boldsymbol{\xi}(m\times 1) = 0(q_i\times 1), \text{ with } i = 1, 2, \dots, k, \quad \dots \quad (12.5.1)$$

or breaking down into submatrices,, we have in place of (12.4.5) the following:

$$\begin{array}{c} s_i \\ q_i-s_i \\ \end{array} \begin{array}{c} \begin{bmatrix} C_{11}^{(i)} & C_{12}^{(i)} \\ C_{21}^{(i)} & C_{22}^{(i)} \end{bmatrix} \\ r \quad m-r \end{array} \begin{array}{c} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} \\ 1 \end{array} \begin{array}{c} r \\ m-r \\ \end{array} = \mathbf{0}\ (q_i\times 1), \text{ with } i = 1, 2, \dots, k. \quad \dots \quad (12.5.2)$$

Now by (12.4.12) the i-th hypothesis of (12.5.1) will have an F_i to go with it, where

$$F_i = (n-r)\mathbf{x}'M_i'M_i\mathbf{x}/s_i\mathbf{x}'L_1'L_1\mathbf{x}, \quad \dots \quad (12.5.3)$$

and is distributed as an F with degrees of freedom s_i and $n-r$. Also $\mathbf{x}'M_i'M_i\mathbf{x}/\sigma^2$ is a χ_i^2 with degrees of freedom s_i and $\mathbf{x}'L_1'L_1\mathbf{x}/\sigma^2$ is a χ^2 with degrees of freedom $n-r$. It is clear from (12.4.10) that each $\chi_i^{2\prime}$ $(i = 1, 2, \dots, k)$ is distributed independently of χ^2. The question is, when are the χ_i^2's themselves mutually independent? If these are so, then the associated linear hypotheses (12.5.1) will be said to be testable in a quasi-independent manner. Going back to (12.4.8) we observe that χ_i^2 and χ_j^2 $(i \neq j)$ will be independent if $M_i M_j' = 0$, that is, if

$$\tilde{V}_i^{-1} C_{11}^{(i)}(\tilde{T}_1 \tilde{T}_1')^{-1}\tilde{T}_1 LL'\tilde{T}_1'(\tilde{T}_1\tilde{T}_1')^{-1}C_{11}^{(j)\prime}\, \tilde{V}_j'^{-1} = 0 \quad \dots \quad (12.5.4)$$

or since $\tilde{V}_i$ and $\tilde{V}_j$ are non-singular, if

$$C_{11}^{(i)}(A_1'A_1)^{-1}(A_1'A_1)(A_1'A_1)^{-1}\ C_{11}^{(j)\prime} = 0$$

or $$C_{11}^{(i)}(A_1'A_1)^{-1}\ C_{11}^{(j)\prime} = 0, \text{ with } i \neq j = 1, 2, \dots, k. \quad \dots \quad (12.5.5)$$

This, therefore, is the set of conditions for the linear hypotheses (12.5.1) being testable in a quasi independent manner. From a practical standpoint it serves, when an appropriate breakdown of the sum of squares is not intuitively evident, exactly the same purpose as Cochran's theorem does when such an appropriate breakdown is intuitively evident.

12.6. *A quasi-multivariate generalization of the problems considered in* (12.1). Suppose that in (12.1) we assume that $\mathbf{x}(n\times 1)$ denote a correlated set with a p.d. dispersion matrix $\sigma^2\Sigma(n\times n)$ where σ^2 is unknown but Σ is supposed to be known and suppose that (12.1.1) is left unchanged. Also in problem III let us assume that $\mathbf{x}$ is $N(E(\mathbf{x}), \sigma^2\Sigma)$ but let us leave (12.1.2) unchanged. Then putting

$$\Sigma(n\times n) = \tilde{T}(n\times n)\tilde{T}'(n\times n) \text{ and } \tilde{T}^{-1}(n\times n)\mathbf{x}(n\times 1) = \mathbf{y}(n\times 1), \quad \ldots \quad (12.6.1)$$

it is easy to check that $\mathbf{y}(n\times 1)$ is a set of uncorrelated variates with a common variance σ^2, and also that if $\mathbf{x}$ is $N(E(\mathbf{x}), \sigma^2\Sigma)$, $\mathbf{y}$ is $N(E(\mathbf{y}), \sigma^2 I(n))$. Also in terms of $\mathbf{y}$, (12.1.1) reduces to

$$E(\mathbf{y}) = \tilde{T}^{-1}A\boldsymbol{\xi}. \quad \ldots \quad (12.6.2)$$

It is now easy to check that (12.2.12) reduces to

$$\mathbf{c}_1'(A_1'\tilde{T}'^{-1}\tilde{T}^{-1}A_1)^{-1}A_1'\tilde{T}'^{-1}\mathbf{y} = \mathbf{c}_1'(A_1'\Sigma^{-1}A_1)^{-1}A_1'\tilde{T}'^{-1}\mathbf{y} = \mathbf{c}_1'(A_1'\Sigma^{-1}A_1)^{-1}A_1'\Sigma^{-1}\mathbf{x}. \quad \ldots \quad (12.6.3)$$

Also (12.4.13) similarly reduces to

$$\frac{(n-r)\mathbf{x}'\Sigma^{-1}A_1(A_1'\Sigma^{-1}A_1)^{-1}C_{11}'[C_{11}(A_1'\Sigma^{-1}A_1)^{-1}C_{11}']^{-1}\,C_{11}(A_1'\Sigma^{-1}A_1)^{-1}A_1'\Sigma^{-1}\mathbf{x}}{s[\mathbf{x}'\Sigma^{-1}\mathbf{x}-\mathbf{x}'\Sigma^{-1}A_1(A_1'\Sigma^{-1}A_1)^{-1}A_1'\Sigma^{-1}\mathbf{x}]} \quad \ldots \quad (12.6.4)$$

Itis easy to verify that, under this model, the 'estimability' condition (12.2.13) and the 'testability' condition (12.4.6) will stay unchanged. The necessary modifications in the other expressions will also follow in an obvious manner.

12.7. *Multivariate generalization.* The set-up for multivariate analysis of variance and covariance, i.e., for a test of the general multivariate linear hypothesis is an easy and direct extension of what has been considered so far in this section.

In place of the set-up of section (12.1), consider the more general set-up of (iiic) of chapter 5, which is the following. Let $X(p\times n)$ (with p and $p(p+1)/2<n$) be n independently distributed column vectors, the r-th vector $\mathbf{x}_r$ $(p\times 1)$ being $N(E(\mathbf{x}_r), \Sigma)$ $(r = 1, 2, \ldots, n)$. In place of (12.1.1) we have

$$E(X') \quad (n\times p) = A(n\times m)\xi(m\times p), \quad \ldots \quad (12.7.1)$$

where ξ is a matrix of unknown parameters and $A(n\times m)$ is given by the design of the experiment such that it is of rank $r\leqslant m<n$. We recall the observations in connection with (12.1.1)and note that here also, for the usual type of experiments where they do not involve regression, A is a matrix whose elements are ordinarily 0 or 1. For experiments which involve regression on the so-called "concomitant variates", A is a matrix, some of whose elements involve these "concomitant variates" or non-stochastic observations, the rest of the elements being pure constants mostly 0 or 1. Also setting

$$A'(m\times n) = \begin{array}{c} \left[\begin{array}{c} A_1' \\ \\ A_2' \end{array}\right] \\ n \end{array} \begin{array}{l} r \\ \\ m-r \end{array}, \quad \ldots \quad (12.7.2)$$

let A_1' be a basis of A', i.e., of A.

In place of the hypothesis (12.1.2) we shall have the hypothesis

$$C(q\times m) \;\; \xi(m\times p) = 0(q \times p), \quad \ldots \quad (12.7.3)$$

where, as before,

$$C(q\times m) = \begin{array}{c} \left[\begin{array}{cc} C_{11} & C_{12} \\ C_{21} & C_{22} \end{array}\right] \\ r \quad m-r \end{array} \begin{array}{l} s \\ q-s \end{array} \quad \ldots \quad (12.7.4)$$

such that C is of rank s, $[C_{11} : C_{12}]$ forms a basis, and also that C satisfies the testability condition of the nature of (12.2) which will be here

$$\begin{array}{cc} \begin{bmatrix} C_{12} \\ C_{22} \end{bmatrix} & \begin{array}{c} s \\ q-s \end{array} \\ m-r & \end{array} = \begin{array}{c} s \\ q-s \end{array} \begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix} [A_1'(r \times n)A_1(n \times r)]^{-1}A_1'(r \times n)A_2(n \times \overline{m-r}). \quad \ldots \quad (12.7.5)$$

Just as, following (12.4.5) we observed that the hypothesis $C(q \times m)\, \xi(m \times 1) = \mathbf{0}(q \times 1)$ $\Longleftrightarrow s[\underset{r}{C_{11}} : \underset{m-r}{C_{12}}]\, \boldsymbol{\xi}(m \times 1) = \mathbf{0}(s \times 1)$, so also it is easy to check that $C(q \times m)\, \xi(m \times p)$ $= 0(q \times p) \Longleftrightarrow s[\underset{r}{C_{11}} : \underset{m-r}{C_{12}}]\, \boldsymbol{\xi}(m \times p) = 0(s \times p)$

or
$$H_0 : \xi'(p \times m) \begin{array}{cc} \begin{bmatrix} C_{11}' \\ C_{12}' \end{bmatrix} & \begin{array}{c} r \\ m-r \end{array} \\ s & \end{array} = 0(p \times s). \quad \ldots \quad (12.7.6)$$

We also observe that

$$H_0 \text{ of } (12.7.6) \Longleftrightarrow \bigcap_{\mathbf{a}} H_{0\mathbf{a}} = \bigcap_{\mathbf{a}} [\mathbf{a}'(1 \times p)\, \xi'(p \times m) \begin{array}{cc} \begin{bmatrix} C_{11}' \\ C_{12}' \end{bmatrix} & \begin{array}{c} r \\ m-r \end{array} \\ s & \end{array} = \mathbf{0}'(1 \times s)], \quad \ldots \quad (12.7.7)$$

where $\bigcap_{\mathbf{a}}$ is taken over all non-null $\mathbf{a}(p \times 1)$.

Now, using (12.4.13), we have for $H_{0\mathbf{a}}$ a critical region of size, say β, given by

$$\frac{(n-r)\mathbf{a}'XA_1(A_1'A_1)^{-1}\, C_{11}'[C_{11}(A_1'A_1)^{-1}C_{11}']^{-1}C_{11}(A_1'A_1)^{-1}A_1'X'\mathbf{a}}{s[\mathbf{a}'XX'\mathbf{a}-\mathbf{a}'XA_1(A_1'A_1)^{-1}A_1'X'\mathbf{a}]} \leqslant F_\beta(s, n-r), \quad \ldots \quad (12.7.8)$$

where $F_\beta(s, n-r)$ is the β-point of the F-distribution with degrees of freedom s and $n-r$. Now, as in (iii) of section (6.4) of Chapter 6, using the extended type I principle, we have for $H_0 = \bigcap_{\mathbf{a}} H_{0\mathbf{a}}$, the critical region of size $\alpha (> \beta)$ formed by the union of the regions (12.7.8) over all non-null $\mathbf{a}(p \times 1)$, the region being given by

$$c_t \geqslant c_\alpha(p,\ s,\ n-r), \quad \ldots \quad (12.7.9)$$

where $t = \min(p, s)$, c_t is the largest characteristic root of S^*S^{-1}, and

$$sS^* = XA_1(A_1'A_1)^{-1}C'_{11}[C_{11}(A'_1A_1)^{-1}C'_{11}]^{-1}C_{11}(A'_1A_1)^{-1}A_1'X' \quad \ldots \quad (12.7.10)$$

and
$$(n-r)S = [XX'-XA_1(A_1'A_1)^{-1}A'_1X']. \quad \ldots \quad (12.7.11)$$

This largest characteristic root has the same central distribution as that of the largest characteristic root that figured in (iii) of (6.4), with degrees of freedom p, t and $n-r$.

The development given above really subsumes an apparently more general development in which the hypothesis (12.7.3) is replaced by

$$C(q \times m)\, \xi\, (m \times p)M(p \times u) = 0(q \times u), \quad \ldots \quad (12.7.11.1)$$

where $u \leqslant p$, M is a given matrix of rank u and C has the same structure as before. The equation (12.7.6) will now be replaced by

$$H_0 : M'(u \times p)\, \xi'\, (p \times m) \begin{array}{cc} \begin{bmatrix} C_{11}' \\ C_{12}' \end{bmatrix} & \begin{array}{c} r \\ m-r \end{array} \\ s & \end{array} = 0(u \times s). \quad \ldots \quad (12.7.11.2)$$

That the development of this case is really subsumed under the one already discussed can be shown in the following way. We note that if $X(p\times n)$ (with $p < n$) be n independently distributed column vectors, the r-th vector $\mathbf{x}_r(p\times 1)$ being $N(E(\mathbf{x}_r), \Sigma)(r = 1, \dots, n)$ then $M'(u\times p)X(p\times n)$ will be n independently distributed column vectors, the r-th row vector being $N(E(M'\mathbf{x}_r),\ M'\Sigma M)$. Putting $\xi(m\times p)M(p\times u) = \xi^*\,(m\times u)$, we can now replace (12.7.1) by

$$E(X'M)(n\times u) = A(n\times m)\ \xi^*\,(m\times u), \qquad \dots \quad (12.7.11.3)$$

and (12.7.11.1) and (12.7.11.2) respectively by

$$C(q\times m)\ \xi^*(m\times u) = 0(q\times u) \qquad \dots \quad (12.7.11.4)$$

and

$$\xi^{*\prime}(u\times m) \begin{array}{cc} \begin{bmatrix} C'_1 \\ \\ C'_{12} \end{bmatrix} & \begin{array}{l} r \\ \\ m-r \end{array} \\ s & \end{array} = 0(u\times s). \qquad \dots \quad (12.7.11.5)$$

It is thus easy to see that for the hypothesis (12.7.11.4) or (12.7.11.5) we can, in exactly the same way as before, work out, step by step, a test of the same nature. In (12.7.8), $X(p\times n)$ is to be replaced by $M'(u\times p)X(p\times n)$; in $c_\alpha(p, s, n-r)$ of (12.7.9), p is to be replaced by u; and in c_t of the same equation t wil now stand for $\min(u, s)$; also in (12.7.10) and (12.7.11), X will have to be replaced by $M'(u\times p)X(p\times n)$. In subsequent developments (specially in connection with confidence bounds related to multivariate linear hypotheses on means) it will be understood that we can always switch over from $H_0 : C\xi = 0$ to $H_0 : C\xi M = 0$, and back and forth. Thus the mathematical treatment given there will suffice for this apparently more general case.

The direct reduction to the canonical form of the problem of the joint distribution of the roots $c_1 \leqslant c_2 \leqslant \dots \leqslant c_t$, and hence of that of c_t is of some interest here, and it also proves incidentally the statement made after (12.7.11), which of course can also be proved otherwise. For this reduction we proceed as follows:

$$\begin{aligned} P(X) &= \text{Const} \exp\left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(X-E(X))(X'-E(X'))\right]dX, \qquad \dots \quad (12.7.12) \\ &= \text{Const} \exp\left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(X-\xi'A')(X'-A\xi)\right]dX, \end{aligned}$$

using (12.7.1).

Next, using the factorization (12.2.1) and the completion (12.2.2), we have

$$A(n\times m) = L'(n\times r)[\underset{r}{\tilde{T}'_1} : \underset{m-r}{T'_2}]r, \text{ and an } \perp \begin{array}{cc} \begin{bmatrix} L \\ \\ L_1 \end{bmatrix} & \begin{array}{l} r \\ \\ n-r \end{array} \\ n & \end{array} \qquad \dots \quad (12.7.13)$$

Now use the orthogonal transformation

$$Z(p\times n) = X[L' : L'_1] = [\underset{r}{Z_1} : \underset{n-r}{Y}]p \text{ (say), i.e.,}$$

$$X = [Z_1 : Y] \begin{bmatrix} L \\ \\ L_1 \end{bmatrix} = Z_1L + YL_1 \qquad \dots \quad (12.7.14)$$

to obtain

$$P(Z_1,\ Y) = \text{Const}\ \exp\left[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}\{(Z_1 - \xi' \begin{bmatrix} \tilde{T}_1 \\ \\ T \end{bmatrix})(Z'_1 - [\tilde{T}'_1 : T'_2]\ \xi) + YY'\}\right]dZ_1\,dY.$$

$$\dots \quad (12.7.15)$$

Notice that the unbiased minimum variance estimate of $[C_{11} : C_{12}]\xi(m \times p)$ is $C_{11}(A_1'A_1)^{-1} \times A_1'X'$, so that under $H_0 : E\,[C_{11}(A_1'A_1)^{-1}A_1'X'] = 0$. Also we have $C_{11}(A_1'A_1)^{-1}A_1 = C_{11}(\tilde{T}_1\tilde{T}_1')^{-1}\tilde{T}_1L = C_{11}\tilde{T}_1'^{-1}L$. Now put

$$C_{11}(s \times r)\tilde{T}_1'^{-1}(r \times r) = V(s \times s)M_1(s \times r), \text{ where } M_1M_1' = I(s). \quad \dots \quad (12.7.16)$$

Also complete M_1 into an $\perp$ $\begin{array}{cc} \begin{bmatrix} M_1 \\ \\ M_2 \end{bmatrix} & \begin{matrix} s \\ \\ r-s \end{matrix} \\ r & \end{array}$

Next use the orthogonal transformation

$$Z_2(p \times r) = Z_1[M_1' : M_2'] = [\underset{s}{Y^*} : \underset{r-s}{Y_1^*}]\, p(\text{ say}), \text{ i.e.,}$$

$$Z_1 = Y^*M_1 + Y_1^*M_2, \text{ and notice that}$$
$$Y^* = Z_1M_1' \text{ and } Y_1^* = Z_1M_2'. \quad \dots \quad (12.7.17)$$

Similarly put

$$\eta^* = \xi' \begin{bmatrix} \tilde{T}_1 \\ \\ T_2 \end{bmatrix} M_1' \text{ and } \eta_1^* = \xi' \begin{bmatrix} \tilde{T}_1 \\ \\ T_2 \end{bmatrix} M_2'. \quad \dots \quad (12.7.18)$$

We now have for $(Y^*,\ Y_1^*,\ Y)$ the distribution

$$P(Y^*,\ Y_1^*,\ Y) = \text{Const} \exp[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}\{(Y^*M_1 + Y_1^*M_2) - \xi' \begin{bmatrix} \tilde{T}_1 \\ \\ T_2 \end{bmatrix})$$
$$\times (M_1'Y^{*\prime} + M_2'Y_1^{*\prime}) - [\tilde{T}_1' : T_2']\xi) + YY'\}]dY^*dY_1^*\,dY$$
$$= \text{Const} \exp\,[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}\{(Y^* - \eta^*)(Y^{*\prime} - \eta^{*\prime}) + (Y_1^* - \eta_1^*)(Y_1^{*\prime} - \eta_1^{*\prime}) + YY'\}]dY^*dY_1^*dY.$$
$$\dots \quad (12.7.19)$$

Integrating out over Y_1^*, we have for (Y^*, Y) the joint distribution

$$P(Y^*,\ Y) = \text{Const} \exp\,[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}\{(Y^* - \eta^*)(Y^{*\prime} - \eta^{*\prime}) + YY'\}]dY^*dY. \quad \dots \quad (12.7.20)$$

Notice that $Y(p \times \overline{n-r})\ Y'(\overline{n-r} \times p)$ is, a.e., p.d. (assuming of course that $p \leqslant n-r$) and $Y^*(p \times s)Y^{*\prime}(s \times p)$ is, a.e., at least p.s.d. of rank $t = \min(\text{p, s})$. Also check that $YY' = XL_1'L_1X =$ the right side of (12.7.11) and $Y^*Y^{*\prime} = Z_1M_1'M_1Z_1' = Z_1\tilde{T}_1^{-1} \times C_{11}'\tilde{V}'^{-1}\tilde{V}^{-1}\ C_{11}\tilde{T}_1'^{-1}Z_1' = XL'\tilde{T}_1'^{-1}\ C_{11}'(\tilde{V}\tilde{V}')^{-1}\ C_{11}\tilde{T}_1^{-1}LX' =$ the right side of (12.7.10). Furthermore, check that

$$\boldsymbol{\eta}^* = \boldsymbol{\xi}' \begin{bmatrix} \tilde{T}_1 \\ \\ T_2 \end{bmatrix} \quad M_1' = \boldsymbol{\xi}' \begin{bmatrix} \tilde{T}_1 \\ \\ T_2 \end{bmatrix} \tilde{T}_1^{-1}C_{11}'\tilde{V}'^{-1} = (\xi_1'\tilde{T}_1 + \xi_2'T_2)\tilde{T}_1^{-1}C_{11}'\tilde{V}'^{-1} \text{ (say)}$$
$$= (\boldsymbol{\xi}_1'C_{11}' + \boldsymbol{\xi}_2'C_{12}')\tilde{V}'^{-1},$$

so that, if $C\boldsymbol{\xi} = 0$, i.e., if $\boldsymbol{\xi}_1'C_{11}' + \boldsymbol{\xi}_2'C_{12}' = 0$, then $\eta^* = 0$. Also notice, after some calculations, that

$$\eta^*\eta^{*\prime} = (\boldsymbol{\xi}_1'C_{11}' + \boldsymbol{\xi}_2'C_{12}')[C_{11}(A_1'A_1)^{-1}C_{11}']^{-1}(C_{11}\boldsymbol{\xi}_1 + C_{12}\boldsymbol{\xi}_2). \quad \dots \quad (12.7.21)$$

Thus, (12.7.20) may be regarded as a quasi-canonical form of the joint distribution problem of the roots, and now, using the same technique as in (A.7.5) and denoting by Y_0^* and Y_0 the transforms of Y^* and Y (recall that the roots are invariant under this transformation), we have for $Y_0^*(p\times s)$ and $Y_0(p\times \overline{n-r})$ the probability law

$$\begin{aligned} P(Y_0^*,\ Y_0) = \text{Const} \exp \Big[-\tfrac{1}{2} \operatorname{tr} \Big\{ & Y_0 Y_0' + \begin{array}{cc} \begin{bmatrix} D_\gamma & 0 \\ 0 & 0 \end{bmatrix} & \begin{matrix} t \\ p-t \end{matrix} \\ \begin{matrix} t & p-t \end{matrix} & \end{array} + Y_0^* Y_0^{*\prime} \\ & -2Y_0^*\ (p\times s) \begin{array}{cc} \begin{bmatrix} D_\gamma & 0 \\ 0 & 0 \end{bmatrix} & \begin{matrix} t \\ s-t \end{matrix} \\ \begin{matrix} t & p-t \end{matrix} & \end{array} \Big\} \Big] dY_0^* dY_0, \qquad \dots \quad (12.7.22) \end{aligned}$$

where, as usual, D_γ stands for a diagonal matrix whose diagonal elements are the roots γ's of the matrix $\eta^*\ \eta^*\Sigma^{-1}$ (some of which may be zero). Now notice that

$$\operatorname{tr} \begin{bmatrix} D_\gamma & 0 \\ 0 & 0 \end{bmatrix} = \sum_{i=1}^{t} \gamma_i \text{ and } \operatorname{tr} Y_0^* \begin{bmatrix} D_\gamma & 0 \\ 0 & 0 \end{bmatrix} = \sum_{i=1}^{t} (Y_0^*)_{ii} \gamma_i^{\frac{1}{2}}, \qquad \dots \quad (12.7.23)$$

and rewrite (12.7.22) as

$$\text{Const} \exp \left[-\tfrac{1}{2} \operatorname{tr} (Y_0^* Y_0^{*\prime} + Y_0 Y_0') + \sum_{i=1}^{t} \gamma_i - 2 \sum_{i=1}^{t} (Y_0^*)_{ii} \gamma_i^{\frac{1}{2}} \right] dY_0^* dY_0, \qquad \dots \quad (12.7.24)$$

which is, therefore, the canonical form for the joint distribution of the roots in the general case.

In another monograph under the title "Least squares and analysis of variance and covariance" which will be a sequel to this one, use will be made of the formulae of this chapter to obtain the customary tests and estimates relating to the standard classes of designs in the context of what is called model I. Adjustment of the general theory given here to the situations of the other models and the derivation of some actual formulae involved in the analysis of some concrete situations there, will also be discussed in that sequel. In actual application we repeatedly run into the problem of inverting matrices which have certain kinds of pattern. Methods will be discussed of obtaining inverses of these patterned matrices in a very simple manner without having recourse to Doolittle's method or any other such method. These latter while extremely useful for general matrices, can be luckily dispensed with so far as these particular patterned matrices are concerned.

The above set up is specially useful for a *general* discussion of linear estimation or testing of linear hypothesis, although it also leads, without much calculation, to the formulae for the different customary designs. However, there is another set-up tobe discussed in the later monograph, which gives the different customary formulae in an even easier manner, although this is not so suitable for a *general* discussion.

CHAPTER THIRTEEN

Some Univariate and Bivariate Confidence Bounds*

13.1. *Some general observations.* The *general* theory (to which nothing is added here) of confidence bounds like the general theory of testing of hypotheses and tests of significance (with a part of which the previous sections have been concerned) has been worked out in a series of papers, now classic. This is readily available not only in papers but in standard books as well and need not be explained here. However, except for some comparatively recent work, most of the earlier *applications* have been concerned with confidence bounds on a single parameter or a single function of the parameters. Simultaneous confidence bounds on several parameters or parametric functions offer nothing new in principle, being already inherent in the general theory and will not, therefore, be discussed here from the point of view of the general theory. In this chapter several examples from univariate normal populations and one from a bivariate normal population will be discussed (some of them simultaneous and some of them "single") which will prepare the ground for the multivariate examples (all of them simultaneous) to be discussed in chapter 14. In this chapter, in every case except one we shall start from a current test of the corresponding hypothesis (having a number of optimum properties in respect of power) and obtain by inversion single" or "simultaneous" confidence bounds which, therefore, by the general theory, will have similar optimum properties in respect of *shortness*, i.e., the probability of covering wrong values of the parameters or parametric functions.

13.2. *Means of normal populations.*

(i) For $N(\xi, \sigma^2)$ we have, in terms of a sample of size n with sample mean $\bar{x}$ and sample standard deviation s, the following well known confidence interval for ξ (with a confidence coefficient $1-\alpha$)

$$\bar{x}-st_{\alpha/2}(n-1)/\sqrt{n} \leqslant \xi \leqslant \bar{x}+st_{\alpha/2}(n-1)/\sqrt{n}, \qquad \dots \quad (13.2.1)$$

where $t_{\alpha/2}(n-1)$ is the upper $\alpha/2$ point of the ordinary t-distribution with d.f. $(n-1)$.

(ii) For $N(\xi_h, \sigma^2)$ $(h = 1, 2)$ we have, in terms of two samples of sizes n_h with sample means and sample standard deviations $\bar{x}_h$ and s_h $(h = 1, 2)$, the following well known confidence interval for $\xi_1-\xi_2$ (with a confidence coefficient $1-\alpha$)

$$(\bar{x}_1-\bar{x}_2)-st_{\alpha/2}(n-2)/\sqrt{n_{12}} \leqslant \xi_1-\xi_2 \leqslant (\bar{x}_1-\bar{x}_2)+st_{\alpha/2}(n-2)/\sqrt{n_{12}}, \quad \dots \quad (13.2.2)$$

where $n = n_1+n_2$, $s^2 = [(n_1-1)s_1^2+(n_2-1)s_2^2]/(n-2)$, $n_{12} = n_1 n_2/n$ and $t_{\alpha/2}(n-2)$ is the upper $\alpha/2$ point of the ordinary t-distribution with d.f. $n-2$, i.e., n_1+n_2-2.

(iii) For confidence bounds relating to ξ_h's of $N(\xi_h, \sigma^2)$ $(h = 1, 2, \dots, k,$ where $k > 2)$ we proceed as follows. Suppose we have random samples of sizes n_h, sample means $\bar{x}_h$, sample standard deviations s_h $(h = 1, 2, \dots, k)$. Put

$$n = \sum_{h=1}^{k} n_h,\ s^2 = \sum_{h=1}^{k} (n_h-1)s_h^2/(n-k),\ \bar{x} = \sum_{h=1}^{k} n_h\bar{x}_h/n,\ s^{*2} = \sum_{h=1}^{k} n_h(\bar{x}_h-\bar{x})^2/(k-1),$$

$$\bar{\xi} = \sum_{h=1}^{k} n_h\xi_h/n. \qquad \dots \quad (13.2.3)$$

---

* See references [27, 28, 29, 44, 49, 51] in this connection.

For the hypothesis $H_0 : \xi_1 = \xi_2 = \ldots = \xi_k$, i.e., $\xi_h = \xi$ $(h = 1, \ldots, k)$, we have at a level of significance, say α, the current F-test with a critical region

$$F = s^{*2}/s^2 \geqslant F_\alpha(k-1,\ n-k), \qquad \ldots \quad (13.2.4)$$

where $F_\alpha(k-1,\ n-k)$ stands for the upper α point of the central F-distribution with d.f. $(k-1)$ and $(n-k)$ (we recall the well known result that when H_0 is true s^{*2}/s^2 is distributed as the central F). When H_0 is not true, it is easy enough to see that s^{**2}/s^2 is distributed as the central F, where s^{**} is given by

$$s^{**2} = \sum_{h=1}^{k} n_h(\bar{x}_h-\bar{x}-\xi_h+\bar{\xi})^2/(k-1) \qquad \ldots \quad (13.2.5)$$

Suppose that we now start from a statement with probability $1-\alpha$, namely

$$s^{**2}/s^2 \leqslant F_\alpha(k-1,\ n-k), \text{ i.e., } \sum_{h=1}^{k} n_h(\bar{x}_h-\bar{x}-\xi_h+\bar{\xi})^2/(k-1)s^2 \leqslant F_\alpha(k-1,\ n-k).$$
$$\ldots \quad (13.2.6)$$

It is easy to check (see (A.2.7)) that the statement (13.2.6) $\Longleftrightarrow$ the following statement.

$$-s[(k-1)\ F_\alpha(k-1,\ n-k)]^{\frac{1}{2}} \leqslant \sum_{h=1}^{k} a_h n_h^{\frac{1}{2}}(\bar{x}_h-\bar{x}-\xi_h+\xi) \leqslant s[(k-1)\ F_\alpha(k-1,\ n-k)]^{\frac{1}{2}}$$

$$\text{or } \sum_{h=1}^{k} a_h n_h^{1/2}(\bar{x}_h-\bar{x})-s[(k-1)\ F_\alpha(k-1,\ n-k)]^{\frac{1}{2}} \leqslant \sum_{h=1}^{k} a_h n_h^{1/2}(\xi_h-\xi) \leqslant \sum_{h=1}^{k} a_h n_h^{1/2}(\bar{x}_h-\bar{x})+$$
$$+s[(k-1)\ F_\alpha(k-1,\ n-k)]^{\frac{1}{2}}, \qquad \ldots \quad (13.2.7)$$

for all arbitrary a_h's subject to $\sum_{h=1}^{k} a_h^2 = 1$. (13.2.7) is obviously a set of simultaneous confidence bounds on all arbitrary linear compounds of $n_h^{1/2}(\xi_h-\xi)$ $(h = 1, 2, \ldots, k)$, the compounding coefficients a_h's being subject to $\sum_{h=1}^{k} a_h^2 = 1$. It is also easy to verify that the set of such linear compounds could be otherwise written as

$$\sum_{h=1}^{k} a_h n_h^{1/2}\xi_h, \text{ for all } a_h\text{'s subject to } \sum_{h=1}^{k} a_h^2 = 1 \text{ and } \sum_{h=1}^{k} a_h n_h^{1/2} = 0. \qquad \ldots \quad (13.2.8)$$

(iv) For confidence bounds in the case of the general linear hypothesis we proceed as follows from the set-up of chapter (12). Suppose we have x_h's $(h = 1, 2, \ldots, n)$ which are n independent $N(E(x_h), \sigma^2)$ such that, putting $\mathbf{x}'(1\times n) = (x_1, x_2, \ldots, x_n)$, we have

$$E(\mathbf{x})(n\times 1) = A(n\times m)\boldsymbol{\xi}(m\times 1), \qquad \ldots \quad (13.2.9)$$

where $m < n$, A is a matrix of rank, say $r \leqslant m < n$, given by the experimental situation and $\boldsymbol{\xi}\ (m\times 1)$ is a set of unknown parameters.

Putting $A'(m\times n) = \underset{n}{\begin{bmatrix} A_1' \\ A_2' \end{bmatrix}}\begin{matrix} r \\ m-r \end{matrix}$, let us assume. as we can without any loss of generality, that $A_1'(r\times n)$ is a set of independent row vectors which might be taken to be a basis of $A'(m\times n)$. Suppose now that it is required to test a "testable" hypothesis

$$C(q\times m)\boldsymbol{\xi}(m\times 1) = 0, \qquad \dots \quad (13.2.10)$$

where C is of rank $s \leqslant \min(q, r) \leqslant m < n$.

Putting

$$C(q\times m)\boldsymbol{\xi}(m\times 1) = \begin{array}{ccc} s & C_{11} & C_{12} \\ q-s & C_{21} & C_{22} \\ & r & m-r \end{array} \begin{array}{cc} \boldsymbol{\xi}_1 & r \\ \boldsymbol{\xi}_2 & m-r \\ 1 & \end{array}, \qquad \dots \quad (13.2.11)$$

assume, without any loss of generality, that $[C_{11}|C_{12}]$ can be taken as the basis of C and notice also from chapter (12) that for 'testability' we should have the further condition

$$\begin{array}{cc} C_{12} & s \\ C_{22} & q-s \\ m-r & \end{array} = \begin{array}{cc} s & C_{11} \\ q-s & C_{21} \\ & r \end{array} [A_1'(r\times n)A_1(n\times r)]^{-1}A_1'(r\times n)A_2(n\times\overline{m-r}).$$

$$\dots \quad (13.2.12)$$

We recall from (12.4.13) that the current F-test for (13.2.10) (at a level, say α) has a critical region given by

$$\frac{(n-r)\mathbf{x}'A_1(A_1'A_1)^{-1}C_{11}'[C_{11}(A_1'A_1)^{-1}C_{11}']^{-1}\, C_{11}(A_1'A_1)^{-1}A_1'\mathbf{x}}{s[\mathbf{x}'\mathbf{x}-\mathbf{x}'A_1(A_1'A_1)^{-1}A_1'\mathbf{x}]} \geqslant F_\alpha(s, n-r).$$

$$\dots \quad (13.2.13)$$

Recall that when (13.2.10) is true, the left hand side of (13.2.13) has the central F-distribution with d.f. s and $n-r$. Assume next that (13.2.10) is not true, but what is true is

$$C(q\times m)\boldsymbol{\xi}(m\times 1) = \boldsymbol{\eta}(q\times 1) \quad (\boldsymbol{\eta} \text{ being given}), \text{ or say}$$

$$\begin{array}{ccc} s & C_{11} & C_{12} \\ q-s & C_{21} & C_{22} \\ & r & m-r \end{array} \begin{array}{cc} \boldsymbol{\xi}_1 & r \\ \boldsymbol{\xi}_2 & m-r \\ 1 & \end{array} = \begin{array}{cc} \boldsymbol{\eta}_1 & s \\ \boldsymbol{\eta}_2 & q-s \\ 1 & \end{array} \qquad \dots \quad (13.2.14)$$

It is evident from the theory of linear equations that $\boldsymbol{\eta}_2$ could not be just arbitrary but that it must be related to $\boldsymbol{\eta}_1$ through the same matrix prefactor through which C_{21} is related to C_{11}, and C_{22} to C_{12}.

Then proceeding exactly as in chapter (12) we check that if in the left side of (13.2.13) we replace $\mathbf{x}(n\times 1)$ by $\mathbf{x}(n\times 1)-B(n\times s)\ \boldsymbol{\eta}_1(s\times 1)$, the resulting expression is distributed as a central F with d.f. s and $n-r$, B being given by

$$B(n\times s) = A_1(n\times r)(A_1'A_1)^{-1}(r\times r)C_{11}'(r\times s)[C_{11}(A_1'A_1)^{-1}C_{11}']^{-1}(s\times s). \quad \dots \quad (13.2.15)$$

If in (13.2.13) we now replace $\mathbf{x}$ by $\mathbf{x}-B\boldsymbol{\eta}_1$ and also put

$$[C_{11}(A_1'A_1)^{-1}C_{11}']^{-1} = \tilde{U}(s\times s)\tilde{U}'(s\times s) \qquad .. \quad (13.2.16)$$

(notice that by (A.3.7) $\tilde{U}$ is determinate and also unique), then it is easy to see exactly in the same way as in the previous case that the resulting statement $\Longleftrightarrow$ the following

$$\mathbf{x}'A_1(A_1'A_1)^{-1}C_{11}'\tilde{U}\mathbf{a}(s\times 1)-(EV)^{\frac{1}{2}}[sF_{\alpha}(s, n-r)]^{\frac{1}{2}}$$

$$\leqslant \boldsymbol{\eta}_1'(1\times s)\, B'(s\times n)A_1\,(A_1'A_1)^{-1}C_{11}'\tilde{U}\mathbf{a} \leqslant \mathbf{x}'A_1(A_1'A_1)^{-1}C_{11}'\tilde{U}\mathbf{a}+[EV]^{\frac{1}{2}}[sF_{\alpha}(s, n-s)]^{\frac{1}{2}}, \quad \ldots \quad (13.2.17)$$

for all $\mathbf{a}$ subject to $\mathbf{a}'(1\times s)\mathbf{a}(s\times 1) = 1$, where B is given by (13.2.15) and $\tilde{U}$ by (13.2.16) and the error variance EV by

$$EV = [\mathbf{x}'\mathbf{x}-\mathbf{x}'A_1(A_1'A_1)^{-1}A_1'\mathbf{x}]/(n-r). \quad \ldots \quad (13.2.18)$$

Substituting for B' from (13.2.15) we check that

$$B'A_1(A_1'A_1)^{-1}C_{11}' = I. \quad \ldots \quad (13.2.19)$$

Also putting $\tilde{U}(s\times s)\mathbf{a}(s\times 1) = \mathbf{b}(s\times 1)$ and using (13.2.16) we note that

$$1 = \mathbf{a}'\mathbf{a} = \mathbf{b}'\tilde{U}'^{-1}\tilde{U}^{-1}\,\mathbf{b} = \mathbf{b}'(\tilde{U}\tilde{U}')^{-1}\mathbf{b} = \mathbf{b}'\,[C_{11}(A_1'A_1)^{-1}C_{11}']\;\mathbf{b}. \quad \ldots \quad (13.2.20)$$

The statement (13.2.17) thus reduces to the following:

$$\mathbf{x}'A_1(A_1'A_1)^{-1}C_{11}'\mathbf{b}-(EV)^{\frac{1}{2}}[s\,F_{\alpha}(s, n-r)]^{\frac{1}{2}} \leqslant \boldsymbol{\eta}_1'\,\mathbf{b}$$

$$\leqslant \mathbf{x}'A_1(A_1'\,A_1)^{-1}C_{11}'\mathbf{b}+(EV)^{\frac{1}{2}}[sF_{\alpha}(s, n-r)]^{\frac{1}{2}} \quad \ldots \quad (13.2.21)$$

for all $\mathbf{b}$ subject to (13.2.20).

If we go back to (13.2.20) and reason as in (iii), it is easy to check that (13.2.21) implies

$$s^{\frac{1}{2}}s^{*}-(EV)^{\frac{1}{2}}[sF_{\alpha}(s, n-r)]^{\frac{1}{2}} \leqslant \{\boldsymbol{\eta}_1'[C_{11}(A_1'A_1)^{-1}C_{11}']^{-1}\boldsymbol{\eta}_1\}^{\frac{1}{2}}$$

$$\leqslant s^{\frac{1}{2}}s^{*}+(EV)^{\frac{1}{2}}[sF_{\alpha}(s, n-r)]^{\frac{1}{2}}, \quad \ldots \quad (13.2.21.1)$$

where ss^{*2} is the "sum of squares due to the hypothesis", given by the numerator of (12.4.13) with the factor $(n-r)$ taken out. (13.2.21.1) is thus a confidence statement with a confidence coefficient $\geqslant 1-\alpha$.

This, therefore, is a set of simultaneous confidence bounds (with a joint confidence coefficient $1-\alpha$) on all arbitrary linear functions of $\boldsymbol{\eta}_1$. It is easy to see that (13.2.21) subsumes as special cases, the confidence statements (13.2.1), (13.2.2) and (13.2.7). Nevertheless, for expository purposes, it is worthwhile to discuss separately the simpler cases first.

Two other particular examples of (13.2.17), of special practical interest are also discussed, separately, in (v) and (vi). (v) Suppose we have y_h's $(h = 1, 2, \ldots, n)$ each being an $N(\theta_h, \sigma^2)$ such that $\text{cov}(y_h, y_{h'}) = \rho\sigma^2(h \neq h' = 1, 2, \ldots, n)$, where ρ is known, but θ_h and σ^2 are unknown, but an independent estimate s^2 of σ^2 based on n' degrees of freedom is available. It is required to obtain a set of simultaneous confidence bounds on the mean differences

$$\theta_h-\theta_{h'}, \text{ with } h, h' = 1, 2, \ldots, n, \quad h \neq h'. \quad \ldots \quad (13.2.22)$$

We have now a finite set of parametric functions. Let $z_h+\chi\bar{\theta} = y_h+\chi\bar{y}$ where $\bar{y} = (y_1+y_2+\ldots+y_n)/n$, $\bar{\theta} = (\theta_1+\theta_2+\ldots+\theta_n)/n$ and the disposable constant χ is so adjusted that the z_h's are uncorrelated. Then

$$E(z_h) = \theta_h,\ \operatorname{var}(z_h) = \sigma^2(1-\rho),\ \text{with } h = 1, 2, \ldots, n. \qquad \ldots \ (13.2.23)$$

Let

$$\psi_{hh'} = \frac{(z_h-\theta_h)-(z_{h'}-\theta_{h'})}{s\sqrt{(1-\rho)}},\ \text{with } h, h' = 1, 2, \ldots, n, h \neq h'. \qquad \ldots \ (13.2.24)$$

Then

$$\psi_{hh'} \leqslant d, \qquad \ldots \ (13.2.25)$$

implies

$$y_h-y_{h'}-sd\sqrt{1-\rho} \leqslant \theta_h-\theta_{h'} \leqslant y_h-y_{h'}+sd\sqrt{1-\rho}. \qquad \ldots \ (13.2.26)$$

Let W_θ be the intersection of the regions (13.2.25). Then clearly the necessary and sufficient condition for the sample point to lie in W_θ is that

$$q = \frac{w}{s\sqrt{1-\rho}} \leqslant d, \qquad \ldots \ (13.2.27)$$

where

$$w = \sup_{hh'}(z_h-\theta_h)-(z_{h'}-\theta_{h'}),\ \text{with } h, h' = 1, 2, \ldots, n;\ h \neq h' \quad \ldots(13.2.28)$$

Thus if we set $d = q_\alpha(n, n')$, where $q_\alpha(n, n')$ is the upper α-point of the distribution of the studentized range with n, n' degrees of freedom, that is the ratio of the range of n independent normal variates with zero mean to the square root of an independent estimate of their common variance based on n' degrees of freedom, then the required simultaneous confidence intervals for the parametric functions (13.2.22) are

$$y_h-y_{h'}-sq_\alpha(n, n')\sqrt{1-\rho} \leqslant \theta_h-\theta_{h'} \leqslant y_h-y_{h'}+sq_\alpha(n, n')\sqrt{1-\rho}. \ \ldots \ (13.2.29)$$

In particular $y_1, y_2, \ldots, y_n$ may be the means of n random samples of equal size drawn from normal populations with a common (unknown) variance, or may be the estimated treatment effects in a randomized block or a balanced incomplete block experiment.
(vi) In factorial experiments we are usually interested in estimating linear functions of treatment effects whose estimates are independently and normally distributed with a common variance which can be independently estimated by an appropriate multiple of the error mean square in the analysis of variance. The distribution needed for simultaneous estimation in this case is slightly different from that occurring in (v).

Suppose, for example, that we have observations for a $2\times2\times2\times2$ factorial experiment with factors A, B, C, D, and that we are interested in simultaneously estimating the main effects and two factor interactions only. We shall suppose that the experiment is so laid out that none of these is confounded in any replication. Let $t_{11}, t_{22}, t_{33}, t_{44}$ denote the true main effects and $t_{12}, t_{13}, t_{14}, t_{23}, t_{24}, t_{34}$ the true two factor interactions. The order of the subscripts in t_{ij} is immaterial, that is, $t_{ij} = t_{ji}$. We can then write in the usual notation,

$$t_{11} = (1/8)(a-1)(b+1)(c+1)(d+1), \quad \ldots \quad (13.2.30)$$

$$t_{12} = (1/8)(a-1)(b-1)(c+1)(d+1), \quad \ldots \quad (13.2.31)$$

with similar expressions for other main effects and interactions. Let y_{ij} be the estimate of t_{ij}. Then reasoning as before we get the following simultaneous confidence intervals for t_{ij}:

$$y_{ij}-sx_{\alpha}(n, n') \leqslant t_{ij} \leqslant y_{ij}+sx_{\alpha}(n, n'), \quad \ldots \quad (13.2.32)$$

where s^2 is an estimate of $V(y_{ij})$, based on n' degrees of freedom available for the estimate of error, and where n, which is 10 in this particular example, is the number of linear functions to be estimated.

The meaning of $x_\alpha(n, n')$ is as follows. Let $x_1, x_2, \ldots, x_n$ be independent normal variates with zero mean and variance σ^2. Let $|x|$ be the maximum of $|x_1|$, $|x_2|, \ldots, |x_n|$ and let s^2 be an independent estimate of σ^2 based on n' degrees of freedom. Then $x_\alpha(n, n')$ is the upper α-point of the distribution of $|x|/s$.

In a factorial experiment in which each factor is at more than two levels, the above will still apply if the n linear functions to be simultaneously estimated (or tested for vanishing) are so chosen that their estimates are independently distributed with a common variance.

13.3. *Variances of one or two normal populations.*

(i) Given a random sample of size $n+1$ (mean: $\bar{x}$ and *s.d.*:s) from an $N(\xi, \sigma^2)$, we take over from (6.3.1) the following statement with probability $1-\alpha$:

$$\chi^2_{1\alpha}(n) \leqslant ns^2/\sigma^2 \leqslant \chi^2_{2\alpha}(n), \quad \ldots \quad (13.3.1)$$

where $\chi^2_{2\alpha}(n)$ and $\chi^2_{1\alpha}(n)$ are the upper α_1 and lower α_2 point of χ^2-distribution with d.f. n and α is partitioned into α_1 and α_2 such that (*a*) $\alpha_1+\alpha_2 = \alpha$ and (*b*) the complement of (13.3.1), i.e., the critical region is locally unbiassed (in the neighbourhood of σ) in which case it has also been shown to have the monotonicity property. We now rewrite (13.3.1) as

$$ns^2/\chi^2_{2\alpha}(n) \leqslant \sigma^2 \leqslant ns^2/\chi^2_{1\alpha}(n), \quad \ldots \quad (13.3.2)$$

which gives confidence bounds on σ^2 with a confidence coefficient $1-\alpha$ and having properties in terms of *shortness* similar to those possessed by (13.3.1) in terms of the second kind of error, already discussed.

(ii) Given two random samples of sizes n_h+1 (mean: $\bar{x}_h$ and *s.d.* :s_h) ($h =$ 1, 2) from two $N(\xi_h, \sigma_h^2)$, we take over from (6.3.2) the following statement with probability $1-\alpha$:

$$F_{1\alpha}(n_1, n_2) \leqslant \frac{s_1^2}{s_2^2} \Big/ \frac{\sigma_1^2}{\sigma_2^2} \leqslant F_{2\alpha}(n_1, n_2), \quad \ldots \quad (13.3.3)$$

where $F_{1\alpha}(n_1, n_2)$ and $F_{2\alpha}(n_1, n_2)$ are the upper α_1 and lower α_2 points of F-distribution with d.f. n_1 and n_2 and α is partitioned into α_1 and α_2 such that (*a*) $\alpha_1+\alpha_2 = \alpha$ and (b) the complement of (13.3.3), i.e. the critical region is locally unbiassed (in the neighbourhood of σ_1/σ_2) in which case it has also been shown to have the monotonicity property. We now rewrite (13.3.3) as

$$\frac{s_1^2}{s_2^2} \Big/ F_{2\alpha}(n_1, n_2) \leqslant \frac{\sigma_1^2}{\sigma_2^2} \leqslant \frac{s_1^2}{s_2^2} \Big/ F_{1\alpha}(n_1, n_2), \qquad \ldots \quad (13.3.4)$$

which gives confidence bounds on σ_1^2/σ_2^2 with a confidence coefficient $1-\alpha$ and having properties in terms of *shortness* similar to those possessed by (13.3.3) in terms of the second kind of error, already discussed.

13.4. *Coefficient of regression for a bivariate normal population.* Let x_1 and x_2 be distributed as a bivariate normal with variances σ_1^2 and σ_2^2 and correlation coefficient ρ, and let the sample variances (on a sample of size $n+2$) be denoted by s_1^2 and s_2^2, and the sample correlation coefficient by r. Also let $b_{12} = s_1r/s_2$ and $\beta_{12}= \sigma_1\rho/\sigma_2$. It is easy to check that the variates $(x_1-\beta_{12}x_2)$ and x_2 are uncorrelated, so that when the population parameters are σ_1, σ_2 and ρ, $n^{\frac{1}{2}}r^*/(1-r^{*2})^{\frac{1}{2}}$ has the t-distribution with n d.f. Here r^* stands for the sample correlation between $(x_1-\beta_{12}x_2)$ and x_2, that is,

$$\begin{aligned} r^* &= (s_1\, s_2\, r-\beta_{12}s_2^2)/(s_1^2-2\beta_{12}s_1s_2r+\beta_1^2s_2^2)^{\frac{1}{2}}s_2 \\ &= (s_1r-\beta_{12}s_2)/[(s_1r-\beta_{12}s_2)^2+(1-r^2)s_1^2]^{\frac{1}{2}} = (b_{12}-\beta_{12})/[(b_{12}-\beta_{12})^2+(1-r^2)s_1^2/s_2^2]^{\frac{1}{2}}, \end{aligned} \qquad \ldots \quad (13.4.1)$$

and, therefore,

$$r^*/\sqrt{1-r^{*2}} = \frac{s_2}{s_1}\;\frac{b_{12}-\beta_{12}}{(1-r^2)^{\frac{1}{2}}}\,. \qquad \ldots \quad (13.4.2)$$

Now consider the statement

$$-t_\alpha(n) \leqslant n^{\frac{1}{2}}r^*/(1-r^{*2})^{\frac{1}{2}} \leqslant t_\alpha(n), \qquad \ldots \quad (13.4.3)$$

where $t_\alpha(n)$ gives the upper $\alpha/2$-point of the t-distribution with n d.f. This is easily seen to reduce to the following confidence statement on β_{12} (with a confidence coefficient $1-\alpha$):

$$b_{12}-\frac{t_\alpha(n)}{\sqrt{n}}\,(1-r^2)^{\frac{1}{2}}\,\frac{s_1}{s_2} \leqslant \beta_{12} \leqslant b_{12}+\frac{t_\alpha(n)}{\sqrt{n}}\,(1-r^2)^{\frac{1}{2}}\,\frac{s_1}{s_2}\,. \qquad \ldots \quad (13.4.4)$$

By inversion of (13.4.4) the test that we obtain for the associated hypothesis H_0: $\beta_{12} = 0$, that is $\rho = 0$, is easily checked to be the customary test based on "r" and hence just the t-test. Similar procedures would go through for "partial regressions" or "multiple regressions". The interesting point here is that it would be far more difficult to give corresponding confidence bounds to ρ, because this would have to be done by inverting the distribution of the noncentral r, which is quite complicated.

13.5. *Difference in mean values between two variates having a bivariate normal distribution.* Let

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ be } N \left[\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix},\ \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \right]$$

Then since (x_1-x_2) is $N[(\xi_1-\xi_2),\ (\sigma_1^2+\sigma_2^2-2\sigma_1\sigma_2\rho)]$, it is well known and easy to check that in terms of a sample of size n with sample means $\bar{x}_1$ and $\bar{x}_2$, sample variances s_1^2 and s_2^2, and sample correlation coefficient r, we have the following confidence interval for $\xi_1-\xi_2$ (with a confidence coefficient $1-\alpha$)

$$\bar{x}_1-\bar{x}_2-st_{\alpha/2}(n-1)/\sqrt{n} \leqslant \xi_1-\xi_2 \leqslant \bar{x}_1-\bar{x}_2+st_{\alpha/2}(n-1)/\sqrt{n}, \quad \ldots \quad (13.5.1)$$

where $t_{\alpha/2}(n-1)$ is the upper $\alpha/2$ point of the ordinary t-distribution with d.f. $(n-1)$ and $s^2 = s_1^2+s_2^2-2rs_1s_2$. Mathematically this is no doubt deducible from (13.2.1) and is also a special case of (13.2.7) but this is so very important in practice that a separate and explicit statement may not be out of place.

13.6. *Ratio of variances of two variates having a bivariate normal distribution.* Let

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ be } N\left[\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix},\ \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}\right].$$

Then for any constant λ, it is easy to check that covariance $(x_1-\lambda x_2,\ x_1+\lambda x_2) = V(x_1)-\lambda^2 V(x_2)$. Thus this will be $= 0$ if $\lambda^2 = V(x_1)/V(x_2) = \sigma_1^2/\sigma_2^2$. Thus, with a positive $\lambda = \sigma_1/\sigma_2$, $(x_1-\lambda x_2)$ and $(x_1+\lambda x_2)$ will be uncorrelated and hence for a sample of size n, with sample variances s_1^2 and s_2^2, and sample correlation coefficient r, $\sqrt{n-2}\ r^*/(1-r^{*2})^{\frac{1}{2}}$ has the t-distribution with d.f. $n-2$ where

$$\begin{aligned} r^* &= \text{sample correlation between } (x_1-\lambda x_2) \text{ and } (x_1+\lambda x_2) \\ &= (s_1^2-\lambda^2 s_2^2)/[(s_1^2+\lambda^2 s_2^2+2\lambda s_1 s_2 r)(s_1^2+\lambda^2 s_2^2-2\lambda s_1 s_2 r)]^{\frac{1}{2}} \\ &= (s_1^2-\lambda^2 s_2^2)/[s_1^4+\lambda^4 s_2^4+2\lambda^2 s_1^2 s_2^2(1-2r^2)]^{\frac{1}{2}} \qquad \ldots \quad (13.6.1) \end{aligned}$$

Thus starting from the statement with probability $1-\alpha$

$$|\sqrt{n-2}r^*/(1-r^{*2})^{\frac{1}{2}}| \leqslant t_{\alpha/2}(n-2), \quad \ldots \quad (13.6.2)$$

and remembering that $\lambda = \sigma_1/\sigma_2$ and substituting (13.6.1) for r^* in terms of s_1, s_2, r we have for σ_1^2/σ_2^2 the following confidence bounds (with a confidence coeficient $1-\alpha$)

$$\begin{aligned} &\frac{s_1^2}{s_2^2}\left[\left(1+\frac{2}{n-2}\, t_{\alpha/2}^2\, \overline{1-r^2}\right)-\left\{\left(1+\frac{2}{n-2}\, t_{\alpha/2}^2\, \overline{1-r^2}\right)^2-1\right\}^{\frac{1}{2}}\right] \leqslant \frac{\sigma_1^2}{\sigma_2^2} \\ &\leqslant \frac{s_1^2}{s_2^2}\left[\left(1+\frac{2}{n-2}\, t_{\alpha/2}^2\, \overline{1-r^2}\right)+\left\{\left(1+\frac{2}{n-2}\, t_{\alpha/2}^2\, \overline{1-r^2}\right)^2-1\right\}^{\frac{1}{2}}\right]. \qquad \ldots \quad (13.6.3) \end{aligned}$$

The cases discussed in 13.5 and 13.6 are relevant from a physical standpoint where we have two *comparable* correlated variates, for example, the measurements on the same characteristic of a set of individuals before and after the administration of a drug.

CHAPTER FOURTEEN

Multivariate Confidence Bounds*

14.1. *A convenient notation.* From now on we shall make use of a rather convenient notation. A random sample of size n from a p-variate normal population, i.e., an $X(p\times n)$ having the p.d.f.

$$(2\pi)^{-pn/2}|\Sigma|^{-n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(X-\xi)\quad(X'-\xi')\right]$$

where $\xi(p\times n)$stands for a $p\times n$ matrix each column of which is the same $p\times 1$ vector $\boldsymbol{\xi}$ (with components $\xi_1, \ldots, \xi_p$) will be referred to as $X(p\times n)$: $N^*(\boldsymbol{\xi},, \Sigma)$. *A* matrix $Y(p\times n)$ having the p.d.f.

$$(2\pi)^{-pn/2}|\Sigma|^{-n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}YY'\right],$$

will be referred to as $Y(p\times n)$: $N^*(\mathbf{0}, \Sigma)$. We recall from (4.4)—(4.14) that starting with an $X(p\times(n+1))$: $N^*(\boldsymbol{\xi}, \Sigma)$ and transforming and integrating we can always have an $Y(p\times n)$: $N^*(\mathbf{0}, \Sigma)$, such that

$$\begin{aligned} nS(p\times p) = Y(p\times n)\;Y'(n\times p) &= X(p\times(n+1))\;X'((n+1)\times p)\\ &\quad -(n+1)\;\bar{\mathbf{x}}(p\times 1)\;\bar{\mathbf{x}}'(1\times p), \end{aligned} \qquad \ldots \quad (14.1.1)$$

where $\bar{\mathbf{x}}(p\times 1) = \dfrac{1}{n}\,X(p\times n)\;1(n\times 1)$, $1(n\times 1)$ being an $n\times 1$ column vector with components $(1, 1, \ldots, 1)$.

14.2. *Confidence bounds relating to the mean vector for a multivarite normal distribution.* Given an $X(p\times(n+1))$: $N^*(\boldsymbol{\xi},\Sigma)$, suppose we try to obtain simultaneous confidence bounds on arbitrary linear compounds of the population mean vector $\boldsymbol{\xi}$. Consider the statement that

$$(n+1)^{\frac{1}{2}}|\mathbf{a}'(\bar{\mathbf{x}}-\boldsymbol{\xi})|/(\mathbf{a}'S\mathbf{a})^{\frac{1}{2}} \leqslant c,$$

or

$$(n+1)\;\mathbf{a}'(\bar{\mathbf{x}}-\boldsymbol{\xi})\;(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\mathbf{a}/\mathbf{a}'S\mathbf{a} \leqslant c^2, \qquad \ldots \quad (14.2.1)$$

where $\bar{\mathbf{x}}$ is the sample mean vector and S is the sample covariance matrix, and $\mathbf{a}(p\times 1)$ is an arbitrary non-null nonstochastic column vector and c is a given positive constant. The statement (14.2.1) stems from the customary Student's t-test and the associated confidence interval (both having well known optimum properties) relating to the parameter $\mathbf{a}'\boldsymbol{\xi}$. Now, for a given (positive) c and given $\bar{\mathbf{x}}$, $\boldsymbol{\xi}$, S and of course n, the set of all statements (14.2.1) for all possible non-null vectors $\mathbf{a}$ is exactly equivalent to the statement that

$$\sup_{\mathbf{a}}\;(n+1)\;\mathbf{a}'(\bar{\mathbf{x}}-\boldsymbol{\xi})(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\mathbf{a}/\mathbf{a}'S\mathbf{a} \leqslant c^2. \qquad \ldots \quad (14.2.2)$$

It is well known that this "sup" comes out as $\operatorname{tr}(n+1)\;S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\xi})(\bar{\mathbf{x}}'-\boldsymbol{\xi}')$ or as $\operatorname{tr}(n+1)\times(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\;S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\xi})$ (since $\operatorname{tr} AB = \operatorname{tr} BA$), or simply as $(n+1)(\bar{\mathbf{x}}'-\boldsymbol{\xi}')\,S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\xi})$ (since tr scalar = scalar). It is also well known that under the null hypothesis, this is distributed as the central Hotelling's T^2 with d.f. p and $n+1-p$ and that if in this statistic $\boldsymbol{\xi}$ is replaced by $\boldsymbol{\xi}^*(\neq\boldsymbol{\xi})$, the resulting statistic is distributed as the

---

* See reference [44, 45, 46] in this connection.

non-central Hotelling's T^2 with the same d.f. and with the non-centrality parameter $\gamma^2 \equiv (\boldsymbol{\xi}^{*\prime}-\boldsymbol{\xi}')\Sigma^{-1}\ (\boldsymbol{\xi}^*-\boldsymbol{\xi})$. Going back to (14.2.1) it is thus easy to see that if,

$$P\left[\frac{(n+1)\ \mathbf{a}'(\overline{\mathbf{x}}-\boldsymbol{\xi}^*)(\overline{\mathbf{x}}'-\boldsymbol{\xi}^{*\prime})\mathbf{a}}{\mathbf{a}'S\mathbf{a}} \leqslant c^2 \middle| \boldsymbol{\xi}^* = \boldsymbol{\xi}\right] = 1-\alpha, \qquad \dots \quad (14.2.3)$$

then $c^2 = T^2_\alpha$ is the upper α-point of the central Hotelling's T^2-distribution with d.f. p and $n+1-p$ and can be conveniently written as $T^2_\alpha(p, n+1-p)$. From (14.2.3) we have thus, with a confidence coefficient $1-\alpha$, the set of simultaneous or multiple confidence bounds (for all $\boldsymbol{\xi}$ and all nonnull $\mathbf{a}$):

$$\mathbf{a}'\overline{\mathbf{x}}-[T^2_\alpha(\mathbf{a}'S\mathbf{a})/(n+1)]^{\frac{1}{2}} \leqslant \mathbf{a}'\boldsymbol{\xi} \leqslant \mathbf{a}'\overline{\mathbf{x}}+[T^2_\alpha(\mathbf{a}'S\mathbf{a})/(n+1)]^{\frac{1}{2}}. \qquad \dots \quad (14.2.4)$$

It should be noted that (14.2.4) gives the simultaneous confidence bounds on all arbitrary linear compounds of the p components of the population mean vector $\boldsymbol{\xi}$. The shortness (in the sense of probability) of this set of confidence bounds, that is, the probability of these bounds covering $\boldsymbol{\xi}^*$ when, in fact, $\boldsymbol{\xi}^* \neq \boldsymbol{\xi}$, is obviously

$$1-P\,[\text{noncentral } T^2 \geqslant T^2_\alpha | \tau^2].$$

From the well known fact that the power function of Hotelling's T^2-test is a monotonically increasing function of the nonnegative τ, it follows, therefore, that the shortness of the confidence bound (14.2.4) tends to zero as $\tau \to \infty$.

Let us go back to (14.2.4) and choose $\mathbf{a}'$ so as to maximize $\mathbf{a}'\boldsymbol{\xi}$. Then it is easy to see that (14.2.4) implies that $(\boldsymbol{\xi}'\boldsymbol{\xi})^{\frac{1}{2}} \leqslant (\overline{\mathbf{x}}'\overline{\mathbf{x}})^{\frac{1}{2}}+[T^2_\alpha/(n+1)]^{\frac{1}{2}}c^{\frac{1}{2}}_{\max}(S)$. A similar result follows for the other side of the inequality and thus (14.2.4) should imply

$$(\overline{\mathbf{x}}'\overline{\mathbf{x}})^{\frac{1}{2}}-[T^2_\alpha/(n+1)]^{\frac{1}{2}}c^{\frac{1}{2}}_{\max}(S) \leqslant (\boldsymbol{\xi}'\boldsymbol{\xi})^{\frac{1}{2}} \leqslant (\overline{\mathbf{x}}'\overline{\mathbf{x}})^{\frac{1}{2}}+[T^2_\alpha/(n+1)]^{\frac{1}{2}}c_{\max}(S), \qquad \dots \quad (14.2.5)$$

which, therefore, is a confidence statement with a confidence coefficient $\geqslant 1-\alpha$.

Back in (14.2.4), if we cut out the i-th element of $\mathbf{a}$, the corresponding element of $\overline{\mathbf{x}}$ and $\boldsymbol{\xi}$, and the corresponding row and column of S (with $i = 1,2,...p$) and reason in the same manner as in the case of (14.2.5) we have p truncated confidence statements of exactly the same form as (14.2.5). Likewise, cutting out any two elements say the i-th and the j-th ($i = j = 1,2,...p$), we have $\binom{p}{2}$ truncated confidence statements of this form, and so on. Thus altogether we have 2^p-1 confidence statements, on which the leading one is (14.2.5), all with a joint confidence coefficient $\geqslant 1-\alpha$, which, in a sense, provides a complete analysis of the problem.

14.3. *Confidence bounds relating to mean differences in k multivariate distribution.* Given $X_h(p\times(n_h+1))$: $N(\boldsymbol{\xi}_h, \Sigma)$, ($h = 1, 2, \dots, k$) let us try to obtain a set of simultaneous confidence bounds on all arbitrary double linear compounds of the p-components of the k population mean vectors measured from the weighted grand mean vector. Consider now the statement

$$\left|\sum_{h=1}^{k} b_h\mathbf{a}'(n_h+1)^{\frac{1}{2}}(\overline{\mathbf{x}}_h-\overline{\mathbf{x}}-\boldsymbol{\xi}_h+\boldsymbol{\xi})\right| \leqslant [(k-1)g^2\mathbf{a}'S\mathbf{a}]^{\frac{1}{2}}, \qquad \dots \quad (14.3.1)$$

where $\bar{\mathbf{x}}_h$ is the mean vector for the h-th sample,

$$\bar{\mathbf{x}} = \sum_{h=1}^{k} (n_h+1)\bar{\mathbf{x}}_h / \sum_{h=1}^{k} (n_h+1), \quad \boldsymbol{\xi} = \sum_{h=1}^{k} (n_h+1)\boldsymbol{\xi}_h / \sum_{h=1}^{k} (n_h+1),$$

where S is the pooled "within" covariance matrix of the k-samples, given by

$$\left(\sum_{h=1}^{k} n_h\right) S = \sum_{h=1}^{k} [X_h X_h' - (n_h+1)\bar{\mathbf{x}}_h \bar{\mathbf{x}}_h'],$$

and g is a given positive constant, $\mathbf{a}(p \times 1)$ is an arbitrary non-null non-stochastic column vector and the b_h's are arbitrary coefficients subject to $\sum_{h=1}^{k} b_h^2 = 1$.

If we now use the result that

$$\left| \sum_{h=1}^{k} b_h y_h \right| \leqslant +\sqrt{d^2} \Longleftrightarrow \sum_{h=1}^{k} y_h^2 \leqslant d^2;$$

then it directly follows that, given all the other quantities including $\mathbf{a}$, and under all possible variations of b_h's subject to $\sum_{h=1}^{k} b_h^2 = 1$, the statement (14.3.1) is precisely equivalent to the statement that

$$\sum_{h=1}^{k} [\mathbf{a}'(n_h+1)^{\frac{1}{2}}(\bar{\mathbf{x}}_h - \bar{\mathbf{x}} - \boldsymbol{\xi}_h + \boldsymbol{\xi})]^2 / (k-1)\mathbf{a}'S\mathbf{a} \leqslant g^2,$$

or

$$\sum_{h=1}^{k} \mathbf{a}'(n_h+1)(\bar{\mathbf{x}}_h - \bar{\mathbf{x}} - \boldsymbol{\xi}_h + \boldsymbol{\xi})(\bar{\mathbf{x}}_h' - \bar{\mathbf{x}}' - \boldsymbol{\xi}_h' + \boldsymbol{\xi}')\mathbf{a} / (k-1)\mathbf{a}'S\mathbf{a} \leqslant g^2. \quad \ldots \quad (14.3.2)$$

Letting now $\mathbf{a}$ vary and putting

$$(k-1)S^* = \sum_{h=1}^{k} (n_h+1)(\bar{\mathbf{x}}_h - \bar{\mathbf{x}} - \boldsymbol{\xi}_h + \boldsymbol{\xi})(\bar{\mathbf{x}}_h' - \bar{\mathbf{x}}' - \boldsymbol{\xi}_h' + \boldsymbol{\xi}'), \quad \ldots \quad (14.3.3)$$

the statement (14.3.2), for all possible values of the non-null $\mathbf{a}$, is precisely equivalent to:

$$\sup_{\mathbf{a}} [\mathbf{a}'S^*\mathbf{a}/\mathbf{a}'S\mathbf{a}] \leqslant g^2. \quad \ldots \quad (14.3.4)$$

As observed after (6.4.7) S is, a.e., p.d. and S^* is, a.e., p.s.d. of rank $q = \min(p, k-1)$ (p.s.d. if $p > k-1$ and p.d. if $p \leqslant k-1$) and $\sup_{\mathbf{a}}[\mathbf{a}'S^*\mathbf{a}/\mathbf{a}'S\mathbf{a}]$ is just the largest root c_q of the p-th degree determinantal equation in $c : |S^* - cS| = 0$. Of this equation all roots are non-negative, $p-q$ of them always zero and q are, a.e., positive. Thus (14.3.4) and hence (14.3.2) under all permissible variations of $\mathbf{a}$ and the b_h's, turns out to be equivalent to:

$$c_q \leqslant g^2. \quad \ldots \quad (14.3.5)$$

The distribution of c_q on the null hypothesis is known and relatively easy and involves as parameters p, $k-1$, $\sum_{h=1}^{k} n_h$. Computation of the 5 per cent and 1 per cent points is in progress . Thus if

$$P[c_q \leqslant c_\alpha | \text{null hypothesis}] = 1-\alpha, \qquad \dots \quad (14.3.6)$$

we can write $c_\alpha = c_\alpha(p, k-1, \sum_{h=1}^{k} n_h)$, and now combining (14.3.1)-(14.3.6) we have, with a confidence coefficient $1-\alpha$, the following set of multiple confidence statements (for all $\boldsymbol{\xi}_h$'s, all non-null $\mathbf{a}$'s and all b_h's subject to $\sum_{h=1}^{k} b_h^2 = 1$:

$$\sum_{h=1}^{k} b_h \mathbf{a}'(n_h+1)^{\frac{1}{2}}(\bar{\mathbf{x}}_h-\bar{\mathbf{x}})-[(k-1)c_\alpha \mathbf{a}'S\mathbf{a}]^{\frac{1}{2}}$$

$$\leqslant \sum_{h=1}^{k} b_h \mathbf{a}'(n_h+1)^{\frac{1}{2}}(\boldsymbol{\xi}_h-\boldsymbol{\xi}) \leqslant \sum_{h=1}^{k} b_h \mathbf{a}'(n_h+1)^{\frac{1}{2}}(\bar{\mathbf{x}}_h-\bar{\mathbf{x}})+[(k-1)c_\alpha \mathbf{a}'S\mathbf{a}]^{\frac{1}{2}}, \quad \dots \quad (14.3.7)$$

where $c_\alpha = c_\alpha(p, k-1, \sum_{h=1}^{k} n_h)$.

This gives simultaneous confidence bounds on all arbitrary double linear compounds of the p components of the difference between the k population mean vectors $\boldsymbol{\xi}_k$'s and the weighted grand mean of these which is $\boldsymbol{\xi}$. To discuss the shortness of (14.3.7) consider the non-central distribution of c_q, where c_q is defined after (14.3.4), i.e., c_q is the largest root of the equation in c:

$$|S^*-cS| = 0, \text{ where } S^* \text{ is given by (14.3.3).} \qquad \dots \quad (14.3.8)$$

It is easy to see that the distribution of the non-central c_q is really the distribution of e_q where e_q is the largest root of the equation in e obtained by (i) replacing in (14.3.2), $\boldsymbol{\xi}_h$ and $\boldsymbol{\xi}$ by $\boldsymbol{\xi}_h^*(\neq \boldsymbol{\xi}_h)$ and $\boldsymbol{\xi}^*(\neq \boldsymbol{\xi})$ and (ii) substituting the resulting value of S^* in (14.3.8) and (iii) assuming that the true population parameters are $\boldsymbol{\xi}_h$ and $\boldsymbol{\xi}$. The distribution is extremely difficult but is well known (see section7.6) to involve as parameters the positive roots $\gamma_1, \dots, \gamma_s(s \leqslant \min(p, k-1)$ of the determinantal equation in γ: $|\Sigma^*-\gamma\Sigma| = 0$, where Σ is the common covariance matrix of the k populations and $\Sigma^* = (k-1)^{-1} \sum_{h=1}^{k} (n_h+1)(\boldsymbol{\xi}_h^*-\boldsymbol{\xi}^*-\boldsymbol{\xi}_h+\boldsymbol{\xi})(\boldsymbol{\xi}^{*\prime}-\boldsymbol{\xi}^{*\prime}-\boldsymbol{\xi}_h'+\boldsymbol{\xi}')$. This Σ^* is necessarily at least p.s.d. of rank $\min(p, k-1) = s$(say), so that out of the p roots of the equation in γ, $p-s$ are zero and s positive. Using (9.2.3) and referring to section (11.3) we observe that there is a good upper bound to the shortness of (14.3.7) and that the shortness is a monotonically decreasing function of the deviation parameters and tends to zero as these tend to infinity. With two populations (and samples), we have $q=\min(p, 1)=1$, and thus only one positive sample root, say c, and at the most one positive population root, say γ. It is easy to check that in this case

$$c = \frac{(n_1+1)(n_2+1)}{n_1+n_2+2} \operatorname{tr} S^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2-\boldsymbol{\xi}_1+\boldsymbol{\xi}_2)(\bar{\mathbf{x}}_1'-\bar{\mathbf{x}}_2'-\boldsymbol{\xi}_1'+\boldsymbol{\xi}_2'),$$

$$\gamma = \frac{(n_1+1)(n_2+1)}{n_1+n_{-}+2} \operatorname{tr} \Sigma^{-1}(\boldsymbol{\xi}_1^*-\boldsymbol{\xi}_2^*-\boldsymbol{\xi}_1+\boldsymbol{\xi}_2)(\boldsymbol{\xi}_1^{*\prime}-\boldsymbol{\xi}_2^{*\prime}-\boldsymbol{\xi}_1'+\boldsymbol{\xi}_2'), \qquad \dots \quad (14.3.9)$$

and it is well known that, on the null hypothesis, c is distributed as central Hotelling's T^2 with d.f. p and n_1+n_2+1-p, and on the alternative as non-central Hotelling's T^2 with the same d.f. and with a deviation parameter γ. It is also easy to check that in this case the confidence statement (14.3.7) reduces to

$$\mathbf{a}'(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)-\left[\frac{n_1+n_2+2}{(n_1+1)(n_2+1)}T_\alpha^2\mathbf{a}'S\mathbf{a}\right]^{\frac{1}{2}} \leqslant \mathbf{a}'(\boldsymbol{\xi}_1-\boldsymbol{\xi}_2)$$

$$\mathbf{a}'(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)+\left[\frac{n_1+n_2+2}{(n_1+1)(n_2+1)}T_\alpha^2\mathbf{a}'S\mathbf{a}\right]^{\frac{1}{2}}, \qquad \ldots \quad (14.3.10)$$

where $T_\alpha^2 = T_\alpha^2(p, n_1+n_2+1-p)$ is the upper α-point of Hotelling's T^2. The shortness of (14.3.10) is exactly known and of course tends to zero as $\gamma \to \infty$.

14.4. *An important subset of the set of bounds* (14.3.7). Suppose now that, instead of all contrasts of the type: $\sum\limits_{h=1}^{k} b_h\mathbf{a}'(n_h+1)^{\frac{1}{2}}(\boldsymbol{\xi}_h-\boldsymbol{\xi})$ (with given restrictions on $\mathbf{a}$ and the b's) , we are interested in contrasts of the type :$\mathbf{a}'(\boldsymbol{\xi}_h-\boldsymbol{\xi}_l)$, for all non-null $\mathbf{a}'$ and all $h \neq l = 1, 2, \ldots, k$. It is easy to offer a multiple set of confidence bounds for contrasts of this type, which can be regarded as one kind of multivariate (under unequal sample sizes) analogue of a somewhat similar set given by Tukey for the corresponding univariate situations, and discussed in section (15.2). The proposed set is built up as follows. With the same notation as before, and with $n_{hl} = (n_h+1)\ (n_l+1)/(n_h+n_l+2)$ note that

$$T_{hl}^2 = n_{hl}\,(\bar{\mathbf{x}}_h'-\bar{\mathbf{x}}_l'-\boldsymbol{\xi}_h'+\boldsymbol{\xi}_l')S^{-1}(\bar{\mathbf{x}}_h-\bar{\mathbf{x}}_l-\boldsymbol{\xi}_h+\boldsymbol{\xi}_l) =$$
$$n_{hl}\ \mathrm{Sup}_{\mathbf{a}}\mathbf{a}'(\bar{\mathbf{x}}_h-\bar{\mathbf{x}}_l-\boldsymbol{\xi}_h+\boldsymbol{\xi}_l)(\bar{\mathbf{x}}_h'-\bar{\mathbf{x}}_l'-\boldsymbol{\xi}_h'+\boldsymbol{\xi}_l')\mathbf{a}/\mathbf{a}'S\mathbf{a}.$$

Thus, for a given pair (h, l), the statement that $T_{hl}^2 \leqslant T_\alpha^2$ is exactly equivalent to the statement that, for all non null $\mathbf{a}$'s,

$$\mathbf{a}'(\bar{\mathbf{x}}_h-\bar{\mathbf{x}}_l)-[T_\alpha^2\mathbf{a}'S\mathbf{a}/n_{hl}]^{\frac{1}{2}} \leqslant \mathbf{a}'(\boldsymbol{\xi}_h-\boldsymbol{\xi}_l) \leqslant \mathbf{a}'(\bar{\mathbf{x}}_h-\bar{\mathbf{x}}_l)+[T_\alpha^2\mathbf{a}'S\mathbf{a}/n_{hl}]^{\frac{1}{2}}.$$

We observe that when the true population means are $\boldsymbol{\xi}_h$'s, T_{hl}^2 is distributed as Hotelling's T^2 with d.f. p and $\sum\limits_{h=1}^{k} n_h+1-p$.

Now, considering all pairs (h, l) out of k samples (and k populations), it is easy to see that the statement: all T_{hl}^2's $\leqslant T_\alpha^2$, is precisely equivalent to the statement that the largest T_{hl}^2 out of all pairs is $\leqslant T_\alpha^2$, which again is equivalent to the statement that, for all non-null $\mathbf{a}$'s and all pairs (h, l) out of k,

$$\mathbf{a}'(\bar{\mathbf{x}}_h-\bar{\mathbf{x}}_l)-[T_\alpha^2\mathbf{a}'S\mathbf{a}/n_{hl}]^{\frac{1}{2}} \leqslant \mathbf{a}'(\boldsymbol{\xi}_h-\boldsymbol{\xi}_l) \leqslant \mathbf{a}'(\bar{\mathbf{x}}_h-\bar{\mathbf{x}}_l)+[T_\alpha^2\mathbf{a}'S\mathbf{a}/n_{hl}]^{\frac{1}{2}}. \quad \ldots \quad (14.4.1)$$

If the confidence coefficient of (14.4.1) is to be $1-\alpha$, then $T_\alpha = T_\alpha(p, n_1, n_2, \ldots, n_k)$ will be given by

$$P\ [\text{largest } T_{hl}^2 \text{ out of } \binom{k}{2} \text{ pairs} \geqslant T_\alpha^2|\ \text{null hypothesis}] = \alpha, \quad \ldots \quad (14.4.2)$$

It is obvious that the distribution of the largest T^2_{hl} involves as parameters just p and $n_1, n_2, \ldots n_k$. It is easy to see that the distribution is manageable only when the number of parameters is small. In particular, the case that $n_1 = n_2 = \ldots = n_k$ and $p = 1$, is identical with the one considered in section 13.2. It may also be noted that when $k = 2$, the largest T^2_{hl} will of course be Hotelling's T^2 distributed with d.f. p and n_1+n_2+1-p. Also the shortness of the confidence bounds (14.4.1) can be formally written as

$$P \text{ [largest } T^2_{hl} \text{ out of } \binom{k}{2} \text{ pairs} \leqslant T^2_\alpha(p, n_1, n_2, \ldots, n_k) \,|\, \text{alternative].}$$

It is important to observe that while each T^2_{hl} is individually distributed (on the null hypothesis) as a central Hotelling's T^2 with d.f. p and $\sum\limits_{h=1}^{k} n_h+1-p$, the $\binom{k}{2}$ T^2_{hl}'s are not independent, nor do we know what the distribution of the largest central T^2_{hl} is, to say nothing of the non-central case, so that the confidence statement (14.4.1) has not been reduced to practical terms as was done for the other cases discussed in this section. The distribution problem arising in this situation is now under investigation.

For the associated problem of testing $H_0 : \boldsymbol{\xi}_1 = \ldots = \boldsymbol{\xi}_k$, we set up as before the rule that if, for all nonnull $\mathbf{a}$ and all pairs (h, l), the bounds (14.4.1) include zero, we accept H_0 and reject it otherwise. The properties (including power) of this test are tied up in an obvious manner with those of the multiple confidence interval statement (14.4.1).

Notice that so far, in testing of hypotheses by inversion of confidence statements, we have considered two-decision problems. Suppose, at this point, for purposes of illustration, we offer a multi-decision procedure, namely that, for a given pair (h, l), we accept or reject $H(\boldsymbol{\xi}_h = \boldsymbol{\xi}_l)$ according as all those bounds (14.4.1) which involve $\bar{\mathbf{x}}_h$ and $\bar{\mathbf{x}}_l$ only include or exclude zero. It is obvious that in all other situations considered so far we could set up similar multi-decision procedures.

14.5. *Further observations.* In many situations it might be of greater physical interest to be able to make, instead of (14.3.7) or (14.4.1), a set of just $p \times \binom{k}{2}$ confidence interval statements, each relating to just one variate and difference betwen one of $\binom{k}{2}$ pairs. In other words, if $\boldsymbol{\xi}_h = (\xi_{1h}, \xi_{2h}, \ldots, \xi_{ph})$ $(h = 1, 2, \ldots, k)$ denote the p means for the h-th population, then we would like to make a statement of the form

$$f_{jhh'}(X_1, X_2, \ldots, X_k) \leqslant \xi_{jh} - \xi_{jh'} \leqslant F_{jhh'}(X_1, X_2, \ldots, X_k) \qquad \ldots \quad (14.5.1)$$

(with obvious applications to subsection 13.2), for all $h \neq h' = 1, 2, \ldots, k$ and all $j = 1, 2, \ldots, p$, where $f_{jhh'}$ and $F_{jhh'}$ are supposed to be two different functions of the whole set of $p \times \sum\limits_{h=1}^{k} (n_h+1)$ raw observations. It is clear that (14.5.1) is a sub-set of (14.4.1) which again is a subset of (14.3.7). Whether it is possible to make a statement like (14.5.1) in an elegant and useful way (i.e., with manageable functions $f_{jhh'}$ and $F_{jhh'}$) and with a given joint confidence coefficient $1-\alpha$, that is, free of the nuisance parameters Σ, is still an open question. It may well be that a range (not too wide) for the confidence coefficient itself is called for. Furthermore, whatever set of confidence intervals like (14.5.1) we propose, be it under a fixed confidence coefficient or under a confidence coefficient lying in a short range, the "goodness" of such a set would pose further questions. It is believed that in this situation a more promising approach might be one involving a suitable two-stage procedure.

14.6. *Confidence bounds connected with a general linear hypothesis.* In place of the set-up of section (14.3) consider the more general set-up of (iii c) of chapter 5, which is the following. We have an $X(p\times n)$ whose column vectors are independently distributed, the r-th vector $\mathbf{x}_r$ being $N(E(\mathbf{x}_r), \Sigma)$ $(r = 1, 2, \ldots, n)$. It is also given that $E(X')(n\times p) = A(n\times m)\xi(m\times p)$ where ξ is a set of unknown parameters and $A(n\times m)$ is given by the experimental situation such that it is of rank $r \leqslant m < n$. Also setting

$$A'(m\times n) = \begin{array}{cc} \begin{bmatrix} A_1' \\ \\ A_{,} \end{bmatrix} & \begin{array}{l} r \\ \\ m-r \end{array} \\ n & \end{array}, \qquad \ldots \quad (14.6.1)$$

let A_1' be a basis of A', i.e. of A. Next consider a matrix $C(q\times m)$ of rank $s \leqslant \min\,(q, r) \leqslant m < n$ with a structure given by

$$C(q\times m) = \begin{array}{cc} \begin{bmatrix} C_{11} & C_{12} \\ \\ C_{21} & C_{22} \end{bmatrix} & \begin{array}{l} s \\ \\ q-s \end{array} \\ \begin{array}{cc} r & m-r \end{array} & \end{array} \qquad \ldots \quad (14.6.2)$$

such that $[C_{11} \quad C_{12}]$ forms a basis and also that C satisfies (13.2.12). Then combining the results (13.2.9)-(13.2.21) with the results (14.3.1)-(14.3.7) it is easy to check that the following is a set of simultaneous confidence bounds (with a confidence coefficient $1-\alpha$),

$$\begin{aligned} &\mathbf{a}'(1\times p)X(p\times n)A_1(n\times r)(A_1'A_1)^{-1}(r\times r)C_{11}'(r\times s)\mathbf{b}(s\times 1) \\ &\quad -(\mathbf{a}'S\mathbf{a})^{\frac{1}{2}}[sc_\alpha(p, s, n-r)]^{\frac{1}{2}} \leqslant \mathbf{a}'(1\times p)\eta'(p\times s)\mathbf{b}(s\times 1) \\ &\quad \leqslant \mathbf{a}'XA_1(A_1'A_1)^{-1}C_{11}'\mathbf{b}+(\mathbf{a}'S\mathbf{a})^{\frac{1}{2}}[sc_\alpha(p, s, n-r)]^{\frac{1}{2}} \qquad \ldots \quad (14.6.3) \end{aligned}$$

for all non null $\mathbf{a}'(1\times p)$ and all $\mathbf{b}(s\times 1)$ subject to

$$\mathbf{b}'[C_{11}(A_1'\,A_1)^{-1}\,C_{11}']\;\mathbf{b} = 1, \qquad \ldots \quad (14.6.4)$$

where $\eta(s\times p)$ is given by

$$s[\underset{r}{C_{11}} \quad \underset{m-r}{C_{12}}]\xi(m\times p) = \eta(s\times p), \qquad \ldots \quad (14.6.5)$$

and $c_\alpha(p, s, n-r)$ is the upper α point of the distribution of the largest root of (6.4.7), under (12.7.3) with d.f. $(p, s, n-r)$.

The confidence bounds (14.6.3) are thus seen to be really on arbitrary double linear compounds of $[C_{11} \quad C_{12}]\xi$.

If we go back to (14.3.1)—(14.3.7) again, and normalize $\mathbf{b}(s \times 1)$ into $\tilde{U}\ (s \times s)\mathbf{b}^*(s \times 1)$ where $\mathbf{b}^*$ is a unit vector and $\tilde{U}\tilde{U}' = [C_{11}(A_1'A_1)^{-1}C_{11}']^{-1}$, then it is easy to see that the statement (14.6.3) will imply

$$c^{\frac{1}{2}}_{\max}(sS^*)-[sc_\alpha(p,\ s,\ n-r)]^{\frac{1}{2}}c^{\frac{1}{2}}_{\max}(S) \leqslant c^{\frac{1}{2}}_{\max}[\eta'(C_{11}(A_1'A_1)^{-1}C_{11}')^{-1}\eta]$$
$$\leqslant c^{\frac{1}{2}}_{\max}(sS^*)+[sc_\alpha(p,\ s,\ n-r)]^{\frac{1}{2}}c^{\frac{1}{2}}_{\max}(S). \qquad \ldots \quad (14.6.6)$$

where S^* and S are the dispersion matrices due to the "hypothesis" and due to the "error" and are given respectively by (12.7.10) and (12.7.11). (14.6.6) is thus a confidence statement with a confidence coefficient $\geqslant 1-\alpha$.

Also harking back to the remarks made after (12.7.11.5) we notice that if $C\xi = 0$ were replaced by $C\,\xi\,M = C\xi^* = 0$ and $[C_{11} \quad C_{12}]\,\xi = \eta$ by $[C_{11} \quad C_{12}]\,\xi \times M = [C_{11}\ C_{12}]\xi^* = \eta^*$, then (14.6.3) would be replaced by a statement in which every thing else would stay the same except that under c_α, p would be replaced by u, X would be replaced by $M'(u \times p)X(p \times n)$, S would be replaced by $M'(u \times p)\ S(p \times p)\ M(p \times u)$ and all non-null $\mathbf{a}'(1 \times p)$ would be replaced by all non-null $\mathbf{a}^{*\prime}(1 \times u)$. Similarly, in (14.6.6), in addition, S^* would be replaced by $M'(u \times p)S^*(p \times p)M(p \times u)$ and $\eta(s \times p)$ would be replaced by $\eta^*(s \times u) = \eta(s \times p)M(p \times u)$, and similarly for η'.

With a confidence coefficient $\geqslant 1-\alpha$, (14.6.6) will now be replaced by the confidence statement

$$c^{\frac{1}{2}}_{\max}(M/S^*M)-[\alpha c_\alpha(u,s,n-r)]^{\frac{1}{2}}c^{\frac{1}{2}}_{\max}(M/SM) \leqslant c^{\frac{1}{2}}_{\max}[\eta^*/(c_1(A_1'A_1)^{-1}c_1')^{-1}\eta^*]$$
$$\leqslant c^{\frac{1}{2}}_{\max}(M/S^*M)+[\alpha c_\alpha(u,s,n-r)]^{\frac{1}{2}}c^{\frac{1}{2}}_{\max}(M/SM). \qquad \ldots \quad (14.6.7)$$

This follows from a modified form of (14.6.3) obtained by replacing $\mathbf{a}^*(p \times 1)$ of (14.6.3) by $\mathbf{a}^*(u \times 1)$ and introducing other modifications just mentioned. If now we cut out the i-th element of $\mathbf{a}^*$ and the corresponding row of M' and $\eta^{*\prime}$ and reason in the same manner we should have (for $i = 1,2,\ldots,u$) u truncated confidence statements like (14.6.7). Likewise cutting out any j-th element of $\mathbf{b}$ and the corresponding row of c_{11} and colum of $\eta^{*\prime}$ and reasoning in the same manner as before we should have (for $j = 1,2,\ldots,s$) s truncated confidence statements like (14.6.7). Next, cutting out i, i' $(i \neq i' = 1,2,\ldots,u)$ we have $\binom{u}{2}$ statements like (14.6.7), and cutting out $j, j'(j \neq j' = 1,2,\ldots,s)$ we have $\binom{s}{2}$ statements like (14.6.7), and so on. Thus we have altogether 2^u-1 statements (based on truncation on u) and 2^s-1 statements based on truncotion on s) and, by combination, $(2^u-1)\times(2^s-1)$ statements of which the leading one is (14.6.7), all with a joiut confidence coefficient $\geqslant 1-\alpha$, which, in a sense, provides a complete anaiysis of the problem.

14.7. *Confidence bounds on departures from a particular kind of multicollinearity of means.* For k $(p+q)$-variate $N(\xi_i, \Sigma)$ (with $k > p+q$), where $\Sigma((p+q)\times(p+q))$ is symmetric p.d. with submatrices $\Sigma_{11}(p \times p)$, $\Sigma_{22}(q \times q)$ and $\Sigma_{12}(p \times q)$, and $\xi_i((p+q)\times 1)$

has column subvectors $\xi_{1i}(p\times 1)$ and ξ_{2i} $(q\times 1)$, let us consider the hypothesis H_0: $\xi_{1i}-\Sigma_{12}\,\Sigma_{22}^{-1}\xi_{2i} = 0$ $(i = 1, 2, \ldots, k)$. Notice (from (14.11.12)) that $\Sigma_{12}\,\Sigma_{22}^{-1}$ can be appropriately regarded as the matrix of regression of the set of p variates on the set of q variates. The hypothesis H_0 can thus be stated otherwise as the hypothesis that the matrix of means of the first p variates, viz, $(\xi_{11}\;\xi_{12}\;\ldots\;\xi_{1k})$ is equal to the matrix of means of the remaining q variates, premultiplied by the regression matrix of the p variates on the q variates. We are now interested in setting confidence bounds on

$$\xi_{1i}-\Sigma_{12}\Sigma_{22}^{-1}\,\xi_{2i} \qquad (i = 1, 2, \ldots, k) \qquad \ldots \quad (14.7.1)$$

which, naturally, are departures from H_0. More properly speaking, we shall be interested in setting simultaneous confidence bounds on arbitrary bilinear compounds $\mathbf{a}'(1\times p)\beta(p\times k)\mathbf{b}(k\times 1)$, where β is a $(p\times k)$ matrix with column vectors given by (14.7.1).

Now taking the 'residuals' of the first p variates with respect to the remaining q variates after the manner of (A.3.17) it is easy to check that for the i-th population the residuals will be distributed as a p-variate normal with a covariance matrix $\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$ and about the mean vector $\xi_{1i}-\Sigma_{12}\Sigma_{22}^{-1}\,\xi_{2i}$ (with $i = 1, 2, \ldots, p$). Also the 'within' covariance matrix of the 'residuals', pooled from k samples of size, say n each, will be given by

$$S_{1\cdot 2} = S_{11}-S_{12}S_{22}^{-1}S_{12}', \qquad \ldots \quad (14.7.2)$$

where $S_{11}(p\times p)$, $S_{22}(q\times q)$ and $S_{12}(p\times q)$ stand for the submatrices of the 'within' covariance matrix (of the $p+q$ variates) pooled from the k samples. The mean vector for the i-th sample will be given by

$$\mathbf{x}_{1i}-S_{12}\,S_{22}^{-1}\,\mathbf{x}_{2i}, \qquad \text{with } i = 1, 2, \ldots, k. \qquad \ldots \quad (14.7.3)$$

Let $B(p\times k)$ stand for the $(p\times k)$ matrix with the k column vectors given by (14.7.3).

Thus exactly as in section (14.6) we have with a confidence coefficient, say $1-\alpha$, the following set of simultaneous confidence bounds (for all arbitrary non-null $\mathbf{a}'(1\times p)$ and unit length $\mathbf{b}(k\times 1)$):

$$\mathbf{a}'B\mathbf{b}-[k(\mathbf{a}'S_{1\cdot 2}\mathbf{a})c_\alpha\,(p, k, nk-k)]^{\frac{1}{2}} \leqslant \mathbf{a}'\beta\mathbf{b} \leqslant \mathbf{a}'B\mathbf{b}+[k(\mathbf{a}'S_{1\cdot 2}\mathbf{a})c_\alpha(p, k, nk-k)]^{\frac{1}{2}}, \qquad \ldots \quad (14.7.4)$$

where $c\;(p, k, nk-k)$ is the upper α-point of the distribution of the (central) largest characteristic root based on p, k and $n-k$ degrees of freedom. The test for the associated hypothesis H_0 is also easily obtained, the critical region being given by

$$c_p \geqslant c_\alpha\,(p, k, nk-k), \qquad \ldots \quad (14.7.5)$$

where c_p is the 'largest root of $\frac{1}{k}\,(BB')S_{1.2}^{-1}$. Notice that β and B are each a $(p\times k)$ matrix with k column vectors given respectively by (14.7.1) and (14.7.3).

14.8. *Confidence bounds on departures from another kind of multicollinearity of means.* It seems that when the population covariance matrix Σ is not supposed to be known there are two kinds of multicollinearity of means (and departures from it) which can be properly handled, namely, (i) that the matrix of means of the first p variates is a *constant matrix times*, the matrix of means of the remaining q variates, the constant matrix factor being equal to the regression matrix of the p-set on the q-set, whatever this regression matrix might be and (ii) that the matrix of means of the first p variates is a *constant* (*and given*) *matrix times* the matrix of means of the remaining q variates. Case (i) is the one discussed in section 14.7 while case (ii) belongs to linear hypothesis in multivariate analysis of variance of means and has already been discussed in section 14.6.

14.9. *Confidence bounds connected with the dispersion matrix of a multivariate normal distribution.* Let us start from a $Y(p \times n): N^*(\mathbf{0}, \Sigma)$ where $\Sigma(p \times p)$ is supposed to be p.d. (so that its characteristic roots are all positive). For simplicity we also assume that $p \leqslant n$, so that, a.e., YY', that is, nS is p.d., and hence all its characteristic roots are positive. We now recall the well known result (A.3.3.) that there exists an orthogonal $\Gamma(p \times p)$ such that $\Sigma(p \times p) = \Gamma(p \times p) D_\gamma\ (p \times p)\ \Gamma'(p \times p)$ where the γ's are the characteristic roots of Σ. If the roots are distinct then by a convention, say by taking all the elements of the first row of Γ to be positive, the transformation could be made one-to-one. However, we do not need this for our present purpose. Note that the number of independent elements on both sides is the same. Except for the factor $(-\frac{1}{2})$ the argument under the exponential in the probability density of Y can now be written, if we put $\Delta = \gamma^{-\frac{1}{2}}$, as

$$\operatorname{tr}(\Gamma D_\gamma\ \Gamma')^{-1} YY' = \operatorname{tr} \Gamma D_\Delta\, D_\Delta\ \Gamma' YY' = \operatorname{tr}(D_\Delta\ \Gamma' Y)(D_\Delta \Gamma' Y)'$$

If we put $Z = D_\Delta \Gamma' Y$, it is easy to check that the probability density of Z is

$$[2\pi]^{\frac{pn}{2}} \exp\left[-\tfrac{1}{2} \operatorname{tr}\ ZZ'\right]. \qquad \dots \quad (14.9.1)$$

For all non-null nonstochastic $\mathbf{a}(p \times 1)$ consider now the simultaneous statement that

$$g_1^2 \leqslant \mathbf{a}'ZZ'\mathbf{a}/\mathbf{a}'\mathbf{a} \leqslant g_2^2 \text{ or } g_1^2 \leqslant \mathbf{a}'(D_\Delta\ \Gamma' YY' \Gamma D_\Delta)\mathbf{a}/\mathbf{a}'\mathbf{a} \leqslant g_2^2. \qquad \dots \quad (14.9.2)$$

This statement, for a given Z and g_1^2 and g_2^2, is precisely equivalent to the statement that

$$g_1^2 \leqslant \inf_{\mathbf{a}} \frac{\mathbf{a}'ZZ'\mathbf{a}}{\mathbf{a}'\mathbf{a}} \leqslant \sup_{\mathbf{a}} \frac{\mathbf{a}'ZZ'\mathbf{a}}{\mathbf{a}'\mathbf{a}} \leqslant g_2^2,$$

or that

$$g_1^2 \leqslant c_1 \leqslant c_p \leqslant g_2^2, \qquad \dots \quad (14.9.3)$$

where c_1 and c_p are the smallest and largest characteristic roots of the matrix ZZ', both, a.e., positive. The relevant distribution on the null hypothesis, i.e., when the true population matrix is Σ, is known and we now put

$$g_1^2 = c_{1\alpha}(p, n) \text{ and } g_2^2 = c_{2\alpha}(p, n), \qquad \dots \quad (14.9.4)$$

where $c_{1\alpha}(p,n)$ and $c_{2\alpha}(p,n)$ are constants taken over from (6.4.3). If we now tie up (14.9.2), (14.9.3) and (14.9.4) we have, with a confidence coefficient $1-\alpha$, the set of multiple or simultaneous confidence interval statements for all non-null **a** and all permissible values of the unknown parameters Γ and Δ :

$$\mathbf{a}'\mathbf{a}c_{1\alpha}(p,n) \leqslant \mathbf{a}'(D_\Delta\ \Gamma' YY'\Gamma D_\Delta)\mathbf{a} \leqslant \mathbf{a}'\mathbf{a}c_{2\alpha}(p,n), \qquad \ldots \quad (14.9.5)$$

or, remembering that $nS = YY'$,

$$\mathbf{a}'\mathbf{a}c_{1\alpha}(p,n) \leqslant \mathbf{a}'(D_\Delta\ \Gamma' nS\Gamma D_\Delta)\mathbf{a} \leqslant \mathbf{a}'\mathbf{a}c_{2\alpha}(p,n).$$

The shortness of the confidence bounds (14.9.5) is tied up with the power of the test (6.4.3), which has been already discussed in Chapters 9–11.

Far more meaningful confidence bounds than (14.9.5) can be obtained in the following way, starting from (14.9.5). As before denoting the characteristic roots of a (square) matrix M by $c(M)$ and the largest and the smallest roots of M(if M is at least p.s.d.) by $c_{\max}(M)$ and $c_{\min}(M)$, and remembering that $\Delta = \gamma^{-\frac{1}{2}}$ and finally using (A.2.5) we can rewrite (14.9.5) as

$$\frac{1}{n}c_{1\alpha}(p,n) \leqslant \text{all } c(D_{1/\sqrt{\gamma}}\Gamma' S\Gamma D_{1/\sqrt{\gamma}}) \leqslant \frac{1}{n}c_{2\alpha}(p,n). \qquad \ldots \quad (14.9.6)$$

Now using (A.1.18) we note that

$$c(D_{1/\sqrt{\gamma}}\Gamma' S\Gamma D_{1/\sqrt{\gamma}}) = c(S\Gamma D_{1/\sqrt{\gamma}}\Gamma') = c(S\Sigma^{-1}), \qquad \ldots \quad (14.9.7)$$

and obtain with a confidence coefficient $1-\alpha$, the confidence bounds

$$\frac{1}{n}c_{1\alpha}(p,n) \leqslant \text{all } c(S\Sigma^{-1}) \leqslant \frac{1}{n}c_{2\alpha}(p,n), \text{ or } nc_1^{-1}(p,n) \geqslant \text{all } c(\Sigma S^{-1}) \geqslant nc_{2\alpha}^{-1}(p,n). \ (14.9.8)$$

We now recall (A.1.25) and deduce from it that

$$c_{\min}(B^{-1})c_{\min}(AB) \leqslant \text{all } c(A) \leqslant c_{\max}(B^{-1})c_{\max}(AB). \qquad \ldots \quad (14.9.9)$$

By using (14.9.9) it is easy to see that the statement (14.9.8)$\Longrightarrow$the following

$$nc_{1\alpha}^{-1}(p,n)c_{\max}(S) \geqslant \text{all } c(\Sigma) \geqslant nc_{2\alpha}^{-1}(p,n)c_{\min}(S). \qquad \ldots \quad (14.9.10)$$

We now use the following result of set-theoretic logic, namely that

$$\text{“If } E_1, \text{ then } E_2\text{”} \Longleftrightarrow \text{“}E_2 \text{ is a necessary condition for } E_1\text{”} \Longleftrightarrow E_1 \subset E_2$$

$$\Longrightarrow P(E_1) \leqslant P(E_2), \qquad \ldots \quad (14.9.11)$$

to observe that if the probability of (14.9.8) is $1-\alpha$, the probability of (14.9.10) is $\geqslant 1-\alpha$. Thus (14.9.10) is set of simultaneous confidence bounds with probability $\geqslant 1-\alpha$.

Also using (A.1.21) we observe that (14.9.8)$\Longrightarrow$the following

$$[nc_{1\alpha}^{-1}(p,n)]^t \operatorname{tr}_t(S) \geqslant \operatorname{tr}_t(\Sigma) \geqslant [nc_{2\alpha}^{-1}(p,n)]^t \operatorname{tr}_t(S), \ (t = 1, 2, \ldots, p) \qquad \ldots \quad (14.9.12,$$

which, by using (14.9.11), is thus another set of simultaneous confidence bounds on $c(\Sigma)$'s.

The derivation of the confidence bounds (14.9.10) can be simplified if we start from the canonical p.d.f.

$$\text{Const}\ \exp\left[-\tfrac{1}{2}\ \text{tr}\ D_{1/\gamma}\ YY'\right], \qquad \ldots \quad (14.9.13)$$

and recall that (i) $S = YY'/n$, (ii) $c(YY') = c(AYY'A')$ for any $\perp A$ and (iii) tr $D_{1/\gamma}\ YY' =$ tr $D_{1/\sqrt{\gamma}}\ YY'D_{1/\sqrt{\gamma}}$, so that $c(D_{1/\sqrt{\gamma}}\ YY'D_{1/\sqrt{\gamma}})$'s or $c(D_{1/\gamma}\ nS)$'s are distributed as $c(S)$'s when γ's $= 1$, which distribution has been already used in the above derivation. However, the lengthier derivation is instructive in certain ways.

Going back to (14.9.8) and using Chapter A.2, we note that the formula (14.9.8) $\Longleftrightarrow$

$$nc_{1\alpha}^{-1}(p, n) \geqslant \frac{\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'S\mathbf{a}} \geqslant nc_{2\alpha}^{-1}(p, n),$$

or

$$nc_{1\alpha}^{-1}(p, n)\mathbf{a}'S\mathbf{a} \geqslant \mathbf{a}'\Sigma\mathbf{a} \geqslant nc_{2\alpha}^{-1}(p, n)\mathbf{a}'S\mathbf{a}, \qquad \ldots \quad (14.9.14)$$

which is, therefore, a set of simultaneous confidence bounds on $\mathbf{a}'\Sigma\mathbf{a}$ for all arbitrary non-null $\mathbf{a}'(1\times p)$, and with a confidence coefficient $1-\alpha$.

If we start from (14.9.14) and choose $\mathbf{a}$ so as to maximize $\mathbf{a}'\Sigma\mathbf{a}$, then it is easy to check that (14.9.14) will imply that $nc_{1\alpha}^{-1}(p, n)\, c_{\max}(S) \geqslant c_{\max}(\Sigma)$; also if we choose $\mathbf{a}$ so as to maximize $\mathbf{a}'S\mathbf{a}$, then (14.9.14) will be seen to imply that $c_{\max}(\Sigma) \geqslant nc_{2\alpha}^{-1}(p, n)c_{\max}(S)$. Similarly for the $c_{\min}$'s. Thus (14.9.14) will imply

$$nc_{1\alpha}^{-1}(p, n)\ c_{\max}(S) \geqslant c_{\max}(\Sigma) \geqslant nc_{2\alpha}^{-1}(p, n)\ c_{\max}(S) \qquad \ldots \quad (14.9.15)$$

and

$$nc_{1\alpha}^{-1}(p, n)\ c_{\min}(S) \geqslant c_{\min}(\Sigma) \geqslant nc_{2\alpha}^{-1}(p, n)\ c_{\min}(S),$$

which, therefore, is a pair of confidence statements with a joint confidence coefficient $\geqslant 1-\alpha$. Incidentally, we notice that (14.9.15) implies (14.9.10) and thus provides another derivation of (14.9.10).

Furthermore, there is a lot more to (14.9.14) than just (14.9.15). Since (14.9.14) is supposed to be true for all non-null $\mathbf{a}(p\times 1)$, we can specialize by putting one, two or more components of $\mathbf{a}(p\times 1)$ equal to zero and then we can take arbitrary values for the other components. Now let us use the same argument as before, and denote by $S^{(i)}$, $\Sigma^{(i)}$ the truncated $(p-1)\times(p-1)$ sample and population dispersion matrices formed by cutting out the i-th variate, by $S^{(i,j)}$ and $\Sigma^{(i,j)}$ the truncated $(p-2)\times(p-2)$ sample and population dispersion matrices formed by cutting out the i-th and j-th variates ($i \neq j = 1, 2, \ldots, p$) and so on. Then it is easy to check that (14.9.14) really implies (14.9.15) together with statements

$$nc_{1\alpha}^{-1}(p, n)\ c_{\max}(S^{(i)}) \geqslant c_{\max}(\Sigma^{(i)}) \geqslant nc_{2\alpha}^{-1}(p, n)\ c_{\max}(S^{(i)}) \qquad \ldots \quad (14.9.16)$$

and

$$nc_{1\alpha}^{-1}(p, n)\, c_{\min}(S^{(i)}) \geqslant c_{\min}(\Sigma^{(i)}) \geqslant nc_{2\alpha}^{-1}(p, n)\, c_{\min}(S^{(i)});$$

for $i = 1, 2, \ldots, p$;

$$nc_{1\alpha}^{-1}(p, n)\, c_{\max}(S^{(i,j)}) \geqslant c_{\max}(\Sigma^{(i,j)}) \geqslant nc_{2\alpha}^{-1}\ (p, n)\, c_{\max}(S^{(i,j)})$$

and
$$nc_{1\alpha}^{-1}(p, n)\, c_{\min}(S^{(i,j)}) \geqslant c_{\min}(\Sigma^{(i,j)}) \geqslant nc_{2\alpha}^{-1}(p, n)\, c_{\min}(S^{(i,j)});$$

for $i \neq j = 1, 2, \ldots, p$;

and so on down to the stage of cutting out any $p-1$, i.e., retaining any one variate. All these statements, 2 -1 in number have a joint confidence coefficient $\geqslant 1-\alpha$ and provide one type of complete attack on what the psychologists call the problem of latent structure.

14.10. *Confidence bounds on the characteristic roots of* $\Sigma_1 \Sigma_2^{-1}$. Let us start from $Y_h(p \times n_h)$: $N^*(0, \Sigma_h)$ $(h = 1, 2)$, where we assume that $p \leqslant n_1, n_2$, and Σ_1 and Σ_2 are both p.d. so that the characteristic roots of $\Sigma_1 \Sigma_2^{-1}$ are all positive and those of $Y_1 Y'_1 (Y_2 Y_2')^{-1}$, i.e., of $(n_1/n_2) S_1 S_2^{-1}$ are, a.e., all positive. We recall the results of chapters A.4 and A.7 and start, without any loss of generality, from the canonical probability density (in terms of transformed variates Y_1^*, Y_2^*).

$$\text{Const} \quad \exp\ [-\tfrac{1}{2} \operatorname{tr}\ (D_{1/\gamma} Y_1^* Y_1^{*\prime} + Y_2^* Y_2^{*\prime})]$$

$$= \text{Const} \exp [-\tfrac{1}{2} \operatorname{tr} (D_{1/\sqrt{\gamma}} Y_1^* Y_1^{*\prime} D_{1/\sqrt{\gamma}} + Y_2^* Y_2^{*\prime})]. \qquad \ldots \quad (14.10.1)$$

It is to be noticed that $c[(n_1/n_2)(S_1 S_2^{-1}]$'s i.e., $c[(Y_1 Y_1')(Y_2 Y_2')^{-1}]$'s are the same as $c[(Y_1^* Y_1^{*\prime})(Y_2^* Y_2^{*\prime})^{-1}]$'s and that γ's are $c(\Sigma_1 \Sigma_2^{-1})$'s.

If we now put $Z_1 = D_{1/\sqrt{\gamma}} Y_1^*$ and $Z_2 = Y_2^*$, it is easy to check that the probability density of Z_1 and Z_2 is $(2\pi)^{-p(n_1+n_2)/2} \exp\ [-\frac{1}{2} \operatorname{tr}\ (Z_1 Z_1' + Z_2 Z_2')]$.

For all non-null nonstochastic $\mathbf{a}(p \times 1)$ consider the set of statements

$$g_1^2 \leqslant \mathbf{a}' Z_1 Z_1' \mathbf{a} / \mathbf{a}' Z_2 Z_2' \mathbf{a} \leqslant g_2^2,$$

or
$$g_1^2 \leqslant \mathbf{a}'(D_{1/\sqrt{\gamma}} Y_1^*)(D_{1/\sqrt{\gamma}} Y_1^*)' \mathbf{a} / \mathbf{a}' Y_2^* Y_2^{*\prime} \mathbf{a} \leqslant g_2^2,$$

or
$$\frac{n_2}{n_1} g_1^2 \leqslant \mathbf{a}'(D_{1/\sqrt{\gamma}}\ S_1 D_{1/\sqrt{\gamma}}) \mathbf{a} / \mathbf{a}' S_2 \mathbf{a} \leqslant \frac{n_2}{n_1}\ g_2^2. \qquad \ldots \quad (14.10.2)$$

For given Z_1, Z_2, g_1^2 and g_2^2, this set of statements is precisely equivalent to the statement that

$$g_1^2 \leqslant \inf_{\mathbf{a}} \frac{\mathbf{a}' Z_1 Z_1' \mathbf{a}}{\mathbf{a}' Z_2 Z_2' \mathbf{a}} \leqslant \sup_{\mathbf{a}} \frac{\mathbf{a}' Z_1 Z_1' \mathbf{a}}{\mathbf{a}' Z_2 Z_2' \mathbf{a}} \leqslant g_2^2 \text{ or } g_1^2 \leqslant c_1 \leqslant c_p \leqslant g_2^2, \ \ldots \quad (14.10.3)$$

where c_1 and c_p are the smallest and the largest characteristic roots of $(Z_1Z_1')(Z_2Z_2')^{-1}$, both, a.e., positive. The relevant distribution on the null hypothesis, i.e. when $\Sigma_1 = \Sigma_2$ is known from chapter 8 and we now put

$$g_1^2 = c_{1\alpha}(p, n_1, n_2)$$

and
$$g_2^2 = c_{2\alpha}(p, n_1, n_2), \qquad \dots \quad (14.10.4)$$

where $c_{1\alpha}(p, n_1, n_2)$ and $c_{2\alpha}(p, n_1, n_2)$ are constants taken over from (6.4.6).

Changing back to S_1, S_2 and γ's and putting $c_{1\alpha}$ and $c_{2\alpha}$ for $c_{1\alpha}(p, n_1, n_2)$ and $c_{2\alpha}(p, n_1, n_2)$ for shortness, we can now rewrite the second form in (14.10.3) as

$$\frac{n_1}{n_2}c_{1\alpha}^{-1} \geqslant \text{all } c(S_2D_{\sqrt{\gamma}}S_1^{-1}D_{\sqrt{\gamma}}) \geqslant \frac{n_1}{n_2}c_{2\alpha}^{-1} \qquad \dots \quad (14.10.5)$$

Now noting from (A.1.12) that $D'_{\sqrt{\gamma}}S_1^{-1}D_{\sqrt{\gamma}}$ and of course S_1^{-1} and S_2 are symmetric p.d. and using (A.1.25) we have

$$c_{\max}(S_1S_2^{-1})\, c_{\max}(S_2D_{\sqrt{\gamma}}S_1^{-1}D_{\sqrt{\gamma}}) \geqslant \dots$$

$$\text{all} \quad c(S_1D_{\sqrt{\gamma}}S_1^{-1}D_{\sqrt{\gamma}}) \geqslant c_{\min}(S_1S_2^{-1})c_{\min}(S_2D_{\sqrt{\gamma}}S_1^{-1}D_{\sqrt{\gamma}}). \qquad \dots \quad (14.10.6)$$

Also using (A.3.9) and putting $S_1 = \tilde{T}\tilde{T}'$ and then using (A.1.24) we should have

$$c_{\max}(S_1D_{\sqrt{\gamma}}S_1^{-1}D_{\sqrt{\gamma}}) = c_{\max}(\tilde{T}\tilde{T}'\, D_{\sqrt{\gamma}}\tilde{T}'^{-1}\tilde{T}^{-1}D_{\sqrt{\gamma}}) \qquad \dots \quad (14.10.7)$$

$= c_{\max}(\tilde{T}'D_{\sqrt{\gamma}}\tilde{T}'^{-1}\tilde{T}^{-1}D_{\sqrt{\gamma}}\tilde{T}) \geqslant$ all $c^2(\tilde{T}'D_{\sqrt{\gamma}}\tilde{T}'^{-1})$, that is $\geqslant$ all $c^2(D_{\sqrt{\gamma}})$, that is, $\geqslant$ all γ_i's.

Likewise we have

$$c_{\min}(S_1D_{\sqrt{\gamma}}S_1^{-1}D_{\sqrt{\gamma}}) \leqslant \text{all } \gamma_i\text{'s}. \qquad \dots \quad (14.10.8)$$

Now combining (14.10.6)-(14.10.8), we observe that (14.10.5) $\Longrightarrow$

$$\frac{n_1}{n_2}c_{1\alpha}^{-1}(p, n_1, n_2)\, c_{\max}(S_1S_2^{-1}) \geqslant \text{all } \gamma_i\text{'s} = \text{all } c(\Sigma_1\Sigma_2^{-1})$$

$$\geqslant \frac{n_1}{n_2}c_{2\alpha}^{-1}(p, n_1, n_2)\, c_{\min}(S_1S_2^{-1}) \qquad \dots \quad (14.10.9)$$

which, therefore, by using (14.9.11), is easily seen to be a set of simultaneous confidence bounds with a joint confidence coefficient, say $1-\beta \geqslant 1-\alpha$.

Now using (A.1.22) we have

$c_{\max}(S_1S_2^{-1}) \leqslant c_{\max}(S_1)\, c_{\max}(S_2^{-1})$, that is, $\leqslant c_{\max}(S_1)/c_{\min}(S_2)$ and

$$c_{\min}(S_1S_2^{-1}) \geqslant c_{\min}(S_1)/c_{\max}(S_2). \qquad \dots \quad (14.10.10)$$

Thus (14.10.9) $\Longrightarrow$

$$\frac{n_1}{n_2} c_{1\alpha}^{-1}(p, n_1, n_2) c_{\max}(S_1)/c_{\min}(S_2) \geqslant \text{all } c(\Sigma_1\Sigma_2^{-1})$$

$$\geqslant \frac{n_1}{n_2} c_{2\alpha}^{-1}(p, n_1, n_2)\; c_{\min}(S_1)/c_{\max}(S_2), \qquad \ldots \quad (14.10.11)$$

which, therefore, by using (14.9.11), is a set of simultaneous confidence bounds with a joint confidence coefficient $1-\delta \geqslant 1-\beta \geqslant 1-\alpha$.

Going back to (14.10.9) and using chapter A.2, we note that (14.10.9) $\Longleftrightarrow$

$$\frac{n_1}{n_2} c_{1\alpha}^{-1}(p, n_1, n_2)\, c_{\max}(S_1S_2^{-1}) \geqslant \frac{\mathbf{a}'\Sigma_1\mathbf{a}}{\mathbf{a}'\Sigma_2\mathbf{a}} \geqslant \frac{n_1}{n_2} c_{2\alpha}^{-1}(p, n_1, n_2)\; c_{\min}(S_1S_2^{-1})$$

$$\ldots \quad (14.10.12)$$

which is, therefore, a set of simultaneous confidence bounds on $\mathbf{a}'\Sigma_1\mathbf{a}/\mathbf{a}'\Sigma_2\mathbf{a}$ for all arbitrary non-null $\mathbf{a}'(1\times p)$ and with a confidence coefficient $\geqslant 1-\alpha$. Notice the essential difference between (14.10.9) and (14.10.12).

Let us now go back to (14.10.2), recall that that statement is supposed to be true for all non-null $\mathbf{a}(p\times 1)$ and specialize by putting one, two or more components of $\mathbf{a}$ equal to zero and then use the same kind of argument as from (14.10.2) to (14.10.9). Also use the same notation as in (14.9.16) for the truncated $(p-1)\times(p-1)$, $(p-2)\times(p-2), \ldots,$ sample and population dispersion matrices obtained by cutting out any i-th variate (with $i = 1, 2, \ldots, p$) any i-th and j-th variates (with $i \neq j = 1, 2, \ldots, p$) and so on. Then (14.10.2) will not only imply (14.10.9) but also statements,

$$\lambda_1 c_{\max}(S_1^{(i)} S_2^{(i)-1}) \geqslant \text{all } c(\Sigma_1^{(i)} \Sigma_2^{(i)-1}) \geqslant \lambda_2 c_{\min}(S_1^{(i)} S_2^{(i)-1}), \qquad \ldots \quad (14.10.13)$$

for $i = 1, 2, \ldots, p$;

$$\lambda_1 c_{\max}(S_1^{(i,j)} S_2^{(i,j)-1} \geqslant \text{all } c(\Sigma_1^{(i,j)}\Sigma_2^{(i,j)-1}) \geqslant \lambda_2 c_{\min}(S_1^{(i,j)} S_2^{(i,j)-1})$$

for $i \neq j = 1, 2, \ldots, p$;

and so on, with a joint confidence coefficient $\geqslant 1-\alpha$. The total number of such statements will be 2^p-1. Here $\lambda_1 = \frac{n_1}{} c_{1\alpha}^{-1}(p, n_1, n_2)$ and $\lambda_2 = \frac{n_1}{n_2} c_{2\alpha}^{-1}(p, n_1, n_2)$.

14.11. *Confidence bounds on regression like parameters.*

(*i*) *Some preliminary observations.* We now start with a random sample of size n ($> p+q$; $p \leqslant q$) from a $(p+q)$—variate normal population, and next reduce for the means and set

$$(n-1)\begin{array}{cc} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \begin{array}{c} [Y_1' \vdots Y_2'], \\ \begin{array}{cc} p & q \end{array} \end{array}$$

where S_{11}, S_{22} and S_{12} stand respectively for the sample dispersion sub-matrices of the p-set, the q-set and that between the p-set and the q-set and where Y_1 and Y_2 have the p.d.f.

$$\text{Const exp}\left[-\tfrac{1}{2}\text{tr}\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, [Y_1' \vdots Y_2']\right]. \qquad \ldots \quad (14.11.1)$$

For reasons which will become clear later we shall take $\beta = \Sigma_{12}\Sigma_{22}^{-1}$ and $B = S_{12}S_{22}^{-1}$ respectively as the population and the sample regression matrix of the p-set on the q-set. Suppose we now consider two new variate sets $\mathbf{x}_1(p\times 1)-\beta(p\times q)\,\mathbf{x}_2(q\times 1)$ and $\mathbf{x}_2(q\times 1)$. Then the population covariance matrix $(\mathbf{x}_1-\beta\mathbf{x}_2, \mathbf{x}_2) = 0$, so that $c[(S_{11}-S_{12}\beta'-\beta S'_{12}+\beta S_{22}\beta')^{-1}(S_{12}-\beta S_{22})S_{22}^{-1}(S'_{12}-S_{22}\beta')]$'s are distributed as $c[S_{11}^{-1}S_{12}S_{22}^{-1}S'_{12}]$'s when $\beta = 0$, i.e. when $\Sigma_{12} = 0$.

(ii) *Confidence bounds on the regression matrix β.* Consider the statement

$$\text{all } c_i\text{'s} \leqslant g^2 \text{ or } c_p \leqslant g^2, \qquad \dots \quad (14.11.2)$$

where c_i's ($i = 1, 2, \dots, p$; $0 \leqslant c_1 \leqslant \dots \leqslant c_p \leqslant 1$) are the roots of the determinantal equation in c:

$$|c(S_{11}-S_{12}\beta'-\beta S'_{12}+\beta S_{22}\beta')-(S_{12}-\beta S_{22})\,S_{22}^{-1}(S'_{12}-S_{22}\beta')| = 0. \quad \dots \quad (14.11.3)$$

Now put $e = c/(1-c)$, so that we have from (14.11.3), the determinantal equation in e,

$$|e(S_{11}-S_{12}\,S_{22}^{-1}S'_{12})-(S_{12}\,S_{22}^{-1}-\beta)S_{22}(S_{22}^{-1}S'_{12}-\beta')| = 0. \qquad \dots \quad (14.11.4)$$

Notice that the statement (14.11.2) can now be replaced by the statement that the largest characteristic root $e_p \leqslant g^2/(1-g^2)$, i.e.,

$$\text{all} \quad c[(S_{11}-S_{12}S_{22}^{-1}S'_{12})^{-1}(B-\beta)S_{22}(B'-\beta')] \leqslant g^2/(1-g^2), \qquad \dots \quad (14.11.5)$$

where

$$B(p\times q) = S_{12}S_{22}^{-1}, \qquad \dots \quad (14.11.6)$$

which may be appropriately called the matrix of sample regression of the p-set on the q-set.

We note that (14.11.5) $\Longleftrightarrow$ (14.11.2), and that c_p has the distribution of the largest characteristic root of the matrix $S_{11}^{-1}S_{12}S_{22}^{-1}S'_{12}$, when $\Sigma_{12} = 0$. The joint distribution of these central roots and also of the largest root being known, all that we have to do to make (14.11.5), i.e., (14.11.2), a simultaneous confidence statement with a joint coefficient $1-\alpha$ is to choose $g^2 = c_\alpha(p, q, n-1)$ where the quantity on the right side is defined by

$$P\,[\text{central } c_p \geqslant c_\alpha(p, q, n-1)] = \alpha. \qquad \dots \quad (14.11.7)$$

Substituting now $c_\alpha(p, q, n-1)$ (to be sometimes denoted more simply by c) for g^2 in (14.11.5), we have a simultaneous confidence statement with a joint confidence coefficient $1-\alpha$.

Now applying (A.1.18) and (A.1.22) (in the same manner as in the previous sections), we have from (14.11.5), now with a joint confidence coefficient $\geqslant 1-\alpha$, the following simultaneous confidence statement

$$\text{all} \quad c[(B-\beta)(B'-\beta')] \leqslant \frac{c_\alpha}{1-c_\alpha} \; c_{\max}(S_{11}-S_{12}S_{22}^{-1}S'_{12})c_{\max}(S_{22}^{-1}). \quad \dots \quad (14.11.8)$$

Using (A.1.22) again, we check that (14.11.8) can be replaced by the following wider bounds (with a confidence coefficient $\geqslant 1-\alpha$):

all $c[(B-\beta)(B'-\beta')] \leqslant (c_\alpha/1-c_\alpha)[1-c_{\min}(S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}')]c_{\max}(S_{11})c_{\max}(S_{22}^{-1})$. ... (14.11.9)

We next recall the following two well-known results (A.2.5) and (A.2.7) which we remember for convenience as

all $c(M) \leqslant g$ (for a $p\times p$ real matrix M with real roots) $\Longleftrightarrow$

$\mathbf{d}_1'(1\times p)$ $M(p\times p)$ $\mathbf{d}_1(p\times 1)$ (for all arbitrary unit vectors $\mathbf{d}_1$) $\leqslant g$ ... (14.11.10)

and $\mathbf{x}'(1\times q)\,\mathbf{x}(q\times 1) \leqslant h(>0) \Longleftrightarrow |\mathbf{x}'(1\times q)\,\mathbf{d}_2(q\times 1)| \leqslant \sqrt{h}$ ... (14.11.11)

(for all arbitrary unit vectors $\mathbf{d}_2$).

Applying (14.11.10) and (14.11.11) to (14.11.8) we have (with a joint confidence coefficient $\geqslant 1-\alpha$) the following simultaneous confidence statement (for all arbitrary unit vectors $\mathbf{d}_1(p\times 1)$ and $\mathbf{d}_2(q\times 1)$,

$$\mathbf{d}_1'(B-\beta)\mathbf{d}_2 \leqslant \text{[right side of (14.11.8)]}^{\frac{1}{2}} \qquad \text{... (14.11.12)}$$

or ultimately

$$\mathbf{d}_1'B\mathbf{d}_2-\sqrt{E} \leqslant \mathbf{d}_1'\beta\mathbf{d}_2 \leqslant \mathbf{d}_1'B\mathbf{d}_2+\sqrt{E}, \qquad \text{... (14.11.13)}$$

where

$$E = [c_{\alpha/1-\alpha}]c_{\max}(S_{11}-S_{12}S_{22}^{-1}S_{12}')c_{\max}(S_{22}^{-1}). \qquad \text{... (14.11.14)}$$

A set of simultaneous confidence bounds on just the elements β_{ij} of the β-matrix would be a subset of the bounds on the total set $\mathbf{d}_1'\beta\mathbf{d}_2$. It is worthwhile to check that if $p = q = 1$(14.11.13) reduces, as it should, to (13.4.4). Also if $p = 1$, we should have another special case of (14.11.13) giving a set of simultaneous confidence bounds on all linear functions of the partial regressions of one variate on several others. Thus, in several ways, (14.11.13). seems to be an appropriate generalization of (13.4.4).

As in the derivation of (14.6.6) from (14.6.3) it is easy to check that (14.11.13) will imply

$$c_{\max}^{\frac{1}{2}}(BB')-\sqrt{E} \leqslant c_{\max}^{\frac{1}{2}}(\beta\beta') \leqslant c_{\max}^{\frac{1}{2}}(BB')+\sqrt{E}, \qquad \text{... (14.11.14.1)}$$

where E is defined by (14.11.14). (14.11.14.1) is thus a confidence statement with a confidence coefficient $\geqslant 1-\alpha$.

Furthermore, if we now go back to (14.11.5) and replace it by

$$\frac{\mathbf{a}'(B-\beta)S_{22}(B'-\beta')\mathbf{a}}{\mathbf{a}'(S_{11}-S_{12}S_{22}^{-1}S_{12}')\mathbf{a}} \leqslant \frac{g^2}{1-g^2}, \qquad \text{... (14.11.14.2)}$$

for all non-null $\mathbf{a}(p\times 1)$, then we observe that (14.11.14.2) implies (14.11.13). Now we can specialize $\mathbf{a}(p\times 1)$ by putting one, two or more components equal to zero and then, in each case, take arbitrary values of the other components and reason as from (14.11.5) to (14.11.13). Thus, if, as in the two previous cases, we denote by $S_{11}^{(i)}$, $S_{12}^{(i)}$, $B^{(i)}$, $\beta_i^{(i)}$, $S_{11}^{(i,j)}$, $S_{12}^{(i,j)}$, $B^{(i,j)}$, $\beta^{(i,j)}$, etc., the truncated matrices obtained by cutting out the i-th variate of the p-set (with $i = 1, 2, \ldots, p$) the i-th and j-th variates of the p-set (with $i \neq j = 1, 2, \ldots, p$) and so on, it is easy to check that (14.11.5) will altogether 2^p-1 in number imply not only (14.11.13) but similar statements on the truncated matrices as well, all with a simultaneous confidence coefficient $\geqslant 1-\alpha$. The same applies to a set of statements on the truncated matrices, similar to the statement (14.11.14.1). We also observe that we can extend this still further by doing a similar truncation with regard to the variates of the q-set, but this will not be discussed in the present monograph.

For an alternative set of confidence bounds we proceed as follows. Going back again to (14.11.14.2) we rewrite it as

$$\frac{\mathrm{a}'(B-\beta)S_{22}(B-\beta)'\mathrm{a}}{\mathrm{a}'(S_{11}-S_{12}S_{22}^{-1}S_{12}')\mathrm{a}} \leqslant \frac{c_\alpha}{1-c_\alpha}, \qquad \ldots \quad (14.11.15)$$

for all arbitrary non-null $\mathrm{a}(p\times 1)$.

Now using (A.3.9) and putting $S_{22} = \tilde{T}\tilde{T}'$ and using (14.11.10)., we have (14.11.15) reducing to

$$\mathrm{a}'(B-\beta)\tilde{T}\tilde{T}'(B-\beta)'\mathrm{a} \leqslant \frac{c_\alpha}{1-c_\alpha}\,\mathrm{a}'(S_{11}-S_{12}S_{22}^{-1}S_{12}')\mathrm{a} \qquad \ldots \quad (14.11.16)$$

or

$$\mathrm{a}'B\tilde{T}\mathrm{b}-\left\{\frac{c_\alpha}{1-c_\alpha}\right\}^{\frac{1}{2}}[\mathrm{a}'(S_{11}-S_{12}S_{22}^{-1}S_{12}')\mathrm{a}]^{\frac{1}{2}} \leqslant \mathrm{a}'\beta\tilde{T}\mathrm{b}$$

$$\leqslant \mathrm{a}'B\tilde{T}\mathrm{b}+\left\{\frac{c_\alpha}{1-c_\alpha}\right\}^{\frac{1}{2}}[\mathrm{a}'(S_{11}-S_{12}S_{22}^{-1}S_{12}')\mathrm{a}]^{\frac{1}{2}}, \qquad \ldots \quad (14.11.17)$$

for all arbitrary unit vectors $\mathrm{b}(q\times 1)$. Now put $\tilde{T}(q\times q)\ \mathrm{b}(q\times 1) = \mathrm{c}(q\times 1)$ say, so that (14.11.17) reduces to the following set of simultaneous bounds with a confidence coefficient $1-\alpha$:

$$\mathrm{a}'B\mathrm{c}-\left\{\frac{c_\alpha}{1-c_\alpha}\right\}^{\frac{1}{2}}\left[\mathrm{a}'(S_{11}-S_{12}S_{22}^{-1}S_{12}')\mathrm{a}\right]^{\frac{1}{2}} \leqslant \mathrm{a}'\beta\mathrm{c}$$

$$\leqslant \mathrm{a}'B\mathrm{c}+\left\{\frac{c_\alpha}{1-c_\alpha}\right\}^{\frac{1}{2}}\left[\mathrm{a}'(S_{11}-S_{12}S_{22}^{-1}S_{12}')\mathrm{a}\right]^{\frac{1}{2}}, \qquad \ldots \quad (14.11.18)$$

for all arbitrary non-null $\mathrm{a}(p\times 1)$ and all $\mathrm{c}(q\times 1)$subject to

$$1 = \mathrm{b}'\mathrm{b} = \mathrm{c}'\tilde{T}'^{-1}\tilde{T}^{-1}\mathrm{c} = \mathrm{c}'(\tilde{T}\tilde{T}')^{-1}\mathrm{c} = \mathrm{c}'S_{22}^{-1}\,\mathrm{c}. \qquad \ldots \quad (14.11.19)$$

These confidence bounds are no doubt closer than those of (14.11.13) but these seem to be useful from a physical standpoint only when we have, in the customary sense, a regression problem of a p-set on a q-set such that the p-set is stochastic while the q-set is fixed, so that S_{22} (and hence S_{22}^{-1}) are neither unknown parameters nor stochastic variates but just a set of given constants.

CHAPTER FIFTEEN

Some Non-Parametric Generalizations of Analysis of Variance and Multivariate Analysis*

15.1. *Preliminaries and notation.* In this chapter we shall be concerned with the statistical analysis of data in the form of observed frequencies in discrete (and finite) categories; a typical category being the (ij)-th cell of a lattice, with $i = 1, 2, \ldots, r_j$ and $j = 1, 2, \ldots, s$ and $\sum_{j=1}^{s} r_j = rs$(say). Let n_{ij} be the frequency in the (ij)-th cell, and p_{ij} be the probability of getting an observation in that cell and let us assume that the observations are independent (in probability). Also let $\sum_j n_{ij} = n_{i0}$, $\sum_i n_{ij} = n_{0j}$, $\sum_{i,j} n_{ij} = n_{00} = n$(say), $\sum_j p_{ij} = p_{i0}$, $\sum_j p_{ij} = \sum_i p_{ij} = p_{0j}$, $\sum_{i,j} p_{ij} = p_{00}$. Now let us assume that the sampling scheme is such that n_{0j} is fixed from sample to sample and $p_{0j} = 1$, with $j = 1, 2, \ldots, s$. Then the likelihood function (which, in this case, is the same as the probability) is given by

$$\phi = \prod_{j=1}^{s} \left[\frac{n_{0j}!}{\prod_{i=1}^{r_j} n_{ij}\,!} \prod_{i=1}^{r_j} p_{i_j}^{\;n_{ij}} \right] \qquad \ldots \quad (15.1.1)$$

In most realistic problems, however, $r_j = r$ (i.e., independent of j) which is what will be assumed in the following discussion, although the possibility of a general form (15.1.1) will also be kept in mind. Notice that (15.1.1) or its special case when $r_j = r$, is based on the product of s separate multinomial distributions. Now suppose that i is a multiple (here k—*ple*) subscript $i_1\, i_2 \ldots i_k$ and j also is a multiple (here l—*ple*) subscript $j_1 j_2 \ldots j_l$, with $i_1 = 1, 2, \ldots, r_{1j}$; $i_2 = 1, 2, \ldots, r_{2j}$; $\ldots$; $i_k = 1, 2, \ldots, r_{kj}$ and $j_1 \epsilon (s_1)_{j_2 j_3 \ldots j_l}, j_2 \epsilon (s_2)_{j_3} \; j_l{}', \ldots, j_{l-1} \epsilon (s_{l-1})_{j_l}, j_l = 1, 2, \ldots, s_l$ where $(s_1)_{j_2 \ldots j_l}$ is a subset of s_1 depending on $j_2 \ldots j_l$, and so on up to $(s_{l-1})_{j_l}$. This will be said to be a k-variate body of data arranged in l ways of classification and each of the running subscripts $i_1, i_2, \ldots, i_k$ will be said to be a 'variate' and each of the running subscripts $j_1, j_2, \ldots, j_l$ will be said to be a 'way of classification'. As observed before, it may be noted that in most realistic problems j will drop out of the range of $i_1, i_2, \ldots, i_k$, i.e., that $i_1 = 1, 2, \ldots, r_1$; $i_2 = 1, 2, \ldots, r_2$, and so on. Also, in one class of problems the range of the subscripts $j_1, j_2, \ldots, j_l$ will be as indicated before while in another class of problems the range will be less general, being given by $j_1 = 1, 2, \ldots, s_1$; $j_2 = 1, 2, \ldots, s_2$; $\ldots$; $j_l = 1, 2, \ldots, s_l$.

In this chapter certain types of composite hypothesis on the p's will be considered, the more general types of composite hypothesis and more general decision procedures involving the p's being reserved for a later monograph. As will be observed later, a composite hypothesis, to be physically meaningful, will have a particular slant

---

* See references [1, 2, 8, 9, 10, 13, 25, 30, 33, 47, 48, 55] in this connection.

with regard to the 'variate' 'i' or its components $i_1, i_2, \ldots, i_k$ and another slant with respect to the 'way of classification' 'j' its components $j_1, j_2, \ldots, j_l$.

In many (though not all) problems to be discussed here either (i) $k = 1$ and $l = 1$ or 2, with $i_1 = i = 1, 2, \ldots, r$, and $j_1 = j = 1, 2, \ldots, s$ or $j_1 = 1, 2, \ldots, s_{j_2}$, and $j_2 = 1, 2, \ldots, s_2$ or (ii) $k = 2$ or 3 and $l = 1$, with $i_1 = 1, 2, \ldots, r_1$; $i_2 = 1, 2, \ldots, r_2$ and (when $k = 3$) $i_3 = 1, 2, \ldots, r_3$, and with $j_1 = j = 1$ (which subscript will, therefore, drop out). Case (i) will be called a univariate problem under one or two ways of classification and case (ii) will be called a bivariate or trivariate problem with one population. We shall now discuss some problems under a two-way or three-way frequency table.

15.2. *Problems in a two-way table.* To fix our ideas, consider first a two-way, say $r \times s$ table with observed frequencies n_{ij} in the (ij)-th cell (with $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$). Also let $\sum_i n_{ij} = n_{0j}$, $\sum_j n_{ij} = n_{i0}$ and $\sum_{i,j} n_{ij} = n_{00} = n$(say).

15.2.1. *Both 'i' and 'j' are 'variates.*

Assume that we have a sample of n independent observations such that p_{ij} is the probability of an observation in the (ij)-th cell, and n is fixed from sample to sample. Also let $\sum_j p_{ij} = p_{i0}$, $\sum_i p_{ij} = p_{0j}$, $\sum_{i,j} p_{ij} = p_{00} = 1$. Then the likelihood function is given by

$$\phi = \frac{n!}{\prod_{i,j} n_{ij}!} \prod_{ij} p_{ij}^{n_{ij}} \qquad \ldots \quad (15.2.1.1)$$

The composite hypothesis [47, 48] *we shall be interested in testing is that 'i' and 'j' are independent, that is, that* $H_0 : p_{ij} = p_{i0}p_{0j}$ *against* $H \neq H_0$, where p_{i0}'s and p_{0j}'s are arbitrary positive nuisance parameters subject to $\sum_i p_{i0} = \sum_j p_{0j} = 1$. This is the analogue of the hypothesis of no correlation in a bivariate normal population. Under H_0 we shall have the likelihood functions ϕ_0 given by

$$\phi_0 = \frac{n!}{\prod_{i,j} n_{ij}!} \prod (p_{i0}p_{0j})^{n_{ij}} = \frac{n!}{\prod_{i,j} n_{ij}!} \prod p_{i0}^{n_{i0}} \prod p_{0j}^{n_{0j}}. \qquad (15.2.1.2)$$

15.2.2. *'i' is a 'way of classification' and 'j' is a 'variate'.*

Assume that we have r independent sets of sizes $n_{10}, n_{20}, \ldots, n_{r0}$ of independent observations such that n_{i0} $(i = 1, 2, \ldots, r)$ is fixed from sample to sample and p_{ij} is the probability of an observation in the (ij)-th cell. Also we notice that $\sum_j p_{ij} = p_{i0} = 1$. Then the likelihood function is given by

$$\phi = \prod_i \left[\frac{n_{i0}!}{\prod_j n_{ij}!} \prod_j p_{ij}^{n_{ij}}\right], \qquad \ldots \quad (15.2.2.1)$$

The composite hypothesis we shall be interested in testing is that p_{ij}, for any j, is independent of 'i' or in other words, that $H_0 : p_{ij} = q_{0j}$(say against $H \neq H_0$, where q_{0j}'s are arbitrary positive nuisance parameters subject to $\sum_j q_{0j} = \sum_j p_{ij} = p_{i0} = 1$. This is the analogue of the hypothesis of the equality of means for r homoscedastic univariate normal populations. Under H_0 we shall have

$$\phi_0 = \prod_i \left[\frac{n_{i0}!}{\prod_j n_{ij}!} \prod_j q_{0j}^{n_{ij}}\right] = \frac{\prod n_{i0}!}{\prod_{i\,j} n_{ij}!} \prod_j q_{0j}^{n_{0j}}. \qquad \ldots \quad (15.2.2.2)$$

This ϕ_0 could also have been obtained by starting from the ϕ of (15.2.1.1), then putting $H_0 : p_{ij} = p_{i0}p_{0j}$ and then finding under H_0 the conditional probability subject to n_{i0}'s being fixed. But it seems that physically this is far less realistic than the model here used, although historically this is more or less what has been done so far.

The case of 'i' being a 'variate' and 'j' a 'way of classification' is exactly similar and need not be separately considered.

15.2.3. *Both 'i' and 'j' are 'ways of classification'.* Here we have a sampling scheme in which n_{i0}'s and n_{0j}'s ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$) are supposed to be fixed from sample to sample. In this situation, on the hypothesis of independence between 'i' and 'j', we can write down the likelihood function ϕ_0 without assuming that the observations are independent. For this we start from an urn problem model in which there is an urn containing $n_{10}, n_{20}, \ldots, n_{r0}$ balls of r different colors from which we draw successively without replacement) $n_{01}, n_{02}, \ldots, n_{0s}$ balls (with $\sum_i n_{i0} = \sum_j n_{0j} = n$). The joint probability that the j-th bunch n_{0j} will contain $n_{1j}, n_{2j}, \ldots, n_{rj}$ balls of different colors (with $j = 1, 2, \ldots, s$) will be given by

$$\phi_0 = \prod_i n_{i0}! . \prod_j n_{0j}!/n! \prod_{i,j} n_{ij}!. \qquad \ldots \quad (15.2.3.1)$$

The great advantage of this scheme is that the different observations need not be assumed to be independent and the great disadvantage is that we would not know how to write down ϕ(under a general H as distinct from the null hypothesis H_0 of independence between 'i' and j'). This means that here it is not only that we do not have any idea of the power of a test for H_0 against alternatives but also that it would not be possible to obtain a one tailed χ^2 test for H_0 by using the same kind of heuristic arguments that we shall use for the first two situations. We can use a one-tailed χ^2-test here just by analogy with what we do in the first two cases.

This ϕ_0 could also have been obtained by starting from the ϕ of (15.2.1.1), then putting H_0 : $p_{ij} = p_{i0}p_{0j}$ and then finding under H_0 the conditional probability subject to n_{i0}'s and n_{0j}'s being fixed. But, for one thing, this except for some very special situations, would be less realistic than the model here used and, for another thing, this would deprive (15.2.3.1) of the one great advantage it possesses in that the successive observations do not have to be independent. Notice that (15.2.1.1) is based on the assumption of the observations being all independent.

It will be seen that the apprcach here is not one of conditional probability at all. It will also be seen that there are three different sampling schemes each leading in a natural way to a particular probability model and a particular type of hypothesis to be tested. From a physical standpoint it would not be proper to break this tie and use a particular probability model and test a particular type of hypothesis when the sampling scheme is something different. It will be noticed that in most situations of life the natural sampling schemes are those of (i) or (ii) but there are situations, e.g. Fisher's tea tasting experiments or those connected with the extra-sensory perception experiments or with the claims of astrologers as to prediction, etc. where (iii) might be a natural sampling scheme.

15.3. *Problems in a three-way table.* As a natural extension of a two-way table consider a three-way $r \times s \times t$ table with observed frequencies n_{ijk} in the (ijk)-th cell (with $i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$; $k = 1, 2, \ldots, t$). Also let $\sum_i n_{ijk} = n_{0jk}$ $\sum_j n_{ijk} = n_{i0k}$, $\sum_k n_{ijk} = n_{ij0}$, $\sum_{i,j} n_{ijk} = n_{00k}$, $\sum_{i,k} n_{ijk} = n_{0j0}$, $\sum_{j,k} n_{ijk} = n_{i00}$, $\sum_{i,j,k} n_{ijk} = n_{000}$ $= n$(say).

15.3.1. *'i' 'j' and 'k' all 'variates'.* Assume that we have a sample of n independent observations such that p_{ijk} is the probability of an observation in the (ijk)-th cell and n is fixed from sample to sample. Also let $\sum_i p_{ijk} = p_{0jk}$, $\sum_j p_{ijk} = p_{i0k}$, $\sum_k p_{ijk}$ $= p_{ij0}$, $\sum_{i,j} p_{ijk} = p_{00k}$, $\sum_{i.k} p_{ijk} = p_{0j0}$, $\sum_{j,k} p_{ijk} = p_{i00}$, $\sum_{i,j,k} p_{ijk} = p_{000} = 1$.

The likelihood function will be given by

$$\phi = \frac{n!}{\prod\limits_{i,j,k} n_{ijk}!} \prod_{i,j,k} p_{ijk}^{n_{ijk}}. \qquad \ldots \quad (15.3.1.1)$$

In this case, as indicated in [47, 48] we shall be interested in testing a class of composite hypotheses, a typical one being,

15.3.1a. *Hypothesis of conditional independence between 'i' and 'j'|'k'.* This will be

$$H_0 : \frac{p_{ijk}}{p_{00k}} = \frac{p_{i0k}}{p_{00k}} \cdot \frac{p_{0jk}}{p_{00k}} \text{ or } p_{ijk} = \frac{p_{i0k} p_{0jk}}{p_{00k}}, \qquad \ldots \quad (15.3.1.2)$$

against $H \neq H_0$ (for $i = 1, \ldots, r$; $j = 1, \ldots, s$; $k = 1, \ldots, t$).

This is the analogue of the hypothesis of no partial correlation (between x and $y)|z$ in a three-variate normal population. As shown in [47], if we superimpose on this the *composite hypothesis of independence between 'i' and 'k', and between 'j' and 'k',* i.e.,

$$p_{i0k} = p_{i00} p_{00k} \text{ and } p_{0jk} = p_{0j0}\, p_{00k}, \qquad \ldots \quad (15.3.1.3)$$

(which is the analogue of the hypothesis of no total correlation between (x and z) and (y and z) in a three variate normal population) we should have

$$p_{ijk} = p_{i00}\, p_{0j0}\, p_{00k}, \qquad \ldots \quad (15.3.1.4)$$

which is the condition of complete independence of 'i', 'j' and 'k'.

Following [47] we shall also be interested in another class of composite hypotheses, the typical one being,

15.3.1b. *Hypothesis of independence between '(i, j)' and 'k'.* This will be

$$H_0 : p_{ijk} = p_{ij0}\, p_{00k} \text{against } H \neq H_0 \text{ (for } i = 1, \ldots, r;\ j = 1, \ldots, s;\ k = 1, \ldots, t). \qquad \ldots \quad (15.3.1.5)$$

This is the analogue of the hypothesis of no multiple correlation between (x, y) and z in a three variate normal population.

As shown in [47], (15.3.1.5) implies the composite hypotheses,

$$p_{i0k} = p_{i00}\, p_{00k} \text{ and } p_{0jk} = p_{0j0}\, p_{00k}. \qquad \ldots \quad (15.3.1.6)$$

But (15.3.1.6) will not imply (15.3.1.5). The extra condition that was needed on top of (15.3.1.6) to lead to (15.3.1.5) was shown [47, 48] to be the composite hypothesis

$$H_0 : p_{ijk} = \frac{q_{ij0}\, q_{i0k}\, q_{0jk}}{q_{i00}\, q_{0j0}\, q_{00k}}, \qquad \ldots \quad (15.3.1.7)$$

where q_{ij0} q_{i0k}, q_{0jk} were defined to be arbitrary (positive) functions of (i, j), (i, k) and (j, k) and q_{i00}, q_{0j0}, q_{00k} arbitrary (positive) functions of i, j and k, with no summation convention connecting them as in the case of the p_{ijk}'s. By analogy with analysis of variance this will be called the hypotheses of 'no interaction'.

15.3.2. *'i' and 'j' are 'variates' and 'k' a 'way of classification'.*

Assume that we have t independent sets of sizes $n_{001}, \ldots, n_{00t}$ of independent observations such that $n_{00k}(k = 1, \ldots, t)$ is fixed from sample to sample and p_{ijk} is the probability of an observation in the (ijk)-th cell. Notice that $\sum\limits_{i,j} p_{ijk} = p_{00k} = 1$. The likelihood function will be given by

$$\phi = \prod_{k} \left[\frac{n_{00k}!}{\prod\limits_{i,j} n_{ijk}!} \prod_{i,j} p_{ijk}^{n_{ijk}} \right] \qquad \ldots \quad (15.3.2.1)$$

Here we shall be interested in testing,

15.3.2a. *Hypothesis of independence between 'i' and 'j' for each 'k'*, i.e.,

$$H_0 : p_{ijk} = p_{i0k}\, p_{0jk} \text{ against } H \neq H_0 \text{ (for } i = 1, \ldots, r; j = 1, \ldots s;\ k = 1, \ldots, t). \qquad \ldots \quad (15.3.2.2)$$

If we superimpose on this *the composite hypothesis that the marginal 'i' (obtained by summing over 'j') is independent of 'k' and similarly for 'j'*, i.e., *that* p_{i0k} *is a pure function of 'i' and* p_{0jk} *is a pure fnnction of 'j'*, i.e.,

$$p_{i0k} = q_{i00} \text{(say) and } p_{0jk} = q_{0j0} \text{ (say)}, \qquad \ldots \quad (15.3.2.3)$$

we should have

$$p_{ijk} = q_{i00}\, q_{0j0}. \qquad \ldots \quad (15.3.2.4)$$

Notice from (15.3.2.3) that $\sum_i q_{i00} = \sum_i p_{i0k} = p_{00k} = 1$ and also that $\sum_j q_{0j0} = \sum_j p_{0jk} = p_{00k} = 1$.

We shall also be interested *in the composite* (15.3.2b) *hypothesis that* p_{ijk} *is independent of 'k', i.e.,* p_{ijk} *is a pure function of '(ij)'*, i.e.,

$$H_0 : p_{ijk} = q_{ij0} \text{ (say) against } H \neq H_0 \text{ (for all } i, j \text{ and } k). \quad \ldots \quad (15.3.2.5)$$

This is the anologue of the hypothesis of the equality of t mean vectors (each consisting of 2 components) for t bivariate normal populations, each having the same variance-covariance matrix.

If we sum over 'j' and 'i' separately, this would imply

$$p_{i0k} = \sum_j q_{ij0} = q^{(2)}_{i00} \text{ (say) and } p_{0jk} = \sum_i q_{ij0} = q^{(1)}_{0j0} \text{ (say).} \quad \ldots \quad (15.3.2.6)$$

As in the case where 'i', 'j' and 'k' are all 'variates'', so also here, (15.3.2.5)$\Longrightarrow$(15.3.2.6) but (15.3.2.6) does not imply (15.3.2.5). Exactly in the same way as in [47, 48] it can be shown that the extra condition which when superimposed on (15.3.2.6) will $\Longrightarrow$ (15.3.2.5) is

$$p_{ijk} = \frac{q_{ij0}\, q_{0jk}\, q_{i0k}}{q_{i00}\, q_{0j0}\, q_{00k}}. \quad \ldots \quad (15.3.2.7)$$

15.3.3. *'i' is a 'variate' and 'j' and 'k' are 'ways of classification'.*

Assume that we have $s \times t$ independent sets of sizes n_{0jk} of independent observations such that $n_{0jk} (j = 1, \ldots, s;\ k = 1, \ldots, t)$ is fixed from sample to sample and p_{ijk} is the probability of an observation in the (ijk)-th cell. Notice that $\sum_i p_{ijk} = p_{0jk} = 1$. The likelihood function will be given by

$$\phi = \prod_{j,k} \left[\frac{n_{0jk}!}{\prod_i n_{ijk}!} \prod_i p_{ijk}^{n_{ijk}} \right] \quad \ldots \quad (15.3.3.1)$$

Here we shall be interested in *the composite* 15.3.3a *hypothesis that for any 'k',* p_{ijk} *is independent of 'j'*, i.e., *that* p_{ijk} *is a pure function of 'i' and 'k'*, i.e.,

$$H_0 : p_{ijk} = q_{i0k} \text{ (say) against } H \neq H_0 \text{ (for all } i, j \text{ and } k). \quad \ldots \quad (15.3.3.2)$$

Notice that $\sum_i q_{i0k} = \sum_i p_{ijk} = p_{0jk} = 1$.

We shall be also interested in the other *composite* (15.3.3b) *hypothesis that for any "j"* p_{ijk} *is independent of "k"*, i.e., *that* p_{ijk} *is a pure function of* 'i' and 'j', i.e.,

$$H_0 : p_{ijk} = q_{ij0} \text{(say), against } H \neq H_0 \text{ (for all } i, j \text{ and } k). \quad \ldots \quad (15.3.3.3)$$

Notice that $\sum_i q_{ij0} = \sum_i p_{ijk} = p_{0jk} = 1$. We now observe that (15.3.3.2)+(15.3.3.3) implies that p_{ijk} is a pure function of 'i'; i.e., that

$$p_{ijk} = q_{i00} \text{ (say), for all } i, j \text{ and } k. \quad \ldots \quad (15.3.3.4)$$

If, in a one-way classification in the usual analysis of variance, 'i' corresponds to the 'variate', 'j' to the so-called 'concomitant variate' and 'k' to the 'way of classification', then it will be seen on a little reflection that (15.3.3.1) will be the analogue of the hypothesis of no regression and (15.3.3.2) will be the analogue of the hypothesis of no covariance. On the other hand, suppose we take 'j' and 'k' as just two 'ways of classification', for example, we take 'j' as, say, blocks and 'k' as, say, treatments in a randomized block experiment (with more than one and in general unequal number of replications in each cell). Then (15.3.3.1) will be the analogue of 'no block effect' for each treatment separately and (15.3.3.2) will be the analogue of 'no treatment effect' for each block separately. In other words, in the usual parlance of analysis of variance (15.3.3.1) lumps together one 'no main effect' and 'no interaction', while (15.3.3.2) lumps together another 'no main effect' and 'no interaction'.

15.3.4. *'i' is a variate and 'j' and 'k' are 'ways of classification' in the sense of a 'balanced incomplete' or 'partially balanced incomplete' or a more general type of 'incomplete' block experiment.*

Assume as before that there are r 'i''s, s 'j''s and t 'k''s. Assume further that 'j' is a block and 'k' a treatment and that, for any 'j', there is a set of treatments $(T)_j$ to go with it, of number t_j. In other words, for a given j, k takes on the set of values $(t)_j$ where $(t)_j$ is a set of indices of number t_j out of $1, 2, \ldots, t$. Now assume that we have $\sum\limits_{j=1}^{s} t_j$ independent sets of sizes n_{0jk} of independent observations such that n_{0jk} $(k \epsilon (t)_j\,;\ j = 1, 2, \ldots, s)$ is fixed from sample to sample and p_{ijk} is the probability of an observation in the (ijk)-th cell. As before $\sum\limits_i p_{ijk} = p_{0jk} = 1$. The likelihood function will be given by

$$\phi = \prod_{j=1}^{s} \prod_{k \varepsilon (t)_j} \left[\frac{n_{0jk}!}{\prod\limits_i n_{ijk}} \prod_i p_{ijk}^{\,n_{ijk}} \right]. \qquad \ldots \quad (15.3.4.1)$$

We can take over the hypothesis (15.3.3.2) of 'no block effect for each treatment separately' and (15.3.3.3) of 'no treatment effect for each block separately'. For a 'balanced incomplete design' all t_j's will be equal and there will be a highly symmetrical pattern while for a 'partially balanced design" all t_j's will be equal and there will be a less symmetrical pattern.

15.3.5. *'i' is a 'way of classification' and 'j, k' also are 'ways of classification' in the sense that n_{i00}'s and n_{0jk}'s are fixed from sample to sample.* Following case (iii) of section 2, we can write down ϕ_0 in this case (exactly the same way as we wrote down the ϕ_0 in that case) on the hypothesis of independence between 'i' and '(j, k)'. This will be

$$\phi_0 = \prod_i n_{i0}! \prod_{j,k} n_{0jk}! / n! \prod_{i,j,k} n_{ijk}!. \qquad \ldots \quad (15.3.5.1)$$

Starting from this we can test the hypothesis of independence between 'i' and 'j' 'k'.

The case of 'i', 'j' and 'k' being 'ways of classification' in the sense of n_{i00}'s, n_{0j0}'s and n_{00k}'s being fixed from sample to sample (but not n_{0jk}'s) is also of some interest. But we shall not consider that case in the present paper.

The extension of the problems of the two-way tables of sections 2 to those of the three-way tables of section 3 is a rather big conceptual jump, but the extension from three-way tables to those of higher dimensions involves no such jump and will not be discussed in this chapter except for some remarks toward the end.

15.4. *The derivation of the χ^2-test by the union-intersection principle.*

Let a random sample of size n from some population be classified into $k(< n)$ mutually exclusive and exhaustive categories according to some observable characteristics (qualitative or quantitative) and let the probability of a random observation falling in the i-th category be p_i with $p_i > 0$ and $\sum\limits_{i=1}^{k} p_i = 1$. Let n_i denote the observed frequency in the i-th category with of course $\sum\limits_{i} n_i = n$. Also let $\mathbf{n}'(1\times k) = \mathbf{n}' = (n_1, n_2, \ldots, n_k)$ and $\mathbf{p}'(1\times k) = \mathbf{p}' = (p_1, p_2, \ldots, p_k)$. We have now

$$P\,[\mathbf{n}'\,|\,\mathbf{p}'] = \frac{n!}{\prod\limits_i n_i!} \prod_i p_i^{n_i}\,. \qquad \ldots \quad (15.4.1)$$

15.4.1. A simple hypothesis $H_0 : \mathbf{p}' = \mathbf{p}_0'$ against the composite alternative $H : \mathbf{p}' \neq \mathbf{p}_0'$.

Consider first the most powerful test at a level say β_1 of $H_0 : \mathbf{p}' = \mathbf{p}_0'$ against a *specific* $\mathbf{p}_1' \neq \mathbf{p}_0'$, which, by the Neyman-Pearson lemma, will be as follows:

reject H_0 if

$$P\,[\mathbf{n}'\,|\,\mathbf{p}_1']/P[\mathbf{n}'\,|\,\mathbf{p}_0'] \geqslant \mu, \qquad \ldots \quad (15.4.1.1)$$

and don't reject H_0 otherwise, where, given μ, the size of the critical region (15.4.1.1) under $\mathbf{p}_0'$ should be $\beta(\mu, \mathbf{p}_0', \mathbf{p}_1', n)$. Substituting in (15.4.1.1) from (15.4) and taking logarithms on both sides we see after a little simplification that (15.4.1.1) $\Longleftrightarrow$

$$\frac{\mathbf{a}'(\mathbf{p}_1)(\mathbf{n}-n\mathbf{p}_0)}{\sqrt{\mathbf{a}'(\mathbf{p}_1)\ \wedge^0\ \mathbf{a}(\mathbf{p}_1)}} \geqslant \frac{\log\mu - n\mathbf{a}'(\mathbf{p}_1)\mathbf{p}_0}{\sqrt{\mathbf{a}'(\mathbf{p}_1)\ \wedge^0\ \mathbf{a}(\mathbf{p}_1)}} \qquad \ldots \quad (15.4.1.2)$$

that is, $\geqslant \mathbf{c}(\mathbf{p}_1', \beta_{\mathbf{p}_1'}, n)$ say,

where $\mathbf{a}'(\mathbf{p}_1) = [\log(p_{11}/p_{10}), \log(p_{21}/p_{20}), \ldots, \log(p_{k1}/p_{k0})]$, and $\wedge^0 = (\sigma_{ij}^0)$ is the variance-covariance matrix of $n_1, n_2, \ldots, n_k$ under $H_0 : \mathbf{p}' = \mathbf{p}_0'$, and where $\beta\mathbf{p}_1'$ is supposed to vary with, that is, depend upon $\mathbf{p}_1$. It is thus evident that, for a fixed c, the critical region

$$w(\mathbf{p}_1', c) = \left\{\mathbf{n}' : \frac{\mathbf{a}'(\mathbf{p}_1)[\mathbf{n}-n\mathbf{p}_0]}{\sqrt{\mathbf{a}'(\mathbf{p}_1)\wedge^0\mathbf{a}(\mathbf{p}_1)}} \geqslant \mathbf{c}\right\}, \qquad \ldots \quad (15.4.1.3)$$

is the most powerful critical region for testing $\mathbf{p}' = \mathbf{p}_0'$ against a specific $\mathbf{p}' = \mathbf{p}_1'(\neq \mathbf{p}_0')$ at a level of significance $\beta(\mathbf{p}_1', c, n)$. Since the composite H: $\mathbf{p}' \neq \mathbf{p}_0'$ is the union of all $H_i : \mathbf{p}' = \mathbf{p}_1' \; (\neq \mathbf{p}_0')$ we use the union intersection principle [45] and take for H_0, against the composite H, the critical region

$$w(c) = \underset{\mathbf{p} \neq \mathbf{p}_0}{U} \; w(\mathbf{p}', c). \qquad \ldots \quad (15.4.1.4)$$

Thus we should have

$$\text{Complement } w(c) = \left\{ \mathbf{n} : \sup_{\mathbf{p}' \neq \mathbf{p}'} \frac{\mathbf{a}'(\mathbf{p})[\mathbf{n} - n\mathbf{p}_0]}{\sqrt{\mathbf{a}(\mathbf{p}) \Lambda^0 \mathbf{a}(\mathbf{p})}} < c \right\}. \quad \ldots \quad (15.4.1.5)$$

Since $\sum\limits_i n_i = n$ and $\sum\limits_i p_{i0} = 1$, we can write

$$\frac{\mathbf{a}'(\mathbf{p})[\mathbf{n} - n\mathbf{p}_0]}{\sqrt{\mathbf{a}'(\mathbf{p}) \Lambda^0 \mathbf{a}(\mathbf{p})}} = \frac{\mathbf{b}'(\mathbf{p})[\mathbf{n}^{*\prime} - n\mathbf{p}_0^{*\prime}]}{\sqrt{\mathbf{b}'(\mathbf{p}) \Lambda_{kk}^0 \mathbf{b}(\mathbf{p})}}, \qquad \ldots \quad (15.4.1.6$$

where

$$\mathbf{n}^{*\prime} = (n_1, n_2, \ldots, n_{k-1}), \; \mathbf{p}_0^{*\prime} = (p_{10}, p_{20}, \ldots, p_{k-1,0}),$$

$$\mathbf{b}'(\mathbf{p}) = (b_1(\mathbf{p}), b_2(\mathbf{p}), \ldots, b_{k-1}(\mathbf{p})),$$

$$b_i(\mathbf{p}) = a_i(\mathbf{p}) - a_k(\mathbf{p}) = \log(p_{i1}p_{k0}/p_{k1}p_{i0}) \quad (\text{for } i = 1, \ldots, k-1);$$

and Λ_{kk}^0 is the matrix formed by cutting out the k-th row and the k-th column of Λ^0. Notice that each $b_i(\mathbf{p})$ can assume any value on the real line and conversely, given any real vector $\mathbf{b}_0' = (b_{10}, \ldots, b_{k-1,\,0})$, the equations: $\mathbf{b}'(\mathbf{p}) = \mathbf{b}_0'$, have always a unique solution in $\mathbf{p}$, e.g.

$$p_i/p_k = (p_{i0}/p_{k0})e^{b_{i0}} = \lambda_{i0}, \text{ say}$$

$$\text{or } p_i = \lambda_{i0}/(1 + \sum_{i=1}^{k-1} \lambda_{i0}) \quad (\text{for } i = 1, 2, \ldots, k-1) \text{ and } \underset{k}{p} = 1/(1 + \sum_{i=1}^{k-1} \lambda_{i0}).$$

Hence we have [25]

$$\sup_{\mathbf{p}' \neq \mathbf{p}'_0} \frac{\mathbf{a}'(\mathbf{p})[\mathbf{n} - n\mathbf{p}_0]}{\sqrt{\mathbf{a}'(\mathbf{p}) \Lambda^0 \mathbf{a}(\mathbf{p})}} = \sup_{\mathbf{b}'} \frac{\mathbf{b}'[\mathbf{n}^* - n\mathbf{p}_0^*]}{\sqrt{\mathbf{b}' \Lambda_{kk}^0 \mathbf{b}}} = + \left[n[\mathbf{n}^* - n\mathbf{p}_0^*]' \Lambda_{kk}^{0^{-1}} [\mathbf{n}^* - n\mathbf{p}_0^*] \right]$$

We next observe that

$$\sigma_{ij} = -np_{i0}p_{j0} \text{ if } i \neq j \text{ and } \sigma_{ii} = np_{i0}(1 - p_{i0}).$$

It is easy to check that if $\Lambda_{kk}^{0-1} = (\alpha_{ij})$, then $\alpha_{ij} = 1/np_{k0}$, if $i \neq j$ and $\alpha_{ii} = 1/np_{i0} + 1/np_{k0}$. We have, therefore

$$\sup_{\mathbf{p}' \neq \mathbf{p}'_0} \frac{\mathbf{a}'(\mathbf{p})[\mathbf{n} - n\mathbf{p}_0]}{\sqrt{\mathbf{a}'(\mathbf{p}) \Lambda^0 \mathbf{a}(\mathbf{p})}} = \left[\sum_{i=1}^{k} \frac{(n_i - np_{i0})^2}{np_{i0}} \right]^{\frac{1}{2}}.$$

Going back to (15.4.1.5) and thence to (15.4.1.4) we now notice that (15.4.1.4) reduces to

$$w(c) = \left\{\mathbf{n} : + \left[\sum_{i=1}^{k} \frac{(n_i - np_{i0})^2}{np_{i0}} \right]^{\frac{1}{2}} \geqslant c \right\}. \qquad \ldots \quad (15.4.1.7)$$

Since the left side of the inequality in (15.4.1.7) is essentially non-negative it is easy to see that we obtain a non-trivial solution only when $c > 0$. It is thus seen that the χ^2-critical region is obtained by using the union-intersection with respect to variation over the alternatives $\mathbf{p}'(\neq \mathbf{p}_0')$, keeping fixed a quantity c defined (in terms of $\mathbf{p}_1'$ and n) by the right side of (15.4.1.2). This means of course letting β vary with $\mathbf{p}$ in an appropriate manner. Now if n is large, we go back to (15.4.1.3) and observe from the asymptotic normality of the left side of the inequality (15.4.1.2) that, as $n \to \infty$,

$$\beta(\mathbf{p}_1', c, n) \to \frac{1}{\sqrt{2\pi}} \int_{c}^{\infty} e^{-t^2/2}\, dt. \qquad \ldots \quad (15.4.1.8)$$

In large samples it is thus seen that keeping c fixed means making β the same for all $\mathbf{p}_1$'s, which means that in large samples the χ^2 critical region (15.4.1.7) comes out as a union-intersection critical region of type I [43]. For large n_i's (the approximation would be good enough even for moderately large values of n_i's) it is well known that the left side of the inequality in (15.4.1.7) is asymptotically distributed as a χ^2 with d.f. $(k-1)$. For a satisfactory proof see [10].

15.4.2 *Test of a certain type of composite hypothesis on p's against a certain type of composite alternative.*

Suppose that the composite hypothesis is given by

$$H_0 : \{p_i = p_i(\theta_1, \theta_2, \ldots, \theta_r)\}_{(\theta_1, \ldots, \theta_r) \varepsilon \Omega},$$

where $p_i(\theta_1, \ldots, \theta_r)$ are k known functions of $r(<k)$ unknown parameters. The hypothesis does not specify the values of the parameters except that they belong to a certain parametric space Ω. The (composite) alternative is $H_1 \neq H_0$. For any specific $(\theta_1^0, \theta_2^0, \ldots, \theta_r^0)$, we obtain, as in the previous section, a heuristic test of the hypothesis $H_0 : \{p_i = p_i(\theta_1^0, \ldots, \theta_r^0)\}$ against $H_1 \neq H_0$, which has a critical region

$$w(c, \theta_1^0; \ldots, \theta_r^0) = \left\{\mathbf{n} : \sum_{i=1}^{k} \frac{(n_i - np_i(\theta_1^0, \ldots, \theta_r^0))^2}{np_i(\theta_1^0, \ldots, \theta_r^0)} \geqslant c^2 \right\}. \quad \ldots \quad (15.4.2.1)$$

This critical region is the region of rejection of

$$H_0 : \{p_i = p_i(\theta_1^0, \ldots, \theta_r^0)\} \text{ for a specific } (\theta_1^0, \ldots, \theta_r^0).$$

Now to reject $H_0 : \{p_i = p_i(\theta_1, \ldots, \theta_r)\}_{(\theta_1, \ldots, \theta)\varepsilon\Omega}$ would be to reject

$$H_0 : p_i = p_i(\theta_1^0, \ldots, \theta_r^0) \text{ for every } (\theta_1^0, \ldots, \theta_r^0) \epsilon \Omega$$

and thus, using the union-intersection principle for the second time, we have for H_0 the critical region.

$$w(c, H_0) = \bigcap_{(\theta_1, \ldots, \theta_r)\varepsilon\Omega} w(c, \theta_1, \ldots, \theta_r)$$

$$= \left\{\mathbf{n} : \inf_{(\theta_1, \ldots, \theta_r)\varepsilon\Omega} \sum_{i=1}^{k} \frac{(n_1 - np_i(\theta_1, \ldots, \theta_r))^2}{np_i(\theta_1, \ldots, \theta_r)} > c^2\right\}, \quad \ldots \quad (15.4.2.2)$$

which is precisely the minimum χ^2-critical region. The equations giving θ_j's in terms of n_i's, in the form, say θ_i's for minimum χ^2 are

$$\frac{\partial}{\partial\theta_j} \sum_{i=1}^{k} \frac{(n_i - np_i(\theta_1, \ldots, \theta_r))^2}{np_i(\theta_1, \ldots, \theta_r)} = 0 \quad (\text{for } j = 1, 2, \ldots, r). \quad \ldots \quad (15.4.2.3)$$

It has been shown [10] that for large n_i's the equation (15.4.2.3) can be replaced by the maximum likelihood equation (so much easier to work with), the likelihood function being

$$\phi_0 = \frac{n!}{\prod_i n_i\,!} \prod_i p_i^{n_i} (\theta_1, \ldots, \theta_r). \quad \ldots \quad (15.4.2.4)$$

The maximum likelihood equations can be put in the form

$$0 = \frac{\partial}{\partial\theta_j} \log \phi_0 = \sum_{i=1}^{k} \frac{n_i}{np_i} \frac{\partial p_i}{\partial\theta_j} = \sum_{i=1}^{k} \frac{n_i - np_i}{np_i} \frac{\partial p_i}{\partial\theta_j} \; (j = 1, 2, \ldots, r). \quad \ldots \quad (15.4.2.5)$$

It has been proved (and will be published in a later monograph) that if we start from the more general probability model of (15.1.1), pose a composite hypothesis problem of the type of section 15.4.2, and then use the union-intersection principle in the sense of sections 15.4.1 and 15.4.2, we obtain the corresponding χ^2-critical region with a structure which is just stated in the following sections.

15.5. *Some useful theorems on* χ^2.

We state here several theorems the first two of which are well known. For a careful statement and satisfactory proof see, e.g., [10]. The remaining theorems have been proved, and the proofs will be offered in later papers.

Theorem I. If we start out from a $P(\mathbf{n}/\mathbf{p}^0)$ or ϕ_0 of the form

$$\phi_0 = n! \prod_{i=1}^{r} p_i^{n_i} \Big/ \prod_{i=1}^{r} n_i! \; (\text{under } \sum_i n_i = n \text{ and } \sum_i p_i = 1 \text{ and } p_i > 0), \text{ as } n \to \infty,$$

the sampling distribution (under ϕ_0) of $\sum_i (n_i - np_i^0)^2/np_i^0$ tends to the χ^2-distribution with d.f. $r-1$.

This is used for testing the simple hypothesis $H_0 : \mathbf{p} = \mathbf{p}^0$ against the composite alternative $H : \mathbf{p} \neq \mathbf{p}^0$ under the general model: $\phi = n! \prod_{i=1}^{r} p_i^{n_i} / \prod_{i} n_i!$ associated with a sampling scheme in which only n is fixed from sample to sample but not any of the n_i's.

Theorem II. Let $P(\mathbf{n}|H_0)$ or ϕ_0 be of the form $\phi_0 = n! \prod_{i=1}^{r} p_i^{n_i} (\theta_1, ..., \theta_s) \prod_{i=1}^{r} n_i!$, where $\Sigma_i n_i = n$ and where $p_1(\theta_1, ..., \theta_s), ... p_r(\theta_1, ..., \theta_s)(s < r)$ are r functions of the s parameters such that, for all points of a non-degenerate interval A in the s-dimensional space of the θ_j's, the p_i's satisfy the following conditions: (a) $\sum_{i=1}^{r} p_i(\theta_1, ..., \theta_s) = 1$, (b) $p_i(\theta_1, ..., \theta_s) > c^2 > 0$ for all i, (c) every p_i has continuous derivatives $\frac{\partial p_i}{\partial \theta_j}$ and $\frac{\partial^2 p_i}{\partial \theta_j \partial \theta_j}$ and (d) the $r \times s$ matrix $\left\{ \frac{\partial p_i}{\partial \theta_j} \right\}$ is of rank s. Then, assuming that A is so defined that the true population parameter point $(\theta_1^0, ..., \theta_s^0)$ $(= \boldsymbol{\theta}^{0\prime}$, say) is an interior point of A, (i) there exists one and only one solution $(\hat{\theta}_1, ..., \hat{\theta}_s)(= \hat{\boldsymbol{\theta}}'$, say) of the equation (15.4.2.5) such that $\hat{\boldsymbol{\theta}}' \rightarrow \boldsymbol{\theta}^{0\prime}$ in probability as $n \rightarrow \infty$. (ii) Furthermore, this $\hat{\boldsymbol{\theta}}'$ has the property that, as $n \rightarrow \infty$, the sampling distribution of $\sum_{i=1}^{r} (n_i - np_i(\hat{\theta}_1, ..., \hat{\theta}_s))^2 / np_i(\hat{\theta}_1, ..., \hat{\theta}_s)$ tends to the χ^2-distribution with with d.f. (r−1)−s. The probability measure ϕ_{00} under which (i) and (ii) hold is one which we obtained by sticking into ϕ_0 the true population point $\boldsymbol{\theta}^{0\prime}$.

This is used to test the composite hypothesis. $H_0 : p_i = p_i(\theta_1, ..., \theta_s)$ $(i = 1, ..., r)$ against the composite alternative $H \neq H_0$ under the general model of theorem I associated with the same sampling scheme. Under H_0(and H_0 alone) will ϕ be of the form ϕ_0 of this theorem.

Theorem III. With the same ϕ_0 as in theorem II and with $p_i(\theta_1, ..., \theta_s)$'s being defined as functions of $(\theta_1, ..., \theta_s)$ subject to the same conditions as in theorem II, suppose that we have furthermore, under the null hypothesis $H_0 : f_k(\theta_1, ..., \theta_s) = f_k^0$, where f_k^0's are fixed and $k : 1, 2, ..., t < s$, such that over A, (e) each f_k has continuous derivatives $\frac{\partial f_k}{\partial \theta_j}$ and $\frac{\partial^2 f_k}{\partial \theta_j \partial \theta_k}$ and (f) the $t \times s$ matrix $\left\{ \frac{\partial f_k}{\partial \theta_j} \right\}$ is of rank t. Next, let us write down likelihood equations subject to the given constraints on θ_j's:

$$\sum_{i=1}^{r} \frac{n_i - np_i(\theta_1, ..., \theta_s)}{np_i(\theta_1, ..., \theta_s)} \frac{\partial p_i}{\partial \theta_j} + \sum_{k=1}^{t} \mu_k \frac{\partial f_k}{\partial \theta_j} = 0 (j = 1, ..., s) \qquad \ldots \quad (15.5.1)$$

$$f_k(\theta_1, ..., \theta_s) = f_k^0 (k = 1, 2, ..., t).$$

Then (i) there exists one and only one solution $(\hat{\theta}_1, ..., \hat{\theta}_s)$, $(\hat{\mu}_1, ..., \hat{\mu}_t)$ of these equations such that $\hat{\boldsymbol{\theta}}' \rightarrow \boldsymbol{\theta}^{0\prime}$ in probability as $n \rightarrow \infty$. (ii) This $\hat{\boldsymbol{\theta}}'$ has the further property that, as $n \rightarrow \infty$, this sampling distribution of $\Sigma(n_i - np_i(\hat{\theta}_1, ..., \hat{\theta}_s))^2 / np_i(\hat{\theta}_1, ..., \hat{\theta}_s)$ tends to the χ^2-distribution with d.f. $(r-1)-(s-t)$. As in theorem II, the probability measure ϕ_{00} under which (i) and (ii) are true is of course the one which we obtain by sticking into ϕ_0 the true population point $\boldsymbol{\theta}^{0'}$.

This is used in the same general situation that ties in with theorem II, the only difference being that here the composite hypothesis is given as $H_0 : p_i = p_i(\theta_1, \ldots, \theta_s)$ subject to additional constraints on θ_j's. Theoretically, under certain reasonable restrictions, we might try to use the constraints on θ_j's to eliminate some of the θ_j's and express the p_i's in terms of the proper number of independent θ_j's.

But in doing so we might well obtain the p_i's as functions of these independent θ_j's such that they have one set of functional forms for one domain of values of the eliminated parameters, another set of functional forms over another domain, and so on. We shall of course be concerned with the functional forms assumed over a sufficiently small neighbourhood enclosing the true (but unknown) parameter point. But this is not directly known and hence in general this problem cannot be thrown back directly on theorem II. A simple illustration will make this clear. Suppose that $p_i = p_i(\theta_1, \ldots, \theta_s)(i = 1, 2, \ldots, r > s)$ and we have furthermore the hypothesis: $\theta_1^2 + \theta_2^2 = 1$ and that (θ_1, θ_2) may take all values on the euclidean plane. Then we have $\theta_1 = +\sqrt{1-\theta_2^2}$, so that we have $p_i = p_i(\sqrt{1-\theta_2^2}, \ldots)$ or $p_i = p_i(-\sqrt{1-\theta_2^2}, \theta_2, \ldots, \theta_s)$ according as the eliminated parameter θ_1 is $+$ *ve* or $-$ *ve*. It is this that prevents a direct appeal being made to theorem II.

But in most practical situations, it is far more convenient to use the customary method of Lagrangian multipliers, and this theorem provides the justification for that.

Theorem IV. With a general ϕ of the form $n! \prod_i p_i^{n_i} / \prod_i n_i!$ (under $\sum_i n_i = n$ and $\sum_i p_i = 1, p_i > 0$), suppose that we have the constraints $f_j(p_1, \ldots, p_r) = 0$ $(j = 1, 2, \ldots, s < r$ and $\sum_i p_i = 1$ is one of the s constraints), where f_j's are defined over an interval A in the r—dimensional space of the p_i's such that (a) f_j's have continuous derivatives $\frac{\partial f_j}{\partial p_i}$ and $\frac{\partial^2 f_j}{\partial p_i \partial p_{i'}}$ and (b) the $s \times r$ matrix $\left\{\frac{\partial f_j}{\partial p_i}\right\}$ is of rank s.

Next, properly using the condition $\sum_i p_i = 1$, let us write down the maximum likelihood equations subject to the given constraints on p_i's:

$$\left\{\begin{array}{ll} \dfrac{n_i - np_i}{np_i} + \sum_{j=1}^{s} \mu_j \dfrac{\partial f_j}{\partial p_i} = 0 & (i = 1, 2, \ldots, r) \\ f_j(p_1, \ldots, p_r) = 0 & (j = 1, 2, \ldots, s < r). \end{array}\right. \qquad \ldots \quad (15.5.2)$$

Then, assuming that the true population parameter point $\mathbf{p}^{0'} = (p_1^0, p_2^0, \ldots, p_r^0)$ is an interior point of A, (i) there exists one and only one solution $\hat{\mathbf{p}}' = (\hat{p}_1, \ldots, \hat{p}_r)$ and $(\hat{\mu}_1, \ldots, \hat{\mu}_s)$ of (15.5.2) such that $\hat{\mathbf{p}} \rightarrow \mathbf{p}^{0'}$ in probability as $n \rightarrow \infty$. (ii) This $\hat{\mathbf{p}}'$ has the further property that, as $n \rightarrow \infty$, the sampling distribution of $\sum_{i=1}^{r} (n_i - n\hat{p}_i)^2 / n\hat{p}_i$ tends to the χ^2-distribution with d.f. $r-(r-s) = s$.

This is used in the same general situations that tie in with theorems II or III, the only difference being that here the composite hypothesis is given as H_0 : $f_j(p_1, \ldots, p_r) = 0 (j = 1, \ldots, s)$. Theoretically it is possible (under certain mild and reasonable restrictions) to move back and forth between the set-ups of theorems II, III and IV, but in practice, depending upon the form in which the hypothesis is put forward, it might algebraically be more convenient to follow one method rather than the others.

Theorem V. If, instead of a ϕ corresponding to a single multinomial distribution, we have a ϕ corresponding to the product of several multinomial distributions, given by

$$\prod_{i=1}^{r} \left[n_{i0}! \prod_{j=1}^{s_i} p_{ij}^{n_{ij}} \Big/ \prod_{j=1}^{s_i} n_{ij}! \right], \text{ with } \prod_{j=1}^{s_i} p_{ij} = p_{i0} \text{ (say)} = 1, (i = 1, 2, \ldots, r)$$

then the theorems I-IV will all hold good as n_{i0}'s$\to \infty$ with n_{i0}/n held constant, with the difference that (i) the statistic concerned will be $\sum_{i,j} (n_{ij} - n_{i0}\hat{p}_{ij})^2/n_{i0}\hat{p}_{ij}$ and (ii) the limiting sampling distribution will be a χ^2 with d.f. = total number of cells − number of independent multinomial distributions − number of independent parameters p_{ij}s that are to be estimated from the data. Corresponding to theorems I, II, III and IV there will be respectively (i) a simple hypothesis H_0: $p_{ij} = p_{ij}^0$, (ii) a composite $H_0 : p_{ij} = p_{ij}(\theta_1, \ldots, \theta_u)$ (with $u < s - r$ where $s = \Sigma\, s_i$), (iii) a composite H_0 : $p_{ij} = p_{ij}(\theta_1, \ldots, \theta_u)$ subject to $f_k(\theta_1, \ldots, \theta_u) = f_k^0$ (with $k = 1, 2, \ldots, v < u$) and (iv) a composite H_0 which is defined in terms of constraints on p_{ij}'s of the form: $f_k(p_{ij}\text{'s}) = 0 (k = 1, 2, \ldots, u < s - r)$. It should be remembered that all the hypotheses (i), (ii), (iii) and (iv) must be so framed as not to violate the basic conditions $\sum_j p_{ij} = p_{i0} = 1 (i = 1, 2, \ldots, r)$. Also, even in considering the alternatives, these basic conditions must not be violated. Notice that j may be a double or a multiple subscript like $j_1, j_2, \ldots, j_l$, in other words it may be a 'bivariate' or a 'multivariate' situation. Also i may be a double or a multiple subscript like $i_1, i_2, \ldots, i_k$; in other words, it may be, in addition, a 'two-way classification' or a 'multiway classification'. In practice, the number of categories s_i of j for a given i will, in general, be independent of i, that is, the same for all i's; but it is better to consider a more general theoretical formulation.

This should be used when we have a mixed model with both 'variates' and 'ways of classification' (as e.g. in subsections 15.2.2, 15.3.2, 15.3.3 and 15.3.4).

Theorem VI. Under a hypergeometric ϕ_0(e.g. of the type (15.2.3)), as n_{i0}'s and n_{0j}'s$\to \infty$, with n_{i0}/n and n_{0j}/n held fixed, the sampling distribution of

$$\sum_{i,j} \left(n_{ij} - \frac{n_{i0}\, n_{0j}}{n}\right)^2 \Bigg/ \frac{n_{i0} n_{0j}}{n} \text{ tends to the } \chi^2\text{-distribution with d.f. } (r-1)(s-1).$$

The proof of theorem III that has been constructed is on the same lines as that of theorems I and II, and is rather long. Theorem IV can be thrown back upon either theorem II or III, without much difficulty. Theorem V is proved on the same lines as theorems I and II. More than one proof of theorem VI are available, but a more straight forward and rigorous proof can be constructed on the same lines as in Feller's book [12]. All this material on theorems III-VI will be given in a later monograph.

Going back to theorems III-V, we may note that in each case a part of what is included in the hypothesis might be moved over into the model and the rest retained under the hypothesis. This would mean that under the relevant non-null hypothesis (in connection with the power of the test) the part that comes under the hypothesis could be violated but not the part that has been moved over into the model. This will not affect the distribution on the null-hypothesis, i.e., the significance points, etc., but would of course change the structure of the power function. For example, going back to theorem III, we can, if we wish, make $p_i = p_i(\theta_1, \ldots, \theta_s)$ $(i = 1, 2, \ldots, r > s)$ a part of our model and define the null hypothesis as $f_k(\theta_1, \ldots, \theta_s) = f_k^0$ $(k = 1, 2, \ldots, t < s)$. Under this changed set-up even if we used the same test (for the null hypothesis) as discussed in connection with the original set-up of theorem III, we would have a power function which would be different from the one associated with the original set-up. But of course another test (which is not discussed here) with a greater power would be more appropriate in this situation. It is obvious that we can introduce similar changes in the original set-up of theorems IV and V.

It may be further observed in connection with theorems II—V, that each gives two main results listed as (i) and (ii). For example, if we denote the maximum likelihood estimate of the parameter point $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$, and the true parameter point by $\boldsymbol{\theta}^0$, theorem II gives that under H_0 and for large n, (i) $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}^0$ in probability and (ii) $\sum_{i=1}^{r} (n_i - n\hat{p}_i)^2/n\hat{p}_i$ tends to have the χ^2 distribution with appropriate degrees of freedom. Now result (i) will hold if we take for $\boldsymbol{\theta}$ any *BAN* [30] or best asymptotically normal estimate and not just the maximum likelihood estimate which itself is of course a *BAN* estimate. For example, the minimum χ^2 estimate would be one such and so also the minimum χ_1^2 estimate whose χ_1^2 is defined by $\sum_{i=1}^{r} (n_i - np_i)^2/n_i$. Next, result (ii) can be replaced by the result that under H_0 and large n,

$$\text{(iii)} \qquad \sum_{i=1}^{r} [n_s - np_i(BAN)]^2/n_i \text{ or } \sum_{i=1}^{r} [n_i - np_i(BAN)]^2/np_i(BAN)$$

each tends to have χ^2 distribution with the same degrees of freedom as for (ii). We here denote by $p_i(BAN)$ the value of p_i obtained by substituting in $p_i(\boldsymbol{\theta})$ any *BAN* estimate of $\boldsymbol{\theta}$. We have of course similar results associated with theorems III, IV and V. This gives us a good deal of latitude and leeway so far as large sample tests of the hypothesis associated with theorems II-V are concerned. In a sense this has been adequately proved in [30], but a proof which is more on the lines of the present development will be given in a later monograph.

15.6. *Large sample χ^2 tests of the null hypotheses in a two-way table.*

15.6.1. *The problem of section* 15.2.1. We consider section 15.2.1, start from (15.2.1.1), maximise log ϕ_0 with respect to p_{i0}'s and p_{0j}'s subject to $\sum_i p_{i0} = \sum_j p_{0j} = 1$ (using Lagrangian multipliers) and end up with the maximum likelihood solutions: $\hat{p}_{i0} : n_{i0}/n$ and $\hat{p}_{0j} = n_{0j}/n$. The number of independent parameters estimated from the data is $r+s-2$ and hence by section 15.5 the test of independence here is based on a statistic which has the χ^2-distribution with degrees of freedom $rs-1-(r+s-2) = (r-1)(s-1)$ and whose form is

$$\sum_{i,j} \frac{\left(n_{ij}-n\frac{n_{i0}}{n}\frac{n_{0j}}{n}\right)^2}{n \, . \, \frac{n_{i0}}{n}\frac{n_{0j}}{n}} = \sum_{i,j} \frac{\left(n_{ij}-\frac{n_{i0}\,n_{0j}}{n}\right)^2}{\frac{n_{i0}\,n_{0j}}{n}} \quad \ldots \quad (15.6.1.1)$$

15.6.2. *The problem of section* 15.2.2. We start from (15.2.2.1) and maximise $\log \phi_0$ with respect to q_{0j}'s subject to $\sum_j q_{0j} = 1$ and end up with the maximum likelihood solutions: $q_{0j} = n_{0j}/n$. The number of independent parameters estimated from the data is $s-1$ and hence by section 15.5 the test here is to be based on a statistic which has the χ^2-distribution with degrees of freedom $r(s-1)-(s-1) = (r-1)(s-1)$ and whose form is

$$\sum_i \sum_j \frac{\left(n_{ij}-n_{i0} \, . \, \frac{n_{0j}}{n}\right)^2}{n_{i0}\frac{n_{0j}}{n}} = \sum_{i,j} \frac{\left(n_{ij}-\frac{n_{i0}\,n_{0j}}{n}\right)}{\frac{n_{i0}n_{0j}}{n}} \quad \ldots \quad (15.6.2.1)$$

15.6.3. *The problem of section* 15.2.3. We start from (15.2.3). Here we have already (under the null hypothesis) $p_{ij} = n_{i0}n_{0j}/n^2$ and using the remarks of section 15.5 we note that the test is based on a statistic having the χ^2-distribution with degrees of freedom $rs-(r+s-1) = (r-1)(s-1)$ and the form

$$\sum_{i,j} \left(n_{ij}-\frac{n_{i0}\,n_{0j}}{n}\right)^2 \Big| \frac{n_{i0}\,n_{0j}}{n}. \quad \ldots \quad (15.6.3.1)$$

15.7. *Large sample χ^2 tests of the null hypotheses in a three-way table.*

15.7.1. *The problems of* 15.3.1, *i.e., where 'i', 'j' and 'k' are all 'variates'.*

15.7.1a. *The problems of* 15.3.1a.

Independence between 'i' and 'j' | *'k'.* Under H_0 of (15.3.1.1) we shall have

$$\phi_0 \sim \prod_{i,j,k} (p_{i0k}\, p_{0jk}/p_{00k})^{n_{ijk}}. \quad \ldots \quad (15.7.1.1)$$

To test the hypothesis here we maximise $\log \phi_0$ with respect to p_{i0k}'s, p_{0jk}'s and p_{00k}'s subject to $\sum_i p_{i0k} = \sum_j{}_{0jk} = p_{00k}$ and $\sum_k p_{00k} = 1$, and end up with the maximum likelihood solutions : $\hat{p}_{i0k} = n_{i0k}/n$, $\hat{p}_{0jk} = n_{0jk}/n$ *and* $\hat{p}_{00k} = n_{00k}/n$. The number of independent parameters estimated from the data is $(r-1)t+(s-1)t+(t-1)$. And hence by section 15.4.2 the test of conditional independence is here based on a statistic which has the χ^2-distribution with degrees of freedom $rst-1-t(r-1)-t(s-1)-(t-1) = t(r-1)(s-1)$ and whose form is

$$\sum_{i,j,k} \left(n_{ijk} - \frac{n_{i0k}n_{0jk}}{n_{00k}} \right)^2 \Big/ \frac{n_{i0k}n_{0jk}}{n_{00k}} . \qquad \ldots \quad (15.7.1.2)$$

Independence between 'i' and 'k' and also between 'j' and 'k'. This can be handled exactly on the lines of section 6 and will not be discussed separately.

Independence between 'i', 'j' and 'k'. To test this we start from the hypothesis of (15.3.1.3) giving

$$\phi_0 \sim \prod_{i,j,k} (p_{i00}\, p_{0j0}\, p_{00k})^{n_{ijk}}, \qquad \ldots \quad (15.7.1.3)$$

maximise log ϕ_0 with respect to p_{i00}'s, p_{0j0}'s and p_{00k}'s subject to $\sum_i p_{i00} = \sum_j p_{0j0} = \sum_k p_{00k} = 1$, and end up with the maximum likelihood solutions: $\hat{p}_{i00} = n_{i00}/n$, $\hat{p}_{0j0} = n_{0j0}/n$ and $\hat{p}_{00k} = n_{00k}/n$. The number of independent parameters estimated from the data is $(r+s+t-3)$ and hence by section 15.3.2 the test is here based on a statistic which has the χ^2-distribution with degrees of freedom $rst-1-(r+s+t-3) = rst-r-s-t+2$, and whose form is

$$\sum_{i,j,k} \left(n_{ijk} - \frac{n_{i00}n_{0j0}n_{00k}}{n^2} \right)^2 \Big/ \frac{n_{i00}n_{0j0}n_{00k}}{n^2} . \qquad \ldots \quad (15.7.1.4)$$

15.7.1b. *The problems of* 15.3.1b.

Independence between '(i. j.)' and 'k'. Under (15.3.1.4) we shall have

$$\phi_0 \sim \prod_{i,j,k} (p_{ij0}\, p_{00k})^{n_{ijk}} . \qquad \ldots \quad (15.7.1.5)$$

To test this hypothesis we maximise log ϕ_0 with respect to p_{ij0}'s and p_{00k}'s subject to $\sum_{i,j} p_{ijk} = \sum_k p_{00k} = 1$ and end up with the maximum likelihood solutions: $\hat{p}_{ij0} = n_{ij0}/n$ and $\hat{p}_{00k} = n_{00k}/n$. The number of independent parameters estimated from the data is $(rs-1)+(t-1)$ and hence by section 15.5 the test is based on a statistic having the χ^2-distribution with d.f. $rst-1-[(rs-1)+(t-1)] = (rs-1)(t-1)$ and having the form

$$\sum_{i,j,k} \left\{ n_{ijk} - \frac{n_{ij0}n_{00k}}{n} \right\}^2 \Big/ \frac{n_{ij0}\, n_{00k}}{n} . \qquad \ldots \quad (15.7.1.6)$$

Independence between 'i' and 'k' and between 'j' and 'k'. Since this can be handled on the same lines as in section 15.6, it will not be separately discussed.

The 'no interaction' hypothesis of (15.3.2.6). This has been discussed in detail in [47] and will not be given here. The test will be based on a statistic having the χ^2-distribution with d.f. $(r-1)(s-1)(t-1)$ and having a rather complicated form which will be reproduced in a later monograph.

15.7.2. *The problems of* 15.3.2, *i.e., when 'i' and 'j' are 'variates' and 'k' is a 'way of classification'.*

15.7.2a. *The problems of* 15.3.2a.

Independence between 'i' and 'j' for each k. Under H_0 of (15.3.2.1) we start from

$$\phi_0 \sim \prod_{i,j,k} (p_{i0k}\, p_{0jk})^{n_{ijk}}, \qquad \ldots \quad (15.7.2.1)$$

and maximise log ϕ_0 with respect to p_{ijk}'s and p_{0jk}'s subject to $\sum_i p_{i0k} = \sum_j p_{0jk} = p_{00k} = 1$, and end up with the maximum likelihood solutions: $\hat{p}_{00k} = n_{i0k}/n_{00k}$, $\hat{p}_{0jk} = n_{0jk}/n_{00k}$. The number of independent parameters to be estimated from the data is $t(r-1)+t(s-1)$ and hence by section 15.5 the test here is to be based on a statistic having the χ^2-distribution with d.f. $t(rs-1)-t(r-1)-t(s-1) = t(r-1)(s-1)$ and having the form

$$\sum_k \Big[\sum_{i,j} \Big\{(n_{ijk} - n_{00k} \cdot \frac{n_{i0k}\, n_{0jk}}{n_{00k}^2}\Big\}^2 \Big/ n_{00k} \cdot \frac{n_{i0k}\, n_{0jk}}{n_{00k}^2} \Big]. \qquad \ldots \quad (15.7.2.2)$$

The problems under (15.3.2.2) or (15.3.2.3) will not be discussed separately.

15.7.2b. *The problems of* 15.3.2b.

The hypothesis that p_{ijk} is independent of 'k', i.e., that p_{ijk} is a pure function of '(i, j)'. Under H_0 of (15.2.4) we start from

$$\phi_0 \sim \prod_{i,j,k} q_{ij0}^{n_{ijk}}, \qquad \ldots \quad (15.7.2.3)$$

maximise log ϕ_0 with respect to q_{ij0}'s subject to $\sum_{i,j} q_{ij0} = 1$, and end up with the maximum likelihood solutions: $\hat{q}_{ij0} = n_{ij0}/n$. The number of independent parameters to be estimated from the data is $(rs-1)$ and hence by section 15.5 the test is to be based on a statistic having the χ^2-distribution with d.f. $t(rs-1)-(rs-1) = (rs-1)(t-1)$ and having the form

$$\sum_k \Big[\sum_{i,j} \Big\{n_{ijk} - n_{00k} \cdot \frac{n_{ij0}}{n}\Big\}^2 \Big/ n_{00k} \cdot \frac{n_{ij0}}{n} \Big]. \qquad \ldots \quad (15.7.2.4)$$

The problems under (15.3.2.5) or (15.2.2.6) will not be separately discussed here.

15.7.3. *The problems of* 15.3.3, *i.e., when 'i' is a 'variate' and 'j' and 'k' are 'ways of classification'.*

15.7.3a. *The problem of* 15.3.3a.

The hypothesis that for any 'k', p_{ijk} is independent of 'j', i.e., that p_{ijk} is a pure function of 'i' and 'k'. Under H_0 of (15.3.3.1) we start from

$$\phi_0 \sim \prod_{i,j,k} q_{i0k}^{n_{ijk}}, \qquad \dots \quad (15.7.3.1)$$

and maximise log ϕ_0 with respect to q_{i0k}'s subject to $\sum_i q_{i0k} = 1$, and end up with the maximum likelihood solution $\hat{q}_{i0k} = n_{i0k}/n_{00k}$. The number of independent parameters to be estimated from the data is $t(r-1)$ and hence by section 15.5 the test is to be based on a statistic having the χ^2-distribution with d.f. $st(t-1)-t(r-1) = t(r-1)(s-1)$ and having the form

$$\sum_{i,k}\left[\sum_i \left\{n_{ijk}-n_{0jk}\cdot\frac{n_{i0k}}{n_{00k}}\right\}^2 \Big/ n_{0jk}\frac{n_{i0k}}{n_{00k}}\right]. \qquad \dots \quad (15.7.3.2)$$

15.7.3b. *The problem of* 15.3.3b. This will be exactly on the same lines as the previous case and will not be discussed separately. We shall also omit a discussion of the problem under (15.3.3.3).

15.7.4. *The problems of* 15.3.4, *i.e., when 'i' is a variate and 'j' and 'k' are ways of classification in the sense of an incomplete design.*

The hypothesis that p_{ijk} is independent of 'j'. i.e., that p_{ijk} is a pure function of 'i' and 'k'. We start from (15.3.4), put $p_{ijk} = q_{i0k}$ and thus have $\phi_0 \sim \prod_{i,j.k} q_{i0k}^{n_{i0k}}$, maximise log ϕ_0 with respect to q_{i0k}'s subject to $\sum_i q_{i0k} = 1$ and end up with a solution q_{i0k}'s in terms of n_{ijk}'s which is a set of functions of n_{ijk}'s of the same structure as the corresponding least squares solutions in linear estimation. One or two such solutions for some linked block designs will be discussed in a later paper. However, this solution, stuck into the 'χ^2' functions will have the χ^2-distribution with d.f. $(r-1)\sum_{j=1}^{s} t_j-(rt-t)$.

The hypothesis that p_{ijk} is independent of 'j', i.e., that p_{ijk} is a pure function of 'i' and 'j' can be handled on exactly similar lines and need not be separately considered.

15.8. *Linear hypothesis.* Linear hypothesis in the sense of chapter 12, on the p's or the logarithms of the p's, can be put forward, distinguishing as in chapter 12, between the model and the hypothesis, and such hypothesis can be tested either in terms of χ^2 or in terms of χ_1^2, in either case, substituting for the unknown free or nuisance parameters any *BAN* estimates and in particular, say the maximum likelihood or minimum χ^2 or minimum χ_1^2 estimate. There are theorems in this sector closely analogous to these least squares and analysis of variance theorems in the customary set-up, most of which have been considered in chapter 12, In terms of this it is possible to develop and study the analogoues of most of the things we customarily do in the usual uninormal or multinormal analysis of variance, including contrasts in general and 'main effects' and 'interactions', etc., in particular. If some numerical quantities or measures are attached to the categories we can also, in terms of such numerical measures, study the hypothesis of equality of the population 'means' or other linear hypothesis involving these 'means' or population 'variances' or other 'parameters' of the probability distributions. This will be discussed in a later monograph.

15.9. *Asymptotic independence of test criteria in certain situations.* In many situations in which a particular hypothesis H_0 with an associated χ^2 is the intersetion of several hypothesis H_{01}, H_{02}, etc., with associated χ_1^2, χ_2^2, etc., it so happens that $\chi^2=\chi_1^2+\chi_2^2+$etc. and that χ_1^2, χ_2^2, etc., are also independently distributed but unlike what happens in ordinary least squares analysis of variance set-up, the additivity is not in the usual algebraic sense; it is only in probability and asymptotically as $n \to \infty$ and the independence is also in the asymptotic sense. Take, for example, the hypothesis (15.3.1.1), (15.3.1.2) and (15.3.1.3), and let us call them H_{01}, (H_{02}, H_{03}) and H_0. We note that $H_0 = H_{01} \cap H_{02} \cap H_{03}$.

Let the associated χ^2's be denoted by χ_1^2, χ_2^2 and χ_3^2. Then, in this case, it has been shown [25], that, in large samples and under the null hypothesis H_0, λ_1^2, χ_2^2 and χ_3^2 are independent central χ^2's and $\chi_1^2+\lambda_2^2+\lambda_3^2 \to \chi^2$ in probability. We have an exactly similar situation for the group of hypotheses (15.3.1.4), (15.3.1.5) and (15.3.1.6). These are situations in multivariate analysis. There are similar situations in analysis of variance also, for example, with the group of hypothesis (15.3.2.4), (15.3.2.5) and (15.3.2.6) or with the group (15.3.3.1), (15.3.3.2) and (15.3.3.3). But this will not be true, for example, with a similar group of hypotheses on an incomplete block design or more general types of designs indicated in section 15.3.4. For linear hypotheses on p's or their logarithms, the mathematical conditions for this asymptotic independence and asymptotic additively in probability are strikingly similar to the corresponding conditions for the customary least squares set-up discussed in chapter 14. For more general types of hypotheses under more general types of models these conditions are a little more complicated with no obvious analogue in the usual least squares theory developed so far. All this will be discussed in a later monograph.

15.10. *On asymptotic power functions.* For analogous null hypotheses under different probability models we have, in many situations, eventually the same χ^2 with the same distribution under the respective null hypotheses. This is exemplified in sections 15.6.1, 15.6.2 and 15.6.3, also again in 15.7.1, 15.7.2b and 15.7.3a and so on. But the power, of these tests, which depend upon the distribution on the respective nonnull hypotheses of the corresponding statistics, are not comparable and in that sense different. It is well known [30] that these tests are consistent i.e., that in large samples these powers tend to 1 in each case. But the asymptotic powers in the sense of Pitman and Lehmann can be obtained and compared. The asymptotic power function (for analogous hypotheses under different probability models) have different structures. Some results are given in the following paragraphs. A systematic development including proofs will be given in a later monograph.

15.10.1. *The asymptotic power function connected with theorem I of section* 15.5. Let us consider an alternative H_n (depending on n) given by

$$H_n : p_{in} = p_i^0 + \frac{\delta_i}{\sqrt{n}} (i = 1, 2, \dots, r;\ \sum_i p_{in} = \sum_i p_i^0 = 1), \dots \quad (15.10.1.1)$$

where δ's are fixed. Then, as $n \to \infty$, under H_n, $\chi^2 = \sum_{i=1} (n_i - np_i^0)^2/np_i^0$ tends to have the noncentral χ^2-distribution with d.f. $(r-1)$ and a noncentrality parameter

$$\Delta = \sum_{i=1}^{r} \delta_i^2/p_i^0.$$

15.10.2. *The asymptotic power function connected with theorem II of section* 15.5. Suppose we have an alternative (which has to be simple in this case) H_n given by

$$H_n : p_{in} = p_i(\theta_1^0, \theta_2^0, \ldots, \theta_s^0) + \frac{\delta_i}{\sqrt{n}} (i = 1, \ldots, r; \sum_i p_{in} = \sum_i p_i^0 = 1) \ldots \quad (15.10.2.1)$$

Then, as $n \to \infty$, under H_n

$$\chi^2 = \sum_{i=1}^{r} [n_i - np_i(\hat{\theta}_1, \ldots, \hat{\theta}_s)]^2/np_i(\hat{\theta}_1, \ldots, \hat{\theta}_s)$$

tends to have the noncentral χ^2-distribution with d.f. $(r-1)-s$ and a non-centrality parameter $\Delta = \boldsymbol{\delta}'[I - B(B'\,B)^{-1}\,B']\boldsymbol{\delta}$, where $\boldsymbol{\delta}'(1 \times r)$ is a row vector with elements $\delta_i/\sqrt{p_i(\theta_1^0, \ldots, \theta_s^0)}$ $(i = 1, 2, \ldots, r)$ and B stands for the $r \times s$ matrix $\left\{ \frac{\partial p_i}{\delta \theta_j} \right\}$ $(i = 1, 2, \ldots, r; j = 1, 2, \ldots, s)$.

15.10.3. *The asymptotic power function connected with case* (ii) *of theorem V of section* 15.5. Consider an H_n given by

$$H_n : p_{ijn} = p_{ij}(\theta_1^0, \ldots, \theta_u^0) + \frac{\delta_{ij}}{\sqrt{n}} \text{ with } u < s-r \text{ and } s = \sum_{i=1}^{r} s_i \text{ and also of course}$$

$$\sum_j p_{ijn} = p_{i0n} = 1, \quad i = 1, \ldots, r. \qquad \ldots \quad (15.10.3.1)$$

Then under H_n and as $n \to \infty$ subject to $q_i = n_{i0}/n$ being held constant with $i = 1, 2, \ldots, r$,

$$\chi^2 = \sum_{i,j} [n_{ij} - n_{i0}p_{ij}(\hat{\theta}_1, \ldots, \hat{\theta}_u)]^2/n_{i0}p_{ij}(\hat{\theta}_1, \ldots, \hat{\theta}_u).$$

will tend to have the noncentral χ^2-distribution with the same d.f. as indicated there and with a noncentrality parameter $\Delta = \boldsymbol{\delta}[I - B(B'B)^{-1}B']\boldsymbol{\delta}$, where

$$B(s \times u) = \left\{ \frac{\sqrt{q_i}}{\sqrt{p_{ij}(\theta_1^0, \ldots, \theta_u^0)}} \left[\frac{\partial p_{ij}}{\partial \theta_k}\right]_0 \right\} \text{ with } j = 1, 2, \ldots, s_i; \quad i = 1, 2, \ldots, r \text{ and}$$

$k = 1, 2, \ldots, u$ and where $\boldsymbol{\delta}'(1 \times s)$ is a row vector with elements $\delta_{ij} \ \sqrt{q_i}/\sqrt{p_{ij}(\theta_1^0, \ldots, \theta_u^0)}$.

Another way to compare the relative efficiency of two comparable and consistent tests in a particular situation would be to consider the ratio of the exact probability of the second kind of error for the two tests and study the limiting form of this ratio as $n \to \infty$. This also will be discussed in a later monograph.

15.11. *Remarks on more general decision problems.* The problems discussed in this monograph, whether based on the 'normality' assumption as in the previous chapters or on the nonparametric model as in this chapter, have been either in terms of the Neyman-Pearson testing of hypothesis or, in several cases, in terms of confidence bounds on meaningful sets of parametric functions which might be regarded as natural measures of departure from certain null hypotheses. In many situations, however, it is of considerably more physical interest to consider more general decision problems. For example, in the analysis of variance situations with say t treatments (whether on the 'normal' assumption or on the nonparametric model) we may be likely to be far more interested on a decision rule for picking out the 'best' or ranking the t treatments in terms of some characteristic. The decision rule has to have certain desirable (if not always optimum) properties in terms of some rational criteria. Some such decision rules already developed, both on 'normal' variate data and on 'categorical' data, and on various types of problems including those of factor analysis and classification will be discussed in a later monograph.

15.12. *Some remarks on factorial experiments.* Looking for a possible motivation behind the customary (and mostly 'normal' variate) analysis, one can not help feeling that factorial experiments (whether on the 'normal' variate type of data or more general types of data) present a problem which is essentially different from that of the rest of analysis of variance, e.g., the usual tests of significance of treatment differences. Assume, for simplicity, in the beginning, that there is just one factor at, say, k levels. One might regard these as treatments, and test whether there are significant differences between these, or, in terms of some characteristic, pick out the 'best' among these or rank these in some order. But that does not appear to be the relevant question here. The (second) characteristic in terms of which we have the levels seems to be a continuous variate which is observed at k levels for practical convenience, and what is of interest seems to be to lay down a statistical rule by which we can, in terms of the observations at discrete levels, decide about the 'best' or 'optimum' point, the 'best' or 'optimum' being in relation to the first characteristic. Likewise, taking for example, two factors at k and l levels respectively the problem seems to be not to test whether there are significant differences between these kl combinations regarded as treatments (which would, really, be a linear problem) or to pick out the 'best' among these or to rank these (in terms of some characteristic), which again would be each a really linear problem. It seems that there is a (second) characteristic in terms of which we have the k levels of the first factor, and a third characteristic in terms of which we have the l levels of the second factor, both these (second and third) characteristics being supposed to be continuous variates. The problem is to lay down a statistical rule by which we can decide about the 'best' or 'optimum' point (in relation to the first characteristic) on the plane of the second and third characteristics, regarded as two continuous variates, the decision rule being in terms of observations at the kl discrete level combinations. This of course can be generalized to several factors. The customary analysis into 'main-effects', 'interactions' of various orders, confounding etc., all seem to point very strongly in this direction. Some work has already been done from this standpoint and further work is under way. This will be discussed in a later monograph.

APPENDIX 1

Some Preliminary Results in Matrix Theory

(A.1.1): *Given four matrices* $A(p \times p)$, $B(p \times q)$, $C(q \times p)$ *and* $D(q \times q)$, *if* D *is non-singular, then*

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D| \; |A - BD^{-1}C|$$

Proof:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} A & B \\ C & D \end{vmatrix} \begin{vmatrix} I(p) & O(p \times q) \\ -D^{-1}C & I(q) \end{vmatrix} = \begin{vmatrix} A - BD^{-1}C & B \\ O & D \end{vmatrix} = |D| \, |A - BD^{-1}C|.$$

(A.1.2): $r[A(p \times q)B(q \times s)] \leqslant min\,[r(A),\; r(B)]$, *where* $min\; (x,\, y)$ *denotes the lesser of two real numbers* x *and* y.

(A.1.3): $r[A(p \times q)] = r[B(p \times p)A(p \times q)] = r[A(p \times q)C(q \times q)]$,

if B *and* C *are non-singular.*

(A.1.4): $r[A(p \times q)] = r[A'(q \times p)] = r[A(p \times q)A'(q \times p)]$.

(A.1.5): $tr[A(p \times q)B(q \times p)] = tr[B(q \times p)A(p \times q)]$.

Proof: If $A = (a_{ij})$ and $B = (b_{ij})$, then by the definition of trace we have $\mathrm{tr}(AB) = \sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij}b_{ji} = \sum_{j=1}^{q} \sum_{i=1}^{p} b_{ji}\,a_{ij} = \mathrm{tr}(BA)$.

(A.1.6): $r[A(p \times q)] = r[A(p \times q)B(q \times t)] = r[C(s \times p)A(p \times q)]$,

if $q \leqslant t$, $p \leqslant s$ *and* B *and* C *are respectively of ranks* q *and* p.

Proof: Using (A.1.2)-(A.1.3) we have

$$r[A(p \times q)] = r[A(p \times q)B(q \times t)B'(t \times q)] \leqslant \min\,[r(AB),\, r(B')],$$

i.e., $\leqslant r(AB)$. But $r(AB) \leqslant r(A)$, whence $r(A) = r(AB)$. Likewise, starting with $A'C'$ and noting that $r(CA) = r(A'C')$, we should have, in an exactly similar manner, $r(CA) = r(A)$, which completes the proof of (A.1.6).

(A.1.7): *If* $L_1(p \times n)$ $(p < n)$ *is subject to* $L_1L_1' = I(p)$, *there exists an* $L_2(\overline{n-p} \times n)$ *such that* $\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$ *is* $\perp$. L_2 *will be called an arbitrary completion of* L_1.

(A.1.8): *If* $M(p \times p)$ *is symmetric and at least p.s.d. of rank* $r(\leqslant p)$, *then, out of the* p $c(M)$*'s* (i.e., *roots of the determinantal equation in* c: $|M - cI| = 0$), r *are* *positive and the rest, $p-r$ in number, are zero. If $r = p$, the number of non-zero roots will of course be p.*

(A.1.9): *If $M_1(p\times p)$ is symmetric and at least p.s.d. of rank $r(\leqslant p)$ and $M_2(p\times p)$ is symmetric and p.d., there are exactly r positive roots of the following equation in c: $|M_1-cM_2| = 0$, the rest, $p-r$ in number, being 0. If $r = p$, the number of positive roots will of course be p.*

(A.1.10): *$X(p\times n)\, X'(n\times p)$ will be symmetric and at least p.s.d. of the same rank as X or X', the common rank r being $\leqslant$ min (p, n), where the symbol (which will be frequently used later) denotes the lesser of p and n. It is easy to see that if $p \leqslant n$ and X is of rank p, then XX' is p.d.*

(A.1.11): *If $A(q\times q)$ is symmetric p.d., $B(p\times q)\, A(q\times q)\, B'(q\times p)$ is symmetric and at least p.s.d. of the same rank as B.*

Proof: Since A is symmetric p.d., there exists, by (A.3.9), a non-singular $\tilde{T}(q\times q)$ such that $A = \tilde{T}\tilde{T}'$. Hence $BAB' = (B\tilde{T})(B\tilde{T})'$, which, by (A.1.10), is symmetric and at least p.s.d. of the same rank as $B\tilde{T}$. But $B\tilde{T}$ is of the same rank as B, since $\tilde{T}$ is non-singular, whence the theorem follows.

(A.1.12): *If $A(p\times p)$ is symmetric and at least p.s.d. of rank $r \leqslant p$ and $B(p\times p)$ is non-singular, BAB' is symmetric and at least p.s.d. of rank r.*

Proof: If A is symmetric and at least p.s.d. of rank r, then by (A.3.10), there exist a non-singular $\tilde{T}_1(r\times r)$ and a $T_2(\overline{p-r}\times r)$ such that without any loss of generality we can put $A = \begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix} [\tilde{T}'_1 : T'_2]$. Therefore, $BAB' = B\left(\begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix}\right) B\left(\begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix}\right)'$ which, by (A.1.10), is symmetric and at least p.s.d. of the same rank as $B\begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix}$. But, since B is non-singular and $\begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix}$ is obviously of rank r, therefore, $B\begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix}$ is of rank r and thus BAB' is of rank r.

(A.1.13): *If $M_1(p\times p)$ is symmetric and at least p.s.d. of rank r ($\leqslant p$) and $M_2(p\times p)$ is symmetric p.d., then (i) all the roots of the equation in c: $|M_1-cM_2| = 0$ are zero if and only if $M_1 = 0$, and (ii) all the roots are unity if and only if $M_1 = M_2$.*

Proof: Part (i) of (A.1.13) is a direct consequence of (A.1.9). To prove part (ii), put $c = 1-e$. We have then the equation in e: $|(M_1-M_2)+eM_2| = 0$ whence it follows that all roots of the equation are zero, (i.e., all roots of the equation in c are unity), if and only if $M_1-M_2 = 0$, i.e., $M_1 = M_2$ which proves part (ii) of (A.1.13).

(A.1.14): *If M is a $(p+q)\times(p+q)$ symmetric matrix shown as, say,*

$$M = \begin{array}{cc} \left[\begin{array}{cc} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \;\; q \end{array} & \end{array},$$

and if M_{22} is non-singular, then (i) $M_{11}-M_{12}M_{22}^{-1}M'_{12}$ is symmetric, and (ii) of rank $r-q$, where q is (evidently) the rank of M_{22}, and r denotes the rank of M (evidently satisfying $q \leqslant r \leqslant p+q$).

Proof: Part (i) is obvious if we remember that M_{11}, M_{22} (and thus M_{22}^{-1}) are symmetric and so also $M_{12}M_{22}^{-1}M'_{12}$. For part (ii) we first observe that the rank of M would be unaltered if it were pre-multiplied and/or post-multiplied by two conformable non-singular matrices. Post-multiply M by the conformable non-singular matrix (of rank $p+q$):

$$\begin{array}{cc} \left[\begin{array}{cc} I & O \\ -M_{22}^{-1}M'_{12} & I \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}$$

and premultiply by the transpose of this matrix. Then we have:

$$\text{rank of } M = \text{rank of } \begin{bmatrix} I(p) & -M_{12}M_{22}^{-1} \\ O & I(q) \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{bmatrix} \begin{bmatrix} I(p) & O \\ -M_{22}^{-1}M'_{12} & I(q) \end{bmatrix}$$

i.e., rank of $\begin{bmatrix} M_{11}-M_{12}M_{22}^{-1}M'_{12} & O \\ O & M_{22} \end{bmatrix}$. But the rank of this last martix is evidently the same as that of M_{22} (which is q) plus that of $(M_{11}-M_{12}M_{22}^{-1}M'_{12})$. This proves part (ii) of (A.1.14).

(A.1.15): *If M has the same structure as in* (A.1.14) *and is, in addition, at least p.s.d. of rank* r $(q \leqslant r \leqslant p+q)$, *then* $M_{11}-M_{12}M_{22}^{-1}M'_{12}$ *is also at least p.s.d. of* rank $r-q$.

Proof: Since M is symmetric and at least p.s.d. of rank $r(q \leqslant r \leqslant p+q)$, premultiplying and post-multiplying it by the same conformable non-singular matrices as in the proof of (A.1.14) and using next (A.1.12), we observe that

$$\begin{array}{cc} \left[\begin{array}{cc} M_{11}-M_{12}M_{22}^{-1}M'_{12} & O \\ O & M_{22} \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} \qquad p & \qquad\quad q \end{array} & \end{array}$$

is at least p.s.d. of rank r. Hence $M_{11}-M_{12}M_{22}^{-1}M'_{12}$ is evidently at least p.s.d. and since (A.1.14) shows that it is of rank $r-q$, the theorem (A.1.15) follows.

(A.1.16): *If* $M = \begin{array}{cc} \left[\begin{array}{cc} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}$ *is symmetric and at least p.s.d. of rank* r $(q \leqslant r \leqslant p+q)$, *and if* $p \leqslant q$ *and* M_{11} *and* M_{22} *are both non-singular* (*i.e., in this situation both p.d. of ranks* p *and* q *respectively*) *and if* s *denotes the rank of* $M_{12}(p \times q)$ (evidently $s \leqslant p \leqslant q$) *then the* p *roots of the* p*-th degree equation in* c: $|cM_{11}-M_{12}M_{22}^{-1}M'_{12}| = 0$ *have the following properties*; (*i*) $0 \leqslant$ *all* c's $\leqslant 1$, (*ii*) *out of the* p c's, $r-q$ *are* $\neq 1$ *and the rest,* $p-(r-q)$ $(= p+q-r < p)$ *in number, are* 1; (*iii*) *also out of the* p c's, $s(\leqslant p)$ *are* $\neq 0$ *and the rest,* $p-s(\leqslant p)$ *are* $= 0$.

Proof: We note that since M_{22} and hence M_{22}^{-1} is p.d. of rank q and M_{12} is of rank s, therefore, from (A.1.11), $M_{12}\, M_{22}^{-1}\, M_{12}'$ is symmetric and at least p.s.d. of rank s ($\leqslant p \leqslant q$). Also $M_{11}(p \times p)$ is supposed to be p.d.. Hence by (A.1.9), out of the p roots of the equation in c: $|cM_{11}-M_{12}M_{22}^{-1}M_{12}'| = 0$, s are>0 and the rest, $p-s$ in number, are $= 0$.

Next, putting $c = 1-e$, we have the equation in e: $|eM_{11}-(M_{11}-M_{12}M_{22}^{-1}M_{12}')| = 0$. But M_{11} is symmetric p.d. and, by (A.1.14) and (A.1.15), $M_{11}-M_{12}M_{22}^{-1}M_{12}'$ is symmetric and at least p.s.d. of rank $r-q(\leqslant p)$. Hence, out of the p roots of the equation in e, $r-q$ are < 0 and the rest, $p-(r-q)$ ($= p+q-r \leqslant p$) in number, are $= 0$. Since $c = 1-e$, this means that, out of the p roots of the equations in c, $r-q$ are < 1 and the rest, $p+q-r$ in number, are $= 1$. This completes the proof of (A.1.16).

(A.1.17): *With the same set-up as in* (A.1.16), *the roots of the equation in* c: $|cM_{11}-M_{12}M_{22}^{-1}M_{12}'| = 0$ *are all zero if and only if the rank of* M_{12} *is zero, i.e.,* M_{12} *is the null matrix.*

This is a direct consequence of (A.1.16). With regard to theorems (A.1.16) and (A.1.17) we observe that in statistical applications we shall always be considering the special case, $r = p+q$, that is, the case where M is symmetric p.d.. In this situation we state and prove two theorems on transformations, (A.3.16) and (A.3.17).

(A.1.18): *Every non-zero characteristic root of* $A(p \times q)\ B(q \times p)$ *is a (non-zero) characteristic root of* $B(q \times p)\ A(p \times q)$ *and vice versa.*

Proof: If c is any (non-zero) characteristic root of AB, we have by definition, $|AB-cI| = 0$ or, by using (A.1.1),

(A.1.18.1)
$$\begin{array}{|cc|c} cI & A & p \\ B & I & q \\ \multicolumn{1}{c}{p} & \multicolumn{1}{c}{q} & \end{array} = 0.$$

Since c is non-zero we can obviously rewrite this as

(A.1.18.2)
$$\begin{array}{|cc|c} cI & B & q \\ A & I & p \\ \multicolumn{1}{c}{q} & \multicolumn{1}{c}{p} & \end{array} = 0, \quad \text{or},$$

(A.1.18.3) $|BA-cI| = 0$, which proves (A.1.18).

There is, in fact, a stronger result than (A.1.18), namely that, not only is every nonzero characteristic root of AB a root of BA and vice versa, but that each such root has the same multiplicity in relation to both matrices AB and BA. This follows if we notice that the left sides of (A.1.18.1) and (A.1.18.2) are the characteristic functions of AB and BA and then relate AB to BA by using (A.1.1).

(A.1.19): (*i*) *If* $B(p \times p)$ *is non-singular, the roots of the equation in* c: $|A(p \times p)-cB(p \times p)| = 0$ *are the same as the characteristic roots of* AB^{-1} *or of* $B^{-1}A$; *and* (*ii*) *in* (A.1.16) *the roots of the equation in* c: $|cM_{11}-M_{12}\, M_{22}^{-1}M_{12}'\,| = 0$ *are the same as the characteristic roots of* $M_{11}^{-1}\, M_{12}\, M_{22}^{-1}\, M_{12}'$ *or of* $M_{12}M_{22}^{-1}M_{12}'\, M_{11}^{-1}$ (*with the exception of zero roots in the case where* $p < q$). The proof is obvious.

(A.1.20): *$tr_t[A(p\times p)] = \sum_{i_1\neq i_2\neq\ldots\neq i_t=1}^{p} c_{i_1}(A)c_{i_2}(A)\ldots c_{i_t}(A)$, where $tr_t\,A$ stands for the sum of all $t\times t$ minors (found by the intersection of any t rows of A with t columns bearing the same number), and, in particular,*

$$tr_1 A = \sum_{i=1}^{p} c_i = \sum_{i=1}^{p} a_{ii} \text{ and } tr_p\, A = \prod_{i=1}^{p} c_i = |A|.$$

(A.1.18) coupled with (A.1.20) supplies another proof of the relation:

$$\text{tr }(AB) = \text{tr }(BA) \quad (\text{see (A.1.5)}).$$

(A.1.21): *If* (a) *$d_1 \leqslant$ all $c(AB^{-1}) \leqslant d_2(d_2 > 0)$, then* (b) *$(d_1)^t\, tr_t(B) \leqslant tr_t(A) \leqslant (d_2)^t\, tr_t(B)$ $(t = 1, 2, \ldots, p)$, where A and B are two $p\times p$ p.d. matrices.* Notice that (b) is a necessary (though not a sufficient) condition for (a).

Proof: It is easy to check that "$d_1 \leqslant$ all $c(AB^{-1})$" $\Longleftrightarrow$ $(A-d_1B)$ is p.d. $\Longleftrightarrow$ $(A_t - d_1B_t)$ $(t = 1, \ldots, p)$ is p.d. (where $A_t-d_1B_t$ is a submatrix formed by the intersection of any t rows of $(A-d_1B)$ with t columns bearing the same numbers) $\Longleftrightarrow$ $d_1 <$ all $c(A_tB_t^{-1})$ $(t = 1, \ldots, p)$. Now, if all $c(A_tB_t^{-1}) > d_1$, one consequence is that

(A.1.21.1) $$\prod_{i=1}^{t} c_i(A_tB_t^{-1}) > (d_1)^t, \text{ i.e., } |A_t| \;/\; |\,B_t\,| > (d_1)^t.$$

For a given t, summing over different possible submatrices, we have

(A.1.21.2) $$\text{tr}_t A > (d_1)^t \text{ tr}_t B.$$

Using the same kind of argument for the other half of the inequality and remembering that $t = 1, 2, \ldots, p$, and combining, we have the following result.

(A.1.21.3) If $$d_1 < \text{all } c(AB^{-1}) < d_2, \text{ then } (d_1)^t \text{ tr}_t(B) \leqslant \text{tr}_t(A) \leqslant (d_2)^t \text{tr}_t(B) \quad (t = 1, 2, \ldots, p).$$

By a slight rephrasing (which is obviously permissible here) we have the result (A.1.21).

(A.1.22): *If $A(p\times p)$ is symmetric p.d. and $B(p\times p)$ is symmetric and at least p.s.d., then (i) all $c(AB)$ are non-negative and (ii) $\underset{min}{c(A)}\underset{min}{c(B)} \leqslant$ all $c(AB) \leqslant \underset{max}{c(A)}\;\underset{max}{c(B)}$, where $\underset{min}{c(M)}$ and $\underset{max}{c(M)}$ stand respectively for the largest and smallest roots (both non-negative) of any M which is symmetric and at least p.s.d.* [45]

Proof: By (A.3.3) there are $\perp$ matrices L_A $(p\times p)$ and L_B $(p\times p)$ such that $A = L_AD_{c(A)}L'_A$ and $B = L_B\, D_{c(B)}L'_B$, and thus $AB = L_AD_{c(A)}L'_A\; L_BD_{c(B)}\; L'_B$.

Now using (A.1.18) (and noting that here $p = q$, so that all characteristic roots are the same in both products), we have the two-way relation

$$c(AB) = c(D_{c(A)}L'_AL_BD_{c(B)}\; L'_BL_A) = c(D_{c(A)}MD_{c(B)}M'),$$

where M stands for $L_A\; L_B$. Notice that $MM' = L'_AL_BL'_BL_A = L'_AL_A = I(p)$ (since L_A and L_B are each $\perp$), so that M itself is $\perp$. Also note that $D_{c(A)}M$ is non-singular since M, being $\perp$, is non-singular, and $D_{c(A)}$ is non-singular, because all the $c(A)$'s are positive.

Using (A.1.18) again we find that $c(AB) = c(D_{\sqrt{c}(A)} M D_{c(B)} \quad 'D_{\sqrt{\ }(A)})$, and since $D_{c(B)}$ is obviously symmetric p.s.d. by virtue of B being p.s.d., we notice by using (A.1.11) that $D_{\sqrt{c}(A)} M D_{c(B)} M' D_{\sqrt{c}(A)}$ is symmetric and at least p.s.d., and thus all $c(AB)$ are non-negative. This proves part (i). For part (ii), let us go back to $D_{c(A)} M D_{c(B)} M'$, denote by λ_i and μ_j the characteristic roots of A and B, observe that here all $\lambda_i > 0$ and all $\mu_j \geqslant 0$, and next observe that, if c is to be a characteristic root of AB (here all roots are non-negative), there exists a set of (real) numbers $x_1, x_2, \ldots, x_p$, not all of which are zero, such that the following set of equations are satisfied.

$$\text{(A.1.22.1)} \qquad \sum_{j,k=1}^{p} \lambda_i m_{ij} \mu_j m_{kj} x_k = c x_i \qquad (i = 1, 2, \ldots, p) \text{ (notice that } (M')_{jk} = (M)_{kj}).$$

Remembering that here $\lambda_i > 0$ and $\mu_j \geqslant 0$ (both sets being real), dividing by λ_i, and squaring any member of (A.1.22.1) and summing over $i = 1, 2, \ldots, p$, we have

$$\text{(A.1.22.2)} \qquad \sum_i \sum_{j,j',k,k'} m_{ij} m_{ij'} \mu_j \mu_{j'} m_{kj} m_{k'j'} x_k x_{k'} = c^2 \sum_{i=1}^{p} x_i^2/\lambda_i^2 .$$

Now, since M is $\perp$, we have $\sum_i m_{ij} m_{ij'} = \delta_{jj'}$ (where δ is the Kronecker symbol), so that (A.1.22.2) reduces to

$$\text{(A.1.22.3)} \qquad c^2 \sum_i x_i^2/\lambda_i^2 = \sum_{j,k,k'} \mu_j^2 m_{kj} m_{k'j} x_k x_{k'} .$$

It is easy to check that the coefficients of λ_i^2 on the left hand side and those of μ_j^2 on the right hand side are each non-negative. Hence, if we replace all μ_j's by $\mu_{\max}$ and all λ_i's by $\lambda_{\max}$, the right hand side is increased (or at least not diminished) and the left hand side is diminished (or at least not increased). We have thus

$$\text{(A.1.22.4)} \qquad (c^2/\lambda_{\max}^2) \sum_i x_i^2 \leqslant \mu_{\max}^2 \sum_j \sum_{k,k'} m_{kj}\, m_{k'j}\, x_k\, x_{k'}, \quad \text{i.e.,} \quad \leqslant \mu_{\max}^2 \sum_j \delta_{kk'} x_k\, x_{k'}$$

(since M is $\perp$), i.e., $\leqslant \mu_{\max}^2 \sum_i x_i^2$. Since $\sum_i x_i^2$ is positive, it follows that $c^2 \leqslant \lambda_{\max}^2 \mu_{\max}^2$ i.e., $c \leqslant \lambda_{\max}\ \mu_{\max}$ (taking the positive square root on both sides). Thus we have

$$\text{(A.1.22.5)} \qquad \text{all } c(AB) \leqslant c_{\max}(A)\ c_{\max}(B).$$

Likewise in (A.1.22.3), replacing all λ_i's by $\lambda_{\min}$ and all μ_j's by $\mu_{\min}$ and arguing in a similar manner, we have

$$\text{(A.1.22.6)} \qquad c_{\min}(A)\ c_{\min}(B) \leqslant \text{all } c(AB).$$

Combining (A.1.22.5) and (A.1.22.6) we have part (ii) of (A.1.22).

Replacing A by a complex non-singular A, B by any complex B, remembering that AA^* is hermitian p.d. and BB^* is hermitian and at least p.s.d. we have the following more general theorem, proved elsewhere [45]:

(A.1.23): $c_{min}(AA^*)\, c_{min}(BB^*) \leqslant$ *all* $c(AB){\cdot}\bar{c}(AB) \leqslant c_{max}(AA^*)\, c_{max}(BB^*)$.

However, this result will not be needed in the present monograph, although a special case will be needed.

Put $B = I$ and let A be a real matrix with real roots. If A is real symmetric this will be true but this might be true even if A were real but not symmetric. We can now put $A^* = A'$ and have, as a special case of (A.1.23), the following:

(A.1.24): $c_{min}(AA') \leqslant$ *all* $c^2(A) \leqslant c_{max}(AA')$.

The following matrix lemma is also repeatedly used in the text:

(A.1.25): $c_{min}(AB^{-1})\, c_{min}(BC) \leqslant$ *all* $c(AC) \leqslant c_{max}(AB^{-1})\, c_{max}(BC)$,

where A, C and B(and hence B^{-1}) are real symmetric positive definite matrices of order p each.

Proof: Using (A.3.9), put $B = \tilde{T}\tilde{T}'$. We have now,

(A.1.25.1) $c_{\max}\ (AB^{-1})\ c_{\max}\ (BC) = c_{\max}\ (A\ \tilde{T}'^{-1}\tilde{T}^{-1})\ c_{\max}\ (\tilde{T}\ \tilde{T}'C) = c_{\max}\ (\tilde{T}^{-1}A\tilde{T}'^{-1}) c_{\max}\ (\tilde{T}'C\,\tilde{T})$, using (A.1.18), $\geqslant c_{\max}\ (\tilde{T}^{-1}AC\tilde{T})$, using (A.1.22) and (A.1.12), that is $\geqslant c_{\max}\ (AC)$, using (A.1.18).

The other side of the inequality in (A.1.25) follows in a similar fashion and completes the proof of (A.1.25).

APPENDIX 2

Some Results in Quadratic Forms

(A.2.1): *If* $A(p\times p)$ *is symmetric and at least p.s.d. of rank* $r(\leqslant p)$, *then* $(i)\mathbf{a}'(1\times p)\ A(p\times p)\ \mathbf{a}(p\times 1)$ *is at least a p.s.d. quadratic form in* a_i*'s* $(i = 1, ..., p)$, (ii) *the stationary values of* $\mathbf{a}'A\mathbf{a}/\mathbf{a}'\mathbf{a}$ (*under variation of* $\mathbf{a}$ *over all non-null* $\mathbf{a}$*'s*) *are the characteristic roots of* A (*all non-negative*) *and* (iii) *in particular, the largest and smallest values of* $\mathbf{a}'A\mathbf{a}/\mathbf{a}'\mathbf{a}$ (*under variation of* $\mathbf{a}$) *are the largest and smallest characteristic roots of* A.

Proof: Part (i) is given in all textbooks and need not be proved. For part (ii) putting $\mathbf{a}'A\mathbf{a}/\mathbf{a}'\mathbf{a} = c$ and differentiating c with respect to the elements of $\mathbf{a}$, we have the vector equation giving the stationary values of c : $A\mathbf{a}-c\mathbf{a} = \mathbf{0}$, whence by eliminating $\mathbf{a}$ we have, for the stationary values of c, the p-th degree determinantal equation in c: $|A-cI| = 0$. The roots of this are the so-called characteristic roots of A, which proves part (ii). In this case the proof of part (iii) is obvious and will not be separately discussed.

(A.2.2): *If* $B(p\times p)$ *is symmetric p.d. and* $A(p\times p)$ *is symmetric and at least p.s.d. of rank* $r(\leqslant p)$, *then for all non-null* $\mathbf{a}$*'s* (i) $\mathbf{a}'(1\times p)\,A(p\times p)\mathbf{a}(p\times 1)/\mathbf{a}'(1\times p)\,B(p\times p)\ \mathbf{a}(p\times 1)$ *is non-negative,* (ii) *the stationary values of* $\mathbf{a}'A\mathbf{a}/\mathbf{a}'B\mathbf{a}$ (*under variation of* $\mathbf{a}$) *are the roots of the determinantal equation in* c: $|A-cB| = 0$ *and* (iii) *in particular, the largest and smallest values of* $\mathbf{a}'A\mathbf{a}/\mathbf{a}'B\mathbf{a}$ *are the largest and smallest roots of the determinantal equation.*

Proof: Part (i) is obvious. For part (ii), putting $\mathbf{a}'A\mathbf{a}/\mathbf{a}'B\mathbf{a} = c$ and differentiating c with respect to the elements of $\mathbf{a}$, we have the vector equation giving the stationary values of c: $A\mathbf{a}-cB\mathbf{a} = \mathbf{0}$, whence by eliminating $\mathbf{a}$ we have, for the stationary values of c, the p-th degree determinantal equation in c: $|A-cB| = 0$, which proves part (ii). The proof of part (iii) is now obvious.

(A.2.3): *If* $M = \begin{array}{cc} \left[\begin{array}{cc} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}$ $(p\leqslant q)$ *is symmetric p.d.* (*from which it follows easily that* M_{11} *and* M_{22} *are each symmetric p.d.*), *then, for all non-null* $\mathbf{a}_1(p\times 1)$ *and* $\mathbf{a}_2(q\times 1)$, $(i)[\mathbf{a}_1'M_{12}\mathbf{a}_2]^2/(\mathbf{a}_1'M_{11}\mathbf{a}_1)(\mathbf{a}_2'M_{22}\mathbf{a}_2)$ *is non-negative,* (ii) *the stationary values of this expression are the roots of the equation in* c: $|cM_{11}-M_{12}M_{22}^{-1}M'_{12}|=0$ *and* (iii) *in particular, the largest and smallest values of the expression are the largest and smallest roots of the determinantal equation.*

Proof: Part (i) is obvious. For part (ii), putting $\mathbf{a}_1'M_{12}\mathbf{a}_2=a_{12}$, $\mathbf{a}_1'M_{11}\mathbf{a}_1=a_{11}$ and $\mathbf{a}_2'M_{22}\mathbf{a}_2 = a_{22}$, and $(a_{12})^2/a_{11}a_{22} = c$ (say), and differentiating c with respect to the elements of $\mathbf{a}_1$ and $\mathbf{a}_2$, we have the vector equations giving the stationary values of c: $M_{12}\mathbf{a}_2-(a_{12}/a_{11})M_{11}\mathbf{a}_1=\mathbf{0}$ and $\mathbf{a}_1'M_{12}-(a_{12}/a_{22})\mathbf{a}_2'M_{22}=\mathbf{0}$ or $(a_{12}/a_{22})M_{22}\mathbf{a}_2-M'_{12}\mathbf{a}_1 = \mathbf{0}$.

Eliminating $\mathbf{a}_2$ and $\mathbf{a}_1$ between the two vector equations, we have, for the stationary values of c, the p-th degree determinantal equation in c:

$$(\text{A.2.3.1}) \qquad \begin{vmatrix} M_{12} & (a_{12}/a_{11})M_{11} \\ (a_{12}/a_{22})M_{22} & M_{12}' \end{vmatrix} = 0$$

or, by using (A.1.1) and remembering that $c = a_{12}^2/a_{11}a_{22}$,

$$(\text{A.2.3.2}) \qquad |cM_{11}-M_{12}\,M_{22}^{-1}\,M_{12}'| = 0,$$

which proves part (ii). The proof of part (iii) is now obvious.

(A.2.4): *If* $M = \begin{array}{c} \left[\begin{array}{ccc} M_{11} & M_{12} & M_{13} \\ M_{12}' & M_{22} & M_{23} \\ M_{13}' & M_{23}' & M_{33} \end{array}\right] \begin{array}{c} p \\ q \\ r \end{array} \\ \begin{array}{ccc} p & q & r \end{array} \end{array}$ $(p \leqslant q)$ *is symmetric p.d., then,*

for all non-null $\mathbf{a}_1[(p+r)\times 1]$ *and* $\mathbf{a}_2[(q+r)\times 1]$, (i)

$$[\mathbf{a}_1' \begin{bmatrix} M_{12} & M_{13} \\ M_{23}' & M_{33} \end{bmatrix} \mathbf{a}_2]^2/[\mathbf{a}_1' \begin{bmatrix} M_{11} & M_{13} \\ M_{13}' & M_{33} \end{bmatrix} \mathbf{a}_1][\mathbf{a}_2' \begin{bmatrix} M_{22} & M_{23} \\ M_{23}' & M_{33} \end{bmatrix} \mathbf{a}_2]$$

is non-negative, (ii) the stationary values of this expression are the roots of the equation in c:

$$|c(M_{11}-M_{13}M_{33}^{-1}M_{13}')-(M_{12}-M_{13}M_{33}^{-1}M_{23}') \times$$
$$\times\ (M_{22}-M_{23}M_{33}^{-1}M_{23}')^{-1}(M_{12}'-M_{23}M_{33}^{-1}M_{13}')|=0$$

and (iii) in particular, the largest and the smallest values of the expression are the largest and smallest roots of the determinantal equation.

Proof: As before, part (i) is obvious. For part (ii) putting the expression under (i) $= c$(say) and proceeding in exactly the same manner as in (A.2.3) we have for the stationary values of c, the determinantal equation in c:

$$(\text{A.2.4.1}) \qquad \begin{vmatrix} c\begin{bmatrix} M_{11} & M_{13} \\ M_{13}' & M_{33} \end{bmatrix} & \begin{bmatrix} M_{12} & M_{13} \\ M_{23}' & M_{33} \end{bmatrix} \\ \begin{bmatrix} M_{12}' & M_{23} \\ M_{13}' & M_{33} \end{bmatrix} & \begin{bmatrix} M_{22} & M_{23} \\ M_{23}' & M_{33} \end{bmatrix} \end{vmatrix} = 0.$$

As in (A.1.14)—(A.1.15), premultiply the left hand side of (A.2.4.1) by the determinant of the non-singular matrix F (and postmultiply by its transpose), where F is given by

$$F = \begin{bmatrix} \begin{array}{cc|c} I & -M_{13}M_{33}^{-1} & p \\ O & I & r \\ p & r & \end{array} & 0 \\ 0 & \begin{array}{cc|c} I & -M_{23}M_{33}^{-1} & q \\ O & I & r \\ q & r & \end{array} \end{bmatrix}$$

The equation now reduces to

$$\begin{vmatrix} c\begin{bmatrix} M_{11}-M_{13}M_{33}^{-1}M'_{13} & O \\ O & M_{33} \end{bmatrix} & \begin{bmatrix} M_{12}-M_{13}M_{33}^{-1}M'_{23} & O \\ O & M_{33} \end{bmatrix} \\ \begin{bmatrix} M'_{12}-M_{23}M_{33}^{-1}M'_{13} & O \\ O & M_{33} \end{bmatrix} & \begin{bmatrix} M_{22}-M_{13}M_{33}^{-1}M'_{23} & O \\ O & M_{33} \end{bmatrix} \end{vmatrix} = 0$$

or,

(A.2.4.2)
$$|c(M_{11}-M_{13}M_{33}^{-1}M'_{13})-(M_{12}-M_{13}M_{33}^{-1}M'_{23})(M_{22}-M_{23}M_{33}^{-1}M'_{23})^{-1} \times(M'_{12}-M_{23}M_{33}^{-1}M'_{13})| = 0.$$

Arguing as in (A.1.14)—(A.1.16) it is easy to see that (i) the roots of this p-th degree equation in c all lie between 0 and 1, (ii) if M of (A.2.4) is p.d., then all roots are < 1 and (iii) if $M_{12}-M_{13}M_{33}^{-1}M'_{23}$ is of rank r ($\leqslant p$), then r of these roots are > 0 and the rest, i.e., $p-r$ are $= 0$. All the roots are zero if and only if $M_{12}-M_{13}M_{33}^{-1}M'_{23}=0$.

(A.2.5): *If* $M(p\times p)$ *is symmetric and at least p.s.d., the statement:* "$g_1 \leqslant \mathrm{a}'(1\times p)\, M(p\times p)\mathrm{a}(p\times 1)/\mathrm{a}'\mathrm{a} \leqslant g_2$ *for all non-null* a" *is exactly equivalent to* "$g_1 \leqslant c_1 \leqslant c_p \leqslant g_2$, " *where* c_1 *and* c_p *stand for the smallest and largest characteristic roots* (*both non-negative*) *of* M. Notice that the last statement gives also the lowest permissible value of g_2 and the highest permissible value of g_1, both in terms of the roots of M. The proof is obvious from (A.2.1).

(A.2.6): *If* M_2 $(p\times p)$ *is symmetric p.d. and* M_1 $(p\times p)$ *is symmetric and at least p.s.d., the statement:* "$g_1 \leqslant \mathrm{a}'(1\times p)M_1(p\times p)\mathrm{a}(p\times 1)/\mathrm{a}'(1\times p)M_2(p\times p)\mathrm{a}(p\times 1) \leqslant g_2$

for all non-null **a**" *is exactly equivalent to* "$g_1 \leqslant c_1 \leqslant c_p \leqslant g_2$" *where* c_1 *and* c_p *stand for the smallest and largest roots of the the equation in* c*(all positive):*

$$|M_1 - cM_2| = 0.$$

Notice that the last statement gives also, in terms of the roots of $M_1M_2^{-1}$, the lowest permissible value of g_2 and the highest permissible value of g_1. The theorem is a direct consequence of (A.2.2).

(A.2.7): *The statement:* "$\mathbf{x}'(1\times p)\mathbf{x}(p\times 1) \leqslant g(g > 0)$" *is exactly equivalen to* "$-\sqrt{g} \leqslant \mathbf{x}'(1\times p)\mathbf{a}(p\times 1) \leqslant +\sqrt{g}$ *(for all* **a** *subject o* $\mathbf{a}'\mathbf{a} = 1$)." The proot follows easily from Cauchy's inequality in algebra.

APPENDIX 3

Some Results in Transformations

(A.3.1): *If* $\mathbf{x}(n\times 1) = A(n\times n)\mathbf{y}(n\times 1)$, *where* A *is* $\perp$, *then* $\mathbf{x}'\mathbf{x} = \mathbf{y}'A'A\mathbf{y} = \mathbf{y}'\mathbf{y}$.

(A.3.2): *If* $\mathbf{x}(n\times 1) = A(n\times n)\mathbf{y}(n\times 1)$ *and* $\mathbf{u}(n\times 1) = A(n\times n)\mathbf{v}(n\times 1)$, *where* A *is* $\perp$, *then* $\mathbf{x}'\mathbf{u} = \mathbf{y}'A'A\mathbf{v} = \mathbf{y}'\mathbf{v}$.

(A.3.3): *If* $M(p\times p)$ *is symmetric and at least p.s.d. of rank* $r(\leqslant)p)$, *then denoting by* c *the roots* $c(M)$ *of* (A.1.8), *there exists an orthogonal matrix* $A(p\times p)$ (*not necessarily unique*) *such that* $M = AD_cA'$.

(A.3.4): *Under the conditions of* (A.1.9), *namely that* $M_1'p\times p)$ *is symmetric and at least p.s.d. of rank* $r(\leqslant p)$ *and* $M_2(p\times p)$ *is symmetric p.d., there exists a non-singular* $A(p\times p)$ (*not necessarily unique*) *such that* $M_1 = AD_cA'$ *and* $M_2 = AA'$.

(A.3.5): *The matrix* A *of* (A.3.3) *will be unique, except for a post-factor* D_k, *if* M *is p.d. and all* $c(M)$*'s are distinct* [31].

Proof: Suppose there are two orthogonal A's, say A_1 and A_2, satisfying the condition of (A.3.3). Then we have $A_1D_cA_1' = A_2D_cA_2'$ or $A_2^{-1}A_1D_c = D_cA_2'(A_1')^{-1}$ or $A_2'A_1D_c = D_cA_2'A_1$ (since for an orthogonal A, $A^{-1} = A'$). If we now denote $A_2'A_1$ by B with elements b_{ij}, then the above equation gives

(A.3.5.1) $$b_{ij}c_j = c_ib_{ij} \text{ or } b_{ij}(c_i-c_j) = 0.$$

Thus, if $i \neq j$ and $c_i \neq c_j$, $b_{ij} = 0$, which means that B is a diagonal matrix D_b with elements, say $b_1, \ldots, b_p$.

$$\therefore \text{ since } \quad D_b = B = A_2'A_1, \text{ we have}$$

(A.3.5.2) $$D_bD_b' = D_{b^2} = A_2'A_1A_1'A_2 = I(p),$$

so that $b_i^2 = +1$ and so $b_i = \pm 1$, $(i = 1, 2, \ldots, p)$.

(A.3.5.3) Thus $$D_b = D_k \text{ and hence } A_2'A_1 = D_k \text{ or } A_1 = A_2D_k;$$

this proves (A.3.5). We note that A can thus be made unique by adopting the convention, say, that its first row be positive. It is easy to check that the transformation is now one-to-one.

(A.3.6): *If* $X(p\times n)$ $(p \leqslant n)$ *is of rank* p (*in which case, by* (A.1.10). $X\,X'$ *is symmetric p.d.*), *then there exists a transformation* $X(p\times n) = A(p\times p)\ D_{\sqrt{c}}(p\times p) \times L(p\times n)$, *where* A *is* $\perp$, $LL' = I(p)$ *and where* c*'s are the characteristic roots* (*all positive*) *of the matrix* $X\,X'$. *If all* c*'s are distinct this transformation is unique except for a post-factor* D_k *to go with* A.

Proof: By (A.3.3) there exists an orthogonal $A(p\times p)$, which may not be unique, such that $XX' = AD_cA'$. We now define a $L(p\times n)$ by $X = AD_{\sqrt{c}}L$ and note that, given X and hence c's and A (which we can find but which may not be unique), this is a linear equation in L uniquely solvable in terms of the above elements. Also $LL' = D_{1/\sqrt{c}}A^{-1}XX'A'^{-1}D_{1/\sqrt{c}} = D_{1/\sqrt{c}}A^{-1}AD_cA'A'^{-1}D_{1/\sqrt{c}} = I(p)$. We have thus the transformation $X = AD_{\sqrt{c}}L$, where A is $\perp$ and $LL' = I(p)$. Notice that, if the c's are distinct, A is unique except for a post-factor D_k and that L will go with A, being defined by $L = D_{1/\sqrt{c}}A^{-1}X$. This proves (A.3.6). It is easy to check that for distinct roots the transformation can be made one to one by adopting the convention, say, that the first row of A be positive.

(A.3.7): *The matrix A of* (A.3.4) *will be unique, except for a factor D_k if M_1 is p.d. and all the roots are distinct* [31]

Proof: Suppose there are two non-singular A's, say A_1 and A_2, satisfying the conditions of (A.3.4). Then we have

(A.3.7.1) $$A_1D_cA_1' = A_2D_cA_2' \text{ and } A_1A_1' = A_2A_2'.$$

These lead, after a little reduction, to

(A.3.7.2) $$A_2^{-1}A_1D_c = D_cA_2^{-1}A_1 \text{ or } BD_c = D_cB, \text{ where } A_2^{-1}A_1 = B.$$

If now $B = (b_{ij})$, say, then (A.3.7.2) leads to

(A.3.7.3) $$b_{ij}c_j = c_ib_{ij} \text{ or } b_{ij}(c_i-c_j) = 0 \text{ or } b_{ij} = 0 \text{ if } i \neq j \text{ and } c_i \neq c_j.$$

Thus $B = D_b$ (say) and so we have

(A.3.7.4) $$D_bD_b' = D_{b^2} = BB' = A_2^{-1}A_1A_1'(A_2^{-1})' = A_2^{-1}A_2A_2'(A_2')^{-1}=I(p)$$

so that $b_i = \pm 1$, i.e., $D_b = D_k$.

(A.3.7.5) $$\text{Thus } A_2^{-1}A_1 = D_k \text{ or } A_1 = A_2D_k,$$

which proves (A.3.7). As before, we note that A can be made unique by adopting the convention, say, that its first row be positive. Check that the transformation in this case is one-to-one.

(A.3.8): *If $X_1(p\times n_1)$, $X_2(p\times n_2)$, $(p \leqslant n_1, n_2)$ are each of rank p (in which case, by* (A.1.10), *X_1X_1' and X_2X_2' are both symmetric p.d.), then there exists a transformation $X_1(p\times n_1) = A(p\times p)D_{\sqrt{c}}(p\times p)L_1(p \times n_1)$, and $X_2(p\times n_2) = A(p\times p)\ L_2(p\times n_2)$ where A is non-singular, c's are the roots (all positive) of the equation $|X_1X_1'-cX_2X_2'|=0$, and $L_1L_1' = L_2L_2' = I(p)$. If all c's are distinct then this transformation is unique except for a post-factor D_k to go with A.*

Proof: By (A.3.4) there exists a non-singular A, which may not be unique, such that $X_1X_1' = AD_cA'$ and $X_2X_2' = AA'$. We now define $L_1(p\times n_1)$ and $L_2(p\times n_2)$ by $X_1 = AD_{\sqrt{c}}L_1$ and $X_2 = AL_2$ and note that, given X_1, X_2 and c's and A (which may not be unique), L_1 and L_2 are uniquely solvable in terms of these. Also $L_1L_1' = D_{1/\sqrt{c}}A^{-1}X_1X_1'A'^{-1}D_{1/\sqrt{c}} = I(p)$ and $L_2L_2' = A^{-1}X_2X_2'A'^{-1} = I(p)$. This proves the existence of the transformation (A.3.8). Notice that if all c's are distinct, then by (A.3.7) A is unique except for a post factor D_k and that L_1 and L_2 will go with A being defined by $L_1 = D_{1/\sqrt{c}}A^{-1}X_1$ and $L_2 = A^{-1}X_2$. Check that the transformation in this case is one-to-one if we adopt the convention, say, that the first row of A is to be positive.

(A.3.9): *If* $M(p \times p)$ *is symmetric and p.d., then there exists a non-singular* $\tilde{T}(p \times p)$ *such that* $M = \tilde{T}\tilde{T}'$, *and this* $\tilde{T}$ *is unique except for a post factor* D_k *and so* $\tilde{T}$ *will be called near unique* $\tilde{T}$ *is a triangular matrix.*

Proof of the near uniqueness: Suppose there are two $\tilde{T}$'s, say $\tilde{T}_1$ and $\tilde{T}_2$, satisfying the condition. Notice from (A.1.10) that since M is p.d. $\tilde{T}$ must necessarily be non-singular. Thus we have

$$\tilde{T}_1\tilde{T}'_1 = \tilde{T}_2\tilde{T}'_2 \text{ or } \tilde{T}_2^{-1}\tilde{T}_1 = \tilde{T}'_2(\tilde{T}'_1)^{-1}. \tag{A.3.9.1}$$

Now making use of the remarks made after (1.1) we note that $\tilde{T}_2^{-1}\tilde{T}_1$ is a triangular matrix with the same configuration as $\tilde{T}_1$ and $\tilde{T}'_2(\tilde{T}'_1)^{-1}$ of opposite configuration. Thus it is obvious that

$$\tilde{T}_2^{-1}\tilde{T}_1 = D_a \text{ (say)}, \tag{A.3.9.2}$$

whence $D_aD'_a = D_{a^2} = \tilde{T}_2^{-1}\tilde{T}_1\tilde{T}'_1(\tilde{T}_2^{-1})' = \tilde{T}_2^{-1}\tilde{T}_2\tilde{T}'_2(T'_2)^{-1} = I(p)$, so that $a_i = \pm 1$, i.e., $D_a = D_k$. Thus

$$\tilde{T}_2^{-1}\tilde{T}_1 = D_k \text{ or } \tilde{T}_1 = \tilde{T}_2D_k, \tag{A.3.9.3}$$

which proves the near uniqueness. It is easy to check that $\tilde{T}$ can be made unique by adopting the convention, say, that the diagonal elements of $\tilde{T}$ be positive. The transformation in this case is one-to-one.

(A.3.10): *If* $M = \begin{array}{cc} \begin{bmatrix} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{bmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}$ *is symmetric and p.s.d. of rank* p *and if the first* p *rows can be taken as a basis, then there exists a non-singular* $\tilde{T}(p \times p)$ *and a* $T_2(q \times p)$ *such that*

$$\begin{bmatrix} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{bmatrix} = \begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix} [\tilde{T}'_1 : T'_2],$$

and furthermore $\begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix}$ *is unique except for a post-factor* D_k.

Proof: Since M is symmetric it is evident that, if the first p rows of M can be taken as a basis, then the first p columns of M also can be taken as a basis. Thus no row of M_{11} is a linear function of the other rows of M_{11} and no column of M_{11} is a linear function of the other columns of M_{11}. Hence M_{11} is non-singular. Here, of course, M_{11} is symmetric p.d.. We can also look at the picture in a reverse way, namely that, since M is symmetric p.s.d. of rank p, we can find a non-singular principal minor of order p which is of course symmetric p.d. Renumbering the rows (and the corresponding columns) of that principal minor, we can call it M_{11}. Then it is easy to show in this case that, we can take the first p rows or the first p columns as a basis.

Now notice that if the first p rows can be taken as a basis then there exists a non-singular $A(p\times q)$ such that $M'_{12} = AM_{11}$ and $M_{22} = A\,M_{12}$. Combining the two we have $M_{22} = M'_{12}M_{11}^{-1}M_{12}$ (note that M_{11} is non-singular and thus we can take the inverse). We next observe that in this set-up M_{11} is p.d. Therefore by (A.3.9), we can find a non-singular $\tilde{T}_1(p\times p)$, unique except for a post-factor D_k, such that $M_{11} = \tilde{T}_1\tilde{T}'_1$. Now find a T'_2 defined by $T'_2 = \tilde{T}_1^{-1}M_{12}$ and check that $M'_{12} = (\tilde{T}_1T_2')' = T_2\tilde{T}'_1$ and $M_{22} = M'_{12}M_{11}^{-1}M_{12} = T_2T'_2$, which proves (A.3.10). We observe that, as in (A.3.9), $\begin{bmatrix}\tilde{T}_1\\ T_2\end{bmatrix}$ can be made unique by adopting the convention that the diagonal elements of $\tilde{T}_1$ be positive. Check that the transformation is now one-to-one.

(A.3.11): *If $X(p\times n)$ $(p \leqslant n)$ is of rank $r(\leqslant p)$such that, say, the first r rows of X can be taken as a basis, then there exists a transformation*

$$\begin{array}{cc} \left[\begin{array}{c} X_1 \\ X_2 \end{array}\right] & \begin{array}{l} r \\ p-r \end{array} \\ n & \end{array} = \begin{array}{cc} \begin{array}{r} r \\ p-r \end{array} & \left[\begin{array}{c} \tilde{T}_1 \\ T_2 \end{array}\right] \\ & r \end{array} L(r\times n),$$

where $LL' = I(r)$ and $\tilde{T}_1$ is non-singular and unique except for a post-factor D_k.

Proof: By (A.1.10) $X_1X'_1$ is symmetric p.d. of rank r and by (A.3.9) there exists a (non-singular) $\tilde{T}_1$ (unique except for a post-factor D_k) such that $X_1X'_1 = \tilde{T}_1\tilde{T}'_1$. We now define an L by $L(r\times n) = \tilde{T}_1^{-1}(r\times r)\ X_1(r\times n)$ and note that given X_1 and hence $\tilde{T}_1$ (which is unique except for a post-factor D_k), L is uniquely solvable in terms of these. Also $LL' = \tilde{T}_1^{-1}X_1X'_1(\tilde{T}_1^{-1})' = \tilde{T}_1^{-1}\tilde{T}_1\tilde{T}'_1(\tilde{T}'_1)^{-1} = I(r)$. Next define a T_2 by $T_2L = X_2$ or $T_2LL' = X_2L'$ or $T_2 = X_2L'$ and note that, given X_1, X_2 and hence L (which is near unique), T_2 is also uniquely solvable in terms of these. We note further that now $X_2 = T_2L = T_2\tilde{T}_1^{-1}X_1 = B(\overline{p-r}\times r)X_1$ (say), where $B = T_2\tilde{T}_1^{-1}$. This is obviously the condition that X be of rank r and X_1 be a basis. Hence the transformation is proved to exist with the near uniqueness already stated. By adopting a convention, say that of (A.3.10), the transformation can be checked to be one-to-one.

(A.3.12): *If $X_1(p\times n_1)$, $X_2(p\times n_2)$ $(p \leqslant n_1, n_2)$ are each of rank p (see(A.3.8)), then there exists a transformation: $X_1(p\times n_1) = \tilde{T}(p\times p)\ L(p\times p)D_{\sqrt{c}}(p\times p)L_1(p\times n_1)$ and $X_2(p\times n_2) = \tilde{T}(p\times p)L_3(p\times n_2)$, where $\tilde{T}$ is non-singular, L is $\perp$, $L_1L'_1 = L_3L'_3 = I(p)$ and the c's are the (all positive) roots of the equation in c: $|X_1X'_1-cX_2X'_2| = 0$. If the c's are distinct, the transformation can be made one-to-one by letting $\tilde{T}$ have a positive diagonal.*

Proof: Start from the transformation (A.3.8) and, using (A.3.11), put $A(p\times p) = \tilde{T}(p\times p)\ L(p\times p)$ where L is $\perp$. Next put $LL_2 = L_3$ and note that $L_3L'_3 = LL_2L'_2 \times L' = I(p)$. The proof of near uniqueness in the case of distinct roots follows along the same lines as in (A.3.8) and (A.3.11). This completes the proof of (A.3.12).

(A.3.13): *If $M_2(p\times p)$ is symmetric p.d. and $M_1(p\times p)$ is symmetric p.s.d. of rank $r(< p)$ then there exists a transformation, which, without any loss of generality, we can write as*

(A.3.13.1)
$$M_1(p\times p)=\begin{array}{c@{}c}\begin{array}{c}p-r\\ r\end{array} & \left[\begin{array}{c}A_1\\ A_2\end{array}\right]\\ & r\end{array} D_c^*(r\times r)\ \begin{array}{c@{}c}[A_1'\ \vdots\ A_2'] & r,\\ \begin{array}{cc}p-r & r\end{array} & \end{array}\ \textit{and}$$

(A.3.13.2)
$$M_2(p\times p)=\begin{array}{c@{}c}\begin{array}{c}p-r\\ r\end{array} & \left[\begin{array}{cc}A_1 & \tilde{A}_3\\ A_2 & A_4\end{array}\right]\\ & \begin{array}{cc}r & p-r\end{array}\end{array}\begin{array}{c@{}c}\left[\begin{array}{cc}A_1' & A_2'\\ \tilde{A}_3' & A_4'\end{array}\right] & \begin{array}{c}r\\ p-r\end{array}\\ \begin{array}{cc}p-r & r\end{array} & \end{array},$$

where the c's of D_c^* *stand for the r non-zero roots of the equation in* $c\!:|M_1-cM_2|=0$ *and where* $A=\left[\begin{array}{cc}A_1 & \tilde{A}_3\\ A_2 & A_4\end{array}\right]$ *is non-singular. If the non-zero roots are distinct, the matrix A is unique except for a post-factor* $D_k(p)$.

Proof: Using (A.1.9) and (A.3.4) we can find a non-singular $G(p\times p)$ such that $M_1(p\times p)=G(p\times p)\ D_c(p\times p)\ G'(p\times p)$ and $M_2(p\times p)=G(p\times p)\ G'(p\times p)$, where the c's of D_c are the roots of the equation in c: $|M_1-cM_2|=0$. We recall that under the conditions of the problem r of these roots are positive and the rest zero. Let us call these r positive roots $c_1, c_2, \ldots, c_r$. Then the r columns of the G matrix (and the rows of the G' matrix) that go with these positive c's have to be numbered $1, 2, \ldots, r$.

Now denoting the matrix formed by these r columns of $G(p\times p)$ by $A(p\times r)$, the remaining submatrix of $G(p\times p)$ by $B(p\times\overline{p-r})$ we can set

(A.3.13.3)
$$M_1(p\times p)=A(p\times r)\,D_c^*(r\times r)\,A'(r\times p),$$
$$M_2(p\times p)=\begin{array}{c}p[A\ \vdots\ B]\\ \begin{array}{cc}r & p-r\end{array}\end{array}\begin{array}{c@{}c}\left[\begin{array}{c}A'\\ B'\end{array}\right] & \begin{array}{c}r\\ p-r\end{array}\\ p & \end{array}$$

Since G is non-singular, $[A\ \vdots\ B]p$ is non-singular, and hence $A(p\times r)$ is of rank r and $B(p\times\overline{p-r})$ is of rank $(p-r)$. We can, therefore, choose $p-r$ rows from B to form a non-singular (square) matrix of order $p-r$. Let these rows be numbered $1, 2, \ldots, \overline{p-r}$. Let us denote the matrix formed by these $p-r$ rows of $B(p\times\overline{p-r})$ by $B_3(\overline{p-r}\times\overline{p-r})$ and the remaining submatrix of $B(p\times\overline{p-r})$ by $B_4(r\times\overline{p-r})$. Let us denote the submatrix formed by the corresponding rows of $A(p\times r)$ by $A_1(\overline{p-r}\times r)$ and $A_2(r\times r)$. We can now rewrite (A.3.13.3) as

(A.3.13.4)
$$M_1=\begin{array}{c@{}c}\begin{array}{c}p-r\\ r\end{array} & \left[\begin{array}{c}A_1\\ A_2\end{array}\right]\\ & r\end{array} D_c{}^*(r\times r)\ \begin{array}{c@{}c}[A_1'\ \vdots\ A_2'] & r,\\ \begin{array}{cc}p-r & r\end{array} & \end{array}$$

$$M_2=\begin{array}{c@{}c}\begin{array}{c}p-r\\ r\end{array} & \left[\begin{array}{cc}A_1 & B_3\\ A_2 & B_4\end{array}\right]\\ & \begin{array}{cc}r & p-r\end{array}\end{array}\begin{array}{c@{}c}\left[\begin{array}{cc}A_1' & A^2\\ B_3' & B_4\end{array}\right] & \begin{array}{c}r\\ p-r\end{array}\\ \begin{array}{cc}p-r & r\end{array} & \end{array},$$

where B_3 is non-singular and $\begin{bmatrix}A_1 & B_3\\ A_2 & B_4\end{bmatrix}$ is also non-singular. Notice that with renumbering of the rows of B and of A, i.e., of G, the rows (and the associated) columns of M_1 and M_2 have also to be renumbered. Assuming now that B_3 is non-singular we can, by (A.3.11), find a transformation $B_3(\overline{p-r}\times\overline{p-r}) = \tilde{A}_3(\overline{p-r}\times\overline{p-r})\, L(\overline{p-r}\times\overline{p-r})$ where L is $\perp$. Now put $B_4(r\times\overline{p-r}) = A_4(r\times\overline{p-r})\, L(\overline{p-r}\times\overline{p-r})$ (which defines A_4 in a unique way in terms of B_4 and L). Thus we have

$$\begin{bmatrix}A_1 & B_3\\ A_2 & B_4\end{bmatrix} = \begin{bmatrix}A_1 & \tilde{A}_3\\ A_2 & A_4\end{bmatrix} \begin{array}{c} \begin{bmatrix}I(r) & 0\\ 0 & L\end{bmatrix} \\ \begin{array}{cc} r & p-r\end{array}\end{array} \begin{array}{l} r \\ p-r \\ \end{array}$$

and thus (A.3.13.4) is replaced by

$$\text{(A.3.13.5)} \qquad M_2 = \begin{bmatrix}A_1 & \tilde{A}\\ A_2 & A_4\end{bmatrix} \begin{bmatrix}A_1' & A_2'\\ \tilde{A}_3' & A_4'\end{bmatrix}.$$

(A.3.13.3) and (A.3.13.5) taken together give us (A.3.13.1) and (A.3.13.2). Now for the near uniqueness in the case of distinct roots under D_c^*, remember that

$$D_c = \begin{array}{r} r \\ p-r \\ \end{array} \begin{array}{c}\begin{bmatrix}D_c^* & 0\\ 0 & 0\end{bmatrix} \\ \begin{array}{cc} r & p-r\end{array}\end{array}, \quad \text{put} \quad U = \begin{bmatrix}A_1 & \tilde{A}_3\\ A_2 & A_4\end{bmatrix} \text{ and write } M_1 = UD_cU'$$

and $M_2 = UU'$. If now there is another matrix V satisfying the same conditions, then arguing in the same manner as in (A.3.7) we have $V^{-1}UD_c = D_cV^{-1}U$ or $B_1D_c = D_cB_1$ where $B_1 = V^{-1}U = b_1^{(1)}$, say. This, as in (A.3.7) leads to the equation $b_{ij}^{(1)}c_j = c_ib_{ij}^{(1)}$, whence $b_{ij}^{(1)} = 0$ if $i \neq j$ and $c_i \neq c_j$. Note that here $c_i = 0$ $(i = r+1, \ldots, p)$. This shows that the B_1 matrix is of the form

$$\left[\begin{array}{c|c} D_a & 0\\ \hline 0 & \text{solid}\end{array}\right] \begin{array}{l} r \\ p-r\end{array} = \left[\begin{array}{c|c} D_a & 0\\ \hline 0 & S\end{array}\right] \text{(say)}.$$

Remembering that $B_1 = V^{-1}U$, we have

$$B_1B_1' = \begin{bmatrix}D_a{}^2 & 0\\ 0 & SS'\end{bmatrix} = V^{-1}UU'(V^{-1})' = V^{-1}VV'(V')^{-1} = I(p),$$

whence $D_a = D_k$ and S is seen to be an orthogonal matrix. Using the structure of U and V and the relation $V^{-1}U = B_1$ we have

$$\begin{bmatrix}U_1 & \tilde{U}_3\\ U_2 & U_4\end{bmatrix} = \begin{bmatrix}V_1 & \tilde{V}_3\\ V_2 & V_4\end{bmatrix} \begin{bmatrix}D_k & 0\\ 0 & S\end{bmatrix} = \begin{bmatrix}V_1D_k & \tilde{V}_3S\\ V_2D_k & V_4S\end{bmatrix},$$

where S is an $\perp$ matrix. But, since $\tilde{U}_3 = \tilde{V}_3 S$, therefore, S must also be a triangular matrix. Hence S is necessarily of the form $D_k(p-r)$. Thus B_1 is of the form $D_k(p)$, which proves the near uniqueness in the case of distinct non-zero roots. This completes the proof of (A.3.13). In this case the transformation is easily checked to be one-to-one if we adopt the convention that the first row of A_1 and the diagonal elements of $\tilde{A}_3$ be positive.

(A.3.14): *If $X_1(p\times n_1)$ $(p > n_1)$ be of rank n_1 such that the last n_1 rows form a square matrix which is non-singular and $X_2(p\times n_2)$ $(p \leqslant n_2)$ be of rank p, then there exists a transformation*

$$X_1(p\times n_1) = \begin{array}{r@{}c} \begin{array}{r} p-n_1 \\ n_1 \end{array} & \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \\ & n_1 \end{array} D_{\sqrt{c}}\,(n_1\times n_1) L_1(n_1\times n_1),$$

$$X_2(p\times n_2) = \begin{array}{r@{}c} \begin{array}{r} p-n_1 \\ n_1 \end{array} & \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix} \\ & \begin{array}{cc} n_1 & p-n_1 \end{array} \end{array} L_2(p\times n_2),$$

such that L_1 is $\perp$ and $L_2L_2' = I(p)$ where c's stand for the non-zero roots of the equation in c: $| X_1X_1' - cX_2X_2' | = 0$, *and* $U = \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix}$ *is non-singular. Also if all the non-zero roots c are distinct, U is unique except for a post-factor D_k. Notice that, by* (A.1.9), *all the c's are anyway non-negative and n_1 of them are positive, the rest being zero.*

Proof: By (A.3.13) there exists an $\begin{array}{c@{}l} \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix} & \begin{array}{l} p-n_1 \\ n_1 \end{array} \\ \begin{array}{cc} n_1 & p-n_1 \end{array} & \end{array}$ not necessarily unique such that $X_1X_1' = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} D_c(n_1\times n_1)[U_1' : U_2']$ and $X_2X_2' = \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix} \begin{bmatrix} U_1' & U_2' \\ \tilde{U}' & U_4' \end{bmatrix}$.

Now define an $L_1(n_1\times n_1)$ and $L_2(p\times n_2)$ by

$$X_1(p\times n_1) = \begin{array}{r@{}c} \begin{array}{r} p-n_1 \\ n_1 \end{array} & \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \\ & n_1 \end{array} = \begin{array}{r@{}c} \begin{array}{r} p-n_1 \\ n_1 \end{array} & \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \\ & n_1 \end{array} D_{\sqrt{c}}(n_1\times n_1)\; L_1(n_1\times n_1)$$

$$X_2(p\times n_2) = \begin{array}{r@{}c} \begin{array}{r} p-n_1 \\ n_1 \end{array} & \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix} \\ & \begin{array}{cc} n_1 & p-n_1 \end{array} \end{array} \quad L_2(p\times n_2),$$

and notice that L_2 is given uniquely by $L_2 = U^{-1}X_2$, in terms of X_2 and U which itself may not be unique and similarly L_1 is given uniquely (in the same sense) by $L_1 = D_{1/\sqrt{c}}U^{-1}Y_2$. Next we check that

$$L_2L_2' = U^{-1}X_2X_2'(U^{-1})' = U^{-1}UU'(U')^{-1} = I(p) \text{ and}$$
$$L_1L_1' = D_{1/\sqrt{c}}\, U_2^{-1}Y_2Y_2'(U_2^{-1})'D_{1/\sqrt{c}}$$
$$= D_{1/\sqrt{c}}U_2^{-1}U_2D_cU_2'(U_2')^{-1}D_{1/\sqrt{c}} = I(n_1).$$

We observe also if the non-zero roots are unique, then, by (A.3.13), U is unique except for a post-factor $D_k(p)$ and thus L_1 and L_2 which hang on U are also indeterminate to the same extent. This completes the proof of (A.3.14). As in the case of (A.3.13) also here, for distinct roots the transformation can be made one-to-one by adopting the same convention as at the end of (A.3.13).

(A.3.15): *If* $X_1(p\times n_1)$, $X_2(p\times n_2)$ $(n_1 < p \leqslant n_2)$ *are of ranks* n_1 *and* p *respectively, then there exists a transformation:* $X_1'(n_1\times p) = L(n_1\times n_1)\ D_{\sqrt{c}}(n_1\times n_1) \times L_1(n_1\times p)\tilde{T}'(p\times p)$ *and* $X_2(p\times n_2) = \tilde{T}(p\times p)\ L_2(p\times n_2)$, *where* L is $\perp$, $L_1L'_1 = I(n_1)$, $L_2L_2' = I(p)$ *and* c*'s are the* n_1 *characteristic roots (all positive) of* $X_1'(X_2X_2')^{-1}X_1$. *For distinct roots the transformation can be made one-to-one by letting* $\tilde{T}$ *have a positive diagonal.*

Proof: Using (A.3.11), put $X_2(p\times n_2) = \tilde{T}(p\times p)\ L_2(p\times n_2)$, subject to $L_2L_2' = I(p)$ and now using (A.3.6) put $X_1'(n_1\times p)(\tilde{T}')^{-1}(p\times p) = L(n_1\times n_1)D_{\sqrt{c}}(n_1\times n_1) \times L_1(n_1\times p)$, where L is $\perp$, $L_1L_1' = I(n_1)$ and c's are the roots of $X_1'(\tilde{T}')^{-1}(\tilde{T}^{-1})X_1$, i.e., of $X_1'(\tilde{T}\tilde{T}')^{-1}X_1$, i.e., of $X_1'(X_2X_2')^{-1}X_1$. Postmultiplying both sides by $\tilde{T}'$ we have: $X_1' = LD_{\sqrt{c}}L_1\tilde{T}'$ and for X_2 we already have $X_2 = \tilde{T}L_2$. Near uniqueness, in the case of distinct roots, follows along the same lines as in (A.3.11) and (A.3.14). Check, by using (A.1.18), that these c's of (A.3.15) are the same as the non-zero roots of the equation in c (considered in (A.3.14)): $|X_1X_1' - cX_2X\ | = 0$.

(A.3.16): *If*
$$M = \begin{array}{cc} \begin{bmatrix} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{bmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array} \quad (p \leqslant q) \textit{ is symmetric p.d.}$$

(*note that, in this situation,* M_{11} *and* M_{22} *are both necessarily symmetric p.d.*) *and if* M_{12} *is of rank* $s(\leqslant p \leqslant q)$ *and if* $D_c(s\times s)$ *is the diagonal matrix based on the* s *non-zero roots of the p-th degree equation in c:* $|cM_{11} - M_{12}M_{22}^{-1}M_{12}'| = 0$, *then there exist non-singular* $A(p\times p)$ *and* $B(q\times q)$ *which, without any loss of generality in the sense of* (A.3.13), *we can take to be of the structure*

$$A = \begin{array}{cc} \begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} & \begin{array}{c} p-s \\ s \end{array} \\ \begin{array}{cc} s & \quad p-s \end{array} & \end{array} \quad \textit{and } B = \begin{array}{cc} \begin{bmatrix} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{bmatrix} & \begin{array}{c} q-s \\ s \end{array} \\ \begin{array}{cc} s & \quad q-s \end{array} & \end{array}$$

($\tilde{A}_3$ *and* $\tilde{B}_3$ *being non-singular*), *such that* $M_{11}(p\times p) = A(p\times p)\ A'(q \qquad , M_{22}(q\times q) = B(q\times q)\,B'(q\times q)$ *and* $M_{12}(p\times q) = A(p\times p) \begin{bmatrix} D_{\sqrt{c}}\,(s\times s) & 0(s\times q-s) \\ 0(p-s\times s) & 0(p-s\times q-s) \end{bmatrix} B'(q\times q) = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} D_{\sqrt{c}}[B_1' \ \vdots\ B_2']$; *also, if the c's are distinct, A is unique except for a post-factor*

$D_k(p)$ and, for a given choice of A, B is unique except for a post factor $D_k(q-s)$ to go with $\tilde{B}_3$.

Proof: Since M_{11} is symmetric p.d. and $M_{12}M_{22}^{-1}M_{12}'$ is symmetric and at least p.s.d. of rank $s(\leqslant p)$, there exists, by and in the sense of (A.3.13), a transformation $M_{11} = AA'$ and

$$M_{12}M_{22}^{-1}M_{12}' = A \left[\begin{array}{c|c} D_c & 0 \\ \hline 0 & 0 \end{array}\right] A', \text{ where } A = \begin{array}{cc} \left[\begin{array}{cc} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{array}\right] & \begin{array}{c} p-s \\ s \end{array} \\ \begin{array}{cc} s & p-s \end{array} & \end{array}$$

is non-singular, $\tilde{A}_3$ is non-singular and c's stand for the s non-zero roots of the equation in c, the rest, $p-s$ in number, being zero. Next, since $M_{22}(q\times q)$ is symmetric p.d. it follows from (A.3.3) that there is an orthogonal $E(q\times q)$ such that $M_{22} = ED_eE'$ where $e = (e_1, \ldots, e_q)$ denotes the q characteristic roots (all positive) of the p.d. matrix M_{22}. Substituting this in $M_{12}M_{22}^{-1}M_{12}'$ we have

$$\text{(A.3.16.1)} \qquad M_{12}(ED_eE')^{-1}M_{12}' = A \begin{bmatrix} D_c & 0 \\ 0 & 0 \end{bmatrix} A',$$

or (since E is $\perp$ and A is non-singular),

$$\text{(A.3.16.2).} \qquad A^{-1}M_{12}ED_{1/e}E'M_{12}'(A^{-1})' = \begin{array}{cc} \begin{bmatrix} D_c & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{c} s \\ p-s \end{array} \\ \begin{array}{cc} s & p-s \end{array} & \end{array}$$

We now define a $G_1(s\times q)$ by

(A.3.16.3) $D_{\sqrt{e}}(s\times s)\, G_1(s\times q) =$ the submatrix formed by the first s rows of $(A^{-1}M_{12}ED_{1/\sqrt{e}})$.

It is easy to check that, given the other elements, (A.3.16.3) defines G_1 uniquely and also that $G_1(s\times q)\, G_1'(q\times s) = I(s)$. It is well known that if $G_1(s\times q)$ $(s < \mathrm{q})$ satisfies $G_1G_1' = I(s)$, then we can adjoin a $G_2(\overline{q-s}\times q)$ to G_1 such that $\begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$ is an $\perp$ matrix. With this adjunction we can now write

$$\text{(A.3.16.4)} \qquad \underset{q}{(A^{-1}M_{12}ED_{1/\sqrt{e}})}\ p = \begin{array}{cccc} s & \left[\begin{array}{cc} D_{\sqrt{c}} & 0 \\ 0 & 0 \end{array}\right] & \left[\begin{array}{c} G_1 \\ G_2 \end{array}\right] & \begin{array}{c} s \\ q-s \end{array} \\ p-s & & & \\ & \begin{array}{cc} s & q-s \end{array} & q & \end{array}$$

or

$$\text{(A.3.16.5)} \qquad (A^{-1}M_{12}ED_{1/\sqrt{e}})[G_1' \quad G_2'] = \begin{array}{cc} \begin{bmatrix} D_{\sqrt{c}} & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{c} s \\ p-s \end{array} \\ \begin{array}{cc} s & q-s \end{array} & \end{array}$$

Next put $(ED_{1/\sqrt{e}})[G_1' \quad G_2'] = F'^{-1}$ (say), so that

$$\text{(A.3.16.6)} \qquad F'(q\times q) = \begin{array}{c} s \\ q-s \end{array} \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} D_{\sqrt{e}}E'$$

(remembering that E is $\perp$). Notice that, given the submatrices of the M matrix. we can find a non-singular A, an $\perp$ E and an $\perp$ $\begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$ (none being necessarily unique) and thus a (non-singular but not necessarily unique) F given by (A.3.16.6) Using now (A.3.16.5), (A.3.16.6) and the definition of A (in the beginning of the proof) we check that we have non-singular A and F satisfying

(A.3.16.7)
$$M_{11}(p\times p) = A(p\times p)A'(p\times p),$$

$$M_{12}(p\times q) = A(p\times p)\begin{array}{cc} \left[\begin{array}{cc} D_{\sqrt{c}} & 0 \\ 0 & 0 \end{array}\right] & \begin{array}{l} s \\ p-s \end{array} \\ \begin{array}{cc} s & p-s \end{array} & \end{array} \times F'(q\times q) \text{ and } M_{22}(q\times q) = F(q\times q)F'(q\times q).$$

We next partition F into $\begin{array}{cc} \left[\begin{array}{cc} F_1 & F_3 \\ F_2 & F_4 \end{array}\right] & \begin{array}{l} q-s \\ s \end{array} \\ \begin{array}{cc} s & q-s \end{array} & \end{array}$, assume in the sense of A.3.13 that F_3 is non-singular (as we obviously can), note that F_3 and F_4 do not occur in the factorization of M_{12} and put $F_1 = B_1$, $F_2 = B_2, F_3(\overline{q-s}\times\overline{q-s}) = \tilde{B}_3(\overline{q-s}\times\overline{q-s}) \times L(\overline{q-s}\times\overline{q-s})$ (where L is $\perp$) and $F_4\,(s\times\overline{q-s}) = B_4(s\times\overline{q-s})\,L(\overline{q-s}\times\overline{q-s})$. As in (A.3.13), remembering the structure of A, we now rewrite (A.3.16.7) as

(A.3.16.8)
$$M_{11}(p\times p) = \begin{array}{cc} \begin{array}{r} p-s \\ s \end{array} & \left[\begin{array}{cc} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{array}\right] \\ & \begin{array}{cc} s & p-s \end{array} \end{array} \begin{array}{cc} \left[\begin{array}{cc} A_1' & A_2' \\ \tilde{A}_3' & A' \end{array}\right] & \begin{array}{l} s \\ p-s \end{array} \\ \begin{array}{cc} p-s & s \end{array} & \end{array},$$

$$M_{12}(p\times q) = \begin{array}{cc} \begin{array}{r} p-s \\ s \end{array} & \left[\begin{array}{c} A_1 \\ A_2 \end{array}\right] \end{array} \quad D_{\sqrt{c}}(s\times s)\ \begin{array}{c} [B_1' : B_2'] \\ \begin{array}{cc} q-s & s \end{array} \end{array}$$

and
$$M_{22}(q\times q) = \begin{array}{cc} \begin{array}{r} q-s \\ s \end{array} & \left[\begin{array}{cc} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{array}\right] \\ & \begin{array}{cc} s & q-s \end{array} \end{array} \begin{array}{cc} \left[\begin{array}{cc} B_1' & B_2' \\ \tilde{B}_3' & B_4' \end{array}\right] & \begin{array}{l} s \\ q-s \end{array} \\ \begin{array}{cc} q-s & s \end{array} & \end{array}$$

which establishes the existence of the transformation (A.3.16).

To prove the near unique ^{4} and B where the c's are distinct, we first recall the definition of A and observe, as in the proof of (A.3.13), that A is unique except for a post-factor $D_k(p)$. The second equation of (A.3.16.8) shows that at this stage B_1 and B_2 are unique except for the post-factor that goes with A. Now consider the third equation of (A.3.16.8) and partition M_{22} into four submatrices and rewrite the equation as

(A.3.16.9)
$$\begin{array}{cc} \left[\begin{array}{cc} M_{22}^{(1)} & M_{22}^{(3)} \\ M_{22}^{(2)} & M_{22}^{(4)} \end{array}\right] & \begin{array}{l} q-s \\ s \end{array} \\ \begin{array}{cc} q-s & s \end{array} & \end{array} = \begin{bmatrix} B_1B_1' + \tilde{B}_3\tilde{B}_3' & B_1B_2'+\tilde{B}_3B_4' \\ B_2B_1' + B_4\tilde{B}_3' & B_2B_2'+B_4B \end{bmatrix}$$

whence, from the relation: $B_1B_1'+\tilde{B}_3\tilde{B}_3' = M_{22}^{(1)}$, remembering that B_1 is already known and using (A.3.10), we see that $\tilde{B}_3$ is uniquely determined except for a post-factor $D_k(q-s)$. The equation $B_1B_2'+\tilde{B}_3B_4' = M_{22}^{(3)}$ now uniquely defines B_4 except for the post-factors that go with the other B_i's. This completes the proof of the near uniqueness in the case of distinct c's. If $s = p$, i.e., if M_{12} is of rank p, then all roots become positive, i.e., D_c becomes $p \times p$, A becomes a solid matrix while B retains its own structure with $q-s$ being replaced by $q-p$. If $q = p$, then B itself becomes a solid matrix. As before, for the case of distinct roots, the transformation is checked to be one-to-one by adopting the convention, say, that the first row of A_1 and the diagonal elements of $\tilde{A}_3$ and $\tilde{B}_3$ are to be positive.

(A.3.17): *If* $\begin{array}{cc} \begin{bmatrix} X_1 \\ \\ X_2 \end{bmatrix} & \begin{matrix} p \\ \\ q \end{matrix} \\ n & \end{array}$ $(p \leqslant q, p+q \leqslant n)$ *is of rank* $(p+q)$

and X_1X_2' *is also of rank* p, *then there exists a transformation*

$$X_2(q \times n) = \tilde{T}(q \times q)\, L_2(q \times n)$$

and $$X_1(p \times n) = U(p \times p)[D_{\sqrt{e}}(p \times p)M_1(p \times \overline{n-q}):\ M_2(p \times q)] \times \begin{array}{cc} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} & \begin{matrix} n-q \\ q \end{matrix} \\ n & \end{array}$$

where $\tilde{T}$ *and* U *are non-singular,* $M_1M_1' = M_2M_2' = I(p)$, $L_2L_2' = I(q)$, *and* L_1 *is a completion of* L_2 *(see (A.1.7)) such that* $\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$ *is* $\perp$ *and* $e_i = (1-c_i)/c_i$ *or* $c_i = 1/(1+e_i)$ $(i = 1, \ldots, p)$ *and* c *'s are the roots of the equation* c:

$$|c(X_1X_1')-(X_1X_2')(X_2X_2')^{-1}(X_2X_1')| = 0.$$

Proof: Using (A.3.11) put $X_2(q \times n) = \tilde{T}(q \times q)\, L_2(q \times n)$ where $\tilde{T}$ is non-singular and $L_2L_2' = I(q)$. Complete L_2 by an L_1 such that $\begin{array}{cc} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} & \begin{matrix} n-q \\ q \end{matrix} \\ n & \end{array}$ is $\perp$.

Now using (A.3.8), put

$$X_1(p \times n)[L_1'(n \times \overline{n-q})\ :\ L_2'(n \times q)] = U(p \times p)[D_{\sqrt{e}}(p \times p)\, M_1(p \times n-q)\ :\ M_2(p \times q)],$$

where U is non-singular, $M_1M_1' = M_2M_2' = I(p)$ and e's are the roots of the equation in e: $|(X_1L_1'L_1X_1')-e(X_1L_2'L_2X_1')| = 0$. Multiplying both sides of the X_1-equation by $\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$ and taking into account the X_2-equation we have the transformation (A.3.17), except for the required interpretation of e, which is as follows. $L_2 = (\tilde{T})^{-1}X_2$, so that $L_2'L_2 = X_2'(\tilde{T}\tilde{T}')^{-1}X_2$. Also $L_1'L_1 = I-L_2'L_2 = I-X_2'(\tilde{T}\tilde{T}')^{-1}X_2$. Hence the equation in e becomes: $|X_1[I-X_2'(\tilde{T}\tilde{T}')^{-1}X_2]X_1'-eX_1X_2'(\tilde{T}\tilde{T}')^{-1}X_2X_1'| = 0$ or $\left|\frac{1}{1+e} X_1X_1'-X_1X_2'(X_2X_2')^{-1}X_2X_1'\right| = 0$ (since $X_2X_2' = \tilde{T}\tilde{T}'$) which completes the proof of (A.3.17).

(A.3.18): *If* $X = \begin{array}{cc} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} & \begin{matrix} p \\ q \end{matrix} \\ n & \end{array}$ $(p \leqslant q, p+q \leqslant n;$ *rank* $= p+q)$ *is such that* X_1X_2' *is of rank* $s \leqslant p$ *(in which case it is easy to check that* X_1X_1' *and* X_2X_2' *are each* *symmetric p.d. and* $X_1X_2'(X_2X_2')^{-1}X_2X_1'$ *is symmetric and at least p.s.d. of rank s, so that s roots of the p-th degree equation in c:* $|c(X_1X_1')-(X_1X_2')(X_2X_2')^{-1}(X_2X_1')| = 0$ *are positive, the rest being zero), then there exists a transformation*

$$(\text{A.3.18.1}) \qquad X_1(p\times n) = \begin{array}{r@{}c} \begin{array}{r} p-s \\ s \end{array} & \begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} \\ & \begin{array}{cc} s & p-s \end{array} \end{array} \begin{array}{c@{}l} \begin{bmatrix} D_{\sqrt{1-c}} & 0 \\ 0 & I \end{bmatrix} & \begin{array}{l} s \\ p-s \end{array} \\ \begin{array}{cc} s & p-s \end{array} & \end{array} \begin{array}{c@{}l} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} & \begin{array}{l} s \\ p-s \end{array} \\ n & \end{array}$$

$$+ \begin{array}{r@{}c} \begin{array}{r} p-s \\ s \end{array} & \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \end{array} D_{\sqrt{c}}(s\times s)L_3(s\times n),$$

$$(\text{A.3.18.2}) \qquad X_2(q\times n) = \begin{array}{r@{}c} \begin{array}{r} q-s \\ s \end{array} & \begin{bmatrix} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{bmatrix} \\ & \begin{array}{cc} s & q-s \end{array} \end{array} \begin{array}{c@{}l} \begin{bmatrix} L_3 \\ L_4 \end{bmatrix} & \begin{array}{l} s \\ q-s \end{array} \\ n & \end{array},$$

where the $D_c(s\times s)$ *is based on the s positive roots of the equation already mentioned and where the A and B are non-singular matrices defined after* (A.3.16) *by*

$$(\text{A.3.18.3}) \qquad X_1X_1' = \begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} \begin{bmatrix} A_1' & A_2' \\ \tilde{A}_3' & A_4' \end{bmatrix},$$

$$X_2X_2' = \begin{bmatrix} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{bmatrix} \begin{bmatrix} B_1' & B_2' \\ \tilde{B}_3' & B_4' \end{bmatrix} \quad \textit{and}$$

$$X_1X_2' = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} D_{\sqrt{c}}[B_1' \quad B_2'] = \begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} \begin{array}{c@{}l} \begin{bmatrix} D_{\sqrt{c}} & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{l} s \\ p-s \end{array} \\ \begin{array}{cc} s & q-s \end{array} & \end{array} \begin{bmatrix} B_1' & B_2' \\ \tilde{B}_3' & B_4' \end{bmatrix}$$

and where the L matrices are subject to

$$\begin{array}{r@{}c} \begin{array}{r} s \\ p-s \\ s \\ q-s \end{array} & \begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{bmatrix} \\ & n \end{array} \begin{array}{c} [L_1' \quad L_2' \quad L_3' \quad L_4'] \\ \begin{array}{cccc} s & p-s & s & q-s \end{array} \end{array} n = I(p+q).$$

Proof:

$$(\text{A.3.18.4}) \quad \text{Put} \qquad X_1(p\times n) = \begin{array}{r@{}c} \begin{array}{r} p-s \\ s \end{array} & \begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} \\ & \begin{array}{cc} s & p-s \end{array} \end{array} \begin{array}{c@{}l} \begin{bmatrix} M \\ L_2 \end{bmatrix} & \begin{array}{l} s \\ p-s \end{array} \\ n & \end{array} \quad \text{and}$$

$$(\text{A.3.18.5}) \qquad X_2(q\times n) = \begin{array}{r@{}c} \begin{array}{r} q-s \\ s \end{array} & \begin{bmatrix} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{bmatrix} \\ & \begin{array}{cc} s & q-s \end{array} \end{array} \begin{array}{c@{}l} \begin{bmatrix} L_3 \\ L_4 \end{bmatrix} & \begin{array}{l} s \\ q-s \end{array} \\ n & \end{array}$$

Now check that, since A and B are non-singular, the above equation defines M, L_2, L_3, L_4 uniquely except for the indeterminacy in A and B. Now, using the first two equations of (A.3.18.3), it is easy to check that

(A.3.18.6)
$$\begin{bmatrix} M \\ L_2 \end{bmatrix} [M' : L_2'] = I(p) \quad \text{and} \quad \begin{bmatrix} L_3 \\ L_4 \end{bmatrix} [L_3' : L_4'] = I(q).$$

Substituting for X_1 and X_2 (in terms of the A B and M, L_2, L_3 and L_4) in the third equation of (A.3.18.3) we have

(A.3.18.7)
$$\begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} \begin{bmatrix} D_{\sqrt{c}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} B_1' & B_2' \\ \tilde{B}_3' & B \end{bmatrix}$$
$$= \begin{bmatrix} A_1 & A_3 \\ A_2 & A_4 \end{bmatrix} \begin{bmatrix} M \\ L_2 \end{bmatrix} [L_3' : L_4'] \begin{bmatrix} B_1' & B_2' \\ \tilde{B}_3' & B_4' \end{bmatrix}$$

whence it follows that

(A.3.18.8)
$$\begin{bmatrix} M \\ L_2 \end{bmatrix} [L_3' : L_4'] = \begin{array}{cc} \left[\begin{array}{cc} D_{\sqrt{c}} & 0 \\ 0 & 0 \end{array}\right] & \begin{array}{l} s \\ p-s \end{array} \\ \begin{array}{cc} s & \quad q-s \end{array} & \end{array}$$

Let us now put,

(A.3.18.9)
$$M(s \times n) = D_{\sqrt{c}}(s \times s)\, L_3(s \times n) + M_1(s \times n),$$

which uniquely defines M_1 in terms of M, L_3 and c's.

Now substituting in the equations (A.3.18.6) and (A.3.18.8) for M the right hand side of (A.3.18.9), we have

(A.3.18.10)
$$M_1[L_2' : L_3', : L_4'] = [0, \ 0, \ 0] \text{ and}$$

(A.3.18.11)
$$I(s) = MM' = D_{\sqrt{c}} L_3 L_3' D_{\sqrt{c}} + M_1 M_1' = D_c + M_1 M_1'.$$

If follows from (A.3.18.11) that

(A.3.18.12)
$$M_1 M_1' = I(s) - D_c = D_{1-c},$$

so that if we put

(A.3.18.13)
$$M_1(s \times n) = D_{\sqrt{1-c}}(s \times s) L_1(s \times n),$$

we shall have, from (A.3.18.12) and (A.3.18.10)

(A.3.18.14)
$$L_1 L_1' = I(s) \text{ and } L_1[L_2', \ L_3', \ L_4'] = [0, 0, 0].$$

Substituting from (A.3.18.13) for M_1 in (A.3.18.9) we have

(A.3.18.15) $$M(s\times n) = D_{\sqrt{c}}(s\times s)\, L_3(s\times n)+D_{\sqrt{1-c}}(s\times s)\, L_1(s\times n),$$

where L_1 satisfies (A.3.18.14).

Now substituting for M from (A.3.18.15) in (A.3.18.4) and using (A.3.18.6), (A.3.18.8) and (A.3.18.14) we have

(A.3.18.16) $$X_1=\begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} \begin{bmatrix} (D_{\sqrt{1-c}}\, L_1+D_{\sqrt{c}}\, L_3) \\ L_2 \end{bmatrix}$$

$$=\begin{bmatrix} A_1 & \tilde{A}_3 \\ A_2 & A_4 \end{bmatrix} \begin{bmatrix} D_{\sqrt{\ -c}} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} + \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} D_{\sqrt{c}}L_3 \text{ and}$$

(A.3.18.17) $$X_2=\begin{bmatrix} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{bmatrix} \begin{bmatrix} L_3 \\ L_4 \end{bmatrix},$$

where the L's satisfy

(A.3.18.18) $$\begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{bmatrix} [L_1' : L_2' : L_3' : L_4'] = I(p+q).$$

This proves (A.3.18). If $s = p$ (which is the case that will be actually considered in this monograph), L_2 will be absent, and $q-s = q-p$ and we shall have

(A.3.18.19) $$X_1(p\times n) = A(p\times p)[(\underset{p}{D_{\sqrt{1-c}}} : \underset{p}{D_{\sqrt{c}}})p] \begin{array}{cc} \begin{bmatrix} L_1 \\ L_3 \end{bmatrix} & \begin{array}{c} p \\ p \end{array} \\ n & \end{array}$$

and $$X_2(q\times n) = \begin{array}{cc} \begin{array}{c} q-p \\ p \end{array} & \begin{bmatrix} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{bmatrix} \\ & \begin{array}{cc} p & q-p \end{array} \end{array} \begin{array}{cc} \begin{bmatrix} L_3 \\ L_4 \end{bmatrix} & \begin{array}{c} p \\ q-p \end{array} \\ n & \end{array},$$

where the L's satisfy

(A.3.18.20) $$\begin{array}{cc} \begin{array}{c} p \\ p \\ q-p \end{array} & \begin{bmatrix} L_1 \\ L_3 \\ L_4 \end{bmatrix} \\ & n \end{array} \begin{array}{ccc} [L_1' & L_3' & L_4'] \\ p & p & q-p \end{array} n = I(p+q).$$

As to the indeterminacy on the right hand side of (A.3.18.1) and (A.3.18.2) (for the case $s < p$) and of (A.3.18.19) (for the case $s = p$), it is easy to check that in either case, if the non-zero roots are all distinct, there is near uniqueness in the sense of (A.3.16), the only indeterminacy arising out of a post-factor $D_k(p)$ going with the total A matrix and a post-factor $D_k(q-s)$ going with $\tilde{B}_3$. In this case the transformation can be made one-to-one by adopting the same convention as, say, at the end of (A.3.16).

(A.3.19): *For* $X = \begin{array}{cc} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} & \begin{array}{c} p \\ q \\ r \end{array} \\ n & \end{array}$ $(p \leqslant q, p+q+r \leqslant n,$ *of rank* $p+q+r)$,

(i) *there exists a transformation:* $X_3(r\times n) = \tilde{T}(r\times r)\, L_3(r\times n)$ *subject to* $L_3L_3' = I(r)$ *and*

$$X_1(p\times n) = p\begin{array}{cc}[Z_{11} & Z_{12}] \\ n-r & r\end{array} \begin{array}{cc} \begin{bmatrix} L \\ L_3 \end{bmatrix} & \begin{array}{c} n-r \\ r \end{array} \\ n & \end{array} \quad \text{and } X_2(q\times n) = q\begin{array}{cc}[Z_{21} \,\vdots & Z_{22}] \\ n-r & r\end{array} \begin{array}{cc} \begin{bmatrix} L \\ L_3 \end{bmatrix} & \begin{array}{c} n-r \\ r \end{array} \\ n & \end{array}$$

where L is just a completion of L_3 *so that* $\begin{bmatrix} L \\ L_3 \end{bmatrix}$ *is* $\perp$. (ii) *Putting* $M = XX'$ (*observe that, by* (A.1.10), *M will be symmetric p.d.*), *the roots of the equation in c, namely* (A.2.4.1) or (A.2.4.2) *are the same as the characteristic roots of* $(Z_{11}\,Z_{11}')^{-1}(Z_{11}\,Z_{21}') \times(Z_{21}\,Z_{21}')^{-1}(Z_{21}Z_{11}')$.

Proof: The proof of (i) is obvious from the preceding sections. For (ii) we observe that $L_3 = (\tilde{T})^{-1}X_3$ so that $L_3' = X_3'(\tilde{T}')^{-1}$ whence $L_3'L_3 = X_3'(\tilde{T}\tilde{T}')^{-1}\,X_3 = X_3'(X_3X_3')^{-1}X_3$. Therefore $L'L = I(n)-L_3'L_3 = I(n)-X_3'(X_3X_3')^{-1}X_3$ and thus

$$\begin{aligned} Z_{11}Z_{11}' &= X_1L'LX_1' = X_1X_1'-X_1X_3'(X_3X_3')^{-1}X_3X_1' = M_{11}-M_{13}M_{33}^{-1}M_{13}', \\ Z_{11}Z_{21}' &= X_1L'LX_2' = X_1X_2'-X_1X_3'(X_3X_3')^{-1}X_3X_2' = M_{12}-M_{13}M_{33}^{-1}M_{23}, \\ \text{and} \qquad Z_{21}Z_{21}' &= X_2L'LX_2' = X_2X_2'-X_2X_3'(X_3X_3{}')^{-1}X_3X_2' = M_{22}-M_{23}M_{33}^{-1}M_{23}'. \end{aligned}$$

This completes the proof of (ii)

(A.3.20): *For an M of the structure* (A.2.4) *there exists the transformation*

$$M = \begin{array}{cc} \begin{array}{c} p \\ q \\ r \end{array} & \left[\begin{array}{cc|c} A_1 & 0 & A_3 \\ 0 & A_2 & A_4 \\ \hline 0 & 0 & \tilde{A}_5 \end{array}\right] \\ & \begin{array}{ccc} p & q & r \end{array} \end{array} \begin{array}{cc} \left[\begin{array}{cc|c} I & [D_{\sqrt{\bar{c}}}\;\; 0] & \\ \begin{bmatrix} D_{\sqrt{c}} \\ \\ 0 \end{bmatrix} & I & 0 \\ \hline & 0 & I \end{array}\right] & \begin{array}{c} p \\ p \\ q-p \\ r \end{array} \\ \begin{array}{cccc} p & p & q-p & r \end{array} & \end{array} \left[\begin{array}{cc|c} A_1' & 0 & 0 \\ 0 & A_2' & 0 \\ \hline A_3' & A_4' & \tilde{A}' \end{array}\right]$$

where $\begin{bmatrix} A_1 & 0 & A_3 \\ 0 & A_2 & A_4 \\ 0 & 0 & \tilde{A}_5 \end{bmatrix}$ *is a non-singular matrix and c's are the roots of the equation in c,* (A.2.4.1) *or* (A.2.4.2).

Proof: We can write

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M'_{12} & M_{22} & M_{23} \\ M'_{13} & M'_{23} & M_{33} \end{bmatrix} = \begin{bmatrix} M_{11}-M_{13}M_{33}^{-1}M'_{13} & M_{12}-M_{13}M_{33}^{-1}M'_{23} & 0 \\ M'_{12}-M_{23}M_{33}^{-1}M'_{13} & M_{22}-M_{23}M_{33}^{-1}M'_{23} & 0 \\ 0, & 0 & 0 \end{bmatrix}$$

$$+\begin{bmatrix} M_{13}M_{33}^{-1}M'_{13} & M_{13}M_{33}^{-1}M'_{23} & M_{13} \\ M_{23}M_{33}^{-1}M'_{13} & M_{23}M_{33}^{-1}M'_{23} & M_{23} \\ M'_{13} & M'_{23} & M_{33} \end{bmatrix}$$

Using (A.3.9) and (A.3.16) we can now put $M_{33}=\tilde{A}_5\tilde{A}'_5$, $M_{11}-M_{13}M_{33}^{-1}M'_{13}=A_1A'_1$, $M_{22}-M_{23}M_{33}^{-1}M'_{23} = A_2A'_2$, and $M_{12}-M_{13}M_{33}^{-1}M'_{23} = A_1(p\times p)\ \underset{p}{[D_{\sqrt{c}}}\ \ \underset{q-p}{0]}\ A'_2(q\times q)$.

If we next put $M_{13}(p\times r) = A_3(p\times r)\tilde{A}'_5(r\times r)$ and $M_{23}(q\times r) = A_4(q\times r)\tilde{A}'_5(r\times r)$, we observe that A_3 and A_4 are determinate. We check furthermore that now $M_{13}M_{33}^{-1}M'_{13} = A_3A'_3$, $M_{23}M_{33}^{-1}M'_{23} = A_4A_4{}'$ and $M_{13}M_{33}^{1}M'_{23} = A_3A_4{}'$, so that altogether we have

$$M = \begin{bmatrix} A_1A'_1+A_3A'_3 & A_1[D_{\sqrt{c}}\ \ 0]A'_2+A_3\tilde{A}'_4 & A_3\tilde{A}'_5 \\ A_2\begin{bmatrix} D_{\sqrt{c}} \\ 0 \end{bmatrix}A'_1+A_4A'_3 & A_2A'_2+A_4A'_4 & A_4\tilde{A}'_5 \\ \tilde{A}_5A'_3 & \tilde{A}_5A'_4 & \tilde{A}_5\tilde{A}'_5 \end{bmatrix}$$

$$= \left[\begin{array}{cc|c} A_1 & 0 & A_3 \\ 0 & A_2 & A_4 \\ \hline 0 & 0 & \tilde{A}_5 \end{array}\right] \left[\begin{array}{cc|c} I & [D_{\sqrt{c}}\ \ 0] & 0 \\ \begin{bmatrix} D_{\sqrt{c}} \\ 0 \end{bmatrix} & I & \\ \hline 0 & & I \end{array}\right] \left[\begin{array}{cc|c} A'_1 & 0 & 0 \\ 0 & A'_2 & \\ \hline A'_3 & A'_4 & \tilde{A}'_5 \end{array}\right],$$

which proves (A.3.20).

(A.3.21): *The passage from L matrices to L_I variables.* Consider the transformations (A.3.6), (A.3.8), (A.3.11), (A.3.14), (A.3.15), (A.3.17) and (A.3.18) and (A.3.19) and notice that everywhere we have, on the right hand side, a post-factor of the form $L(p\times n)(p \leqslant n)$ subject to the constraint $LL' = I(p)$. Check that the actual number of independent constraints is just $p(p+1)/2$. Suppose now that instead of transforming to L subject to $LL' = I(p)$, we take a slightly different set of variates in the following way. Putting

$$L(p\times n) = \begin{bmatrix} l_{11} & \cdots & l_{1n} \\ . & \cdots & . \\ l_{p1} & \cdots & l_{pn} \end{bmatrix} = \begin{bmatrix} 1'_1 \\ . \\ 1'_p \end{bmatrix} \text{ (say),}$$

we notice that $LL' = I(p) \Longleftrightarrow \mathbf{l}'_i\mathbf{l}_j = \delta_{ij}$ $(i, j = 1,2, \ldots, p)$, the kronecker delta, so that, by virtue of the $p(p+1)/2$ constraints, L really consists of $pn-p(p+1)/2$ *independent* elements, although the $(p \times n)$ matrix itself is naturally one of pn elements. From L let us choose an independent set, say, $(l_{11}, l_{12}, \ldots l_{1,\,n-1}), (l_{21}, l_{22}, \ldots, l_{2,\,n-2}), \ldots, (l_{p1}, l_{p2} \ldots, l_{p,\,n-p})$ and let us call this set L_I. Throughout this monograph L_I will stand uniformly for this set of variates.

(A.3.22): *It will now be shown that if no elements of L are* 0, *then the correspondence between L_I and L is one-to-2^P.*

Proof: Having regard to the constraint $LL' = I(p)$, under our set-up, we are going to treat l_{ij} $(i = 1, 2, \ldots, p;\ j = 1, 2, \ldots, n-i)$ $(= L_I$ say) as the (so-called) independent variates and l_{ij} $(i = 1, 2, \ldots, p;\ j = n-i+1, \ldots, n)$ $(= L_D$ say) as the (so called) dependent variates. This notation will be uniformly followed. We have now the following equations in the dependent variates (in terms of the independent):

For the first row of the L matrix

$$\text{(A.3.22.1)} \qquad l_{1n}^2 = 1 - \sum_{j=1}^{n-1} l_{ij}^2.$$

For the 2-nd row of the L matrix

$$\text{(A.3.22.2)} \qquad l_{2,n-1}\, l_{1,n-1} + l_{2n} l_{1n} = -\sum_{j=1}^{n-2} l_{1j} l_{2j}\,; \quad l_{2,n-1}^2 + l_{2n}^2 = 1 - \sum_{j=1}^{n-2} l_{2j}^2.$$

And in general for the i-th row of the L matrix (with $i = 1, 2, \ldots, p$)

$$\text{(A.3.22.3)} \qquad \sum_{j=n-i+1}^{p} l_{ij}\, l_{i'j} = -\sum_{j=1}^{n-i} l_{ij} l_{i'j}, \quad \sum_{j=n-i+1}^{p} l_{ij}^2 = 1 - \sum_{j=1}^{n-i} l_{ij}^2,$$

$$\text{for} \quad i' = 1, 2, \ldots, i-1.$$

It is easy to see that, for the first row of L, the equation (A.3.22.1) gives (in this case) two real and distinct values of l_{1n} in terms of $(l_{11}, \ldots, l_{1.\,n-1})$. Next, for the second row of L, the equations (A.3.22.2) give (in this case) two real and distinct pairs of values for $(l_{2,\,n-1}, l_{2n})$ in terms of the first row (now supposed to be given), and so on. In general, for the i-th row of L, the equations (A.3.22.3) give (in this case) two real and distinct sets of values for $(l_{i,\,n-i+1}, \ldots, l_{ip})$ in terms of the $(i-1)$ previous rows (now supposed to be given). This proves (A.3.22).

APPENDIX 4

Invariance of the Characteristic Roots under Certain Linear Transformations

(A.4.1): *If $X(p\times n)$ $(p \leqslant n)$ is of rank p (in which case, by* (A.1.10), *XX is symmetric p.d.), then the characteristic roots of XX' are invariant under the transformation: $X(p\times n) = A(p\times p)\ Y(p\times n)\ B(n\times n)$ where A and B are any two $\perp$ matrices.*

Proof: $c(XX') = c(AYBB'Y'A') = c(AYY'A')$(since B is $\perp$)$= c(YY'A'A)$ (using (A.1.18)) $= c(YY')$ (since A is $\perp$), which completes the proof of (A.4.1).

(A.4.2): *If $X_1(p\times n_1)$, $X_2(p\times n_2)$ $(p \leqslant n_1, n_2)$ are each of rank p(in which case, by* (A.1.10), *X_1X_1' and $X_2X_2'$are both symmetric p.d.), then the characteristic roots of $(X_1X_1')(X_2X_2')^{-1}$ are invariant under the transformation: $X_1(p\times n_1) = A(p\times p)\ Y_1(p\times n_1)B_1(n_1\times n_1)$ and $X_2(p\times n_2) = A(p\times p)\ Y_2(p\times n_2)\ B_2(n_2\times n_2)$, where A is any non-singular matrix and B_1 and B_2 any two $\perp$ matrices.*

Proof: $c[(X_1X_1')(X_2X_2')^{-1}] = c[(AY_1B_1B_1'Y_1'A')(AY_2B_2B_2'Y_2'A')^{-1}]$

$= c[(AY_1Y_1'A')(AY_2Y_2'A')^{-1}]$ (since B_1 and B_2 are $\perp$)

$= c[A(Y_1Y_1')(Y_2Y_2')^{-1}A^{-1}] = c[(Y_1Y_1')(Y_2Y_2')^{-1}A^{-1}A]$

(using (A.1.18)), which completes the proof of (A.4.2).

(A.4.3): *If $X_1(p\times n_1)$ be of rank $n_1(\leqslant p)$ and $X_2(p\times n_2)(p \leqslant n_2)$ of rank p, then the characteristic roots of $(X_1X_1')(X_2X_2')^{-1}$ are invariant under the transformation: $X_1(p\times n_1) = A(p\times p)\ \ Y_1(p\times n_1)\ \ B_1(n_1\times n_1)$ and $X_2(p\times n_2) = A(p\times p) \times Y_2(p\times n_2)\ B_2(n_2\times n_2)$, where A is any non-singular matrix and B_1 and B_2 two arbitrary $\perp$ matrices. The proof is on the lines of that of* (A.4.2) *and is thus obvious.*

(A.4.4): *For* $X = \begin{array}{cc} \left[\begin{array}{c} X_1 \\ \\ X_2 \end{array}\right] & \begin{array}{c} p \\ \\ q \end{array} \\ n & \end{array}$ *$(p \leqslant q,\ p+q \leqslant n$, rank $= p+q)$ the characteristic roots of $(X_1X_1')^{-1}(X_1X_2')(X_2X_2')^{-1}(X_2X_1')$ are invariant under the transformation: $X_1(p\times n) = A_1(p\times p)\ Y_1(p\times n)\ B(n\times n)$ and $X_2(q\times n) = A_2(q\times q)\ Y_2(q\times n) \times B(n\times n)$, where A_1 and A_2 are any two non-singular matrices and B is any $\perp$ matrix.*

Proof: $c[(X_1X_1')^{-1}(X_1X_2')(X_2X_2')^{-1}(X_2X_1')]$

$= c[(A_1Y_1BB'Y_1'A_1')^{-1}(A_1Y_1BB'Y_2'A_2')(A_2Y_2BB'Y_2'A_2')^{-1}(A_2Y_2BB'Y_1'A_1')]$

$= c[(A_1Y_1Y_1'A_1')^{-1}(A_1Y_1Y_2'A_2')(A_2Y_2Y_2'A_2')^{-1}(A_2Y_2Y_1'A_1')]$(since B is $\perp$)

$= c[(A_1')^{-1}(Y_1Y_1')^{-1}(Y_1Y_2')(Y_2Y_2')^{-1}(Y_2Y_1')A_1']$

$= c[(Y_1Y_1')^{-1}(Y_1Y_2')(Y_2Y_2')^{-1}(Y_2Y_1')A_1'(A_1')]^{-1}$ (using (A.1.18))

$= c[(Y_1Y_1')^{-1}(Y_1Y_2')(Y_2Y_2')^{-1}(Y_2Y_2')]$, which proves (A.4.4).

(A.4.5): *For* $X = \begin{array}{cc} \left[\begin{array}{c} X_1 \\ X_2 \\ X_3 \end{array}\right] & \begin{array}{c} p \\ q \\ r \end{array} \\ n & \end{array}$ *$(p \leqslant q, p+q+r \leqslant n$, rank$=p+q+r)$,* *the roots of the equation in c of the form* (A.2.4.1) *i.e., of*

$$\begin{matrix} c & \begin{bmatrix} X_1X_1' & X_1X_3' \\ X_3X_1' & X_3X_3' \end{bmatrix} & \begin{bmatrix} X_1X_2' & X_1X_3' \\ X_3X_2' & X_3X_3' \end{bmatrix} \\ & \begin{bmatrix} X_2X_1' & X_2X_3' \\ X_3X_1' & X_3X_3' \end{bmatrix} & \begin{bmatrix} X_2X_2' & X_2X_3' \\ X_3X_2' & X_3X_3' \end{bmatrix} \end{matrix} = 0$$

i.e., of

(A.4.5.1)
$$\begin{matrix} c & \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} [X_1' \quad X_3'] & \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} [X_2' \quad X_3'] \\ & \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} [X_1' \quad X_3'] & \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} [X_2' \quad X_3'] \end{matrix} = 0$$

are invariant under the transformation

$$\begin{array}{cc} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} & \begin{matrix} p \\ q \\ r \end{matrix} \\ n & \end{array} = \begin{array}{cc} \begin{matrix} p \\ q \\ r \end{matrix} & \begin{bmatrix} A_1 & 0 & A_3 \\ 0 & A_2 & A_4 \\ 0 & 0 & A_5 \end{bmatrix} \\ & \begin{matrix} p & \quad q & \quad r \end{matrix} \end{array} \begin{array}{cc} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} & \begin{matrix} p \\ q \\ r \end{matrix} \\ n & \end{array} \times B(n \times n),$$

where B *is* $\perp$ *and* $A = \begin{bmatrix} A_1 & 0 & A_3 \\ 0 & A_2 & A_4 \\ 0 & 0 & A_5 \end{bmatrix}$ *is any non-singular matrix.*

Proof: The proof follows by noting that B will pass out of the picture and the equation (A.4.5.1) can be written in terms of Y's and A's as

$$\begin{vmatrix} \begin{bmatrix} A_1 & A_3 \\ 0 & A_5 \end{bmatrix} & 0 \\ 0 & \begin{bmatrix} A_2 & A_4 \\ 0 & A_5 \end{bmatrix} \end{vmatrix} \begin{vmatrix} c \begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix} [Y_2' : Y_3'] & \begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix} [Y_2' : Y_3'] \\ \begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix} [Y_1' : Y_3'] & \begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix} [Y_2' : Y_3'] \end{vmatrix}$$

$$\times \begin{vmatrix} \begin{bmatrix} A_1' & 0 \\ A_3' & A_5' \end{bmatrix} & 0 \\ 0 & \begin{bmatrix} A_2' & 0 \\ A_4' & A_5' \end{bmatrix} \end{vmatrix} = 0 \;;$$

then, since A is non-singular, $\begin{bmatrix} A_1 & A_3 \\ 0 & A_5 \end{bmatrix}$ and $\begin{bmatrix} A_2 & A_4 \\ 0 & A_5 \end{bmatrix}$ are both easily checked to be non-singular.

APPENDIX 5

Some General Theorems in Jacobians

(A.5.1): *If* $\mathbf{x}(n \times 1) = A(n \times n)\, \mathbf{y}\, (n \times 1)$, *where* A *is non-singular, then* $J(\mathbf{x} : \mathbf{y}) = |A|$.

(A.5.2): *If* $X(m \times n) = A(m \times m)\ Y(m \times n)$, *where* A *is non-singular, then* $J(X : Y) = |A|^n$.

(A.5.3): *If* $X(m \times m) = A(m \times m)\ Y(m \times n)\ B(n \times n)$, *where* A *and* B *are non-singular, then* $J(X : Y) = |A|^n |B|^m$.

(A.5.4): *If* A *and* B *are each* $\perp$, *then* $|A| = |B| = 1$ *and* (A.5.1) *and* (A.5.2)—(A.5.3) *will reduce respectively to* $J(\mathbf{x} : \mathbf{y}) = 1$ *and* $J(X : Y) = 1$.

(A.5.5): *If* $y_i = f_i(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n})$ $(i = 1, \ldots, m)$ *where* x_j*'s* $(j = 1, 2, \ldots, m+n)$ *are subject to* n *constraints*

$$f_i(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n}) = 0 \qquad (i = m+1, \ldots, m+n),$$

then (under the usual conditions for the existence of the Jacobian, including the non-vanishing of the numerator and the denominator in the following) we have, [42],

$$J(y_1, \ldots, y_m : x_1, \ldots, x_m) = \frac{\partial(f_1, \ldots, f_m, f_{m+1}, \ldots, f_{m+n})}{\partial(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n})} \div \frac{\partial(f_{m+1}, \ldots, f_{m+n})}{\partial(x_{m+1}, \ldots, x_{m+n})}.$$

Proof: Let us denote by $\frac{\partial y_i}{\partial x_j}$, $i, j, = 1, \ldots, m$, the partial differential coefficient of y_i with reespect to x_j after having expressed y_i $(i = 1, \ldots, m)$ in terms of $(x_1, \ldots, x_m)$, that is, after eliminating $(x_{m+1}, \ldots, x_{m+n})$ with the help of the constraints. Next denote by

$$\left|\frac{\partial y_i}{\partial x_j}\right|, \ (i, j = 1, 2, \ldots, m),$$

the absolute value of the determinant of the $m \times m$ (square) matrix

$$\left[\frac{\partial y_i}{\partial x_j}\right], \quad (i, j = 1, 2, \ldots, m).$$

Then we have

$$\text{(A.5.5.1)} \qquad J(y_1, \ldots, y_m : x_1, \ldots, x_m) = \left.\frac{\partial y_i}{\partial x_j}\right|, \quad (i, j = 1, 2, \ldots, m),$$

$$= \left|\frac{\partial f_i}{\partial x_j} + \sum_{k=m+1}^{m+n} \frac{\partial f_i}{\partial x_k} \frac{\partial x_k}{\partial x_j}\right|, \quad (i, j = 1, 2, \ldots, m).$$

Notice that in $\frac{\partial f_i}{\partial x_j}$ or $\frac{\partial f_i}{\partial x_k}$, f_i is supposed to be expressed in terms of all the $(m+n)$ x's and the partial differentiation is supposed to be with respect to x_j or x_k assuming all the other $(m+n-1)$ independent variates to be kept fixed, while in $\frac{\partial y_i}{\partial x_j}$ or $\frac{\partial x_k}{\partial x_j}$ it is supposed that y_i $(i = 1, 2, \ldots, m)$ or x_k $(k = m+1, \ldots, m+n)$ has first been expressed in terms of x_j's $(j = 1, 2, \ldots, m)$ and then the partial differentiation is made with respect to a particular x_j, assuming the other $(m-1)$ 'independent' variates to be kept fixed. Now from the set of n constraints on x_j's $(j = 1, 2, \ldots, m+n)$ given by the conditions of (A.5.5) we have

(A.5.5.2)
$$\frac{\partial f_i}{\partial x_j} + \sum_{k=m+1}^{m+n} \frac{\partial f_i}{\partial x_k} \frac{\partial x_k}{\partial x_j} = 0 \quad (i = m+1, \ldots, m+n, \text{ and } j = 1, \ldots, m),$$

or, in matrix notation,

(A.5.5.3)
$$-\left[\frac{\partial f_i}{\partial x_j}\right] = \left[\frac{\partial f_i}{\partial x_k}\right]\left[\frac{\partial x_k}{\partial x_j}\right] \quad (i, k = m+1, \ldots, m+n; j = 1, \ldots, m), \text{ or}$$

$$\left[\frac{\partial x_k}{\partial x_j}\right] = -\left[\frac{\partial f_i}{\partial x_k}\right]^{-1}\left[\frac{\partial f_i}{\partial x_j}\right]$$

(note that, by the conditions of (A.5.5), $\left[\frac{\partial f_i}{\partial x_k}\right]$ can be assumed to be non-singular).

Substituting from (A.5.5.3) in (A.5.5.1) we have

(A.5.5.4) $J(y_1, \ldots, y_m : x_1, \ldots, x_m)$:

$$= \left| \left[\frac{\partial f_i}{\partial x_j}\right] - \left[\frac{\partial f_i}{\partial x_k}\right]_{\substack{i = 1, \ldots, m \\ k = m+1, \ldots, m+n}} \left[\frac{\partial f_l}{\partial x_k}\right]^{-1} \left[\frac{\partial f_l}{\partial x_j}\right]_{\substack{l = m+1, \ldots, m+n \\ j = 1, \ldots, m}} \right|$$

$$= \left| \begin{matrix} \left[\frac{\partial f_i}{\partial x_j}\right]_{i, j = 1, \ldots, m} & \left[\frac{\partial f_i}{\partial x_k}\right]_{\substack{i = 1, \ldots, m \\ k = m+1, \ldots, m+n}} \\ \left[\frac{\partial f_l}{\partial x_j}\right]_{\substack{l = m+1, \ldots, m+n \\ j = 1, \ldots, m}} & \left[\frac{\partial f_l}{\partial x_k}\right]_{l, k = m+1, \ldots, m+n} \end{matrix} \right| \div$$

$$\div \left| \left[\frac{\partial f_l}{\partial x_k}\right] \right|_{k, l = m+1, \ldots, m+n} \quad \text{(by using (A.1.1))}$$

$$= \left| \left[\frac{\partial f_i}{\partial x_j}\right] \right|_{i, j = 1, \ldots, m+n} \div \left| \left[\frac{\partial f_i}{\partial x_{\cdot}}\right] \right|_{i, j = m+1, \ldots. m+n}$$

which proves (A.5.5).

The real use of this theorem (as also of the next one) is in those situations where it would be difficult to express y_i's in terms of x_j's ($j = 1, \ldots, m$) (after elimination of $x_{m+1}, \ldots, x_{m+n}$ with the help of the constraints), but where it is much easier to express the right hand side of (A.5.5.4) in terms of $(x_1, \ldots, x_m)$, or where even this explicit expression is not directly needed.

(A.5.6): *If* $F_i(y_1, \ldots, y_m; x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n}) = 0$ $(i = 1, 2, \ldots, m+n)$ *are a set of equations solvable in the real domain in the sense that corresponding to real* $(x_1, \ldots, x_m)$ *we can find real* $(y_1, \ldots, y_m)$ *and* $(x_{m+1}, \ldots, x_{m+n})$, *then, under the other usual conditions for the existence of the Jacobian (including the non-vanishing of the numerator and the denominator in the following), we have,* [42]

$$J(y_1, \ldots, y_m : x_1 \ldots, x_m) = \frac{\partial(F_1, \ldots, F_{m+n})}{\partial(x_1, \ldots, x_{m+n})} \div \frac{\partial(F_1, \ldots, F_{m+n})}{\partial(y_1, \ldots, y_m; x_{m+1}, \ldots, x_{m+n})}.$$

Proof: As before we have

$$J(y_1, \ldots, y_m : x_1, \ldots, x_m) = \left|\frac{\partial y_i}{\partial x_j}\right| i, j = 1, \ldots, m.$$

But from the basic conditions of (A.5.6) we have

$$\text{(A.5.6.1)} \qquad \sum_{i=1}^{m} \frac{\partial F_k}{\partial y_i}\frac{\partial y_i}{\partial x_j} + \sum_{l=m+1}^{m+n} \frac{\partial F_k}{\partial x_l}\frac{\partial x_l}{\partial x_j} + \frac{\partial F_k}{\partial x_j} = 0$$

$$(k = 1, \ldots, m+n; j = 1, 2, \ldots, m).$$

Notice that in $\frac{\partial F_k}{\partial y_i}$ and $\frac{\partial F_k}{\partial x_l}$, $\frac{\partial F_k}{\partial x_j}$, F_k is supposed to be expressed in terms of all the $(2m+n)$ variates $(y_1, \ldots, y_m, x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n})$, and the partial differentiation is with respect to y_i or x_l or x_j, keeping all the other $(2m+n)-1$ variates fixed.

Also notice that, in $\frac{\partial y_i}{\partial x_j}$ or $\frac{\partial x_l}{\partial x_j}$, y_i (or x_l) $(i = 1, 2, \ldots, m;\ l = m+1, \ldots, m+n)$ is supposed to be expressed in terms of x_j's ($j = 1, 2. \ldots, m$) and then the partial differentiation is with respect to a particular x_j, keeping all the other $(m-1)$ of the x_j's fixed.

(A.5.6.1) can be written as

$$\text{(A.5.6.2)} \qquad \left[\frac{\partial F_k}{\partial y_i}\right]\left[\frac{\partial y_i}{\partial x_j}\right] + \left[\frac{\partial F_k}{\partial x_l}\right]\left[\frac{\partial x_l}{\partial x_j}\right] + \left[\frac{\partial F_k}{\partial x_j}\right] = 0,$$

where each side is a $(m+n) \times m$ matrix.

Taking, say, the first m rows of this matrix equation (A.5.6.2) we shall have the square $(m \times m)$ matrix equation:

$$\left[\frac{\partial F_k}{\partial y_i}\right]\left[\frac{\partial y_i}{\partial x_j}\right]+\left[\frac{\partial F_{k_1}}{\partial x_l}\right]\left[\frac{\partial x_l}{\partial x_j}\right]+\left[\frac{\partial F_{k_1}}{\partial x_j}\right]=0 \tag{A.5.6.3}$$

where now $i, j, k_1 = 1, 2, \ldots, m$, and $l = m+1, \ldots, m+n$, and $\left[\frac{\partial F_k}{\partial y}\right]$ is square $(m \times m)$.

Again taking the last n rows of the matrix equation (A.5.6.2) we have

$$\left[\frac{\partial F_{k_2}}{\partial y_i}\right]\left[\frac{\partial y_i}{\partial x_j}\right]+\left[\frac{\partial F_{k_2}}{\partial x_l}\right]\left[\frac{\partial x_l}{\partial x_j}\right]+\left[\frac{\partial F_{k_2}}{\partial x_j}\right]=0, \tag{A.5.6.4}$$

where now $i, j = 1, 2, \ldots, m$ and $k_2, l = m+1, \ldots, m+n$, so that $\left[\frac{\partial F_{k_2}}{\partial x_l}\right]$ is now square $(n \times n)$.

Treating (A.5.6.3) and (A.5.6.4) as a pair of simultaneous equations in $\left[\frac{\partial y_i}{\partial x_j}\right]$ $(i, j = 1, \ldots, m)$ and $\left[\frac{\partial x_l}{\partial x_j}\right]$ $(l = m+1, \ldots, m+n$ and $j = 1, \ldots, m)$, and solving for them we have for $\left[\frac{\partial y_i}{\partial x_j}\right]$ the following:

$$\left[\frac{\partial y_i}{\partial x_j}\right]=-\left\{\left[\frac{\partial F_{k_1}}{\partial y_i}\right]-\left[\frac{\partial F_{k_1}}{\partial x_l}\right]\left[\frac{\partial F_{k_2}}{\partial x_l}\right]^{-1}\left[\frac{\partial F_{k_2}}{\partial y_i}\right]\right\}^{-1} \tag{A.5.6.5}$$

$$\times\left\{\left[\frac{\partial F_{k_1}}{\partial x_j}\right]-\left[\frac{\partial F_{k_1}}{\partial x_l}\right]\left[\frac{\partial F_{k_2}}{\partial x_l}\right]^{-1}\left[\frac{\partial F_{k_2}}{\partial x_j}\right]\right\}.$$

Hence

$$\left|\frac{\partial y_i}{\partial x_j}\right|=\left\|\left[\frac{\partial F_{k_1}}{\partial x_j}\right]-\left[\frac{\partial F_{k_1}}{\partial x_l}\right]\left[\frac{\partial F_{k_2}}{\partial x_l}\right]^{-1}\left[\frac{\partial F_k}{\partial x_j}\right]\right\|$$

$$\div\left\|\left[\frac{\partial F_{k_1}}{\partial y_i}\right]-\left[\frac{\partial F_{k_1}}{\partial x_l}\right]\left[\frac{\partial F_l}{\partial x_l}\right]^{-1}\left[\frac{\partial F_{k_2}}{\partial y_i}\right]\right\|.$$

But we have by (A.1.1),

$$\begin{array}{c|cc|} & m & n \\ m & \left[\frac{\partial F_{k_1}}{\partial x_j}\right] & \left[\frac{\partial F_{k_1}}{\partial x_l}\right] \\ n & \left[\frac{\partial F_{k_2}}{\partial x_j}\right] & \left[\frac{\partial F_{k_2}}{\partial x_l}\right] \end{array}=\left|\left[\frac{\partial F_{k_2}}{\partial x_l}\right]\right|\left|\left[\frac{\partial F_{k_1}}{\partial x_j}\right]-\left[\frac{\partial F_{k_1}}{\partial x_l}\right]\left[\frac{\partial F_{k_2}}{\partial x_l}\right]^{-1}\left[\frac{\partial F_{k_2}}{\partial x_j}\right]\right| \tag{A.5.6.6}$$

and

$$\begin{array}{c|cc|} & m & n \\ m & \left[\dfrac{\partial F_{k_1}}{\partial y_i}\right] & \left[\dfrac{\partial F_{k_1}}{\partial x_l}\right] \\ n & \left[\dfrac{\partial F_{k_2}}{\partial y_i}\right] & \left[\dfrac{\partial F_{k_2}}{\partial x_l}\right] \end{array} = \left|\left[\frac{\partial F_{k_2}}{\partial x_l}\right]\right| \left|\left[\frac{\partial F_{k_1}}{\partial y_i}\right] - \left[\frac{\partial F_{k_1}}{\partial x_l}\right]\left[\frac{\partial F_{k_2}}{\partial x_l}\right]^{-1}\left[\frac{\partial F_{k_2}}{\partial y_i}\right]\right|.$$

Substituting from (A.5.6.6) in (A.5.6.5) we have

(A.5.6.7) $$J(y_1, \ldots, y_m : x_1, \ldots\ x_m) = \left|\frac{\partial y_i}{\partial x_j}\right|$$

$$= \begin{array}{c|cc|} & m & n \\ m & \left[\dfrac{\partial F_{k_1}}{\partial x_j}\right] & \left[\dfrac{\partial F_{k_1}}{\partial x_l}\right] \\ n & \left[\dfrac{\partial F_{k_2}}{\partial x_j}\right] & \left[\dfrac{\partial F_{k_2}}{\partial x_l}\right] \end{array} \div \begin{array}{c|cc|} & m & n \\ m & \left[\dfrac{\partial F_{k_1}}{\partial y_i}\right] & \left[\dfrac{\partial F_{k_1}}{\partial x_l}\right] \\ n & \left[\dfrac{\partial F_{k_2}}{\partial y_i}\right] & \left[\dfrac{\partial F_{k_2}}{\partial x_l}\right] \end{array}$$

$$= \frac{\partial(F_1, \ldots, F_{m+n})}{\partial(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n})} \div \frac{\partial(F_1, \ldots, F_{m+n})}{\partial(y_1, \ldots, y_m, x_{m+1}, \ldots, x_{m+n})}$$

which proves (A.5.6).

(A.5.5) is really a special case of (A.5.6), which can be shown by putting in (A.5.6), $F_i = y_i - f_i(x_1, \ldots, x_{m+n})$ $(i = 1, 2, \ldots, m)$ and next $F_i = f_i(x_1, \ldots, x_{m+n})$ $(i = m+1, \ldots, m+n)$, that is, by assuming that the last n equations are free from the y_i's. Substituting in the right hand side of (A.5.6.7), we easily check that it goes over into the right hand side of (A.5.5.4).

It seems that (A.5.6) is a very general theorem in Jacobians and yields as special cases practically all the usual well-known Jacobian theorems.

APPENDIX 6

Jacobians of Certain Specific Transformations

We shall consider the transformations (A.3.6), (A.3.8), (A.3.11) with rank $= p$, (A.3.14), (A.3.15), (A.3.17) and (A.3.18.19) and, in each case, pass on to L_I from the postfactor and prefactor of the form L or M (subject to $LL' = I$) and discuss, for the different cases, the respective Jacobians (i) $J(X : M_I, c\text{'s}, L_I)$, (ii) $J(X_1, X_2 : A, c\text{'s}, L_{1I}, L_{2I})$, (iii) $J(X : \tilde{T}, L_I)$, (iv) $J(X_1, X_2 : U_1, U_2, \tilde{U}_3, U_4, c\text{'s}, L_{1I}, L_{2I})$, (v) $J(X_1, X_2 : \tilde{T}, c\text{'s}\ L_I, L_{1I}, L_{2I})$, (vi) $J(X_1, X_2 : \tilde{T}, U, c\text{'s}\ M_{1I}, M_{2I}, L_{2I})$ and (vii) $J(X_1, X_2 : A, B_1, B_2, B_3, B_4, c\text{'s}, L_I)$, where, in (vii), L_I's are respectively the (so-called) independent elements formed, as in section (A.3.21), out of the matrices

$$\begin{array}{c|c} \begin{bmatrix} L_1 \\ L_3 \\ L_4 \end{bmatrix} & \begin{array}{c} p \\ p \\ q-p \end{array} \\ n & \end{array}$$

We shall first obtain the following two Jacobians which will be basic to the derivations of all the other ones.

(A.6.1): *Jacobian of the transformation* (A.3.11) (*with rank*$=p$), *i.e.*, $J(X : \tilde{T}, L_I)$ *where* $\tilde{T}(p \times p)$ *is non-singular with a positive diagonal.* To obtain the Jacobian from X to $\tilde{T}$ and L_I we use (A.5.5), remembering that now $X = \tilde{T}L$ takes the place of $y_i = f_i$ and $LL' - I(p) = 0$ takes the place of $f_i = 0$. We also note that $d(LL' - I(p)) = d(LL')$. We have now, using (A.5.5),

$$\text{(A.6.1.1)} \qquad J(X : \tilde{T}, L_I) = \left|\frac{\partial(X, LL')}{\partial(\tilde{T}, L)}\right| \div \left|\frac{\partial(LL')}{\partial(L_D)}\right| = \left|\frac{\partial(X, LL')}{\partial(\tilde{T}, L)}\right|_{\tilde{T}, L_I} \div \left|\frac{\partial(LL')}{\partial(L_D)}\right|_{L_I},$$

where on the extreme right, for practical usability, everything is expressed in terms of $\tilde{T}$ and L_I. The calculation of the numerator in the Jacobian of (A.6.1), (A.6.2) and (A.6.7) can be considerably abridged by expressing, in each case, that numerator in terms of Kronecker products (and sums) of matrices, otherwise known as direct products and direct sums. However, in this monograph, for expository purposes, a more familiar and straightforward, but lengthier method is given in each case. It is hoped that, for each problem, the reader will have no difficulty in verifying the main steps by spelling out in further detail on a sheet of paper. To calculate the numerator of (A.6.1.1) we proceed as follows.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \cdot & \cdots & \cdot \\ x_{p1} & \cdots & x_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \cdot \\ \mathbf{x}_p' \end{bmatrix} \text{ (say);}$$

$$L = \begin{bmatrix} l_{11} & \cdots & l_{1n} \\ \cdot & \cdots & \cdot \\ l_{p1} & \cdots & l_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{l}_1' \\ \cdot \\ \mathbf{l}_p' \end{bmatrix} \text{ (say).}$$

Also put $LL' = K$ with elements k_{ij} $(i, j = 1, 2, \ldots, p;\ k_{ij} = k_{ji})$. Then (A.3.11) can be written as $\mathbf{x}_i' = (t_{i1} \cdots t_{ii}\ 0 \ldots 0) \times L$ $(i = 1, 2, \ldots, p)$ or

$$\text{(A.6.1.2)} \qquad \mathbf{x}_i = L' \begin{bmatrix} t_{i1} \\ \cdot \\ t_{ii} \\ 0 \\ \cdot \\ 0 \end{bmatrix} = [\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_i] \cdot \begin{bmatrix} t_{i1} \\ t_{i2} \\ \cdot \\ t_{ii} \end{bmatrix} \quad (i = 1, 2, \ldots, p).$$

To calculate $\dfrac{\partial(X, LL')}{\partial(\tilde{T}, L)} = \dfrac{\partial(X, K)}{\partial(\tilde{T}, L)}$ we display below the partial differential coefficients of X and K $(= LL')$ with respect to the elements of $\tilde{T}$ and L (all elements of L being temporarily regarded as independent for purpose of the present differentiation):

$$\begin{array}{r|cccccccccc|cccccc}
 & t_{11} & t_{21} & \cdot & t_{p1} & t_{22} & \cdot & t_{p2} & \cdot & \cdot & t_{pp} & \mathbf{l}_1' & \mathbf{l}_2' & \cdot & \cdot & \mathbf{l}_{p-1}' & \mathbf{l}_p' \\
\hline
\mathbf{x}_1 & \mathbf{l}_1 & 0 & \cdot & 0 & 0 & \cdot & 0 & \cdot & \cdot & 0 & D_{t_{11}} & 0 & \cdot & \cdot & \cdot & 0 \\
\mathbf{x}_2 & 0 & \mathbf{l}_1 & \cdot & 0 & \mathbf{l}_2 & \cdot & 0 & \cdot & \cdot & 0 & D_{t_{21}} & D_{t_{22}} & 0 & \cdot & \cdot & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\mathbf{x}_p & 0 & \cdot & \cdot & \mathbf{l}_1 & 0 & \cdot & \mathbf{l}_2 & \cdot & \cdot & \mathbf{l}_p & D_{t_{p1}} & D_{t_{p2}} & \cdot & \cdot & D_{t_{p,p-1}} & D_{t_{p,p}} \\
\hline
k_{11} & & & & & & & & & & & 2\mathbf{l}_1' & 0 & \cdot & \cdot & \cdot & 0 \\
\cdot & & & & & & & & & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
k_{pp} & & & & & & & & & & & 0 & \cdot & \cdot & \cdot & \cdot & 2\mathbf{l}_p' \\
k_{12} & & & & & & & & & & & \mathbf{l}_2' & \mathbf{l}_1' & 0 & \cdot & \cdot & 0 \\
\cdot & \multicolumn{10}{c|}{0} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
k_{1p} & & & & & & & & & & & \mathbf{l}_p' & 0 & \cdot & \cdot & 0 & \mathbf{l}_1' \\
\cdot & & & & & & & & & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
k_{p-1,p} & & & & & & & & & & & 0 & \cdot & \cdot & \cdot & \mathbf{l}_p' & \mathbf{l}_{p-1}'
\end{array}$$

where D_a will stand for a diagonal matrix with diagonal elements all equal to a. Recall that $\mathbf{x}_i'$ is $1 \times n$, $\mathbf{l}_i'$ is also $1 \times n (i = 1, \ldots, p)$ and $K(p \times p)$ has $p(p+1)/2$

independent elements so that the above is really a $(np+p(p+1)/2)\times(np+p(p+1)/2)$ matrix. Now put

$$M_{11}(p(p+1)/2\times p(p+1)/2) = \mathbf{0}; \tag{A.6.1.3}$$

$$M_{12}(p(p+1)/2\times np) = \begin{bmatrix} \mathbf{1}_1' & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \cdot & \cdot & \cdot & \mathbf{1}_p' \\ \mathbf{1}_2' & \mathbf{1}_1' & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \cdot & \cdot & \mathbf{1}_p' & \mathbf{1}_{p-1}' \end{bmatrix};$$

$$M_{21}(np\times p(p+1)/2) = \left[\begin{array}{cccc|cccc|c|c} \mathbf{1}_1 & \mathbf{0} & \cdot & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_1 & \cdot & \cdot\,\cdot & \mathbf{1}_2 & \mathbf{0} & \cdot & \mathbf{0} & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \mathbf{0} & \cdot & \cdot & \mathbf{1}_1 & \mathbf{0} & \cdot & \cdot & \mathbf{1}_2 & \cdots & \mathbf{1}_p \end{array}\right]; \text{ and}$$

$$M_{22}(np\times np) = \begin{bmatrix} D_{t_{11}} & \mathbf{0} & \cdot & \mathbf{0} & \mathbf{0} \\ D_{t_{21}} & D_{t_{22}} & \cdot & \mathbf{0} & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ D_{t_{p1}} & D_{t_{p2}} & \cdot & \cdot & D_{t_{pp}} \end{bmatrix} \text{(notice that each } D \text{ is } n\times n).$$

By (A.1.1) we shall now have

$$\frac{\partial(X,K)}{\partial(\tilde{T},L)} = 2^p \begin{array}{|cc|l} M_{11} & M_{12} & p(p+1)/2 \\ M_{21} & M_{22} & np \\ \multicolumn{1}{c}{\frac{p(p+1)}{2}} & \multicolumn{1}{c}{np} & \end{array} = 2^p \begin{vmatrix} \mathbf{0} & M_{12} \\ M_{21} & M_{22} \end{vmatrix} \tag{A.6.1.4}$$

$$= 2^p |M_{22}|\ |\mathbf{0}-M_{12}\ M_{22}^{-1}\ M_{21}| = 2^p\ |M_{22}|\ |M_{12}\ M_{22}^{-1}\ M_{21}|.$$

Recalling the structure of $\tilde{T}$ we have

$$\tilde{T}^{-1} = \begin{bmatrix} t^{11} & 0 & . & 0 \\ t^{12} & t^{22} & . & 0 \\ . & . & . & . \\ t^{1p} & t^{2p} & . & t^{pp} \end{bmatrix}, \quad M_{22}^{-1} = \begin{bmatrix} D_{t^{11}} & 0 & . & 0 \\ D_{t^{12}} & D_{t^{22}} & . & 0 \\ . & . & . & . \\ D_{t^{1p}} & D_{t^{2p}} & . & D_{t^{pp}} \end{bmatrix},$$

so that $|M_{22}| = |\tilde{T}|^n$. We have furthermore

(A.6.1.5)
$$M_{22}^{-1}\mathrm{M}_{21} = \left[\begin{array}{cccc|ccc|c|c} \mathbf{l}_1 t^{11} & 0 & . & 0 & 0 & . & 0 & . & 0 \\ \mathbf{l}_1 t^{12} & \mathbf{l}_1 t^{22} & . & 0 & \mathbf{l}_2 t^{22} & . & 0 & . & 0 \\ . & . & . & . & . & . & . & . & . \\ \mathbf{l}_1 t^{1p} & \mathbf{l}_1 t^{2p} & . & \mathbf{l}_1 t^{pp} & \mathbf{l}_2 t^{2p} & . & \mathbf{l}_2 t^{pp} & . & \mathbf{l}_p t^{pp} \end{array}\right]$$

$$= \left[\begin{array}{cccc|ccc|c|c} \mathbf{l}_1 & 0 & . & 0 & 0 & . & 0 & . & 0 \\ 0 & \mathbf{l}_1 & . & 0 & \mathbf{l}_2 & . & 0 & . & . \\ . & . & . & . & . & . & . & . & . \, . \\ 0 & . & . & \mathbf{l}_1 & 0 & . & \mathbf{l}_2 & . & \mathbf{l}_p . \end{array}\right] \left[\begin{array}{ccc|ccc|c|c} t^{11} & . & 0 & & & & & \\ . & . & . & & 0 & & . & 0 \\ t^{1} & {}^{p}. & t^{pp} & & & & & \\ \hline & & & t^{22} & . & 0 & & \\ & 0 & & . & . & . & . & 0 \\ & & & t^{2p} & . & t^{pp} & & \\ \hline & . & & . & & & . & 0 \\ \hline & & 0 & & & & & t^{pp} \end{array}\right]\begin{array}{l} \\ p \\ \\ \\ p-1 \\ \\ \\ 1 \end{array}$$

$= M_{21}N_{22}$ (say) where N_{22} stands for the right matrix factor. We note that N_{22} is $p(p+1)/2 \times p(p+1)/2$ and is non-singular if $\tilde{T}$ is non-singular. We note also that $M_{12}(p(p+1)/2 \times pn)\ M_{21}(pn \times p(p+1)/2)$ is $p(p+1)/2 \times p(p+1)/2$ and non-singular, so that

(A.6.1.6)
$$|M_{12}M_{22}^{-1}M_{21}| = |M_{12}M_{21}|\ |N_{22}|.$$

It is easy to check that

(A.6.1.7)
$$|N_{22}| = \prod_{i=1}^{p} t_{ii}^{p-i} / |\tilde{T}|^p \text{ and } |\tilde{T}| = \prod_{i=1}^{p} t_{ii}.$$

It is also easy to verify, by using the condition $LL' = I(p)$, that

(A.6.1.8)
$$|M_{12} \quad M_{21}|_{LL'=I(p)} = 1.$$

Hence (A.6.1) will now reduce to, [31, 32],

$$(\text{A.6.1.9}) \qquad J(X:\tilde{T}, L_I) = \left|\frac{\partial(X, LL')}{\partial(\tilde{T}, L)}\right|_{\tilde{T},\, L_I} \div \left|\frac{\partial(LL')}{\partial(L_D)}\right|_{L_I}$$

$$= 2^p \prod_{i=1}^{p} t_{ii}^{n-i} \div \left|\frac{\partial(LL')}{\partial(L_D)}\right|_{L_I},$$

so that we have

$$(\text{A.6.1.10}) \qquad dX \rightarrow J(X:\tilde{T}, L_I) d\tilde{T}\, dL_I,$$

where J is given by (A.6.1.9).

It is easy to check that. with $nS = \tilde{T}\tilde{T}'$, we have

$$(\text{A.6.1.11}) \qquad J(S:\tilde{T}) = 2^p \prod_{i=1}^{p} t_{ii}^{p-i+1} / n^{p(p+1)/2},$$

so that

$$(\text{A.6.1.12}) \qquad d\tilde{T} \rightarrow \left[n^{p(p+1)/2} \div \left(2^p \prod_{i=1}^{p} t_{ii}^{p-i+1}\right)\right] dS.$$

Another transformation (together with its Jacobian) that is useful and interesting is the following:

$$(\text{A.6.1.13}) \qquad X_{r+1,p}(\overline{p-r} \times n) = \begin{bmatrix} \mathbf{x}'_{r+1} \\ \cdot \\ \cdot \\ \mathbf{x}'_p \end{bmatrix}$$

$$= \begin{bmatrix} t_{r+1,1} & \cdots & t_{r+1,\,r+1} & 0 & \cdots & 0 \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ t_{p,1} & \cdots & t_{p,\,r+1} & \cdot & \cdots & t_{p,p} \end{bmatrix} \underset{n}{\begin{bmatrix} \mathbf{l}'_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{l}'_p \end{bmatrix}}.$$

This transformation is obtained if we start out from the transformation (A.3.11), cut out the first r rows of X and the first r rows of $\tilde{T}$ and assume that the first r rows of L are given constants, in other words, that the transformation is from the variable and truncated X to the variable and truncated $\tilde{T}$ and the variable $\mathbf{l}'_{r+1}, \ldots, \mathbf{l}'_p$ and that $\mathbf{l}'_1, \ldots, \mathbf{l}'_r$ are assumed to be given constants such that the whole $LL' = I(p)$. It is easy to show that, with $t_{pp}, \ldots, t_{r+1,\, r+1}$ being, say, positive, this transformation is also one-to-one. Let us denote the truncated $\tilde{T}$ by $\tilde{T}_{r+1,p}$, the truncated L by $L_{r+1,p}$ and the initial block of (constant) L by $L_{1,r}$. Then (A.6.1.13) can be rewritten as

(A.6.1.14) $$X_{r+1,p} = \tilde{T}_{r+1,p} \begin{bmatrix} L_{1,r} \\ L_{r+1,p} \end{bmatrix},$$

where the variable $L_{r+1,p}$ is subject to

(A.6.1.15) $$L_{r+1,p}\ L'_{r+1,p} = I(p-r) \text{ and } L_{r+1,p}\ L'_{1,r} = 0,$$

$L'_{1,r}$ being a given matrix of constants, subject itself to $L_{1,r}\ L'_{1,r} = I(r)$.

It is easy to see that the independent elements of the variable truncated L(i.e., of $L_{r+1,p}$) in this situation can be taken to be the same as of the truncated $L_{r+1,p}$ in the original set-up. Let us denote this by $L_{r+1,pI}$ consisting of $\mathbf{l}_{r+1,1}, \ldots, \mathbf{l}_{r+1,n-r-1}, \ldots, \mathbf{l}_{p,1}, \ldots, \mathbf{l}_{p,n-p}$ as elements and the dependent part by $L_{r+1,pD}$. We are now interested in the Jacobian $J(X_{r+1,p} : \tilde{T}_{r+1,p}, L_{r+1,pI})$, which, by using (A.5.5) subject to (A.6.1.15) comes out to be

(A.6.1.16) $$J(X_{r+1,p} : \tilde{T}_{r+1,p},\ L_{r+1,pI}) = \left| \frac{\partial(X_{r+1,p}, L_{r+1,p}\,L'_{r+1,p}, L_{r+1,p}L'_{1,r}}{\partial(\tilde{T}_{r+1,p}, L_{r+1,p})} \right|_{\tilde{T}_{r+1,p}, L_{r+1,pI}}$$

$$\div \left| \frac{\partial(L_{r+1,p}\,L'_{r+1,p}, L_{r+1,p}L'_{1,r}}{\partial(L_{r+1,pD})} \right|_{L_{r+1,pI}}$$

To calculate the numerator on the right side of (A.6.1.16) we proceed in the same manner as in the beginning of this section, go back to the scheme of partial differentiation shown after (A.6.1.2) and observe that the same scheme will serve, subject to the following modifications. Omit all columns below $t_{11}, \ldots, t_{r1}, \ldots, t_{rr}$ and below $\mathbf{l}'_1, \ldots, \mathbf{l}'_r$ and all rows along $\mathbf{x}_1, \ldots, \mathbf{x}_r$ and along $k_{11}, k_{12}, k_{22}, \ldots, k_{1r}, k_{2r}, \ldots, k_{rr}$. If now we make the same kind of calculation as from (A.6.1.3) to (A.6.1.9) we can verify that the numerator on the right side of (A.6.1.16) will reduce to

(A.6.1.17) $$2^{p-r} \prod_{i=r+1}^{p} t_{ii}^{n-i},$$

and thus we have

(A.6.1.18) $$dX_{r+1,p} \longrightarrow 2^{p-r} \prod_{i=r+1}^{p} t_{ii}^{n-i}\ d\tilde{T}_{r+1,p}\, dL_{r+1,pI}$$

$$\div \left| \frac{\partial(L_{r+1,p}\ \ L'_{r+1,p}, L_{r+1,p}\ \ L'_{1,r}}{\partial(L_{r+1,pD})\quad ,} \right|_{L_{r+1,pI}}$$

(A.6.2): *Jacobian of the transformation* (A.3.8), i.e., $J(X_1, X_2 : A, c\text{'s}, L_{1I}, L_{2I})$, *where* $X_1(p \times n_1)$, $X_2(p \times n_2)$ $(p \leqslant n_1, n_2)$ *are each of rank* p, *c's are distinct, and* A *is solid* $p \times p$ *non-singular with a positive first row.* Putting $c_i^{\frac{1}{2}} = t_i$ $(i = 1, 2, \ldots, p)$ and using (A.1.1) we have

$$J(X_1, X_2 : A, t\text{'s}, L_{1I}, L_{2I}) = \left|\frac{\partial(X_1, X_2, L_1L_1', L_2L_2')}{\partial(A, t_i\text{'s}, L_1, L_2)}\right|_{A,\, t\text{'s},\, L_{1I},\, L_{2I}} \div \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}} \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}. \tag{A.6.2.1}$$

To evaluate the numerator we proceed as follows. Denote, as before, the row vectors of L_1, L_2, X_1, X_2 by $\mathbf{l}_{1i}', \mathbf{l}_{2i}', \mathbf{x}_{1i}', \mathbf{x}_{2i}'$ $(i = 1, 2, \ldots, p)$ and L_1L_1' by (k_{1ij}) and L_2L_2' by (k_{2ij}). Then the transformation can be written as

$$\mathbf{x}_{1i} = [\mathbf{l}_{11} \ldots \mathbf{l}_{1p}] \begin{bmatrix} a_{i1} t_1 \\ \vdots \\ a_{ip} t_p \end{bmatrix}; \mathbf{x}_{2i} = [\mathbf{l}_{21} \ldots, \mathbf{l}_{2p}] \begin{bmatrix} a_{i1} \\ \vdots \\ a_{ip} \end{bmatrix} \tag{A.6.2.2}$$

$(i = 1, 2, \ldots, p)$, or in full,

$$\begin{bmatrix} \mathbf{x}_{11} \\ \cdot \\ \mathbf{x}_{1p} \\ \mathbf{x}_{21} \\ \cdot \\ \mathbf{x}_{2p} \end{bmatrix} = \left[\begin{array}{c|c} \begin{array}{ccccccc} \mathbf{l}_{11} & \cdot & \mathbf{l}_{1p} & \cdot & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \mathbf{l}_{11} & \cdot & \mathbf{l}_{1p} \end{array} & 0 \\ \hline 0 & \begin{array}{ccccccc} \mathbf{l}_{21} & \cdot & \mathbf{l}_{2p} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \mathbf{l}_{21} & \cdot & \mathbf{l}_{2p} \end{array} \end{array}\right] \begin{bmatrix} a_{11}t_1 \\ \cdot \\ a_{1p}t_p \\ \cdot\ \cdot \\ a_{p1}t_1 \\ \cdots \\ a_{pp}t_p \\ a_{11} \\ \cdot \\ a_{1p} \\ \cdot \\ a_{p1} \\ \cdot \\ a_{pp} \end{bmatrix}$$

The scheme of partial differentiation is given below.

	$\mathbf{a}'$	$\mathbf{t}'$	$\mathbf{l}_1'$	$\mathbf{l}_2'$
$\mathbf{x}_1$	(l_3, t)	(a_1, l_3)	(a_1, t)	0
$\mathbf{x}_2$	(l_4)	0	0	(a_2)
$\mathbf{k}_1$	0	0	(l_5)	0
$\mathbf{k}_2$	0	0	0	(l_6)

where $\mathbf{a}' = (a_{11} \,.\, a_{1p} \,.\, a_{p1} \,.\, a_{pp})$, $\mathbf{t}' = (t_1 \,.\, t_p)$, $\mathbf{l}_1' = (\mathbf{l}_{11}' \, \mathbf{l}_{12}' \,.\, \mathbf{l}_{1,\, p-1}' \, \mathbf{l}_{1p}')$,

$$\mathbf{l}_2' = (\mathbf{l}_{21}' \, \mathbf{l}_{22}' \,.\, \mathbf{l}_{2,\, p-1}' \, \mathbf{l}_{2p}'),\ \mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_{11} \\ . \\ \mathbf{x}_{1p} \end{bmatrix},\ \mathbf{x}_2 = \begin{bmatrix} \mathbf{x}_{21} \\ . \\ \mathbf{x}_{2p} \end{bmatrix},$$

$$\mathbf{k}_1 = \begin{bmatrix} k_{111} \\ . \\ k_{1pp} \\ k_{112} \\ . \\ k_{11p} \\ . \\ k_{1p-1,\, p} \end{bmatrix},\ \mathbf{k}_2 = \begin{bmatrix} k_{211} \\ . \\ k_{2pp} \\ k_{212} \\ . \\ k_{21p} \\ . \\ k_{2p-1,\, p} \end{bmatrix},\ (l_3, t) = \begin{bmatrix} \mathbf{l}_{11}t_1 & . & \mathbf{l}_{1p}t_p & . & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & . & \mathbf{l}_{11}t_1 & . & \mathbf{l}_{1p}t_p \end{bmatrix},$$

$$(a_1, l_3) = \begin{bmatrix} a_{11}\mathbf{l}_{11} & . & a_{1p}\mathbf{l}_{1p} \\ . & . & . \\ a_{p1}\mathbf{l}_{11} & . & a_{pp}\mathbf{l}_{1p} \end{bmatrix},\ (a_1, t) = \begin{bmatrix} D_{a_{11}t_1}(n_1) & . & . & . & D_{a_{1p}t_p}(n_1) \\ . & . & . & . & . \\ D_{a_{p1}t_1}(n_1) & . & . & . & D_{a_{pp}t_p}(n_1) \end{bmatrix},$$

$$(l_4) = \begin{bmatrix} \mathbf{l}_{21} & . & \mathbf{l}_{2p} & . & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & . & \mathbf{l}_{21} & . & \mathbf{l}_{2p} \end{bmatrix},\ (a_2) = \begin{bmatrix} D_{a_{11}}(n_2) & . & . & . & D_{a_{1p}}(n_2) \\ . & . & . & . & . \\ D_{a_{p1}}(n_2) & . & . & . & D_{a_{pp}}(n_2) \end{bmatrix},$$

$$(l_5) = \begin{bmatrix} 2\mathbf{l}'_{11} & 0 & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & 2\mathbf{l}'_{1p} \\ \mathbf{l}'_{12} & \mathbf{l}'_{11} & 0 & . & 0 \\ . & . & . & . & . \\ \mathbf{l}'_{1p} & 0 & . & . & \mathbf{l}'_{11} \\ . & . & . & . & . \\ 0 & 0 & . & \mathbf{l}'_{1p} & \mathbf{l}'_{1,\,p-1} \end{bmatrix}, \quad (l_6) = \begin{bmatrix} 2\mathbf{l}'_{21} & 0 & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & 2\mathbf{l}'_{2p} \\ \mathbf{l}'_{22} & \mathbf{l}'_{21} & 0 & . & 0 \\ . & . & . & . & . \\ \mathbf{l}'_{2p} & 0 & . & 0 & \mathbf{l}'_{21} \\ . & . & . & . & . \\ 0 & . & . & \mathbf{l}'_{2p} & \mathbf{l}'_{2,\,p-1} \end{bmatrix}$$

We are interested in the absolute value of the determinant of the above matrix (which is really the numerator in the Jacobian) and which is: $\{p^2+p+(n_1+n_2)p\}\times\{p^2+p+(n_1+n_2)p\}$. After some obvious manipulations we can take out a factor $2^{2p}\prod_{i=1}^{p} t_i^{n_1-p}$ so that we have the whole determinant reducing to

$$\text{(A.6.2.3)} \qquad 2^{2p}\prod_{i=1}^{p} t_i^{n_1-p} \begin{array}{|cccc|l} M_{11} & M_{12} & M_{13} & 0 & pn_1 \\ M_{21} & 0 & 0 & M_{24} & pn_2 \\ 0 & 0 & M_{33} & 0 & p(p+1)/2 \\ 0 & 0 & 0 & M_{44} & p(p+1)/2 \\ \multicolumn{1}{c}{p^2} & p & pn_1 & \multicolumn{1}{c}{pn_2} & \end{array},$$

where

$$M_{11}(pn_1\times p^2) = \begin{bmatrix} \mathbf{l}_{11}t_1 & . & \mathbf{l}_{1p}t_p & . & 0 & 0 \\ . & . & . & . & . & . \\ 0 & . & 0 & . & \mathbf{l}_{11}t_1 & \mathbf{l}_{1p}t_p \end{bmatrix}; \quad M_{21}(pn_2\times p^2) = \begin{bmatrix} \mathbf{l}_{21} & . & \mathbf{l}_{2p} & . & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & . & \mathbf{l}_{21} & . & \mathbf{l}_{2p} \end{bmatrix};$$

$$M_{12}(pn_1\times p) = \begin{bmatrix} a_{11}\mathbf{l}_{11} & . & a_{1p}\mathbf{l}_{1p} \\ . & . & . \\ a_{p1}\mathbf{l}_{11} & . & a_{pp}\mathbf{l}_{1p} \end{bmatrix}; \quad M_{13}(pn_1\times pn_1) = \begin{bmatrix} D_{a_{11}}(n_1) & . & D_{a_{1p}}(n_1) \\ . & . & . \\ D_{a_{p1}}(n_1) & . & D_{a_{pp}}(n_1) \end{bmatrix};$$

$$M_{24}(pn_2\times pn_2) = \begin{bmatrix} D_{a_{11}}(n_2) & . & D_{a_{1p}}(n_2) \\ . & . & . \\ D_{a_{p1}}(n_2) & . & D_{a_{pp}}(n_2) \end{bmatrix}$$

$$M_{33}\left(\frac{p(p+1)}{2}\times pn_1\right)=\begin{bmatrix} 1'_{11} & 0 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & . & . & . & 1'_{1p} \\ 1'_{12}t_2 & 1'_{11}t_1 & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & 1'_{1p}t_p & 1'_{1,\,p-1}t_{p-1} \end{bmatrix}$$

$$M_{44}\left(\frac{p(p+1)}{2}\times pn_2\right)=\begin{bmatrix} 1'_{21} & 0 & . & 0 & 0 \\ . & . & . & . & \\ 1'_{22} & 1'_{21} & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & 1'_{2p} & 1'_{2,p-1} \end{bmatrix}$$

Hence we should have

$$\text{(A.6.2.4)}\qquad \begin{vmatrix} M_{11} & M_{12} & M_{13} & 0 \\ M_{21} & 0 & 0 & M_{24} \\ 0 & 0 & M_{33} & 0 \\ 0 & 0 & 0 & M_{44} \end{vmatrix} = \begin{vmatrix} M_{13} & 0 \\ 0 & M_{24} \end{vmatrix}$$

$$\times\left|\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}_{\frac{p(p+1)}{2}}^{\frac{p(p+1)}{2}} - \begin{bmatrix} M_{33} & 0 \\ 0 & M_{44} \end{bmatrix} \begin{bmatrix} M_{13} & 0 \\ 0 & M_{24} \end{bmatrix}^{-1} \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & 0 \end{bmatrix}\right|$$

$$=\begin{vmatrix} M_{13} & 0 \\ 0 & M_{24} \end{vmatrix} \left|\begin{bmatrix} M_{33} & 0 \\ 0 & M_{44} \end{bmatrix} \begin{bmatrix} M_{13}^{-1} & 0 \\ 0 & M_{24}^{-1} \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & 0 \end{bmatrix}\right|,$$

$$=|M_{13}|\,|M_{24}| \begin{vmatrix} M_{33}M_{13}^{-1} & [M_{11} : M_{12}] \\ M_{44}M_{24}^{-1} & [M_{21} : \ 0\ ] \end{vmatrix}$$

It is now easy to check that

$$\text{(A.6.2.5)}\quad |M_{13}|=|A|^{n_1},\ |M_{24}|=|A|^{n_2}\text{ and also } |M_{13}^{-1}|=\begin{array}{ccc} D_{a^{11}}(n_1) & . & D_{a^{p1}}(n_1) \\ & & \\ D_{a^{1p}}(n_1) & . & D_{a^{pp}}(n_1) \end{array}$$

and M_{24}^{-1} is exactly of this form, each D being of n_2 dimensions. Hence we shall have

$$\text{(A.6.2.6)}\quad M_{13}^{-1}M_{11}=\begin{bmatrix} a^{11}\mathbf{l}_{11}t_1 & . & a^{11}\mathbf{l}_{1p}t_p & . & a^{p1}\mathbf{l}_{11}t_1 & . & a^{p1}\mathbf{l}_{1p}t_p \\ . & . & . & . & . & . & . \\ a^{1p}\mathbf{l}_{11}t_1 & . & a^{1p}\mathbf{l}_{1p}t_p & . & a^{pp}\mathbf{l}_{11}t_1 & . & a^{pp}\mathbf{l}_{1p}t_p \end{bmatrix}$$

$$=\begin{bmatrix} \mathbf{l}_{11}t_1 & . & \mathbf{l}_{1p}t_p & . & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & . & 0 & . & \mathbf{l}_{11}t_1 & . & \mathbf{l}_{1p}t_p \end{bmatrix}\begin{bmatrix} D_{a^{11}}(p) & . & D_{a^{p1}}(p) \\ . & . & . \\ D_{a^{1p}}(p) & . & D_{a^{pp}}(p) \end{bmatrix}$$

$= M_{11}D$(suppose) where we denote the right hand matrix factor by D. In an exactly similar manner we have $M_{24}^{-1}M_{21} = M_{21}D$. Next we have

$$\text{(A.6.2.7)}\quad M_{13}^{-1}M_{12}=\begin{bmatrix} \mathbf{l}_{11} & . & 0 \\ . & . & . \\ 0 & . & \mathbf{l}_{1p} \end{bmatrix}= N_{12}\ \text{(say)}$$

(using $UU^{-1} = I(p)$). Thus we have

$$\text{(A.6.2.8)}\quad M_{13}^{-1}[M_{11} \vdots M_{12}] = [M_{11}D \vdots N_{12}] = pn_1[\underset{p^2}{M_{11}} \vdots \underset{p}{N_{12}}]\begin{array}{c} \begin{bmatrix} D & 0 \\ 0 & I \end{bmatrix}\begin{array}{c} p^2 \\ p \end{array} \\ \begin{array}{cc} p^2 & p \end{array} \end{array}$$

and

$$M_{24}^{-1}[M_{21} \vdots 0] = pn_2[M_{21} \vdots 0]\begin{array}{c} \begin{bmatrix} D & 0 \\ 0 & I \end{bmatrix}\begin{array}{c} p^2 \\ p \end{array} \\ \begin{array}{cc} p^2 & p \end{array} \end{array},$$

so that (A.6.2.4) now reduces to

$$\text{(A.6.2.9)}\quad |A|^{n_1+n_2}\left\|\begin{bmatrix} M_{33}\times M_{11} & \vdots & M_{33}\times N_{12} \\ M_{44}\times M_{21} & \vdots & 0 \end{bmatrix}\begin{bmatrix} D & 0 \\ 0 & I(p) \end{bmatrix}\right\|$$

$$= |A|^{n_2+n_1-p}\left\|\begin{bmatrix} M_{33}\times M_{11} & \vdots & M_{33}\times N_{12} \\ M_{44}\times M_{21} & \vdots & 0 \end{bmatrix}\right\| \text{(Remembering that } |D|=|A|^{-p}$$

and $|I(p)| = 1$).

Using $L_1L_1' = L_2L_2' = I(p)$ the structure and reduction of the 2nd factor (which is a determinant) can now be displayed and visualized by considering $p = 3$ which will make immediately obvious the corresponding structure and mechanism of reduction for the general case. Below is given the case of $p = 3$.

$$\begin{vmatrix} t_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & t_2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & t_3 & 0 & 0 & 1 \\ 0 & t_2^2 & 0 & t_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & t_3^2 & 0 & 0 & 0 & t_1^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & t_3^2 & 0 & t_2^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{vmatrix} = \operatorname{mod} (t_1^2-t_2^2)(t_1^2-t_3^2)(t_2^2-t_3^2).$$

In the general case this is easily checked to be replaceable by $\operatorname{mod} \prod_{i<j=1}^{p-1} (t_i^2-t_j^2)$, so that, substituting in (A.6.2.3) and noting that $t_i^2 = c_i$, we have

$$\text{(A.6.2.10)} \quad \left|\frac{\partial(X_1, X_2, L_1L_1', L_2L_2')}{\partial(A, t, L_1, L_2)}\right|_{A, t, L_{1I}, L_{2I}} = 2^{2p} \prod_{i=1}^{p} t_i^{n_1-p} |A|^{n_1+n_2-p} \operatorname{mod} \prod_{i<j=1}^{p-1} (t_i^2-t_j^2).$$

so that

$$\text{(A.6.2.11)} \qquad J(X_1, X_2 : A, c, L_{1I}, L_{2I}) = \left|\frac{\partial(X_1, X_2, L_1L_I', L_2L_2')}{\partial(A, c, L_1, L_2)}\right|$$

$$\div \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right| \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|$$

$$= 2^p |A|^{n_1+n_2-p} \prod_{i=1}^{p} c_i^{\frac{n_1-p-1}{2}} \operatorname{mod} \prod_{i<j=1}^{p-1} (c_i-c_j) \div \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}} \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}.$$

It may be noticed now that (A.6.2.11) is the Jacobian (ii) mentioned in the beginning of (A.6), i.e., $J(X_1, X_2 : A, c\text{'s}, L_{1I}, L_{2I})$ and (A.6.1.9) is the Jacobian (iii) mentioned there, i.e., $J(X : \tilde{T}, L_I)$, [31, 32].

(A.6.3): *Jacobian of the transformation* (A.3.6), *i.e.*, $J(X : M_I, c\text{'s}, L_I)$, *where* $X(p \times n)$ $(p < n)$ *is of rank* p, *c's are distinct and* M *has a positive first row.* By straightforward methods of exactly the kind used in the preceding subsection (A.6.1) which is rather lengthy it can be shown that

$$J(X : M_I, c\text{'s}, L_I) = 2^p \prod_{i=1}^{p} c_i^{\frac{n-p-1}{2}} \bmod \prod_{i<j=1}^{p-1} (c_i - c_j) \tag{A.6.3.1}$$

$$\div \left|\frac{\partial(MM')}{\partial(M_D)}\right|_{M_I} \left|\frac{\partial(LL')}{\partial(L_D)}\right|_{L_I}.$$

But a shorter proof of this result can be given by combining (A.6.1.9) and (A.6.2.11) in the following way. By (A.6.2.11) we have

$$J(X_1, X_2 : A, c\text{'s}, L_{1I}, L_{2I}) = 2^p |A|^{n+n_2-p} \prod_{i=1}^{p} c_i^{\frac{n-p-1}{2}} \bmod \prod_{i<j=1}^{p} (c_i - c_j) \tag{A.6.3.2}$$

$$\div \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}} \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}},$$

where $X_1(p \times n) = AD_{\sqrt{c}}L_1$ and $X_2(p \times n_2) = AL_2$.

Also, using (A.3.11), we put

$$A(p \times p) = \tilde{T}(p \times p)M(p \times p) \tag{A.6.3.3}$$

where M is $\perp$, and using (A.6.1.9), we have

$$J(A : \tilde{T}, M_I) = 2^p \prod_{i=1}^{p} t_{ii}^{p-i} \div \left|\frac{\partial(MM)}{\partial(M_D)}\right|_{M_I}. \tag{A.6.3.4}$$

Next put

$$X_1(p \times n) = \tilde{T}(p \times p)\, X(p \times n),\ M(p \times p)L_2(p \times n_2) = M_2(p \times n_2)(\text{say}) \tag{A.6.3.5}$$

(so that $X = MD_{\sqrt{c}}L_1$, $X_2 = \tilde{T}M_2$) and note from orthogonality of M that

$$|A| = |\tilde{T}| = \prod_{i=1}^{p} t_{ii} \text{ and } M_2M_2' = ML_2L_2'M' = I(p). \tag{A.6.3.6}$$

We thus have

$$\begin{aligned} J(X_1, X_2 : A, c\text{'s}, L_{1I}, L_{2I}) &= J(X_1, X_2 : \tilde{T}, M_I, c\text{'s}, L_{1I}, L_{2I}) \div J(A : \tilde{T}, M_I) \\ &= J(X_1 : X)\, J(X, X_2 : \tilde{T}, M_I, c\text{'s}, L_{1I}, M_{2I})\ J(M_{2I} : L_{2I})/J(A : \tilde{T}, M_I) \\ &= J(X_1 : X)\ J(X : M_I, c\text{'s}, L_{1I})\ J(X_2 : \tilde{T}, M_{2I})\ J(M_{2I} : L_{2I})/J(A : \tilde{T}, M_I). \end{aligned} \tag{A.6.3.7}$$

Now notice that

$$\text{(A.6.3.8)} \qquad J(X_1 : X) = |\tilde{T}|^n = \prod_{i=1}^{p} t_{ii}^n,\ J(X_2 : \tilde{T}, M_{2I}) = 2^p \prod_{i=1}^{p} t_{ii}^{n_2-i} \Big/ \left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right|_{M_{2I}}$$

and
$$J(A : \tilde{T}, M_I) = 2^p \prod_{i=1}^{p} t_{ii}^{p-i} \div \left|\frac{\partial(MM')}{\partial(M_D)}\right|_{M_I}.$$

Now to evaluate $J(M_{2I} : L_{2I})$ we temporarily regard M as a constant but $\perp$ matrix, notice that $L_2L_2' = I(p)$ is equivalent to $M_2M_2' = I(p)$ and now using (A.5.5) we find

$$\text{(A.6.3.9)} \qquad J(M_{2I} : L_{2I}) = \left|\frac{\partial(M_2-ML_2,\ M_2M_2')}{\partial(M_{2D},\ L_2)}\right| \div \left|\frac{\partial(M_2-ML_2,\ M_2M_2')}{\partial(L_{2D},\ M_2)}\right|$$

$$= \left|\frac{\partial(M_2-ML_2,\ M_2M_2')}{\partial(M_{2D},\ L_2)}\right| \div \left|\frac{\partial(M_2-ML_2,\ L_2L_2')}{\partial(L_{2D},\ M_2)}\right| = \left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right| \div \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|.$$

Now substituting in the left hand side of (A.6.3.7) from (A.6.3.2) and (A.6.3.6) and in the right hand side from (A.6.3.8) and (A.6.3.9) and putting $L_1 = L$ (say), we have the Jacobian (A.6.3.1).

(A.6.4): *Jacobian of the transformation* (A.3.15), *i.e.*, $J(X_1, X_2 : \tilde{T}, c\text{'s}, L_I, L_{1I}, L_{2I})$ *where* $\tilde{T}$ *is non-singular with a positive diagonal and c's are distinct.* Using (A.6.1.9) we have $J(X_2 : \tilde{T}, L_{2I}) = 2^p \prod_{i=1}^{p} t_{ii}^{n_2-i} \Big/ \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}$.

Next, using (A.6.3.1) we have

$$J(X_1'\tilde{T}'^{-1} : L_I, c\text{'s}, L_{1I}) = 2^{n_1} \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j) \div \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I} \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}}.$$

From these it is easy to check that

$$\text{(A.6.4.1)} \qquad J(X_1, X_2 : \tilde{T}, c\text{'s}, L_I, L_{1I}, L_{2I}) = 2^{p+n_1} \prod_{i=1}^{p} t_{ii}^{n_1+n_2-i} \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}}$$

$$\times \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j) \Big/ \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I} \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}} \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}.$$

(A.6.5): *Jacobian of the transformation* (A.3.17), *i.e.*, $J(X_1, X_2 : \tilde{T}, U, c\text{'s}, M_{1I}, M_{2I}, L_{2I})$, *when the c's are distinct,* $\tilde{T}$ *is non-singular with a positive diagonal and* U *is non-singular solid with a positive first row.* Using (A.6.1.9) we have $J(X_2: \tilde{T}, L_{2I})$ $= 2^q \prod_{i=1}^{q} t_{ii}^{n-i} \Big/ \left|\frac{\partial(L_2L_2')}{(\partial L_{2D})}\right|_{L_{2I}}$. Next, notice that $J(X_1 : X_1[L_1' \vdots L_2']) = 1$ (since $[L_1' \vdots L_2']$ is $\perp$).

Next, using (A.6.2.12), we have $J(X_1[L_1' : L_2'] : U,\ e\text{'s},\ M_{1I},\ M_{2I})$

$$= 2^p |U|^{n-p} \prod_{i=1}^{p} e_i^{\frac{n-q-p-1}{2}} \bmod \prod_{i<j=1}^{p-1} (e_i - e_j) \Big/ \left|\frac{\partial(M_1M_1')}{\partial(M_{1D})}\right|_{M_{1I}} \left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right|_{M_{2I}} .$$

It is easy to check by combining the three Jacobians, that

$$\text{(A.6.5.1)} \quad J(X_1, X_2 : \tilde{T}, U, e\text{'s}, M_{1I}, M_{2I}, L_{2I}) = 2^{p+q} \prod_{i=1}^{q} t_{ii}^{n-i} |U|^{n-p} \prod_{i=1}^{p} e_i^{\frac{n-q-p-1}{2}}$$

$$\times \bmod \prod_{i<j=1}^{p-1} (e_i - e_j) \Big/ \left|\frac{\partial(M_1M_1')}{\partial(M_{1D})}\right|_{M_{1I}} \left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right|_{M_{2I}} \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}} .$$

We recall from (A.3.17) that if $e_i = (1-c_i)/c_i (i = 1, \ldots, p)$, then the c_i's are the roots of the equation in c: $|c(X_1X_1') - (X_1X_2')(X_2X_2')^{-1}(X_2X_1')| = 0$. In terms of the c's, therefore, we should have the Jacobian given by

$$\text{(A.6.5.2)} \qquad J(X_1, X_2 : \tilde{T}, U, c\text{'s } M_{1I}, M_{2I}, L_{2I}) = 2^{p+q} \prod_{i=1}^{q} t_{ii}^{n-i} |U|^{n-p}$$

$$\times \left[\prod_{i=1}^{p} (1-c_i)^{\frac{n-p-q-1}{2}} \Big/ c_i^{\frac{n-q+2}{2}} \right] \bmod \prod_{i<j=1}^{p-1} (c_i - c_j) \div \left|\frac{\partial(M_1M_1')}{\partial(M_{1D})}\right|_{M_{1I}} \left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right|_{M_{2I}}$$

$$\times \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}} .$$

(A.6.6): *Jacobian of the transformation* (A.3.19), *i.e.*, $J(X_1, X_2, X_3 : Z_{11}, Z_{12}, Z_{21}, Z_{22}, \tilde{T}, L_{3I})$. Using (A.6.1.9) we have $J(X_3 : \tilde{T}, L_{3I}) = 2^r \prod_{i=1}^{r} t_{ii}^{n-i} \div \left|\frac{\partial(L_3L_3')}{\partial(L_{3D})}\right|_{L_{3I}}$.

Next we notice that $J(X_1, X_2 : Z_{11}, Z_{12}, Z_{21}, Z_{22}) = J\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} : \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \begin{pmatrix} L \\ L_3 \end{pmatrix}\right] = 1$, since $\begin{bmatrix} L \\ L_3 \end{bmatrix}$ is $\perp$. Therefore it is easily checked that the total Jacobian

$$= 2^r \prod_{i=1}^{r} t_{ii}^{n-i} \div \left|\frac{\partial(L_3L_3')}{\partial(L_{3D})}\right|_{L_{3I}} .$$

(A.6.7): *Jacobian of the transformation* (A.3.14), *i.e.*, $J(X_1, X_2 : U_1, U_2, \tilde{U}_3, U_4, c\text{'s}, L_{1I}, L_{2I})$, *where* $X_1(p \times n_1)$ $(p > n_1)$ *is of rank* n_1, $X_2(p \times n_2)$ $(p < n_2)$ *is of rank* p, *the* c*'s are distinct and* U_1 *has a positive first row and* $\tilde{U}_3$ *a positive diagonal.* We start with the transformation (A.3.15) and use the Jacobian result (A.6.4.1) and rename the symbols. The transformation is $X_1(p \times n_1) = \tilde{T}(p \times p)\ L(p \times n_1)\ D_{\sqrt{c}}(n_1 \times n_1)$

$\times L_1(n_1\times n_1)$, $X_2(p\times n_2) = \tilde{T}(p\times p)\, M_2(p\times n_2)$ subject to L_1 being $\perp$, $M_2M_2' = I(p)$, $L'L = I(n_1)$ and $\tilde{T}$ being non-singular, and the Jacobian being given by

$$\text{(A.6.7.1)} \qquad J(X_1, X_2 : \tilde{T}, c\text{'s}, L_I, L_{1I}, M_{2I}) = 2^{p+n_1} \prod_{i=1}^{p} t_{ii}^{n_1+n_2-i} \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}}$$

$$\times \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j) \div \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I} \left|\frac{\partial(L_1L_1')}{\partial(L_{1D})}\right|_{L_{1I}} \left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right|_{M_{2I}}.$$

Let us write

$$L(p\times n_1) = \begin{array}{cc} \left[\begin{array}{c} K_1 \\ K_2 \end{array}\right] & \begin{array}{l} p-n_1 \\ n_1 \end{array} \\ n_1 & \end{array}.$$

To this L now, if we adjoin, as we could, a matrix $\begin{array}{cc} \left[\begin{array}{c} \tilde{K}_3 \\ K_4 \end{array}\right] & \begin{array}{l} p-n_1 \\ n_1 \end{array} \\ p-n_1 & \end{array}$ such that $K(p\times p)$ $= \begin{array}{cc} \left[\begin{array}{cc} K_1 & \tilde{K}_3 \\ K_2 & K_4 \end{array}\right] & \begin{array}{l} p-n_1 \\ n_1 \end{array} \\ n_1 \;\; p-n_1 & \end{array}$ is orthogonal (note that this could be done since $L'L = I(n_1)$), it will be seen that the number of independent elements in K is the same as in L. This is verified as follows:

In L (by virtue of $L'L = I(n_1)$) the number of independent elements are $pn_1-n_1(n_1+1)/2$. In K the total number of elements is $p^2-(p-n_1)(p-n_1-1)/2$ and by virtue of $KK' = I(p)$, the number of constraints is $p(p+1)/2$, so that the number of independent elements is $p^2-(p-n_1)(p-n_1-1)/2-p(p+1)/2 = pn_1-n_1(n_1+1)/2$. If we now put

$$\text{(A.6.7.2)} \qquad U(p\times p) = \begin{array}{cc} \left[\begin{array}{cc} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{array}\right] & \begin{array}{l} p-n_1 \\ n_1 \end{array} \\ n_1 \;\; p-n_1 & \end{array} = \begin{array}{cc} \begin{array}{r} p-n_1 \\ n_1 \end{array} & \left[\begin{array}{cc} \tilde{T}_1 & 0 \\ T_2 & \tilde{T}_4 \end{array}\right] \\ & p-n_1 \;\; n_1 \end{array} \begin{array}{cc} \left[\begin{array}{cc} K_1 & \tilde{K}_3 \\ K_2 & K_4 \end{array}\right] & \begin{array}{l} p-n_1 \\ n_1 \end{array} \\ n_1 \;\; p-n_1 & \end{array}$$

(by examining the right hand side we note that the left hand side is really of the structure indicated), we observe that the number of independent elements in U which is $p^2-(p-n_1)(p-n_1-1)/2$, is the same as in $(\tilde{T}, K)$, i.e., as in $(\tilde{T}, K_1, K_2)$, which is $p(p+1)/2+pn_1-n_1(n_1+1)/2$. It will be shown in the next article (and we assume the result here) that

$$\text{(A.6.7.3)} \qquad J(U : \tilde{T}, K_I) = J(U : \tilde{T}, L_I) = 2^{n_1} \prod_{i=1}^{p} t_{ii}^{p-i} \div \prod_{i=1}^{p-n_1} (u_{3ii})^{p-n_1-i} \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I},$$

so that, by taking the inverse, we should have

$$\text{(A.6.7.4)} \qquad J(\tilde{T}, K_I : U) = J(\tilde{T}, L_I: U) = \prod_{i=1}^{p-n_1} (u_{3ii})^{p-n_1-i} \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I}$$

$$\div\; 2^{n_1} \prod_{i=1}^{p} t_{ii}^{p-i}.$$

Also if we put

$$\text{(A.6.7.5)} \qquad L_2(p\times n_2) = K(p\times p)\ M_2(p\times n_2)$$

(where by virtue of $KK' = K'K = M_2M_2' = I(p)$ we have $L_2L_2' = I(p)$), then exactly as in (A.6.3) (treating K as a constant $\perp$ matrix) we have

$$\text{(A.6.7.6)} \qquad J(M_{2I} : L_{2I}) = \left|\frac{\partial(M_2M_2')}{\partial(M_{2D})}\right|_{M_{2I}} \div \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}.$$

Thus we have

$$\text{(A.6.7.7.)} \qquad J(\tilde{T}, L_I, M_{2I} : U.\ L_{2I}) = \left|\frac{\partial(M_2M_2')}{\partial(L_D)}\right|_{M_{2I}} \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I} \prod_{i=1}^{p-n_1} (u_{3ii})^{p-n_1-i}$$

$$\div\ 2^{n_1} \prod_{i=1}^{p} t_{ii}^{p-i} \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}}.$$

Using these and remembering that $|U| = |\tilde{T}| = \prod_{i=1}^{p} t_{ii}$, we have

$$\text{(A.6.7.8)} \qquad J(X_1, X_2 :\ c, U, L_{1I}, L_{2I}) = J(X_1, X_2 :\ c, \tilde{T}, L_I, L_{1I}, M_{2I})$$
$$\times J(\tilde{T}, L_I, M_{2I} :\ U, L_{2I})$$

$$= 2^p |U|^{n_1+n_2-p} \prod_{i=1}^{n_1} c_i^{\frac{p-n_1-1}{2}} \bmod \prod_{i<j=1}^{n_1-1} (c_i-c_j) \prod_{i=1}^{p-n_1} (u_{3ii})^{p-n_1-i} \div \left|\frac{\partial(L_1L_1'}{\partial(L_{1D})}\right|_{L_{1I}}$$

$$\times \left|\frac{\partial(L_2L_2')}{\partial(L_{2D})}\right|_{L_{2I}},$$

which gives $J(X_1, X_2 : c, U, L_{1I}, L_{1I})$.

Now for the proof of (A.6.7.3) with a transformation of the form (A.3.14) we proceed as follows. We start from (A.6.7.2), postmultiply both sides by the $p\times p$ matrix $\begin{array}{cc} \left[\begin{array}{cc} I & 0 \\ 0 & M \end{array}\right] & \begin{array}{c} n_1 \\ p-n_1 \end{array} \\ \begin{array}{cc} n_1 & p-n_1 \end{array} & \end{array}$, where M is $\perp$ and then write

$$\text{(A.6.7.9)} \qquad U = \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} = \begin{bmatrix} U_1 & \tilde{U}_3M \\ U_2 & U_4M \end{bmatrix} = \begin{bmatrix} V_1 & V_3 \\ V_2 & V_4 \end{bmatrix} \quad \text{(say)}$$
$$= V\text{(say)}$$
$$= \tilde{T}K \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} = \tilde{T}N\ \text{(say)}.$$

Then using (A.6.1.9) and (A.5.5) and (A.5.6) we have

$$\text{(A.6.7.10)} \qquad J(V:\tilde{T}, K_I) = 2^p \prod_{i=1}^{p} t_{ii}^{p-i} \Big/ \left|\frac{\partial(NN')}{\partial(N_D)}\right|_{N_I} = 2^p \prod_{i=1}^{p} t_{ii}^{p-i} \Big/ \left|\frac{\partial(KK')}{\partial(K_D)}\right|_{K_I} \times \left|\frac{\partial(MM')}{\partial(M_D)}\right|_{M_I},$$

and also

$$\text{(A.6.7.11)} \qquad J(V:U, M_I) = 2^{p-n_1} \prod_{i=1}^{p-n_1} (u_{3ii})^{p-n_1-i} \Big/ \left|\frac{\partial(MM')}{\partial(M_D)}\right|_{M_I}$$

Using (A.5.6) and taking account of the remarks after (A.6.6.1) it is easy to check that

$$\text{(A.6.7.12)} \qquad \left|\frac{\partial(KK')}{\partial(K_D)}\right|_{K_I} = \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I}.$$

Now combining (A.6.7.9), (A.6.7.10), (A.6.7.11) and (A.6.7.12), we have

$$dV \to 2^{p-n_1} \prod_{i=1}^{p-n_1} (u_{3ii})^{p-n_1-i} dU\, dM_I \Big/ \left|\frac{\partial(MM')}{\partial(M_D)}\right|_{M_I}$$

$$\to 2^p \prod_{i=1}^{p} t_{ii}^{p-i} d\tilde{T}\, dL_I dM_I \div \left|\frac{\partial(L'L)}{\partial(L_D)}\right|_{L_I} \times \left|\frac{\partial(MM')}{\partial(M_D)}\right|_{M_I},$$ which proves (A.6.7.3).

(A.6.8): *Jacobian of the transformation* (A.3.18.19), *i.e.*, $J(X_1, X_2 : A, B_1, B_2, \tilde{B}_3, B_4$, *c's*, $L_I)$, *where the c's are distinct, A is non-sigular with a positive first row, $\tilde{B}_3$ has a positive diagonal and* $B = \begin{bmatrix} B_1 & \tilde{B}_3 \\ B_2 & B_4 \end{bmatrix}$ is *non-singular.* This Jacobian can be derived in the same manner as in sub-section (A.6.2). We shall not need it in this monograph and so will not derive it. We merely state without proof that

$$\text{(A.6.8.1)} \qquad J(X_1, X_2 : A, \mathrm{B}, c\text{'s}, L_I) = 2^p |A|^{n-p} |B|^{n-q} \prod_{i=1}^{q-p} (\tilde{B}_3)_{ii}^{q-p-i}$$

$$\times \prod_{i=1}^{p} (1-c_i)^{\frac{n-p-q-1}{2}} c_i^{\frac{-p-1}{2}} \bmod \prod_{i<j=1}^{p-1} (c_i-c_j) \Big/ \left|\frac{\partial(LL')}{\partial(L_D)}\right|_{L_I},$$

where $L = \begin{array}{c} \left[\begin{array}{c} L_1 \\ L_3 \\ L_4 \end{array}\right] \begin{array}{l} p \\ q \\ q-p \end{array} \\ n \end{array}$ *and is subject to* $LL' = I(p+q)$, [31].

APPENDIX 7

Canonical Reduction of Certain Distribution Problems

(A.7.1): *If* $X(p\times n)$ $(p \leqslant n)$ *has the probability law* (4.13):

$$\{1/(2\pi)^{\overline{2}}|\Sigma|^{\frac{n}{2}}\} \times \exp\,[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}XX']dX,$$

then the distribution of the characteristic roots of XX' *(to be called c's) could not involve as parameters anything except the characteristic roots of* Σ *(to be called* γ*'s).*

Proof: Note that, a.e., XX' is p.d. so that, a.e., all roots $c(XX')$ are positive. Notice also that, a.e., they are also distinct. It is of course assumed that Σ is symmetric p.d., so that all $c(\Sigma)$'s i.e., γ's are positive. Using (A.3.3), set $\Sigma = \mu D_\gamma \mu'$, where μ is $\perp$. We have now $\operatorname{tr}\Sigma^{-1}XX' = \operatorname{tr}(\mu D_\gamma\mu')^{-1}XX' = \operatorname{tr} D_{\sqrt{1/\gamma}}\,\mu'XX'\mu D_{\sqrt{1/\gamma}}$ (using (A.1.5) and the orthogonality of μ). Now put $\mu'X = Y$ or $X(p\times n) = \mu(p\times p) \times Y(p\times n)$ and observe that, by (A.4.1). $c(XX') = c(YY')$. and $\operatorname{tr} D_{1/\gamma}\,\mu'XX'\mu = \operatorname{tr} D_{1/\gamma}YY'$. Also by (A.5.2), $J(X : Y) = |\mu|^n = 1$.

Remembering further that $|\Sigma| = |\mu|^2 \prod\limits_{i=1}^{p} \gamma_i$. it is easy to check that Y has the probability law:

$$\text{(A.7.1.1)} \qquad \left[1/(2\pi)^{\frac{pn}{2}} \prod_{i=1}^{p} \gamma_i^{\frac{n}{2}}\right] \exp\left[-\tfrac{1}{2}\operatorname{tr} D_{1/\gamma}\;YY'\right]dY$$

which, in view of the fact that $c(XX') = c(YY')$, proves (A.7.1). For the distribution of $c(XX')$, therefore we can, without any loss of generality, start directly from the above form of probability law which is accordingly a canonical law for this purpose.

(A.7.2): *If* $X_1(p\times n_1)$, $X_2(p\times n_2)$ $(p \leqslant n_1, n_2)$ *have the joint probability law:*

$$\left[1/(2\pi)^{\frac{p(n_1+n_2)}{2}}|\Sigma_1|^{\frac{n_1}{2}}|\Sigma_2|^{\frac{n_2}{2}}\right]\exp\left[-\tfrac{1}{2}\operatorname{tr}(\Sigma_1^{-1}X_1X_1'+\Sigma_2^{-1}X_2X_2')\right]dX_1\,dX_2,$$

Σ_1 *and* Σ_2 *being each symmetric p.d., then the distribution of* $c((X_1X_1')(X_2X_2')^{-1})$ *(to be called c's) could not involve as parameters anything except the* $c(\Sigma_1\Sigma_2^{-1})$*'s (to be called* γ*'s).*

Proof: Notice that, a.e., $c(X_1X_1'(X_2X_2')^{-1})$ are positive and distinct. Since Σ_1 and Σ_2 are each p.d., use (A.3.4) to set $\Sigma_1 = \mu D_\gamma \mu'$ and $\Sigma_2 = \mu\mu'$, where μ is non-singular and all γ's are positive. We have now, using (A.1.5), $\operatorname{tr}\Sigma_1^{-1}X_1X_1' = \operatorname{tr} D_{1/\gamma}\mu^{-1}X_1X_1'\mu'^{-1}$ and $\operatorname{tr}\Sigma_2^{-1}X_2X_2' = \operatorname{tr}\mu^{-1}X_2X_2'\mu'^{-1}$. Now put $\mu^{-1}X_1 = Y_1$ and $\mu^{-1}X_2 = Y_2$, i.e., $X_1(p\times n_1) = \mu(p\times p)\;Y_1(p\times n)$ and $X_2(p\times n_2) = \mu(p\times p)\;Y_2(p\times n_2)$ and observe that, by (A.4.2), $c(X_1X_1'(X_2X_2')^{-1}) = c(Y_1Y_1'(Y_2Y_2')^{-1})$. Also, by (A.5.2), $J(X_1, X_2 : Y_1, Y_2) = |\mu|^{n_1+n_2}$.

Remembering further that $|\Sigma_1|=|\mu|^2\prod_{i=1}^{p}\gamma_i$ and $|\Sigma_2|=|\mu|^2$, we check that Y_1 and Y_2 have the joint probability law:

$$\text{(A.7.2.1)}\qquad \left[1/(2\pi)^{\frac{p(n_1+n_2)}{2}}\cdot\prod_{i=1}^{p}\gamma_i^{\frac{n_1}{2}}\right]\times\exp\left[-\tfrac{1}{2}\operatorname{tr}(D_{1/\gamma}Y_1Y_1'+Y_2Y_2')\right]dY_1\,dY_2,$$

which, in view of the fact that $c(X_1X_1'(X_2X_2')^{-1}) = c(Y_1Y_1'(Y_2Y_2')^{-1})$, proves (A.7.2). If we are interested in the distribution of these roots, i.e., of the c's we can, without any loss of generality, start right away from the above form which, for the purpose of this problem, will thus be called a canonical distribution law.

(A.7.3): *If* $X = \begin{array}{c}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\begin{matrix} p \\ q \end{matrix} \\ n \end{array}$ $(p \leqslant q,\ p+q \leqslant n)$ *has the probability law* (4.15):

$$[1/(2\pi)^{(p+q)\frac{n}{2}}\,|\Sigma|^{\frac{n}{2}}]\exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{bmatrix}^{-1}\begin{bmatrix}X_1\\ X_2\end{bmatrix}[X_1' \quad X_2']\right]\right\}dX_1\,dX_2,$$

where $\Sigma = \begin{array}{cc}\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{bmatrix} & \begin{matrix} p \\ q\end{matrix}\\ \begin{matrix} p \qquad q\end{matrix} & \end{array}$ *is supposed to be symmetric p.d., then the distribution of* $c[(X_1X_1')^{-1}(X_1X_2')(X_2X_2')^{-1}(X_2X_1')]$, *(to be called c's) could not involve as parameters anything except* $c(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}')$ *(to be called γ's).*

Proof: Notice that, a.e., the p c's are positive and distinct and also that γ's are all non-negative. Use (A.3.16) to set $\Sigma_{11}(p\times p) = \mu_1(p\times p)\mu_1'(p\times p)$,

$\Sigma_{22}(q\times q) = \mu_2(q\times q)\,\mu_2'(q\times q)$ and $\Sigma_{12}(p\times q) = \mu_1(p\times p)[\underset{p}{D_{\sqrt{\gamma}}} \quad \underset{q-p}{0}]\ (p)\mu_2'(q\times q)$,

where μ_1 and μ_2 are non-singular. We have now

$$\text{(A.7.3.1)}\qquad \Sigma^{-1} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{bmatrix}^{-1} = \begin{bmatrix}\mu_1'^{-1} & 0\\ 0 & \mu_2'^{-1}\end{bmatrix}\begin{bmatrix} I(p) & [D_{\sqrt{\gamma}}\ 0]\\ \begin{bmatrix}D_{\sqrt{\gamma}}\\ 0\end{bmatrix} & I(q)\end{bmatrix}^{-1}\begin{bmatrix}\mu_1^{-1} & 0\\ 0 & \mu_2^{-1}\end{bmatrix}$$

Also

$$\text{(A.7.3.2)}\qquad \begin{bmatrix} I(p) & [D_{\sqrt{\gamma}} \quad 0]\\ \begin{bmatrix}D_{\sqrt{\gamma}}\\ 0\end{bmatrix} & I(q)\end{bmatrix} = \begin{bmatrix} I(p) & 0\\ \begin{bmatrix}D_{\sqrt{\gamma}}\\ 0\end{bmatrix} & \begin{bmatrix}D_{\sqrt{1-\gamma}} & 0\\ 0 & I(q-p)\end{bmatrix}\end{bmatrix}$$

$$\times\begin{bmatrix} I(p) & [D_{\sqrt{\gamma}} \quad 0]\\ 0 & \begin{bmatrix}D_{\sqrt{1-\gamma}} & 0\\ 0 & I(q-p)\end{bmatrix}\end{bmatrix}$$

where we notice that, on the right hand side, one matrix factor is the transpose of the other matrix factor. Taking the inverse on both sides of (A.7.3.2) we have

$$\text{(A.7.3.3)} \qquad \begin{bmatrix} I(p) & [D_{\sqrt{\gamma}} \quad 0] \\ \begin{bmatrix} D_{\sqrt{\gamma}} \\ 0 \end{bmatrix} & I(q) \end{bmatrix}^{-1} = \begin{bmatrix} I(p) & -[D_{\sqrt{\gamma 1-\gamma}} \quad 0] \\ 0 & \begin{bmatrix} D_{\sqrt{1/1-\gamma}} & 0 \\ 0 & I(q-p) \end{bmatrix} \end{bmatrix}$$

$$\times \begin{bmatrix} I(p) & 0 \\ -\begin{bmatrix} D_{\sqrt{\gamma/1-\gamma}} \\ 0 \end{bmatrix} & \begin{bmatrix} D_{\sqrt{1/1-\gamma}} & 0 \\ 0 & I(q-p) \end{bmatrix} \end{bmatrix} = M(\gamma)M'(\gamma) \text{ (say).}$$

Taking into account (A.7.3.1), (A.7.3.2) and (A.7.3.3) and using (A.1.5) we have

$$\text{(A.7.3.4)} \qquad \operatorname{tr} \Sigma^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} [X_1' \quad X_2'] = \operatorname{tr} M'(\gamma) \begin{bmatrix} \mu_1^{-1} & 0 \\ 0 & \mu_2^{-1} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\times [X_1' \quad X_2'] \begin{bmatrix} \mu_1'^{-1} & 0 \\ 0 & \mu_2'^{-1} \end{bmatrix} M(\gamma).$$

Now put $\mu_1^{-1}X_1 = Y_1$ and $\mu_2^{-1} X_2 = Y_2$, i.e., $X_1(p\times n) = \mu_1(p\times p)Y_1(p\times n)$ and $X_2(q\times n) = \mu_2(q\times q)Y_2(q\times n)$ and observe that, by (A.4.4), $c(X_1X_1')^{-1}(X_1X_2')(X_2X_2')^{-1} \times (X_2X_1') = c(Y_1Y_1')^{-1}(Y_1Y_2')(Y_2Y_2')^{-1}(Y_2Y_1')$. Also, by (A.5.2), $J(X_1 . X_2 : Y_1, Y_2) = |\mu_1|^n|\mu_2|^n$.

Next check that $|\Sigma|^{\frac{n}{2}} = |\mu_1|^n|\mu_2|^n \begin{vmatrix} I(p) & [D_{\sqrt{\gamma}} \quad 0] \\ \begin{bmatrix} D_{\sqrt{\gamma}} \\ 0 \end{bmatrix} & I(q) \end{vmatrix}^{\frac{n}{2}} = |\mu_1|^n|\mu_2|^n \prod_{i=1}^{p} [1-\gamma_i]^{\frac{n}{2}}$, and

finally check that (Y_1, Y_2) have the probability law:

$$\text{(A.7.3.5)} \qquad \left[1/(2\pi)^{\frac{(q+p)n}{2}} \prod_{i=1}^{p} [1-\gamma_i]^{\frac{n}{2}} \right] \exp \left[-\tfrac{1}{2} \operatorname{tr} M'(\gamma) \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \right.$$

$$\left. \times [Y_1' \quad Y_2'] M(\gamma) \right] dY_1 \; dY_2.$$

In view of the fact that $c[(X_1X_1')^{-1}(X_1X_2')(X_2X_2')^{-1}(X_2X_1')] = c[(Y_1Y_1')^{-1}(Y_1Y_2')(Y_2Y_2')^{-1} \times (Y_2Y_1')]$ the probability law (A.7.3.5) proves (A.7.3). Thus, as in the two previous sections, if we are interested in the distribution of these roots, i.e., the c's, we can, without any loss of generality, start from the probability law (A.7.3.5), which is thus a canonical form for this purpose.

(A.7.4): *If* $X = \begin{array}{c c} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} & \begin{array}{l} p \\ q\,(p \leqslant q,\, p+q+r \leqslant n) \\ r \end{array} \\ n & \end{array}$ *has the probability law* (4.15):

$$\left[1/(2\pi)^{\frac{(p+q+r)n}{2}}\,|\Sigma|^{\frac{n}{2}}\right] \exp\left[-\tfrac{1}{2}\,\Sigma^{-1}XX'\right]dX, \text{ where } \Sigma = \begin{array}{cc} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12}' & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13}' & \Sigma_{23}' & \Sigma_{33} \end{bmatrix} & \begin{array}{l} p \\ q \\ r \end{array} \\ \begin{array}{ccc} p \quad & q \quad & r \end{array} & \end{array}$$

is supposed to be symmetric p.d., then the distribution of the c's could not involve as parameters anything except γ's where c's and γ's are respectively the characteristic roots of $[X_1X_1' - X_1X_3'(X_3X_3')^{-1}X_3X_1']^{-1}[X_1X_2' - X_1X_3'(X_3X_3')^{-1}X_3X_2'][X_2X_2' - X_2X_3' \times (X_3X_3')^{-1}X_3X_2']^{-1}[X_2X_1' - X_2X_3'(X_3X_3')^{-1}X_3X_1']$ and $[\Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{13}']^{-1}[\Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{23}'] \times [\Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{23}']^{-1}[\Sigma_{12}' - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{13}']$.

Proof: Notice as in previous section that, a.e., the c's are positive and distinct and that the γ's are all non-negative. Use (A.3.20) to set

$$\text{(A.7.4.1)} \quad \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12}' & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13}' & \Sigma_{23}' & \Sigma_{33} \end{bmatrix} = \left[\begin{array}{cc|c} \mu_1 & 0 & \mu_3 \\ 0 & \mu_2 & \mu_4 \\ \hline 0 & 0 & \tilde{\mu}_5 \end{array}\right] \left[\begin{array}{cc|c} I & [D_{\sqrt{\gamma}} \vdots 0] & 0 \\ \begin{bmatrix} D_{\sqrt{\gamma}} \\ 0 \end{bmatrix} & I & \\ \hline \multicolumn{2}{c|}{0} & I \end{array}\right]$$

$$\times \left[\begin{array}{cc|c} \mu_1' & 0 & 0 \\ 0 & \mu_2' & 0 \\ \hline \mu_3' & \mu_4' & \tilde{\mu}_5' \end{array}\right],$$

and check that

$$\text{(A.7.4.2)} \quad \left[\begin{array}{cc|c} \mu_1 & 0 & \mu_3 \\ 0 & \mu_2 & \mu_4 \\ \hline \multicolumn{2}{c|}{0} & \tilde{\mu}_5 \end{array}\right]^{-1} \text{ is of the form } \left[\begin{array}{cc|c} \nu_1 & 0 & \nu_3 \\ 0 & \nu_2 & \nu_4 \\ \hline \multicolumn{2}{c|}{0} & \tilde{\nu}_5 \end{array}\right].$$

Proceeding as in the previous section we have now

$$(A.7.4.3) \qquad \Sigma^{-1} = \left[\begin{array}{cc|c} \mu_1' & 0 & 0 \\ 0 & \mu_2' & 0 \\ \hline \mu_3' & \mu_4' & \tilde{}_5 \end{array}\right]^{-1} \left[\begin{array}{cc|c} I & -[D_{\sqrt{\gamma/1-\gamma}} \quad 0] & 0 \\ 0 & \begin{bmatrix} D_{\sqrt{\gamma/1-\gamma}} & 0 \\ 0 & I \end{bmatrix} & \\ \hline \multicolumn{2}{c|}{0} & I \end{array}\right]$$

$$\times \left[\begin{array}{c:c|c} I & 0 & \\ -\begin{bmatrix} D_{\sqrt{\gamma/1-\gamma}} \\ 0 \end{bmatrix} & \begin{bmatrix} D_{\sqrt{1/1-\gamma}} & 0 \\ 0 & I \end{bmatrix} & 0 \\ \hline \multicolumn{2}{c|}{0} & I \end{array}\right] \left[\begin{array}{cc|c} \mu_1 & 0 & \mu_3 \\ 0 & \mu_2 & \mu_4 \\ \hline 0 & 0 & \tilde{\mu}_5 \end{array}\right]^{-1} = \mu'^{-1} M(\gamma) M'(\gamma) \mu^{-1} \text{(say)},$$

and thus, as in the previous section,

$$(A.7.4.4) \qquad \operatorname{tr} \Sigma^{-1} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} [X_1' \quad X_2' \quad X_3']$$

$$= \operatorname{tr} [X_1' \quad X_2' \quad X_3'] \mu'^{-1} M(\gamma) M'(\gamma) \mu^{-1} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

Now set $\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} \mu_1 & 0 & \mu_3 \\ 0 & \mu_2 & \mu_4 \\ 0 & 0 & \tilde{\mu}_5 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$ and note that the c's are invariant under this transformation. Also $J(X:Y) = |\mu|^n$ and $|\Sigma|^{\frac{n}{2}} = |\mu|^n \prod_{i=1}^{p} (1-\gamma_i)^{\frac{n}{2}}$. Thus, finally Y has the probability law

$$(A.7.4.5) \qquad \left[1/(2\pi)^{\frac{n(p+q+r)}{2}} \prod_{i}^{p} (1-\gamma_i)^{\frac{n}{2}} \right]$$

$$\times \exp \left[-\tfrac{1}{2} \operatorname{tr} M'(\gamma) \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} [Y_1' \quad Y_2' \quad Y_3'] \; M(\gamma) \right] \quad dY_1 dY_2 dY_3$$

which proves (A.7.4) which we take to be a canonical form.

(A.7.5): *If* $X_1(p\times n_1)$ *and* $X_2(p\times n_2)$ $(p \leqslant n_2$ *but might be* $\leqslant$ *or* $> n_1)$ *have he joint probability law* (4.21): $[1/(2\pi)^{p(n_1+n_2)/2}|\Sigma|^{(n_1+n_2)/2}] \exp [-\frac{1}{2} \text{ tr } \Sigma^{-1}\{X_2X_2' +(X_1-\xi)(X_1'-\xi')\}]dX_1dX_2$, *where* $\Sigma(p\times p)$ *is symmetric p.d. and* ξ *is* $p\times n_1$, *then the distribution of* $c(X_1X_1'(X_2X_2')^{-1})$, *to be called c's, could not involve as parameters anything except* $c(\xi\xi'\Sigma^{-1})$, *to be called* γ*'s.*

Proof: Notice that, a.e., out of the p c's, r are positive and $p-r$ are zero, where $r = \min(p, n_1)$ and also that γ's are all non-negative, and, out of them s are positive and $p-s$ are zero, where $s \leqslant \min(p, n_1)$ is the rank of $\xi\xi'$, i.e., of $\xi(p\times n_1)$.

Assuming, as we can without any loss of generality, that the last s rows of ξ, i.e., the last s rows of $\xi\xi'$ can be taken as the basis, use (A.3.13) to set

$$(\text{A.7.5.1}) \qquad (\xi\xi')(p\times p) = \begin{array}{c} \begin{array}{c} p-s \\ s \end{array} \left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right] \\ s \end{array} D_\gamma^* \ (s\times s) \ \begin{array}{c} [\mu_1' \quad \mu_2'] \\ \begin{array}{cc} p-s & s \end{array} \end{array} \text{ and}$$

$$\Sigma(p\times p) = \left[\begin{array}{cc} \mu_1 & \tilde{\mu}_3 \\ \mu_2 & \mu_4 \end{array}\right] \begin{array}{c} \left[\begin{array}{cc} \mu_1' & \mu_2' \\ \tilde{\mu}_3' & \mu_4' \end{array}\right] \begin{array}{c} s \\ p-s \end{array} \\ \begin{array}{cc} p-s & s \end{array} \end{array},$$

where $\mu = \left[\begin{array}{cc} \mu_1 & \tilde{\mu}_3 \\ \mu_2 & \mu_4 \end{array}\right]$ and $\tilde{\mu}_3$ are non-singular and D^* stands for the diagonal matrix with s (non-zero and here positive) roots. If we now set

$$(\text{A.7.5.2}) \qquad \xi(p\times n_1) = \begin{array}{c} \left[\begin{array}{c} \xi_1 \\ \xi_2 \end{array}\right] \begin{array}{c} p-s \\ s \end{array} \\ n_1 \end{array} = \begin{array}{c} \begin{array}{c} p-s \\ s \end{array} \left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right] \\ s \end{array} D^*_{\sqrt{\bar{\gamma}}}(s\times s)\nu(s\times n_1),$$

it is easy to check that ν is determined by $\nu = D^{*-1}_{\sqrt{\bar{\gamma}}}\mu_2^{-1}\,\xi_2$ and that $\nu\nu' = I(s)$. Let

$$D_\gamma(p\times p) = \begin{array}{c} \left[\begin{array}{cc} D_\gamma^* & 0 \\ 0 & 0 \end{array}\right] \begin{array}{c} s \\ p-s \end{array} \\ \begin{array}{cc} s & p-s \end{array} \end{array}.$$

Recall that $s \leqslant \min(p, n_1)$. Recalling now that $tr\ X_1\xi' = \text{tr }\xi'X_1$ and using (A.1.4), (A.7.5.1) and (A.7.5.2) we have

$$(\text{A.7.5.3}) \qquad \text{tr }\Sigma^{-1}\{X_2X_2'+(X_1-\xi)(X_1'-\xi')\} = \text{tr }\mu^{-1}\{X_2X_2'+X_1X_1'-2X_1\nu' \\ \times \quad D^*_{\sqrt{\bar{\gamma}}}[\mu_1' : \mu_2']+\mu D_\gamma\mu'\}\mu'^{-1}.$$

Now using (A.1.7) complete $\nu'(n_1\times s)$ into an $\perp$ $\delta'(n_1\times n_1)$ and rewrite the right hand side of (A.7.5.3) as tr $\mu^{-1}\{X_2X_2'+X_1X_1'-2X_1(p\times n_1)\ \delta'(n_1\times n_1)\times \begin{array}{c} \left[\begin{array}{cc} D^*_{\sqrt{\bar{\gamma}}} & 0 \\ 0 & 0 \end{array}\right] \begin{array}{c} s \\ n_1-s \end{array} \\ \begin{array}{cc} s & s-p \end{array} \end{array}$ $\times\mu'(p\times p)+\mu D_\gamma\mu'\}\mu'^{-1}$. Put now $\mu^{-1}X_2 = Y_2$ and $\mu^{-1}X_1\,\delta' = Y_1$, i.e., $X_1(p\times n_1)$

$= \mu(p\times p)\ Y_1(p\times n_1)\ \delta(n_1\times n_1)$ and $X_2(p\times n_2) = \mu(p\times p)\ Y_2(p\times n_2)$ and observe that by (A.4.3), $c(X_1X_1'(X_2X_2')^{-1}) = c(Y_1Y_1'(Y_2Y_2')^{-1})$ since μ is non-singular and δ' is $\perp$. Also, by (A.5.2), $J(X_1, X_2 : Y_1, Y_2) = |\mu|^{n_1+n_2}$. Also observe that $|\Sigma| = |\mu|^2$. Finaly check that (Y_1, Y_2) has the probability law

$$\text{(A.7.5.4)} \qquad \left[1/2\pi^{\frac{p(n_1+n_2)}{2}}\right] \exp\left[-\tfrac{1}{2}\operatorname{tr}\left\{Y_2Y_2'+D_\gamma+Y_1Y_1'-2Y_1(p\times n_1)\right.\right.$$

$$\left.\left.\times\begin{bmatrix} D^*{}_{\sqrt{\gamma}} & 0 \\ 0 & 0\end{bmatrix}\right\}\right]\ dY_1\ \ dY_2,$$

which, in view of the fact that $c(X_1X_1'(X_2X_2')^{-1}) = c(Y_1Y_1'(Y_2Y_2')^{-1})$, completes the proof of (A.7.5). We also note as before that for the purpose of any discussion of the distribution of c's the probability law (A.7.5.4) can be taken as a canonical form. In (A.7.5.4) notice that

$$\text{(A.7.5.5)} \qquad \operatorname{tr}\ Y_1(p\times n_1) \begin{array}{c} \begin{bmatrix} D^*{}_{\sqrt{\gamma}} & 0 \\ 0 & 0\end{bmatrix}\begin{matrix} s \\ n_1-s\end{matrix} \\ \begin{matrix} s & p-s\end{matrix}\end{array} = \sum_{i=1}^{s} (Y_1)_{ii}\gamma_i^{\frac{1}{2}};\ \operatorname{tr}\ D_\gamma = \sum_{i=1}^{s}\gamma_i.$$

Using (A.7.5.5) the canonical form (A.7.5.4) can thus be reduced to the more convenient form

$$\text{(A.7.5.6)} \quad [1/2\pi]^{\overline{p(n_1+n_2)}} \exp\left[-\tfrac{1}{2}\left\{\operatorname{tr}(Y_1Y_1'+Y_2Y_2')+\sum_{i=1}^{s}\gamma_i-2\Sigma(Y_1)_{ii}\gamma_i^{\frac{1}{2}}\right\}\right]dY_1dY_2.$$

The reader must be cautioned against stretching any further the theorems (A.7.1), (A.7.2), (A.7.3), (A.7.4) and (A.7.5). For example, taking (A.7.1), suppose that $0 \leqslant c_1 \leqslant c_2 \leqslant \dots \leqslant c_p < \infty$ and $0 \leqslant \gamma_1 \dots \leqslant \gamma_p < \infty$. The joint distribution of c_i's could not involve as parameters anything except γ_i's and, in fact, it does involve all these parameters. But it must not be inferred from this (and it is not true either) that the distribution of c_i involves just γ_i, with $i = 1, 2, \dots, p$. In fact, the distribution of any c_i will involve as parameters all γ_i's. Nor do the distributions of the usual symmetric functions of the c_i's involve, in general, the same functions of the γ_i's. The same thing is also true for (A.7.2) — (A.7.5).

APPENDIX 8

Some Results in Integration

$$\text{(A.8.1):}\quad \int\limits_{\Sigma(x_i/a_i)^{q_i}\leqslant 1} \prod_{i=1}^{n} x_i^{p_i-1}dx_i = \prod_{i=1}^{n} \Gamma\left(\frac{p_i}{q_i}\right) a_i^{p_i} \Big/ \Gamma\left(\sum_{i=1}^{n} \frac{p_i}{q_i}+1\right)\prod_{i=1}^{n} q_i.$$

where $x_i \geqslant 0$ and $p_i, q_i, a_i > 0$, $i = 1, 2, \ldots, n$. An important special case is where $a_i = r, p_i = 1$ and $q_i = 2$, in which case we have

$$\text{(A.8.2):}\quad \int\limits_{\sum_{i=1}^{n} x_i^2 \leqslant r^2 (x_i\geqslant 0)} \prod_{i=1}^{n} dx_i = \left[\Gamma\left(\frac{1}{2}\right)\right]^n r^n/2^n\Gamma\left(\frac{n}{2}+1\right).$$

If, however, we integrate over x_i's in the domain $\sum_{i=1}^{n} x_i^2 \leqslant r^2$, after dropping the restriction that $x_i \geqslant 0$, i.e., if x_i's could take both $+ve$ and $-ve$ values, subject to $\sum_{i=1}^{n} x_i^2 \leqslant r^2$, then from considerations of symmetry we shall have

$$\text{(A.8.3):}\quad \int\limits_{\sum_{i=1}^{n} x_i^2 \leqslant r^2} \prod_{i=1}^{n} dx_i = \left[\Gamma\left(\frac{1}{2}\right)\right]^n r^n/\Gamma\left(\frac{n}{2}+1\right).$$

Differentiating the above on both sides w.r.t. r we have

$$\text{(A.8.4):}\quad \int\limits_{r \leqslant \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \leqslant r+dr(r\geqslant 0)} \prod_{i=1}^{n} dx_i = n\left[\Gamma\left(\frac{1}{2}\right)\right]^n r^{n-1}dr/\Gamma\left(\frac{n}{2}+1\right).$$

$$\text{(A.8.5):}\quad \int\limits_{\substack{r \leqslant \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \leqslant r+dr(r \geqslant 0)\\ \theta \leqslant \cos^{-1}\left[\sum_{i=1}^{n} x_i a_i/\left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n} a_i^2\right)^{\frac{1}{2}}\right] \leqslant \theta+d\theta}} \prod_{i=1}^{n} dx_i = \frac{(n-1)\left[\Gamma\left(\frac{1}{2}\right)\right]^{n-1} r^{n-1}}{\Gamma\left(\frac{n-1}{2}+1\right)}\, dr(\sin\theta)^{n-2}d\theta.$$

Proof: Make a transformation $y_1 = \sum_{i=1}^{n} x_i a_i / (\sum_{i=1}^{n} a_i^2)^{\frac{1}{2}} = r^* \cos\theta^*$, say, and

$y_i = \sum_{j=1}^{n} \mu_{ij} x_j (i = 2, 3, \ldots, n)$ such that $\begin{bmatrix} a_1/a & a_2/a & \ldots & a_n/a \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nn} \end{bmatrix}$ $(a^2 = \sum_{i=1}^{n} a_i^2 \text{ and } a \geqslant 0)$

is an orthogonal matrix.

Then using (A.5.4) and remembering that $J(\mathbf{x} : \mathbf{y}) = 1/J(\mathbf{y} : \mathbf{x})$,

we shall have $\prod_{i=1}^{n} dx_i \rightarrow \prod_{i=1}^{n} dy_i$, i.e., $\rightarrow dy_1 \prod_{i=2}^{n} dy_i$. We have furthermore $\sum_{i=1}^{n} y_i^2 = y_1^2 + \sum_{i=2}^{n} y_i^2 = \sum_{i=1}^{n} x_i^2 = r^{*2}$, so that $\sum_{i=2}^{n} y_i^2 = r^{*2} - y_1^2 = r^{*2} \sin^2\theta^*$, whence $(\sum_{i=2}^{n} y_i^2)^{\frac{1}{2}} = r^* \sin\theta^* = u^*$ (say). It is easy to see that the domain: $r \leqslant r^* \leqslant r+dr$ and $\theta \leqslant \theta^* \leqslant \theta+d\theta$, is exactly equivalent to $u \leqslant u^* \leqslant u+du$ and $v \leqslant y_1 \leqslant v+dv$, so that

(A.8.5.1)
$$\int_{\substack{r \leqslant r^* \leqslant r+dr \\ \theta \leqslant \theta^* \leqslant \theta+d\theta}} \prod_{i=1}^{n} dx_i = \int_{\substack{v \leqslant y_1 \leqslant v+dv \\ u \leqslant u^* \leqslant u+du}} \prod_{i=1}^{n} dy_i$$

$$= dv \int_{u \leqslant (\sum_{i=2}^{n} y_i^2)^{\frac{1}{2}} \leqslant u+du} \prod_{i=2}^{n} dy_i$$

$$= dv(n-1)[\Gamma(\tfrac{1}{2})]^{n-1} u^{n-2} du / \Gamma\left(\frac{n-1}{2} + 1\right) \text{ (using (A.8.4))} = (n-1)[\Gamma(\tfrac{1}{2})]^{n-1} r^{n-1} dr$$

$$\times \ (\sin\theta)^{n-2} d\theta / \Gamma\left(\frac{n-1}{2} + 1\right),$$

which proves (A.8.5). Notice that $y_1 = r^* \cos\theta^*$ and $u^* = r^* \sin\theta^*$, whence $J(y_1, u^* : r^*, \theta^*) = r^*$, so that the $dy_1 du^* \rightarrow r^* \, dr^* \, d\theta^*$.

(A.8.6): *The integral* $\int_{L' = I(p)} dL \Big/ \left| \frac{\partial(LL')}{\partial(L_D)} \right|_{L_I} = F(p, n)$ *(say), where* L *is* $p \times n$ *with* $p \leqslant n$. This can be evaluated directly but we shall use an artifice to derive this. Consider the integral

(A.8.6.1)
$$\int_{Y} [1/(2\pi)^{\frac{pn}{2}}] \exp[-\tfrac{1}{2} \operatorname{tr} YY'] \, dY,$$

where the elements of $Y(p \times n)$ $(p \leqslant n)$ vary from $-\infty$ to ∞. It is of course known that this integral is equal to 1.

Using now the transformation (A.3.11) we have $Y(p \times n) = \tilde{T}(p \times p)\, L(p \times n)$ under $LL' = I(p)$. Notice that almost everywhere YY', and so $\tilde{T}$ will be non-singular. The t_{ii}'s vary from 0 to ∞ and t_{ij}'s $(i \neq j)$ vary from $-\infty$ to ∞. Observe that $YY' = \tilde{T}\tilde{T}'$ and tr $YY' = \sum\limits_{i,j=1}^{p} t_{ij}^2$. Using now the result (A.6.1.9) we have

$$\text{(A.8.6.2)} \qquad 1 = \int\limits_{Y} \left[1/(2\pi)^{\frac{pn}{2}}\right] \exp\left[-\tfrac{1}{2} \operatorname{tr} YY'\right] dY$$

$$= \int\limits_{LL'=I(p)} \left[dL_I \left| \left| \frac{\partial(LL')}{\partial(L_D)} \right|_{L_I} \right] 2^p [1/2\pi]^{\frac{pn}{2}}$$

$$\times \int\limits_{\substack{0 \leqslant t_{ii}\text{'s} < \infty \\ -\infty < t_{ij}\text{'s} < \infty \\ (i > j).}} \exp\left[-\tfrac{1}{2} \sum_{i \geqslant j=1}^{p} t_{ij}^2\right] \prod_{i=1}^{p} t_{ii}^{n-i} \prod_{i \geqslant j=1}^{p} dt_{ij}.$$

But the last integral on the right hand side of (A.8.6.2) is easily evaluated to be $2^{-p+pn/2}\pi^{p(p-1)/4} \prod\limits_{i=1}^{p} \Gamma\left[\frac{n-i+1}{2}\right]$. Hence we have the following result (to be repeatedly used):

$$\text{(A.8.6.3)} \qquad F(p, n) = \int\limits_{LL'=I(p)} dL_I \left| \left| \frac{\partial(LL')}{\partial(L_D)} \right|_{L_I} = \pi^{pn/2-p(p-1)/4} \Big/ \prod_{i=1}^{p} \Gamma\left[\frac{n-i+1}{2}\right].$$

Another integral that is useful is

$$\text{(A.8.6.4)} \qquad \int\limits_{LL'=I(p)} dL_I \div \left| \frac{\partial(LL')}{\partial(L_D)} \right|_{L_I} = F_1(p, n) \quad \text{(say)},$$

where L is $p \times n$, $p \leqslant n$ and where the first row of L is to be non-negative. To evaluate this we consider the integral

$$\text{(A.8.6.5)} \qquad \int\limits_{Y} \left[1/(2\pi)^{\frac{pn}{2}}\right] \exp\left[-\tfrac{1}{2} \operatorname{tr} YY'\right] dY$$

where $Y(p \times n)(p \leqslant n)$ is such that the elements of the first row vary from 0 to ∞ and the other elements vary from $-\infty$ to ∞. Now using the transformation (A.3.11) we have $Y(p\times n) = \tilde{T}(p\times p)\, L(p\times n)$, where $LL' = I(p)$ and the first row of L is to be positive. Then proceeding exactly in the same manner as in the previous case we should have

$$F_1(p,\, n) = 2^{-n}\pi^{pn/2-p(p-1)/4} \Big/ \prod_{i=1}^{p} \Gamma\left\{\frac{n-i+1}{2}\right\}, \tag{A.8.6.6}$$

or in other words,

$$F_1(p,n) = 2^{-n}F(p,\, n), \tag{A.8.6.7}$$

and hence

$$F_1(p,\, p) = 2^{-p}F(p,\, p). \tag{A.8.6.8}$$

Another integral that is useful is the one that arises out of (A.6.1.18), namely,

$$\int\limits_{L_{r+1,pI}} dL_{r+1,pI} \Bigg/ \left| \frac{\partial(L_{r+1,p}L'_{r+1,p},\ L_{r+1,p}L'_{1,r})}{\partial(L_{r+1,pD})} \right|_{L_I}, \tag{A.8.6.9}$$

where the variables $L_{r+1,p}$ are subject to constraints among themselves and also in relation to the constants $L_{1,r}$ which are described by (A.6.1.15). The integral can be obtained by going back to (A.6.1.18) and equating the integral of the left side over $(\mathrm{x}_i\, \mathrm{x}_i')^{\frac{1}{2}} \leqslant 1 (i = r+1, \ldots, p)$, and the right side over $(\sum\limits_{j=1}^{i} t_{ij}^2)^{\frac{1}{2}} \leqslant 1$, subject to $t_{ii} \geqslant 0\, (i = r+1, \ldots, p)$ and over $L_{r+1,pI}$. By using (A.8.1) modified by using a factor of 2 to allow for each $t_{ij}(j = 1, 2, \ldots, i-1 \text{ and } i = r+1, \ldots, p)$ to be both positive and negative we observe that this leads to the equation

$$\pi^{\frac{n(p-r)}{2}} \Big/ \Gamma^{p-r}\left(\frac{n}{2}+1\right) = \pi^{\frac{(p-r)(p+r-1)}{4}} \prod_{i=r+1}^{p} \Gamma\left(\frac{n-i+1}{2}\right) \Big/ \Gamma^{p-r}\left(\frac{n}{2}+1\right) \tag{A.8.6.10}$$

$$\times \text{ the integral (A.8.6.9),}$$

whence it follows that

$$\int\limits_{L_{r+1,pI}} dL_{r+1,pI} \Bigg/ \left| \frac{\partial(L_{r+1,p}L'_{r+1,p}, L_{r+1,p}L'_{1,r})}{\partial(L_{r+1,pD})} \right|_{L_I} = \pi^{\frac{n(p-r)}{2} - \frac{(p-r)(p-r-1)}{4}} \Bigg/ \prod_{i=r+1}^{p} \Gamma\left(\frac{n-i+1}{2}\right). \tag{A.8.6.11}$$

It is worth a careful notice that this result is independent of the set of given constants $L_{1,r}$.

(A.8.7): *The integral* $\int_U \exp[-\frac{1}{2} \operatorname{tr} UU']|U|^q \, dU$, *where* $U(p \times p)$ *has its elements varying from* $-\infty$ *to* ∞. Using the transformation (A.3.11) we have $U(p \times p) = \tilde{T}(p \times p) L(p \times p)$ with an orthogonal L. Notice that $|U| = |\tilde{T}| = \prod_{i=1}^{p} t_{ii}$ and that almost everywhere $\tilde{T}$ is non-singular. Also, as before, $\operatorname{tr} UU' = \sum_{i \geqslant j=1}^{p} t_{ij}^2$. Hence we have, by using (A.6.1.9) and (A.8.6.3), [31, 32],

$$\text{(A.8.7.1)} \qquad \int_U \exp[-\tfrac{1}{2} \operatorname{tr} UU']|U|^q dU = 2^p \int_{LL'=I(p)} dL_I \Big/ \left| \frac{\partial(LL')}{\partial(L_D)} \right|_{L_I}$$

$$\times \int_{\substack{0 \leqslant t_{ii} < \infty, \\ -\infty < t_{ij} < \infty \\ (i > j)}} \exp[-\tfrac{1}{2} \sum_{i \geqslant j=1}^{p} t_{ij}^2] \prod_{i=1}^{p} t_{ii}^{p+q-i} \prod_{i \geqslant j=1}^{p} dt_{ij}$$

$$= 2^{p(q+p)/2} \pi^{p^2/2} \prod_{i=1}^{p} \Gamma\left(\frac{q+p-i+1}{2}\right) \Big/ \prod_{i=1}^{p} \Gamma\left(\frac{p-i+1}{2}\right).$$

Since $U(p \times p)$, (with a positive first row) $= \tilde{T}(p \times p) L(p \times p)$ (where $LL' = I(p)$ and L has a positive first row), therefore, from (A.8.6.8) and (A.8.7.1) we have

$$\text{(A.8.7.2)} \qquad \int_{U(\textit{with a positive first row})} \exp[-\tfrac{1}{2} \operatorname{tr} UU']|U|^q \, dU$$

$$= 2^{p(q+p-2)/2} \pi^{p^2/2} \prod_{i=1}^{p} \Gamma\left(\frac{q+p-i+1}{2}\right) \Big/ \prod_{i=1}^{p} \Gamma\left(\frac{p-i+1}{2}\right).$$

(A.8.8): *The integral* $\int_U \exp[-\frac{1}{2} \operatorname{tr} UU']|U|^q \prod_{i=1}^{p-n} (\tilde{U}_3)_{ii}^{p-n-i} \, dU$; *where* $U(p \times p) = \begin{array}{cc} \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix} & \begin{array}{c} p-n \\ n \end{array} \\ \begin{array}{cc} n & p-n \end{array} & \end{array}$, *where the first row of* U_1 *and the diagonal of* $\tilde{U}_3$ *vary from* 0 *to* ∞ *and the rest from* $-\infty$ *to* ∞.

To evaluate this integral, let $V = \begin{array}{cc} \begin{bmatrix} V_1 & V_3 \\ V_2 & V_4 \end{bmatrix} & \begin{array}{c} p-n \\ n \end{array} \\ \begin{array}{cc} n & p-n \end{array} & \end{array} = \begin{bmatrix} U_1 & \tilde{U}_3 M \\ U_2 & U_4 M \end{bmatrix}$

$= \begin{array}{cc} \begin{array}{c} p-n \\ n \end{array} & \begin{bmatrix} U_1 & \tilde{U}_3 \\ U_2 & U_4 \end{bmatrix} \\ & \begin{array}{cc} n & p-n \end{array} \end{array} \begin{array}{cc} \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} & \begin{array}{c} n \\ p-n \end{array} \\ \begin{array}{cc} n & p-n \end{array} & \end{array}$, where M is $\perp$ and with a positive first row.

Then we have $VV' = UU'$ and $|V| = |U|$ and

$$J(V : U) = J(V_3, V_4 : \tilde{U}_3, M_I, U_4) = J(V_3 : \tilde{U}_3, M_I)\, J(V_4 : U_4)$$

$$= 2^{p-n} \prod_{i=1}^{p-n} (\tilde{U}_3)_{ii}^{p-n-i} |M|^n \Big/ \left| \frac{\partial(MM')}{\partial(M_D)} \right|_{M_I} = 2^{p-n} \prod_{i=1}^{p-n} (\tilde{U}_3)_{ii}^{p-n-i} \Big/ \left| \frac{\partial(MM')}{\partial(M_D)} \right|_{M_I}.$$

Thus we have

$$\text{(A.8.8.1)} \qquad \int_V \exp[-\tfrac{1}{2} \operatorname{tr} VV'] |V|^q \, dV$$

$$= 2^{p-n} \int_U \exp[-\tfrac{1}{2} \operatorname{tr} UU'] |U|^q \prod_{i=1}^{p-n} (\tilde{U}_3)_{ii}^{p-n-i} \, dU \times \int_{M_I} dM_I \Big/ \left| \frac{\partial(MM')}{\partial(M\;\;)} \right|_{M_I}.$$

Now substituting from (A.8.7.2) in the left hand side of (A.8.8.1) and from (A.8.6.3) and (A.8.6.8) in the right hand side of (A.8.8.1) we have, [31],

$$\text{(A.8.8.2)} \qquad \int_U \exp[-\tfrac{1}{2}\, tr\; UU'] |U|^q \prod_{i=1}^{p-n} (\tilde{U}_3)_{ii}^{p-n-i} \, dU$$

$$= 2^{p(q+p)/2} \pi^{\frac{p^2}{2} - \frac{(p-n)(p-n+1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{q+p-i+1}{2}\right) \prod_{i=1}^{p-n} \Gamma\left(\frac{p-n-i+1}{2}\right) \div \prod_{i=1}^{p} \Gamma\left(\frac{p-i+1}{2}\right).$$

APPENDIX 9

Some Results in Integration Connected with the Distribution of the Largest and the Smallest Roots

(A.9.1): *Evaluation of the integral*

$$\int_{x_s=0}^{x}\int_{x_{s-1}=0}^{x_s}\cdots\int_{x_2=0}^{x_3}\int_{x_1=0}^{x_2}\prod_{i=1}^{s}dx_i\begin{vmatrix} x_s^{m_s}(1-x_s)^{n_s} & x_{s-1}^{m_s}(1-x_{s-1})^{n_s} & \cdots & x_1^{m_s}(1-x_1)^{n_s} \\ x_s^{m_{s-1}}(1-x_s)^{n_{s-1}} & x_{s-1}^{m_{s-1}}(1-x_{s-1})^{n_{s-1}} & \cdots & x_1^{m_{s-1}}(1-x_1)^{n_{s-1}} \\ \vdots & \vdots & \cdots & . \\ x_s^{m_1}(1-x_s)^{n_1} & x_{s-1}^{m_1}(1-x_{s-1})^{n_1} & \cdots & x_1^{m_1}(1-x_1)^{n_1} \end{vmatrix}$$

$$=\beta[x;\, m_s,\, n_s;\, m_{s-1},\, n_{s-1};\, \ldots;\, m_1,\, n_1]\ (\text{say})=\beta\left[x;\begin{pmatrix} m_s,n_s & m_{s-1},n_{s-1} & \cdots & m_1,n_1 \\ m_s,n_s & m_{s-1},n_{s-1} & \cdots & m_1,n_1 \\ . & . & \cdots & . \\ m_s,n_s & m_{s-1},n_{s-1} & \cdots & m_1,n_1 \end{pmatrix}\right]$$

(say).

The last expression is in the form of a pseudo-determinant whose meaning is made clear by considering, for illustration, the case of $s = 3$, for which

(A.9.1.1)
$$\beta\left[x;\begin{pmatrix} m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \\ m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \\ m_3,\, n_3 & m_2,\, n_2 & m_1,\, n_1 \end{pmatrix}\right]$$

$$= \int_0^x x_3^{m_3}(1-x_3)^{n_3}\,dx_3\left[\int_0^{x_3} x_2^{m_2}(1-x_2)^{n_2}\,dx_2\int_0^{x_2} x_1^{m_1}(1-x_1)^{n_1}\,dx_1\right.$$

$$\left.-\int_0^{x_3} x_2^{m_1}(1-x_2)^{n_1}\,dx_2\int_0^{x_2} x_1^{m_2}(1-x_1)^{n_2}\,dx_1\right]$$

$$-\int_0^x x_3^{m_2}(1-x_3)^{n_2}\,dx_3\left[\int_0^{x_3} x_2^{m_3}(1-x_2)^{n_3}\,dx_2\int_0^{x_2} x_1^{m_1}(1-x_1)^{n_1}\,dx_1\right.$$

$$\left.\int_0^{x_3} x_2^{m_1}(1-x_2)^{n_1}\,dx_2\int_0^{x_2} x_1^{m_3}(1-x_1)^{n_3}\,dx_1\right]$$

$$+\int_0^x x_3^{m_1}(1-x_3)^{n_1}\,dx_3\left[\int_0^{x_3} x_2^{m_3}(1-x_2)^{n_3}\,dx_2\int_0^{x_2} x_1^{m_2}(1-x_1)^{n_2}\,dx_1\right.$$

$$\left.\int_0^{x_3} x_2^{m_2}(1-x_2)^{n_2}\,dx_2\int_0^{x_2} x_1^{m_3}(1-x_1)^{n_3}\,dx_1\right].$$

In opening out the pseudo-determinant it is very important to stick to the order of the factors, indicated in the expansion on the right side of (A.9.1.1) and to keep in mind that the factors are non-commutative. It is also clear that the whole expression will be zero if any two columns become equal in

$$\begin{pmatrix} m_s, n_s & m_{s-1}, n_{s-1} & \cdots & m_1, n_1 \\ m_s, n_s & m_{s-1}, n_{s-1} & \cdots & m_1, n_1 \\ \cdot & \cdot & \cdots & \cdot \\ m_s, n_s & m_{s-1}, n_{s-1} & \cdots & m_1, n_1 \end{pmatrix}$$

Next we use the notation

(A.9.1.2)
$$\beta(x; m_s, n_s; m_{s-1}, n_{s-1}; \ldots ; m_1, n_1)$$

$$= \int_0^x x_s^{m_s}(1-x_s) \quad dx_s \int_0^{x_s} x_{s-1}^{m_{s-1}} (1-x_{s-1})^{n_{s-1}} dx_{s-1}$$

$$\cdots \int_0^{x_3} x_2^{m_2}(1-x_2)^{n_2}\, dx_2 \int_0 x_1^{m_1}(1-x_1)^{n_1}\, dx_1 ,$$

so that $\beta(x; m, n) = \int_0^x x_1^m(1-x_1)^n dx_1 =$ the incomplete β-function, using a slightly different notation from the usual one. Also let $x^m(1-x)^n = \beta_0(x; m\ n)$. In terms of (A.9.1.2), the expression (A. 9.1.1) can be rewritten as

(A.9.1.3)
$$\begin{aligned} &\beta(x; m_3, n_3; m_2, n_2; m_1, n_1)-\beta(x; m_3, n_3; m_1, n_1; m_2, n_2) \\ &-\beta(x; m_2, n_2; m_3, n_3; m_1, n_1)+\beta(x; m_2, n_2; m_1, n_1; m_3, n_3) \\ &+\beta(x; m_1, n_1; m_3, n_3; m_2, n_2)-\beta(x; m_1, n_1; m_2, n_2; m_3, n_3) \end{aligned}$$

and (A.9.1) can be rewritten as

(A.9.1.4)
$$\Sigma\pm\beta(x; m_s', n_s'; m_{s-1}', n_{s-1}'; \ldots; m_1', n_1'),$$

where (m_s', n_s') $(m_{s-1}', n_{s-1}'), \ldots, (m_1', n_1')$ is any permutation of (m_s, n_s), $(m_{s-1}, n_{s-1}), \ldots,$ (m_1, n_1), the summation is taken over all such permutations, the positive or negative sign is taken exactly as in the usual expansion of a determinant, care being taken to preserve the order of factorization from x_s through x_{s-1}, x_{s-2} down to x_1.

(A.9.2): Lemma:

$$\int_0^{x_0} x^m(1-x)^n f(x)dx$$

$$=\frac{1}{m+n+1}\left[-x_0^m(1-x_0)^{n+1}f(x_0)+\int_0^{x_0} x^m(1-x)^{n+1}f'(x)dx+m\int_0^{x_0} x^{m-1}(1-x)^n f(x)dx\right],$$

where m, $n > -1$, $x_0 \leqslant 1$, *and* $f(x)$ *is such that* $f'(x)$ *and the three integrals on two sides of* (A.9.2) *exist.*

Proof: The proof can be carried out by integration by parts, the integration being with respect to the function $(1-x)^{m+n}$ and the differentiation being with respect to the function $f(x)x^m/(1-x)^m$.

(A.9.3) Lemma: $\Sigma\ \beta(x;\ m_s',\ n_s';\ m_{s-1}',\ n_{s-1}';\ \ldots;\ m_1',\ n_1') = \prod_{i=1}^{s} \beta(x;\ m_i,\ n_i)$, *where on the left hand side* $(m_s', n_s'), \ldots, (m_1', n_1')$ *is any permutation of* $(m_s, n_s), \ldots, (m_1, n_1)$, *the summation is taken over all such permutations and where the factors on the right hand side have been already defined.*

Proof: The nature of the proof will be evident by considering, for simplicity of algebra, the case of $s = 2$. We have

$$\text{(A.9.3.1)} \qquad \int_0^x x_2^{m_2}(1-x_2)^{n_2}\,dx_2 \int_0^{x_2} x_1^{m_1}(1-x_1)^{n_1}\,dx_1$$

$$+\int_0^x x_2^{m_1}(1-x_2)^{n_1}\,dx_2 \int_0^{x_2} x_1^{m_2}(1-x_1)^{n_2}\,dx_1$$

$$=\int_0^x x_2^{m_2}(1-x_2)^{n_2}\,dx_2 \int_0^{x_2} x_1^{m_1}(1-x_1)^{n_1}\,dx_1+\int_0^x x_2^{m_2}(1-x_2)^{n_2}\,dx_2 \int_{x_2}^x x_1^{m_1}(1-x_1)^{n_1}dx_1,$$

(which is obtained by interchanging, in the second term on the left side of (A.9.3.1) the variables x_2 and x_1 and rewriting the domain of integration in the appropriate manner)

$$=\int_0^x x_2^{m_2}(1-x_2)^{n_2}\,dx_2 \int_0^x x_1^{m_1}(1-x_1)^{n_1}\,dx_1 = \beta(x;\ m_2,\ n_2)\ \beta(x;\ m_1,\ n_1).$$

(A.9.4): Lemma: $\sum_r \beta_r(x;\ m_{s-1},\ n_{s-1};\ \ldots;\ m_r,\ n_r;\ m,\ n;\ m_{r-1},\ n_{r-1};\ \ldots;\ m_1,\ n_1)$ $= \beta(x;\ m,\ n)\ \beta(x;\ m_{s-1},\ n_{s-1};\ \ldots;\ m_1,\ n_1)$, *where* β_r *is the result of putting* (m, n) *in the* r^{th} *place and filling up the other positions with* $(m_{s-1}, n_{s-1}), (m_{s-2}, n_{s-2}), \ldots, (m_1, n_1)$, r *running from* 1 *to* s. *Notice that each* β_r *is an s-fold integral, while* $\beta(x;\ m_{s-1},\ n_{s-1};\ \ldots,\ m_1,\ n_1)$ *is an* $(s-1)$*-fold integral.*

Proof: The mechanism of the proof is brought out by considering, in particular, the case $s = 3$, where we have

$$\text{(A.9.4.1)} \qquad \beta_1(x;\, m_2, n_2; m_1, n_1;\, m, n) + \beta_2(x;\, m_2, n_2;\, m, n;\, m_1, n_1) + \beta_3(x;\, m, n;\, m_2, n_2;\, m_1, n_1)$$

$$= \int_0^x x_3^{m_2}(1-x_3)^{n_2}\, dx_3 \int_0^{x_3} x_2^{m_1}(1-x_2)^{n_1}\, dx_2 \int_0^{x_2} x_1^{m}(1-x_1)^{n} dx_1$$

$$+ \int_0^x x_3^{m_2}(1-x_3)^{n_2}\, dx_3 \int_0^{x_3} x_2^{m_1}(1-x_2)^{n_1}\, dx_2 \int_{x_2}^{x_3} x_1^{m}(1-x_1)^{n} dx$$

$$+ \int_0^x x_3^{m_2}(1-x_3)^{n_2}\, dx_3 \int_0^{x_3} x_2^{m_1}(1-x_2)^{n_1}\, dx_2 \int_{x_3}^{x} x_1^{m}(1-x_1)^{n} dx$$

(by interchanging the variables and suitably adjusting the domain of integration

$$= \beta(x;\, m, n)\ \beta(x;\, m_2, n_2;\ m_1, n_1).$$

(A.9.5): Lemma:

$$\sum_r (-1)^{r-1} \beta_r \left[x; \begin{pmatrix} m_s, n_s & \cdots & m_1, n_1 \\ \cdot & \cdots & \cdot \\ m'_s, n'_s & \cdots & m'_1, n'_1 \\ \cdot & \cdots & \cdot \\ m_s, n_s & \cdots & m_1, n_1 \end{pmatrix} \right]$$

$$= \sum_r (-1)^{r-1} \beta(x;\, m'_{s-r+1}, n'_{s-r+1})\ \beta_{rr} \left[x; \begin{pmatrix} m_s, n_s & \cdots & m_1, n_1 \\ \cdot & \cdots & \cdot \\ m_s, n_s & \cdots & m_1, n_1 \end{pmatrix} \right],$$

where $\beta_r[x\,; \ldots]$ on the left side is the result of replacing the r^{th} row of $\beta[x\,; \ldots]$ by $(m'_s, n'_s), \ldots, (m'_1, n'_1)$ and $\beta_{rr}[x\,; \ldots]$ on the right side is the result of suppressing the r^{th} row and r^{th} column of $\beta[x\,; \ldots]$. Notice that $\beta_r\,[x\,; \ldots\,]$ is an $s \times s$ and $\beta_{rr}\,[x\,; \ldots]$ an $(s-1)\times(s-1)$ pseudo-determinant.

Proof: The mechanism of the proof will be made clear by considering for simplicity the case of $s = 3$ and picking out from the expansion of each pseudo-determinant on the left side of (A.9.5) (for the case $s = 3$) the term involving the index, say, (m_3', n_3') and putting together all such terms (with index m_3', n_3'). We have thus the following contribution from such terms

(A.9.5.1) $$\beta(x; m_3', n_3'; m_2, n_2,; m_1, n_1) - \beta(x; m_3', n_3'; m_1, n_1; m_2, n_2)$$
$$+\beta(x; m_2,n_2; m_3',n_3'; m_1, n_1) - \beta(x; m_1, n_1; m_3', n_3'; m_2, n_2) + \beta(x; m_2,n_2; m_1,n_1; m_3', n_3')$$
$$-\beta(x; m_1, n_1; m_2, n_2; m_3',n_3') = \beta(x; m_3', n_3')[\beta(x; m_2, n_2; m_1, n_1) - \beta(x; m_1, n_1; m_2, n_2)]$$

(using (A.9.4))

$$= \beta(x;\ m_3', n_3')\ \beta\left[x;\begin{pmatrix} m_2, n_2 & m_1, n_1 \\ m_2, n_2 & m_1, n_1 \end{pmatrix}\right]$$

$$= \beta(x;\ m_3', n_3')\ \beta_{11}\left[x;\begin{pmatrix} m_3, n_3 & m_2, n_2 & m_1, n_1 \\ m_3, n_3 & m_2, n_2 & m_1, n_1 \\ m_3, n_3 & m_2, n_2 & m_1, n_1 \end{pmatrix}\right]$$

(using the notation introduced in the beginning of lemma (A.9.5)). This immediately shows that if, in the general case, from the expansion of each pseudo-determinant (with the proper sign) on the left side of (A.9.5) we pick out the term with the index (m_s', n_s') and add together such terms (with the same index (m_s', n_s')) we shall have the following contribution

(A.9.5.2) $$\beta(x;\ m_s', n_s')\ \beta_{11}\left[x;\begin{pmatrix} m_s, n_s & \dots & m_1, n_1 \\ \cdot & \dots & \cdot \\ m_s, n_s & \dots & m_1, n_1 \end{pmatrix}\right]$$

whence the proof becomes obvious by combining different expressions like (A.9.5.2) involving the different indices (m_r', n_r') $(r = 1, 2, \dots, s)$.

(A.9.6): *Reduction and evaluation of the integral*

$$\beta\left[x;\begin{pmatrix} m_s, n & m_{s-1}, n & \dots & m_1, n \\ \cdot & \cdot & \dots & \cdot \\ m_s, n & m_{s-1}, n & \dots & m_1, n \end{pmatrix}\right],$$

where $m_s > m_{s-1} > \dots > m_1 > -1$ *and* $n > -1$ *and the m's differ by integers.*

We have already seen from (A.9.1.4) that the pseudo-determinant can be expanded into $\Sigma\pm\beta(x; m_s', n; \ldots; m_1', n)$ where $(m_s', \ldots m_1')$ is any permutation of $(m_s, \ldots, m_1)$, the summation is over all such permutations, $s!$ in number, and the positive or negative sign is to be taken according as it is an even or an odd permutation. Recalling from (A.9.1) that β will be zero if any two columns of the pseudo-determinant are equal, let us try to reduce m_s to m_{s-1} by successive integration by parts. Toward this end consider the typical term in the expansion and in that term let m_r' be the largest exponent $= m_s$ (of course). To reduce this exponent by 1 we proceed as follows. By definition

(A.9.6.1) $$\beta(x; m_s', n; \ldots; m_{r+1}', n; m_s, n; m_{r-1}', n; \ldots, m_1', n) = \int_0^x x_s^{m_s'}(1-x_s)^n\, dx_s \ldots$$

$$\ldots \int_0^{x_{r+2}} x_{r+1}^{m_{r+1}'} (1-x_{r+1})^n\, dx_{r+1} \int_0^{x_{r+1}} x_r^{m_s} (1-x_r)^n\, dx_r \int_0^{x_r} x_{r-1}^{m_{r-1}'}(1-x_{r-1})^n\, dx_{r-1} \ldots$$

$$\ldots \int_0^{x_2} x_1 {}^{\prime 1}(1-x_1)^n\, dx_1.$$

Now using (A.9.2) we have

(A.9.6.2) $$\int_0^{x_{r-1}} x_r^{m_s} (1-x_r)^n\, dx_r \int_0^{x_r} x_{r-1}^{m_{r-1}'}(1-x_{r-1})^n dx_{r-1} \ldots \int_0^{x_2} x_1^{m_1'} (1-x_1)^n\, dx_1$$

$$= \int_0^{x_{r+1}} x_r^{m_s} (1-x_r)^n dx_r \quad \beta(x_r;\ m_{r-}' ,\ n;\ \ldots;\ m_1', n)$$

$$= \frac{1}{m_s+n+1} \Bigg[-x_{r+1}^{m_s} (1-x_{r+1})^{n+1} \beta(x_{r+1}; m_{r-1}', n; \ldots; m_1', n)$$

$$+ \int_0^{x_{r+1}} x_r^{m_s} (1-x_r)^{n+1} \beta'(x_r; m_{r-1}'\, n; \ldots; m_1', n) dx_r$$

$$+ m_s \int_0^{x_{r+1}} x_r^{m_s-1} (1-x_r)^n \beta(x_r; m_{r-1}', n; \ldots; m_1', n) dx_r \Bigg]$$

$$= \frac{1}{m_s+n+1} [-x_{r+1}^{m_s}(1-x_{r+1})^{n+1}\, \beta(x_{r+1}; m_{r-1}', n; \ldots; m_1', n)$$

$$+ \beta(x_{r+1}; m_{r-1}'+m_s, 2n+1; m_{r-2}', n; \ldots; m_1', n)$$

$$+ m_s \beta(x_{r+1}; m_s-1, n;\ m^{r}_{\prime -1}, n; \ldots; m_1', n)]$$

(notice that $\beta'(x_r; m'_{r-1}, n; \ldots; m'_1, n) = x_r^{m'_r-1}(1-x_r)^n \beta(x_r; m'_{r-2}, n; \ldots m'_1, n)$ and also that on the right hand side of (A.9.6.2), the first and second β's are each an $(r-1)$-fold integral while the third β is an r-fold integral).

Now substituting the right hand side of (A.9.6.2) we have (A.9.6.1) reducing to

$$\begin{aligned}\frac{1}{m_s+n+1}&[-\beta(x; m'_s, n; \ldots; m'_{r+1}+m_s, 2n+1; m'_{r-1}, n; \ldots, m'_1, n)\\ &+\beta(x; m'_s, n; \ldots; m'_{r+1}, n; m'_{-r1}+m_s, 2n+1; m'_{r-2}, n; \ldots; m'_1, n)\\ &+m_s\beta(x; m'_s, n; \ldots; m'_{r+1}, n; m_s-1, n; \ldots; m'_1, n)],\end{aligned}\tag{A.9.6.3}$$

where the first and second β's are each an (s-1) fold integral while the third β is an s-fold integral with the index m_s reduced to m_s-1. It is easy to check through (A.9.6.1) to (A.9.6.3) that the reduction to (A.9.6.3) holds for $r = s-1, s-2, \ldots, 2$. If $r = s$, it is easy to see that (A.9.6.3) will be replaced by

$$\begin{aligned}\frac{1}{m_s+n+1}&[-\beta_0(x; m_s, n+1)\,\beta(x; m'_{s-1}, n; \ldots; m'_1, n)\\ &+\beta(x; m'_{s-1}+m_s, 2n+1; m'_{s-2}, n; \ldots; m'_1, n)\\ &+m_s\beta(x; m_s-1, n; m'_{s-1}, n; \ldots; m'_1, n)];\end{aligned}\tag{A.9.6.4}$$

and if $r = 1$, (A.9.6.3) will be replaced by

$$\begin{aligned}\frac{1}{m_s+n+1}&[-\beta(x; m'_s, n; \ldots; m'_3, n; m'_2+m_s, 2n+1)\\ &+m_s\beta(x; m'_s, n; \ldots; m'_2, n; m_s-1, n)]\end{aligned}\tag{A.9.6.5}$$

We can now use the rather convenient notation

$$\begin{aligned}&\beta(x; m'_s, n; \ldots; m'_{r+1}+m_s, 2n+1; m'_{r-1}, n; \ldots; m'_1, n)\\ &=\beta(x; m'_s, n; \ldots; m'_{r+1}, n; \overleftarrow{m_s}, \overleftarrow{n+1}; m'_{r-1}, n; \ldots; m'_1, n)\end{aligned}\tag{A.9.6.6}$$

where $(\overleftarrow{m_s}, \overleftarrow{n+1})$ is supposed to be added to the (m'_{r+1}, n) on the left so as to reduce the integral by one dimension,

$$\begin{aligned}&\beta_0(x; m_s, n+1)\,\beta(x; m'_{s-1}, n; \ldots; m'_1, n)\\ &\qquad= \beta(x; \overleftarrow{m_s}, \overleftarrow{n+1}; m'_{s-1}, n; \ldots; m'_1, n), \quad \text{and}\end{aligned}\tag{A.9.6.7}$$

$$\begin{aligned}&\beta(x; m_s, n; \ldots; m'_{r-1}+m_s, 2n+1; m'_{r-2}, n; \ldots; m'_2, n)\\ &=\beta(x; m'_s, n\ldots; \overrightarrow{m_s}, \overrightarrow{n+1}; m'_{r-1}, n; m'_{r-2}, n; \ldots; m'_2, n),\end{aligned}\tag{A.9.6.8}$$

where $(\overrightarrow{m_s}, \overrightarrow{n+1})$ is supposed to be added to the (m'_{r-1}, n) on the right so as to reduce the integral by one dimension. Using now (A.9.6.1)—(A.9.6.8) we have

$$\beta\left[x;\begin{pmatrix} m_s, n & m_{s-1}, n & \cdots & m_1, n \\ \cdot & \cdot & \cdots & \cdot \\ m_s, n & m_{s-1}, n & \cdots & m_1, n \end{pmatrix}\right] \tag{A.9.6.9}$$

$$= -\frac{1}{m_s+n+1}\,\beta\left[x;\begin{pmatrix} \overleftarrow{m_s}, \overleftarrow{n+1} & m_{s-1}, n & \cdots & m_1, n \\ \overleftarrow{m_s}, \overleftarrow{n+1} & m_{s-1}, n & \cdots & m_1, n \\ \overleftarrow{m_s}, \overleftarrow{n+1} & m_{s-1}, n & \cdots & m_1, n \end{pmatrix}\right]$$

$$+\frac{1}{m_s+n+1}\,\beta\left[x;\begin{pmatrix} \overrightarrow{m_s}, \overrightarrow{n+1} & m_{s-1}, n & \cdots & m_1, n \\ \overrightarrow{m_s}, \overrightarrow{n+1} & m_{s-1}, n & \cdots & m_1, n \\ \cdot & \cdot & \cdots & \cdot \\ \square & m_{s-1}, n & \cdots & m_1, n \end{pmatrix}\right]$$

$$+\frac{m_s}{m_s+n+1}\,\beta\left[x;\begin{pmatrix} m_s-1, n & m_{s-1}, n & \cdots & m_1, n \\ \cdot & \cdot & \cdots & \cdot \\ m_s-1, n & m_{s-1}, n & \cdots & m_1, n \end{pmatrix}\right],$$

where, in the second pseudo-determinant, $\square$ indicates that the corresponding terms in the formal expansion are not to be considered at all, $\square$ being introduced merely to write the pseudo-determinant in a complete form. Recalling the notation (A.9.6.6)—(A.9.6.8) and the lemma (A.9.5) it is easy to see that

$$\beta\left[x;\begin{pmatrix} \overleftarrow{m_s}, \overleftarrow{n+1} & m_{s-1}, n & \cdots & m_1, n \\ \cdot & \cdot & \cdots & \cdot \\ \overleftarrow{m_s}, \overleftarrow{n+1} & m_{s-1}, n & \cdots & m_1, n \end{pmatrix}\right] \tag{A.9.6.10}$$

$$= \beta_0(x;\, m_s, n+1)\,\beta\left[x;\begin{pmatrix} m_{s-1}, n & \cdots & m_1, n \\ \cdot & \cdots & \cdot \\ m_{s-1}, n & \cdots & m_1, n \end{pmatrix}\right]$$

$$+\sum_{r=1}^{s-1}(-1)^r\beta_r\left[x;\begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ \cdot & \ldots & \cdot \\ m_s+m_{s-1}, 2n+1 & \ldots & m_s+m_1, 2n+1 \\ \cdot & \ldots & \cdot \\ m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right],$$

where $\beta_r[x;\ \ldots]$ is an $(s-1)$-fold pseudo-determinant obtained by substituting $(m_s+m_{s-1}, 2n+1) \ldots, (m_s+m_1, 2n+1)$ for $(m_{s-1},, n)$, $(m_{s-2}, n), \ldots, (m_1, n)$ in the r^{th} row of

the $(s-1)$-fold pseudo-determinant $\beta\left[x;\begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ \cdot & \ldots & \cdot \\ m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right]$.

$$\text{Thus (A.9.6.10)} = \beta_0(x; m_s, n+1)\ \beta\left[x;\begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ \ldots & \ldots & \cdot \\ m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$+\sum_{r=1}^{s-1}(-1)^r\ \beta(x; m_s+m_{s-r}, 2n+1)\ \beta_{rr}\left[x;\begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ \cdot & \ldots & \cdot \\ m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right],$$

where β_{rr} is the $(s-2)$-fold integral obtained by suppressing the r^{th} row and r^{th} column of the $(s-1)$-fold pseudo-determinant β already referred to. We have likewise

$$\text{(A.9.6.11)}\qquad \beta\left[x;\begin{pmatrix} \vec{m}_s, \overrightarrow{n+1} & m_{s-1}, n & \ldots & m_1, n \\ \cdot & \cdot & \ldots & \cdot \\ \vec{m}_s, \overrightarrow{n+1} & m_{s-1}, n & \ldots & m_1, n \\ \square & m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$=\sum_{r=1}^{s-1}(-1)^{r-1}\beta_r\left[x;\begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ \cdot & \ldots & \cdot \\ m_s+m_{s-1}, 2n+1 & \ldots & m_s+m_1, 2n+1 \\ \cdot & \ldots & \cdot \\ m_{s-1}, n & \ldots & m_1, n \end{pmatrix},\right]$$

$$=\sum_{r=1}^{s-1}(-1)^{r-1}\ \beta(x; m_s+m_{s-r}, 2n+1)\ \beta_{rr}\left[x;\begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ \cdot & \ldots & \\ m_{-1};\, n & \ldots & m_1, n \end{pmatrix}\right].$$

Now substituting from (A.9.6.10) and (A.9.6.11) in the right side of (A.9.6.9) we have

$$\beta\left[x;\begin{pmatrix} m_s, n & m_{s-1}, n & \dots & m_1, n \\ . & . & \dots & . \\ m_s, n & m_{s-1}, n & \dots & m_1, n \end{pmatrix}\right] \tag{A.9.6.12}$$

$$= -\frac{1}{m_s+n+1}\beta_0(x;\, m_s,\, n+1)\ \beta\left[x;\begin{pmatrix} m_{s-1}, n & \dots & m_1, n \\ . & \dots & . \\ m_{s-1}, n & \dots & m_1, n \end{pmatrix}\right]$$

$$+\frac{2}{m_s+n+1}\sum_{r=1}^{s-1}(-1)^{r-1}\,\beta(x;\, m_s+m_{s-r},\, 2n+1)\ \beta_{rr}\left[x;\begin{pmatrix} m_{s-1}, n & \dots & m_1, n \\ . & \dots & . \\ m_{s-1}, n & \dots & m_1, n \end{pmatrix}\right]$$

$$+\frac{m_s}{m_s+n+1}\beta\left[x;\begin{pmatrix} m_s-1, n & m_{s-1}, n & \dots & m_1, n \\ . & . & \dots & . \\ m_s-1, n & m_{s-1}, n & \dots & m_1, n \end{pmatrix}\right].$$

It may be noticed that the left hand side is an s^{th} order pseudo-determinant while, on the right hand side, the first $\beta[\,x;\,\dots]$ is an $(s-1)^{st}$ order pseudo determinant, the second group of terms involves β_{rr}, each such β_{rr} being an $(s-2)^{nd}$ order pseudo-determinant, and the last term has a β which is an s-th order pseudo-determinant with the exponent m_s reduced to m_s-1. It may be also noticed that β_{rr} may be conveniently written as

$$\beta\left[x;\begin{pmatrix} m_{s-1}, n & \dots & m_{s-r+1}, n & m_{s-r-1}, n & \dots & m_1, n \\ . & \dots & . & . & \dots & . \\ m_{s-1}, n & \dots & m_{s-r+1}, n & m_{s-r-1}, n & \dots & m_1, n \end{pmatrix}\right].$$

(A.9.6.12) thus gives us a recurrence relation, whereby, proceeding along the chain and reducing m_s to m_{s-1} (in which case the pseudo-determinant will be zero) we have the following reduction of the integral by one dimension.

$$\text{(A.9.6.13)} \qquad \beta\left[x;\begin{pmatrix} m_s, n & \ldots & m_1, n \\ . & \ldots & . \\ m_s, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$= -\beta\left[x;\begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ . & \ldots & . \\ m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$\times \sum_{r'=1}^{m_s-m_{s-1}} \beta_0(x; m_s-r'+1, n+1)(m_s)_{r'-1}/(m_s+n+1)_{r'}$$

$$+2\sum_{r=1}^{s-1} (-1)^{r-1} \sum_{r'=1}^{m_s-m_{s-1}} \beta\left[x;\begin{pmatrix} m_{s-1}; n & \ldots & m_{s-r+1}, n & m_{s-r-1}, n & \ldots & m_1, n \\ . & \ldots & . & . & \ldots & . \\ m_{s-1}, n & \ldots & m_{s-r+1}, n & m_{s-r-1}, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$\times \beta(x; m_s+m_{s-r}-r'+1, 2n+1)(m_s)_{r'-1}/(m_s+n+1)_{r'},$$

where $(m)_p$ stands for $m(m-1)\ldots(m-p+1)$. The s-th order pseudo-determinant is thus thrown back on $(s-1)^{st}$ and $(s-2)^{nd}$ order pseudo-determinants, and these again on $(s-2)^{nd}$ and $(s-3)^{rd}$ order ones and so on till we get to first order pseudo-determinants which are easily evaluated from the incomplete β-function tables.

(A.9.7): *Evaluation of the integral*

$$\int_{x_s=x_0}^{x} \int_{x_{s-1}=x_0}^{x_s} \cdots \int_{x_2=x_0}^{x_3} \int_{x_1=x_0}^{x_2} M \prod_{i=1}^{s} dx_i = \beta[x, x_0; m_s, n; \ldots, m_1, n] \text{ (say)}$$

$$= \beta\left[x, x_0;\begin{pmatrix} m_s, n & m_{s-1}, n & \ldots & m_1, n \\ m_s, n & m_{s-1}, n & \ldots & m_1, n \\ . & . & \ldots & . \\ m_s, n & m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right] \text{ (say)}$$

where M stands for the determinant under the integration sign in (A.9.1). *We shall also use the notation*

$$(\text{A.9.7.1}) \qquad \beta(x, x_0;\ m_s, n_s;\ m_{s-1}, n_{s-1};\ \ldots;\ m_1, n_1)$$

$$= \int_{x_0}^{x} x_s^{m_s}(1-x_s)^{n_s}\,dx_s \int_{x_0}^{x_s} x_{s-1}^{m_{s-1}}(1-x_{s-1})^{n_{s-1}}\,dx_{s-1}$$

$$\ldots \int_{x_0}^{x_3} x_2^{m_2}(1-x_2)^{n_2}\,dx_2 \int_{x_0}^{x_2} x_1^{m_1}(1-x_1)^{n_1}dx_1.$$

Proceeding now exactly as in sections (A.9.1), (A.9.2), (A.9.3), (A.9.4) and (A.9.5) with obvious modifications at each stage we have in place of (A.9.6.12) the following result

$$(\text{A.9.7.2}) \qquad \beta\left[x, x_0;\ \begin{pmatrix} m_s, n & m_{s-1}, n & \ldots & m_1, n \\ m_s, n & m_{s-1}, n & \ldots & m_1, n \\ . & . & \ldots & . \\ m_s, n & m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$= -\beta\left[x, x_0;\ \begin{pmatrix} m_{s-1}, n & \ldots & m_1, n \\ . & \ldots & . \\ . & \ldots & . \\ m_{s-1}, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$\times \sum_{r'=1}^{m_s-m_{s-1}} [(m_s)_{r'-1}/(m_s+n+1)_{r'}][\beta_0(x;\ m_s-r'+1, n+1)-(-1)\,{}^{s}\beta_0(x_0;\ m_s-r'+1,\ n+1)]$$

$$+2\sum_{r=1}^{s-1}\ \sum_{r'=1}^{m_s-m_{s-1}} (-1)^{r-1}\,\beta\left[x, x_0;\ \begin{pmatrix} m_{s-1}, n & \ldots & m_{s-r+1}, n & m_{s-r-1}, n & \ldots & m_1, n \\ m_{s-1}, n & \ldots & m_{s-r+1}, n & m_{s-r-1}, n & \ldots & m_1, n \\ . & \ldots & . & . & \ldots & . \\ m_{s-1}, n & \ldots & m_{s-r+1}, n & m_{s-r-1}, n & \ldots & m_1, n \end{pmatrix}\right]$$

$$\times \frac{(m_s)_{r'-1}}{(m_s+n+1)_{r'}}\ \beta(x, x_0;\ m_s+m_{s-r}-r'+1,\ 2n+1),$$

where $(m)_p = m(m-1)\ldots(m-p+1)$ and $\beta_0(x;\ m, n)$ stands for $x^m(1-x)^n$. The s-th order pseudo-determinant is thus thrown back on the $(s-1)^{st}$ and $(s-2)^{nd}$ order pseudo-determinants, and so on till we get to first order pseudo-determinants which can be easily evaluated from the incomplete beta function tables.

REFERENCES

1. Barnard, G. A. (1947): Significance tests for 2×2 tables. *Biometrika*, XXXIV, 123-138.
2. Bartlett, M. S. (1935): Contingency table interaction. *Suppl. J.R.S.S.*, II, 248-252.
3. Bose, P. K. (1942-1944): On confluent hypergeometric series *Sankhyā*, **6**, 407-412.
4. Bose, P. K. (1947): Parametric relations in multivariate distributions *Sankhyā*, **8**, 167-171.
5. Bose, P. K. (1949): Incomplete probability integral tables connected with Studentized D^2-Statistics. *Calcutta Statistical Association Bulletin*, **2**, 131-137.
6. Bose, P. K. ((1951): Remarks on computing the incomplete probability integral in multivariate distribution functions. *Bull. Int. Stat. Institute*, XXXIII, II, 55-64.
7. Bose R. C. and Roy S. N. (1938): The distribution of the studentized D^2-statistic. *Sankhyā*, **4**, 10-38.
8. Cochran, W. G. (1952): The χ^2 test of goodness of fit. *Ann. Math. Stat.*, **23**, 314-345.
9. Cochran, W. G. (1954): Some methods of strengthening the commom χ^2 tests. *Biometrics*, X, 417-451.
10. Cramér, H. (1946): *Mathematical Methods of Statistics*, Princeton University Press.
11. David, F. N. (1938): *Tables of the distribution of the correlation coefficient*. Biometrika office, London.
12. Feller, W. (1950): *An introduction to probability theory and its applicalions*, **1**, John Wiley and Sons, Inc., New York.
13. Fisher, R. A. (1922): On the interpretation of chi-square from contingency tables and the calculation of p. *J.R.S.S*, LXXXV, 87-94.
14. Fisher, R. A. (1928): The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc. London*, Ser A, **121**, 654-673.
15. Fisher, R. A. (1939): The sampling distribution of some statistics obtained from nonlinear equations. *Ann. Eugenics*, **9**, 238-249.
16. Girshick, M. A. (1939): On the sampling distribution of the roots of determinantal equations. *Ann. Math. Stat.*, **10**, 203-224.
17. Hotelling, H. (1931): The generalization of Student's ratio. *Ann. Math. Stat.*, **2**, 360-378.
18. Hotelling, H. (1936): Relations between two sets of variates. *Biometrika*, **28**, 321-377.
19. Hsu, P. L. (1939): On the distribution of the roots of certain determinantal equations. *Ann. Eugenics* **9**, 250-258.
20. Kendall, M. G. (1946): *Advanced theory of statistics*, Charles Griffin and Company, London.
21. Lehmann E. L and Stein, C. (1948): Most powerful tests of composite hypotheses. *Ann. Math. stat.*, **19**, 495-516.
22. Lehmer, Emme (1944): Inverse tables of probabilities of errors of the second kind. *Ann. Math. Stat.*, **15**, 388-398.
23. Mahalanobis, P. C. (1930): On tests and measures of group divergence. *Journ. Roy. Asiatic Soc. of Bengal* (New series), **26**, 541-588.
24. Mahalanobis, P. C., Bose R. C. and Roy S. N. (1937): Normalisation of statistical variates, and the use of rectangular co-ordinates in the theory of sampling distributions. *Sankhyā*, **3**, 1-40.
25. Mitra, S. K. (1955): Contributions to the statistical analysis of categorical data. North Carolina Institute of Statistics Mimeograph Series No. 142.
26. Nanda, D. N. (1948): Distribution of a root of a determinantal equation. *Ann. Math. Stat.*, **19**, 485-516.
27. Nandi, H. K. (1947): On the average power of a test criteria. *Sankhyā*, **8**, 67-72.
28. Nandi, H. K. (1951): On analysis of variance test. *Cal. Stat. Assn. Bull.*, **3**, 103-114.
29. Nandi, H. K. (1951): On the efficiency of experimental designs. *Cal. Stat. Assn. Bull.*, **3**, 167-171.

30. Neyman, J. (1949): Contributions to the theory of χ^2 test. Proceedings of the Berkeley symposium on mathematical statistics and probability. University of California Press. 239-273.

31. Olkin, I. (1951): On distribution problems in multivariate analysis. North Carolina Insitute of Statistics Mimeograph Series, No. 43, 1-126.

32. Olkin, I. and Roy, S. N. (1954): On multivariate distribution theory. *Ann. Math. Stat.*, **25**, 329-339.

33. Pearson, E. S. (1947): The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, **34**, 139-167.

34. Pillai, K. C. S. (1954): On some distribution problems in multivariate analysis. North Carolina Institute of Statistics Mimeograph Series, No. 88, 1-136.

35. Ramachandran, K. V. (1954): On certain tests and the monotonicity of their powers. North Carolina Institute of Statistics Mimeograph Series, No. 120, 1-119.

36. Rao, C. R. (1952): *Advanced statistical methods in biometric research*, John Wiley and Sons, Inc., New York.

37. Roy, S. N. (1939): p-statistics or some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā*, **4**, 381-396.

38. Roy, S. N. (1942): The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis. *Sankhyā*, **6**, 15-34.

39. Roy, S N. (1945): The individual sampling distribution of the maximum, the minimum and any intermediate one of the p-statistics on the null hypothesis. *Sankhyā*, **7**, 133-158.

40. Roy, S. N. (1948): Notes on testing of composite hypotheses—II. *Sankhyā*, **9**, 19-38.

41. Roy, S. N. (1950): Univariate and multivariate analysis as problems in testing of composite hypotheses—I *Sankhyā*, **10**, 29-80.

42. Roy, S. N. (1952): Some useful results in Jacobians. *Cal. Stat. Assn. Bull.*, **4**, 117-122.

43. Roy, S. N. (1953): On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Stat.*, **24**, 220-238.

44. Roy, S. N. and Bose R. C. (1953): Simultaneous confidence interval estimation. *Ann. Math. Stat.*, **24**, 513-536.

45. Roy, S. N. (1954): A useful theorem in matrix theory. *Proc. Amer. Math. Soc.*, **5** 635-638.

46. Roy, S. N. (1954): Some further results in simultaneous confidence interval estimation. *Ann. Math. Stat.*, **25**, 752-761.

47. Roy, S. N. and Kastenbaum, M. (1955): A generalisation of analysis of variance and multivariate analysis to data based on frequencies in qualitative categories or class intervals. North Carolina Institute of Statistics Mimeograph Series No. 131.

48. Roy S. N. and Mitra, S. K. (1956): An introduction to some nonparametric generalisations of analysis of variance and multivariate analysis. *Biometrika*, **43**, 361-376.

49. Scheffe, H. (1953) A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87-104.

50. Tang, P. C. (1938): The power function of analysis of variance tests with tables and illustrations of their use. *Stat. Res. Mem.*, **2**, 126-149.

51. Tukey, J. W. (1952): Allowances for various types of error rates. (unpublished invited address, Blacksburg meeting of the Institute of Mathematical Statistics), March, 1952.

52. Watson, G. N. (1945): *The theory of Bessel Functions*, Cambridge University Press.

53. Wilks, S. S. (1943): *Mathematical Statistics*, Princeton University Press.

54. Wishart, John (1928): The generalised product moment distribution in samples drawn from a normal multivariate population. *Biometrika*, **20**, 38-40.

55. Yule, U. (1911): *An introduction to the theory of statistics*, Charles Griffin and Co. Ltd. London.