International Statistical Review (1988), **56**, 1, pp. 49–62. Printed in Great Britain

Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis

A. Shapiro

Department of Mathematics and Applied Mathematics, University of South Africa, P.O. Box 392, Pretoria 0001, South Africa

Summary

In this paper a distributional theory of test statistics in various problems of multivariate analysis involving inequality constraints is examined. A unified point of view based on geometrical properties of convex cones is presented. Chi-bar-squared and E-bar-squared test statistics are introduced. Their applications to hypothesis testing problems are discussed.

Key words: Chi-bar-squared statistic; Convex cone; E-bar-squared statistic; Inequality constraints; Isotonic inference; Likelihood ratio test; Linear regression; Order restrictions; Orthogonal projection; Structural model.

1 Introduction

Statistical inference for equality constrained problems in multivariate analysis is well established and widely known. An account of basic results and ideas are given, for example, by Silvey (1970) and Rao (1973, § 4a9). In contrast a relevant theory for *inequality* constraints is scattered in the literature under various names such as order restricted inference, isotonic regression, one-sided testing, etc. An exposition of some achievements in this direction and a historical background prior to the seventies are given in the seminal book of Barlow et al. (1972).

The aim of this paper is to present a unified approach to hypothesis testing involving inequality constraints. In this process we hope to clarify and simplify the understanding of known results occasionally introducing possible extensions and relevant applications. Throughout the paper we make use of a few simple geometrical facts associated with convex cones. A brief description of required results and definitions will be given in § 2. As a reference book for the corresponding theory we mention Stoer & Witzgall (1970).

In § 3 we introduce the basic concept of chi-bar-squared statistics. In order to illustrate general ideas an example of linear regression under inequality constraints will be considered in detail in § 4. Section 5 is concerned with the weights associated with chi-bar-squared distributions. Some asymptotic results are described in § 6. In § 7 we introduce E-bar-squared statistics and discuss their applications, while § 8 gives a summary and conclusions.

2 Convex cones and orthogonal projections

A subset C of $\mathbb{R}^m$ is called a cone (or a positively homogeneous set) if $\mathbf{x} \in C$ implies that $t\mathbf{x} \in C$ for every positive scalar t. We suppose throughout that cones considered are closed and *convex*. Convex cones share many useful properties of linear spaces. In particular it is meaningful to consider an orthogonal (minimum norm, metric) projection onto a convex cone C. Let $\mathbf{U}$ be a given $m \times m$ positive-definite symmetric matrix. We denote by $(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{U}\mathbf{y}$ and $\|\mathbf{x}\| = (\mathbf{x}'\mathbf{U}\mathbf{x})^{\frac{1}{2}}$ the inner product and the norm, respectively, associated with $\mathbf{U}$. The orthogonal projection $P(., C)$ onto C assigns to each $\mathbf{x}$ the closest point in C, that is $\bar{\boldsymbol{\eta}} = P(\mathbf{x}, C)$ is the solution of the program:

$$\text{minimize } (\mathbf{x} - \boldsymbol{\eta})'\mathbf{U}(\mathbf{x} - \boldsymbol{\eta}) \text{ subject to } \boldsymbol{\eta} \in C. \tag{2.1}$$

The corresponding distance from $\mathbf{x}$ to C is given by $\|\mathbf{x} - P(\mathbf{x}, C)\|$.

The following formula is, in a sense, a version of Pythagoras' theorem,

$$\|\mathbf{x}\|^2 = \|P(\mathbf{x}, C)\|^2 + \|\mathbf{x} - P(\mathbf{x}, C)\|^2. \tag{2.2}$$

Furthermore, suppose that C is contained in a convex cone K and either C or K is a *linear* space. Then

$$\|\mathbf{x} - P(\mathbf{x}, C)\|^2 = \|\mathbf{x} - P(\mathbf{x}, K)\|^2 + \|P(\mathbf{x}, K) - P(\mathbf{x}, C)\|^2. \tag{2.3}$$

Note that (2.3) does not hold in general for two convex cones $C \subset K$.

With the cone C is associated the so-called polar (or dual) cone C^0,

$$C^0 = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \leq 0 \text{ for all } \mathbf{x} \in C\}.$$

If $C = L$ is a linear subspace of $\mathbb{R}^m$, then $C^0 = L^\perp$ is the usual orthogonal complement of L,

$$L^\perp = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) = 0 \text{ for all } \mathbf{x} \in L\}.$$

Similarly to the case of linear spaces, it follows that, if C is closed and convex, then $(C^0)^0 = C$ and, for all $\mathbf{x}$,

$$\mathbf{x} - P(\mathbf{x}, C) = P(\mathbf{x}, C^0). \tag{2.4}$$

In order to obtain the orthogonal projection $P(\mathbf{x}, C)$ corresponding to a given vector $\mathbf{x}$, one must solve the mathematical programming problem (2.1). In applications the cone C is often defined by a number of equality and inequality linear constraints

$$C = \{\mathbf{x} : \mathbf{a}_i'\mathbf{x} = 0, i = 1, \ldots, s; \mathbf{a}_i'\mathbf{x} \leq 0, i = s+1, \ldots, t\}. \tag{2.5}$$

Then (2.1) becomes a quadratic programming problem. An interesting algorithm exploiting specific features of problem (2.1) has been proposed by Dykstra & Robertson (1982b) and Dykstra (1983). For a closely related problem of finding the nearest point in a polytope see Mitchell, Demyanov & Malozemov (1974) and Wolfe (1976).

Finally we note that, if C is given by (2.5), then the polar cone C^0 is generated by linear combinations of vectors $\mathbf{U}^{-1}\mathbf{a}_i$ $(i = s+1, \ldots, t)$, with nonnegative coefficients, and vectors $\mathbf{U}^{-1}\mathbf{a}_i$ $(i = 1, \ldots, s)$, with unrestricted coefficients.

3 Chi-bar-squared statistics

In this section we discuss the so-called chi-bar-squared statistics which will play a major role in our considerations. Let $\mathbf{y} \sim N(\mathbf{0}, \mathbf{V})$ be an $m \times 1$ normal variable, C be a convex cone and consider

$$\bar{\chi}^2 = \mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - \min_{\boldsymbol{\eta} \in C} (\mathbf{y} - \boldsymbol{\eta})'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\eta}). \tag{3.1}$$

It follows from (2.2) that the chi-bar-squared statistic defined in (3.1) can be written as

$\bar{\chi}^2 = \|P(\mathbf{y}, C)\|^2$, where the norm and inner product are taken with respect to the matrix $\mathbf{V}^{-1}$. The basic distributional result concerning $\bar{\chi}^2$ is that it is distributed as a mixture of chi-squared distributions. That is

$$\Pr\{\bar{\chi}^2 \geq c\} = \sum_{i=0}^{m} w_i \Pr\{\chi_i^2 \geq c\}, \tag{3.2}$$

where χ_i^2 is a chi-squared random variable with i degrees of freedom, $\chi_0^2 \equiv 0$, and w_i are nonnegative weights such that $w_0 + \ldots + w_m = 1$. The weights $w_i = w_i(m, \mathbf{V}, C)$ depend on $\mathbf{V}$ and C and will be discussed later. The distribution of $\bar{\chi}^2$ is determined by $\mathbf{V}$ and C and will be denoted by $\bar{\mathscr{X}}^2(\mathbf{V}, C)$. We also write $\bar{\chi}^2 \sim \bar{\mathscr{X}}^2(\mathbf{V}, C)$.

The distributional result (3.2) has a long history. Particular cases of it appeared in earlier works of Bartholomew (1959, 1961). A decisive step was made by Kudô (1963) and independently by Nüesch (1964, 1966) where this result was proved for the nonnegative orthant $C = \mathbb{R}_+^m$. Later their approach was extended by Kudô & Choi (1975), and recently Shapiro (1985a) proposed a simple proof for any convex cone C.

Often we will be interested in the following form of chi-bar-squared statistics

$$\bar{\chi}^2 = \min_{\boldsymbol{\eta} \in D} (\mathbf{y} - \boldsymbol{\eta})'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\eta}), \tag{3.3}$$

where D is a closed convex cone. Formulae (2.2) and (2.4) imply that, if $D = C^0$, then $\bar{\chi}^2$ in (3.1) and (3.3) are identical. Since $(C^0)^0 = C$ we have that $\bar{\chi}^2$ given in (3.3) has the chi-bar-squared distribution $\bar{\mathscr{X}}^2(\mathbf{V}, D^0)$. It follows that every result about $\bar{\chi}^2$ in the form (3.3) has its dual counterpart for $\bar{\chi}^2$ defined by (3.1). Another source of duality is provided by the following equality connecting the weights corresponding to polar cones (Shapiro, 1985a)

$$w_i(m, \mathbf{V}, C^0) = w_{m-i}(m, \mathbf{V}, C), \quad (i = 0, \ldots, m). \tag{3.4}$$

Consider the $\bar{\chi}^2$ statistic given in (3.3) and suppose that the cone D is contained in a linear space L generated by column vectors of an $m \times k$ matrix $\boldsymbol{\Delta}$ of rank k. Then D can be represented in the form $D = \boldsymbol{\Delta}K$, where K is a convex cone in $\mathbb{R}^k$ and

$$\boldsymbol{\Delta}K = \{\boldsymbol{\eta} : \boldsymbol{\eta} = \boldsymbol{\Delta}\boldsymbol{\beta},\ \boldsymbol{\beta} \in K\}.$$

Consequently we can write $\bar{\chi}^2$ as follows

$$\bar{\chi}^2 = \min_{\boldsymbol{\beta} \in K} (\mathbf{y} - \boldsymbol{\Delta}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\Delta}\boldsymbol{\beta}). \tag{3.5}$$

Then by applying (2.3) and using the matrix form of the orthogonal projection onto the linear space L one can show that the right-hand side of (3.5) is the sum of two independent variables $\|\mathbf{y} - P(\mathbf{y}, L)\|^2$ and

$$\min_{\boldsymbol{\beta} \in K} (\mathbf{z} - \boldsymbol{\beta})'\mathbf{W}^{-1}(\mathbf{z} - \boldsymbol{\beta}), \tag{3.6}$$

where $\mathbf{W} = (\boldsymbol{\Delta}'\mathbf{V}^{-1}\boldsymbol{\Delta})^{-1}$ and $\mathbf{z} = (\boldsymbol{\Delta}'\mathbf{V}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{V}^{-1}\mathbf{y}$, $\mathbf{z} \sim N(\mathbf{0}, \mathbf{W})$ (Shapiro, 1985a, Th. 2.1). It is well known that $\|\mathbf{y} - P(\mathbf{y}, L)\|^2$ has the (central) chi-squared distribution with $m - k$ degrees of freedom. Of course the second variable given by (3.6) is $\bar{\mathscr{X}}^2(\mathbf{W}, K^0)$.

The result above can be extended to the noncentral case as follows. Let $\mathbf{y}$ be $N(\boldsymbol{\mu}, \mathbf{V})$ with the mean vector $\boldsymbol{\mu}$ being *orthogonal* to L; that is $P(\boldsymbol{\mu}, L) = \mathbf{0}$ or equivalently $\boldsymbol{\Delta}'\mathbf{V}^{-1}\boldsymbol{\mu} = \mathbf{0}$. Then the normal variable $\mathbf{z}$ in (3.6) remains $N(\mathbf{0}, \mathbf{W})$, and hence the distribution of the second variable (3.6) remains unchanged, while the first variable becomes noncentral chi-squared with $m-k$ degrees of freedom and noncentrality parameter

$$\delta = \|\boldsymbol{\mu} - P(\boldsymbol{\mu}, L)\|^2 = \|\boldsymbol{\mu}\|^2.$$

Therefore in this case $\bar{\chi}^2$ is distributed as a mixture of noncentral chi-squared distributions with the same noncentrality parameter δ.

Finally let us consider the case where C contains a linear space M. Then C is representable as the direct sum of M and the cone $C^* = C \cap M^{\perp}$, where $M^{\perp}$ is the orthogonal complement of M with respect to the matrix $\mathbf{V}^{-1}$. This implies that

$$\|P(\mathbf{x}, C)\|^2 = \|P(\mathbf{x}, M)\|^2 + \|P(\mathbf{x}, C^*)\|^2$$

and then

$$\|\mathbf{y} - P(\mathbf{y}, M)\|^2 - \|\mathbf{y} - P(\mathbf{y}, C)\|^2 = \|P(\mathbf{y}, C^*)\|^2.$$

It follows that, if $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ and $\boldsymbol{\mu} \in M$, then the statistic

$$\min_{\boldsymbol{\eta} \in M} (\mathbf{y} - \boldsymbol{\eta})'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\eta}) - \min_{\boldsymbol{\eta} \in C} (\mathbf{y} - \boldsymbol{\eta})'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\eta}) \tag{3.7}$$

has the distribution $\bar{\mathscr{X}}^2(\mathbf{V}, C^*)$.

4. Applications to linear models

In this section we discuss in detail an illustrative example of the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{4.1}$$

where $\mathbf{X}$ is a known $N \times k$ matrix of rank k, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters and $\mathbf{e}$ is $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Throughout the paper we denote by $\boldsymbol{\beta}_0$ the true value of the parameter vector $\boldsymbol{\beta}$ corresponding to a given model. In this section we consider the case where the covariance matrix $\boldsymbol{\Sigma}$ is *known*. The mean vector $\boldsymbol{\mu}$ will be specified later.

Suppose that we wish to test the null hypothesis that $\boldsymbol{\beta}_0$ satisfies a set of equality constraints

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r} \tag{4.2}$$

against the restricted alternative

$$\mathbf{R}_1\boldsymbol{\beta} \geqslant \mathbf{r}_1, \tag{4.3}$$

where $\mathbf{R}_1$ is a $q \times k$ submatrix of the $p \times k$ matrix $\mathbf{R}$ and $\mathbf{r}_1$ is the corresponding subvector of $\mathbf{r}$. (The inequality sign between two vectors is understood to be applied componentwise.) It will be assumed that equations (4.2) are *consistent*, i.e. they define a *nonempty* (affine) subspace of $\mathbb{R}^k$. Of course this assumption holds if $\mathbf{R}$ has full row rank p. The testing problem above has been discussed, at various degrees of generality, by several authors. Gouriéroux, Holly & Monfort (1982) considered the case where $\mathbf{R}$ is equal to $\mathbf{R}_1$ and has the full rank p. A discussion of a more general situation and additional references is given by Farebrother (1986). Subsequently we rederive and extend some of their results. In particular it will be shown that the null distribution of an appropriate test statistic is chi-bar-squared, which gives a confirmative answer to the question raised by Farebrother (1986, p. 29).

The likelihood ratio test rejects the null hypothesis for large values of the following statistic

$$\min_{\mathbf{R}\boldsymbol{\beta}=\mathbf{r}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \min_{\mathbf{R}_1\boldsymbol{\beta}\geqslant\mathbf{r}_1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{4.4}$$

Suppose that the mean vector $\boldsymbol{\mu}$ is *orthogonal* to the linear space L generated by the column vectors of $\mathbf{X}$; that is $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{X}=\mathbf{0}$. This is a natural generalization of the usual assumption $\boldsymbol{\mu}=\mathbf{0}$. Then by applying formula (2.3) in a similar manner to (3.5) and (3.6), we obtain the test statistic (4.4) in the form

$$\min_{\mathbf{R}\boldsymbol{\beta}=\mathbf{r}} (\mathbf{z}-\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\beta}) - \min_{\mathbf{R}_1\boldsymbol{\beta}\geqslant\mathbf{r}_1} (\mathbf{z}-\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\beta}), \tag{4.5}$$

where $\boldsymbol{\Omega}=(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$ and $\mathbf{z}=(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}\sim N(\boldsymbol{\beta}_0, \boldsymbol{\Omega})$. Consider the linear space

$$M=\{\boldsymbol{\beta}:\mathbf{R}\boldsymbol{\beta}=\mathbf{0}\}.$$

Clearly M is contained in the cone

$$C=\{\boldsymbol{\beta}:\mathbf{R}_1\boldsymbol{\beta}\geqslant\mathbf{0}\}.$$

Moreover, subtracting from $\mathbf{z}$ and $\boldsymbol{\beta}$ the true value $\boldsymbol{\beta}_0$ we come, under the null hypothesis, to the following form of the test statistic, compare Gouriéroux et al. (1982, p. 67),

$$\min_{\boldsymbol{\beta}\in M} (\mathbf{z}-\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\beta}) - \min_{\boldsymbol{\beta}\in C} (\mathbf{z}-\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\beta}), \tag{4.6}$$

with $\mathbf{z}\sim N(\mathbf{0}, \boldsymbol{\Omega})$. By (3.7) this statistic is $\bar{\mathscr{X}}^2(\boldsymbol{\Omega}, C^*)$, where $C^*=C\cap M^{\perp}$. Consequently its null distribution is a mixture of chi-squared distributions. An approach to the evaluation of the corresponding weights $w_i(k, \boldsymbol{\Omega}, C^*)$ will be discussed in § 5. Note that under the null hypothesis the distribution of the test statistic is *independent* of a particular population value $\boldsymbol{\beta}_0$ of the parameter vector.

Let us consider another problem associated with the linear model (4.1). Now we want to test equality and inequality linear constraints

$$\mathbf{R}_1\boldsymbol{\beta}\geqslant\mathbf{r}_1, \quad \mathbf{R}_2\boldsymbol{\beta}=\mathbf{r}_2, \tag{4.7}$$

against the unrestricted alternative. Here the matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ are of order $s\times k$ and $t\times k$, respectively, and it is assumed that restrictions (4.7) are consistent; i.e. define a nonempty subset of $\mathbb{R}^k$. The likelihood ratio test rejects the null hypothesis for large values of the test statistic ζ which is given by the difference between the minimum of $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$ subject to constraints (4.7) and the unrestricted minimum $\|\mathbf{y}-P(\mathbf{y}, L)\|^2$. It can be shown that the *least favorable distribution* (Lehmann, 1959, Ch. 3, § 8) of ζ occurs when all inequality constraints are active at $\boldsymbol{\beta}_0$, that is $\mathbf{R}_1\boldsymbol{\beta}_0=\mathbf{r}_1$ (Robertson & Wegman, 1978, Th. 2.1, 2.2). Therefore we suppose that $\mathbf{R}_1\boldsymbol{\beta}_0=\mathbf{r}_1$ and $\mathbf{R}_2\boldsymbol{\beta}_0=\mathbf{r}_2$.

It will be assumed again that $\boldsymbol{\mu}$ is orthogonal to the linear space L. Then by derivations similar to those of (4.5) and (4.6) one can show that ζ is distributed as

$$\min_{\boldsymbol{\beta}\in K} (\mathbf{z}-\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\beta}), \tag{4.8}$$

where $\mathbf{z}$ and $\boldsymbol{\Omega}$ are defined as in (4.6) and K is the convex cone

$$K=\{\boldsymbol{\beta}:\mathbf{R}_1\boldsymbol{\beta}\geqslant\mathbf{0}, \mathbf{R}_2\boldsymbol{\beta}=\mathbf{0}\}. \tag{4.9}$$

Consequently the least favorable null distribution of ζ is $\bar{\mathscr{X}}^2(\boldsymbol{\Omega}, K^0)$. Moreover, as we have mentioned earlier the unrestricted statistic $\|\mathbf{y}-P(\mathbf{y}, L)\|^2$ and ζ are independent.

5 The weights of chi-bar-squared distributions

Practical applications of the chi-bar-squared statistics require the numerical calculation of the weights $w_i(m, \mathbf{V}, C)$. In some particular situations these weights are expressible in closed form, while in others their evaluation may represent a quite difficult problem. It is easy to show that, if $\mathbf{V}$ is the identity matrix $\mathbf{I}$ and C is the nonnegative orthant $\mathbb{R}_+^m = \{\mathbf{x} : \mathbf{x} \geq \mathbf{0}\}$, then the corresponding weights are given by

$$w_i(m, \mathbf{I}, \mathbb{R}_+^m) = \binom{m}{i} 2^{-m} \quad (i = 0, \ldots, m)$$

(Gouriéroux et al., 1982, p. 79). Another important much discussed case can be described as follows. Let $\mathbf{y} = (y_1, \ldots, y_k)'$ be a normal $N(\mathbf{0}, \mathbf{V})$ variable with the diagonal covariance matrix $\mathbf{V}$, $\mathbf{V}^{-1} = \text{diag}(v_i)$ where v_i $(i = 1, \ldots, k)$ are some positive numbers. Consider the linear space $M = \{\mathbf{x} : x_1 = x_2 = \ldots = x_k\}$, the convex cone

$$C = \{\mathbf{x} : x_1 \leq x_2 \leq \ldots \leq x_k\}$$

and let $\bar{\chi}_k^2$ be the corresponding chi-bar-squared statistic given by formula (3.7). As we know, $\bar{\chi}_k^2 \sim \bar{\mathscr{X}}^2(\mathbf{V}, C^*)$. The corresponding weights $w_i(k, \mathbf{V}, C^*)$ $(i = 0, \ldots, k-1)$ are usually denoted by $P(i+1, k; \mathbf{v})$ and studied in detail by Barlow et al. (1972, § 3.3). Since the cone C^* is contained in the $(k-1)$-dimensional space $M^\perp$, the last weight $w_k(k, \mathbf{V}, C^*)$ is zero. An additional discussion of $P(i, k; \mathbf{v})$ and relevant references are given by Siskind (1976), Robertson & Wright (1982, 1983) and Kudô & Yao (1982).

In the case of $C = \mathbb{R}_+^m$, Kudô (1963, p. 414) proposed a formula for the weights $w_i(m, \mathbf{V}, \mathbb{R}_+^m)$, denoted subsequently by $w_i(m, \mathbf{V})$. The formula can be written as follows

$$w_i(m, \mathbf{V}) = \sum_{|\alpha| = i} p\{(\mathbf{V}_{\alpha'})^{-1}\} p\{\mathbf{V}_{\alpha;\alpha'}\}, \tag{5.1}$$

where the summation runs over all subsets α of $\{1, \ldots, m\}$ having i elements, α' is the complement of α, $\mathbf{V}_\alpha$ is the covariance matrix corresponding to the normal variables y_i $(i \in \alpha)$, $\mathbf{V}_{\alpha;\alpha'}$ is the same under the condition $y_j = 0$ $(j \in \alpha')$, and $p(\mathbf{\Lambda})$ denotes the probability that $\mathbf{z} \geq \mathbf{0}$ for a normal variable $\mathbf{z} \sim N(\mathbf{0}, \mathbf{\Lambda})$.

Of course, formula (5.1) involves evaluation of the probabilities $p(.)$. For a small number of variables these probabilities can be calculated explicitly which leads to a closed form expression of $w_i(m, \mathbf{V})$ for $m \leq 4$ (Kudô, 1963, p. 415); see also Shapiro (1985a, p. 141). A computer program based on (5.1) for calculation of $w_i(m, \mathbf{V})$ with moderate values of m has been proposed by Bohrer & Chow (1978). In the remainder of this section we assume that the weights $w_i(m, \mathbf{V})$ are computationally available, and concentrate on showing how the situation of a general cone C can be reduced to the standard case $C = \mathbb{R}_+^m$.

Suppose that the cone C is defined by a number of linear inequality constraints

$$C = \{\boldsymbol{\eta} : \mathbf{R}\boldsymbol{\eta} \geq \mathbf{0}\}, \tag{5.2}$$

where $\mathbf{R}$ is an $m \times m$ nonsingular matrix. Then by linear transformations (Perlman, 1969, p. 561) we obtain

$$w_i(m, \mathbf{V}, C) = w_i(m, \mathbf{RVR}'). \tag{5.3}$$

Consider the polar cone D^0 of the nonnegative orthant $D = \mathbb{R}_+^m$. We have that $D^0 = \{\boldsymbol{\eta} : \mathbf{V}^{-1}\boldsymbol{\eta} \leq \mathbf{0}\}$, and hence, by (5.2), $w_i(m, \mathbf{V}, D^0) = w_i(m, \mathbf{V}^{-1})$. Together with (3.4) this implies that

$$w_{m-i}(m, \mathbf{V}) = w_i(m, \mathbf{V}^{-1}), \tag{5.4}$$

for all $i = 0, \ldots, m$.

Suppose that C is defined by (5.2) with $\mathbf{R}$ being a $k \times m$ matrix of rank k. Then the polar cone C^0 is given by $C^0 = \mathbf{\Delta}K$, where $\mathbf{\Delta} = -\mathbf{VR}'$ and $K = \mathbb{R}_+^m$. The corresponding $\bar{\chi}^2$ variable defined in (3.5) is the sum of independent χ_{m-k}^2 and a chi-bar-squared variable

$\bar{\mathscr{X}}^2((\mathbf{RVR}')^{-1}, K^0)$. Then it follows from (3.4) and (5.4) that

$$w_{m-k+j}(m, \mathbf{V}, C) = w_j(k, \mathbf{RVR}') \quad (j = 0, \ldots, k) \tag{5.5}$$

while the remaining weights vanish.

Now suppose that C is given by equality and inequality linear constraints

$$C = \{\boldsymbol{\eta} : \mathbf{R}_1\boldsymbol{\eta} \geqslant \mathbf{0}, \mathbf{R}_2\boldsymbol{\eta} = \mathbf{0}\}, \tag{5.6}$$

where $\mathbf{R}_1$ is $s \times m$ and $\mathbf{R}_2$ is $t \times m$. We assume that the composed $k \times m$ matrix $\mathbf{R}$, $\mathbf{R}' = [\mathbf{R}_1', \mathbf{R}_2']$, is of full rank $k = s + t$. Then it can be shown as above that for $i < m - k$ the corresponding weights vanish, while for $i \geqslant m - k$ the weights are associated with the matrix $\mathbf{RVR}'$ and the cone

$$\{\boldsymbol{\beta} \in \mathbb{R}^k : \beta_i \geqslant 0, i = 1, \ldots, s; \beta_i = 0, i = s + 1, \ldots, k\}. \tag{5.7}$$

Let $\mathbf{Z}$ be the $s \times s$ upper-left submatrix of $(\mathbf{RVR}')^{-1}$. Then by calculating the polar of cone (5.7), and using (3.4) and (5.5) one can verify that

$$w_{m-k+j}(m, \mathbf{V}, C) = w_{s-j}(s, \mathbf{Z}) \quad (j = 0, \ldots, s), \tag{5.8}$$

and the remaining weights are zeros. Also by (5.4) we can write (5.8) in the form

$$w_{m-k+j}(m, \mathbf{V}, C) = w_j(s, \mathbf{Z}^{-1}). \tag{5.9}$$

Finally consider the cone $C^* = C \cap M^\perp$, where $M = \{\boldsymbol{\beta} : \mathbf{R}\boldsymbol{\beta} = \mathbf{0}\}$, $C = \{\boldsymbol{\beta} : \mathbf{R}_1\boldsymbol{\beta} \geqslant \mathbf{0}\}$ and $\mathbf{R}_1$ is a $q \times k$ submatrix of the $p \times k$ matrix $\mathbf{R}$. The corresponding weights $w_i(k, \boldsymbol{\Omega}, C^*)$ appeared in connection with the investigation of statistic (4.5). Assume that $\mathbf{R}$ is of full row rank p and consider a $(k - p) \times k$ matrix $\mathbf{A}$ of rank $k - p$ such that $\mathbf{R\Omega A}' = \mathbf{0}$. Then the cone C^* is given by

$$C^* = \{\boldsymbol{\beta} : \mathbf{R}_1\boldsymbol{\beta} \geqslant \mathbf{0}, \mathbf{A}\boldsymbol{\beta} = \mathbf{0}\}.$$

The corresponding matrix $\mathbf{B\Omega B}'$, where $\mathbf{B}' = [\mathbf{R}_1', \mathbf{A}']$, is block diagonal where the $q \times q$ upper-left block is given by $\mathbf{R}_1\boldsymbol{\Omega}\mathbf{R}_1'$. Consequently we obtain by (5.9) that

$$w_{p-q+j}(k, \boldsymbol{\Omega}, C^*) = w_j(q, \mathbf{R}_1\boldsymbol{\Omega}\mathbf{R}_1') \quad (j = 0, \ldots, q) \tag{5.10}$$

and the remaining weights vanish. It is interesting to note that here the weights depend on the dimensionality p of the matrix $\mathbf{R}$ but not on $\mathbf{R}$ itself. Therefore it is sufficient to make the full row rank assumption apply to the matrix $\mathbf{R}_1$ rather than $\mathbf{R}$, and subsequently to replace the number p in (5.10) by the rank of $\mathbf{R}$.

6 Asymptotic results

The distributional results associated with linear model (4.1) we have discussed in § 4 are exact. Often similar results hold asymptotically. A relevant theory for the maximum likelihood method is given by Chernoff (1954); see also Feder (1968) for some extensions to the noncentral case. Basic asymptotics can be described as follows. Suppose we are interested in testing a null hypothesis that a $k \times 1$ parameter vector $\boldsymbol{\theta}$ belongs to a subset ω of $\mathbb{R}^k$ against an alternative $\boldsymbol{\theta} \in \tau$. Furthermore, let the true value $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$ be a *boundary* point of ω and (or) τ. More specifically we assume that the sets ω and τ are approximated at $\boldsymbol{\theta}_0$ by *cones* C_ω and C_τ, respectively; for a detailed discussion and characterizations of cone approximations see Shapiro (1987). Then under some regularity conditions the likelihood ratio statistic $-2\log\lambda$ is asymptotically distributed as

$$v = \min_{\boldsymbol{\theta}\in C_\omega} (\mathbf{z}-\boldsymbol{\theta})'\mathbf{J}(\mathbf{z}-\boldsymbol{\theta}) - \min_{\boldsymbol{\theta}\in C_\tau} (\mathbf{z}-\boldsymbol{\theta})'\mathbf{J}(\mathbf{z}-\boldsymbol{\theta}), \tag{6.1}$$

where $\mathbf{J}$ is the information matrix and $\mathbf{z} \sim N(\mathbf{0}, \mathbf{J}^{-1})$ (Chernoff, 1954).

For arbitrary, even convex, cones C_ω and C_τ the variable v defined in (6.1) is not $\bar{\chi}^2$. It becomes chi-bar-squared if C_ω is contained in C_τ, C_ω and C_τ are convex and at least one of them is a linear space. Two important cases are worth mentioning. If $\omega = \{\boldsymbol{\theta}_0\}$, that is the null hypothesis is simple, then $C_\omega = \{\mathbf{0}\}$ and hence the first term in (6.1) is $\mathbf{z}'\mathbf{J}\mathbf{z}$. Consequently $-2\log\lambda$ is asymptotically $\bar{\mathscr{X}}^2(\mathbf{J}^{-1}, C_\tau)$. When $\boldsymbol{\theta}_0$ is an interior point of τ, the approximating cone C_τ coincides with the space $\mathbb{R}^k$ and hence the second term in (6.1) is identically zero. In this case $-2\log\lambda$ is asymptotically $\bar{\mathscr{X}}^2(\mathbf{J}^{-1}, C_\omega^0)$.

The situation of boundary solutions often happens in the analysis of structural models. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)'$ be a vector variable giving a parameter vector of some statistical population. For example, in multinomial models ξ_i are the corresponding cell probabilities and in the analysis of covariance structures $\boldsymbol{\xi}$ represents the $p^2 \times 1$ vector $\operatorname{vec}(\boldsymbol{\Sigma})$ formed by stacking columns of a $p \times p$ covariance matrix $\boldsymbol{\Sigma}$. A structural model for $\boldsymbol{\xi}$ is an $m \times 1$ vector-valued function $\mathbf{g}(\boldsymbol{\theta})$ which relates the $k \times 1$ parameter vector $\boldsymbol{\theta}$, from a specified parameter set Θ, to $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} = \mathbf{g}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta. \tag{6.2}$$

Let $\hat{\mathbf{x}}$ be a sample estimate, based on a sample of size n, of the true (population) value $\boldsymbol{\xi}_0$ of the parameter vector $\boldsymbol{\xi}$. For example, in multinomial models $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_m)'$ is given by observed frequencies $\hat{x}_i = n_i/n$ $(i = 1, \ldots, m)$, and in the analysis of covariance structures $\hat{\mathbf{x}} = \operatorname{vec}(\mathbf{S})$, where $\mathbf{S}$ is the sample covariance matrix.

Given $\hat{\mathbf{x}}$ one fits the model (6.2) by minimizing the discrepancy between $\hat{\mathbf{x}}$ and $\boldsymbol{\xi} = \mathbf{g}(\boldsymbol{\theta})$, which is measured by means of a certain real-valued (discrepancy) function $F(\mathbf{x}, \boldsymbol{\xi})$ of two $m \times 1$ vector variables. For instance, the maximum likelihood approach leads to the following discrepancy functions:

$$F = 2\sum_{i=1}^{m} (n_i/n)(\log(n_i/n) - \log\xi_i) \tag{6.3}$$

in the multinomial case, for example, Rao (1973, § 5e), and

$$F = \log|\boldsymbol{\Sigma}| - \log|\mathbf{S}| + \operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - p \tag{6.4}$$

in the analysis of covariance structures (Jöreskog, 1967, 1981). Many other examples of discrepancy functions relevant to multinomial and covariance structural models are given by Rao (1973, p. 352) and Swain (1975) and Browne (1984), respectively. The associated test statistic for testing the null hypothesis that the population value of $\boldsymbol{\xi}$ satisfies the model is given by $n\hat{F}$, where $\hat{F}$ is the minimal value of the discrepancy function

$$\hat{F} = \min\{F(\hat{\mathbf{x}}, \mathbf{g}(\boldsymbol{\theta})) : \boldsymbol{\theta} \in \Theta\}.$$

We assume that $n^{\frac{1}{2}}(\hat{\mathbf{x}} - \boldsymbol{\xi}_0)$ is asymptotically normal with a null mean vector and a covariance matrix $\boldsymbol{\Gamma}$. Usually this property is ensured by an application of the Central Limit Theorem. Please note that in both cases considered above the matrix $\boldsymbol{\Gamma}$ is singular. Therefore we replace the nonsingularity assumption by a *maximal rank* condition. That is we assume that $\boldsymbol{\Gamma}$ has the maximal rank possible, for example $\operatorname{rank}(\boldsymbol{\Gamma}) = m - 1$ in the multinomial case and the rank of $\boldsymbol{\Gamma}$ is given by the number $p(p+1)/2$ of nonduplicated elements of $\boldsymbol{\Sigma}$ in covariance structural models. It can be shown that if a few simple conditions are satisfied (Bishop, Fienberg & Holland, 1975, p. 504; Browne, 1982, p. 81), then the discrepancy function is approximated near the point $(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$ by a quadratic function $(\mathbf{x}-\boldsymbol{\xi})'\mathbf{V}_0(\mathbf{x}-\boldsymbol{\xi})$, where $\mathbf{V}_0$ is a symmetric nonnegative-definite matrix given by $\frac{1}{2}(\partial^2/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}')F(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0)$ (Shapiro, 1985b). We suppose that the weight matrix $\mathbf{V}_0$ is a *generalized inverse* $\boldsymbol{\Gamma}^-$ of the matrix $\boldsymbol{\Gamma}$. Discrepancy functions satisfying this condition are said to be *correctly specified* (Browne, 1984). It is well known that the discrepancy function (6.3) is correctly specified and the discrepancy function (6.4) is correctly specified if the data is drawn from a normally distributed population.

It can be shown that under the null hypothesis $\boldsymbol{\xi}_0 = \mathbf{g}(\boldsymbol{\theta}_0)$, $\boldsymbol{\theta}_0 \in \Theta$, and some regularity conditions the test statistic $n\hat{F}$ is asymptotically distributed as

$$\kappa = \min_{\boldsymbol{\beta}\in K} (\mathbf{y}-\boldsymbol{\Delta\beta})'\mathbf{V}_0(\mathbf{y}-\boldsymbol{\Delta\beta}), \tag{6.5}$$

where $\mathbf{y}$ is a normal variable $N(\mathbf{0}, \boldsymbol{\Gamma})$, K is a cone giving an approximation to the parameter set Θ at the point $\boldsymbol{\theta}_0$ and $\boldsymbol{\Delta} = (\partial/\partial\boldsymbol{\theta}')\mathbf{g}(\boldsymbol{\theta}_0)$ is the $m \times k$ Jacobian matrix of $\mathbf{g}(\boldsymbol{\theta})$ (Shapiro, 1985a). We assume that the column space of $\boldsymbol{\Delta}$ is *included* in the column space of the matrix $\boldsymbol{\Gamma}$. This is ensured by the common structure of the model $\mathbf{g}(\boldsymbol{\theta})$ and the parameter vector $\boldsymbol{\xi}$ and the maximal rank condition.

If $\boldsymbol{\theta}_0$ is an interior point of Θ, then $K = \mathbb{R}^k$ and κ is a chi-squared variable with $t = \text{rank}(\boldsymbol{\Gamma}) - k$ degrees of freedom. On the other hand if $\boldsymbol{\theta}_0$ is a boundary point of Θ and the cone $K \subset \mathbb{R}^k$ is convex, then κ is the sum of independent χ^2_t and a chi-bar-squared variable $\bar{\mathscr{X}}^2(\boldsymbol{\Pi}, K^0)$ where $\boldsymbol{\Pi} = (\boldsymbol{\Delta}'\boldsymbol{\Gamma}^-\boldsymbol{\Delta})^{-1}$. Note that, because of the column space assumption, the matrix $\boldsymbol{\Pi}$ is independent of a particular choice of the generalized inverse of $\boldsymbol{\Gamma}$; see, for example Rao & Mitra (1971, Lemma 2.2.4). This result can be extended to a noncentral case as follows. Suppose that $\{\boldsymbol{\xi}_{0,n}\}$ represents a *sequence* of population values of $\boldsymbol{\xi}$ converging to a point $\boldsymbol{\xi}_0 = \mathbf{g}(\boldsymbol{\theta}_0)$ satisfying the model, i.e. the population drift (Stroud, 1972; Kendall & Stuart, 1979, p. 247). Furthermore, suppose that $n^{\frac{1}{2}}(\boldsymbol{\xi}_{0,n} - \boldsymbol{\xi}_0)$ tends to a vector $\boldsymbol{\mu}$ such that $\boldsymbol{\mu}$ belongs to the column space of $\boldsymbol{\Gamma}$ and $\boldsymbol{\mu}'\mathbf{V}_0\boldsymbol{\Delta} = \mathbf{0}$. Then $n\hat{F}$ is asymptotically distributed as the sum of a chi-bar-squared variable $\bar{\mathscr{X}}^2(\boldsymbol{\Pi}, K^0)$ and a noncentral chi-squared variable $\chi^2_t(\delta)$, where the noncentrality parameter δ is given by $\boldsymbol{\mu}'\mathbf{V}_0\boldsymbol{\mu}$.

It can be seen that asymptotically the situation here is quite analogous to the one of § 4. Therefore, similarly to (4.2) and (4.3), it is possible to test (linear) equality constraints against the associated inequality constrained alternative. Or, as in (4.7), to test inequality and equality constraints against the unrestricted alternative. Of course, the corresponding distributional results will hold asymptotically (Chacko, 1966; Robertson, 1978; Dykstra & Robertson, 1982a).

7 E-bar-squared statistics

In this section we study the case where the covariance matrix $\mathbf{V}$ is equal to $\sigma^2\mathbf{U}$ where the matrix $\mathbf{U}$ is completely *known* but the scalar σ^2 is *unknown*. Let C be a convex cone, $\mathbf{y} \sim N(\mathbf{0}, \sigma^2\mathbf{U})$ and $\bar{\boldsymbol{\eta}}$ be the corresponding minimizer $\bar{\boldsymbol{\eta}} = P(\mathbf{y}, C)$ in the right-hand side of (3.1). Of course the minimizer $\bar{\boldsymbol{\eta}}$ will not be altered if we replace the unknown matrix $\mathbf{V}^{-1}$ by $\mathbf{U}^{-1}$. Then the E-bar-squared statistic is defined as follows

$$\bar{E}^2 = \frac{\bar{\boldsymbol{\eta}}'\mathbf{U}^{-1}\bar{\boldsymbol{\eta}}}{\mathbf{y}'\mathbf{U}^{-1}\mathbf{y}}. \tag{7.1}$$

Some particular cases of the $\bar{E}^2$ statistic have appeared in works of Bartholomew (1959, 1961), Shorack (1967) and have been discussed extensively by Barlow et al. (1972).

Because $\mathbf{V} = \sigma^2\mathbf{U}$ and by (2.2), the $\bar{E}^2$ statistic can be written in the form

$$\bar{E}^2 = \frac{\bar{\boldsymbol{\eta}}'\mathbf{V}^{-1}\bar{\boldsymbol{\eta}}}{\bar{\boldsymbol{\eta}}'\mathbf{V}^{-1}\bar{\boldsymbol{\eta}} + (\mathbf{y} - \bar{\boldsymbol{\eta}})'\mathbf{V}^{-1}(\mathbf{y} - \bar{\boldsymbol{\eta}})}. \tag{7.2}$$

Then it can be shown that $\bar{E}^2$ is distributed as a mixture of distributions associated with variables $\chi_i^2/(\chi_i^2 + \chi_{m-i}^2)$. Since the variable $\chi_i^2/(\chi_i^2 + \chi_{m-i}^2)$ has a (central) beta distribution with (shape) parameters $i/2$ and $(m-i)/2$ we obtain that

$$\Pr\{\bar{E}^2 \geqslant c\} = \sum_{i=0}^{m} w_i \Pr\{\beta_{i/2,(m-i)/2} \geqslant c\}, \tag{7.3}$$

where $\beta_{s,t}$ denotes a beta variable with parameters s and t. It should be noted that the weights $w_i = w_i(m, \mathbf{U}, C)$ in (7.3) are the same as in the case of chi-bar-squared statistics and hence the results of § 5 apply. In all its generality the distributional result (7.3) can be proved along the same lines as Shapiro (1985a, Th. 3.1); see also Barlow et al. (1972, Th. 3.2).

Now let us consider the situation where C is contained in a linear space L of dimension k with $k < m$. Let $\mathbf{y}$ be a normal $N(\boldsymbol{\mu}, \sigma^2\mathbf{U})$ variable and suppose that the mean vector $\boldsymbol{\mu}$ is orthogonal to L with respect to the weight matrix $\mathbf{U}^{-1}$. Then similarly to the case of chi-bar-squared statistics, $\bar{E}^2$ is distributed as a mixture of *noncentral* beta distributions. That is

$$\Pr\{\bar{E}^2 \geqslant c\} = \sum_{i=0}^{k} w_i \Pr\{\beta_{i/2,(m-i)/2}(0, \delta) \geqslant c\}, \tag{7.4}$$

where the second noncentrality parameter δ is given by $\sigma^{-2}\boldsymbol{\mu}'\mathbf{U}^{-1}\boldsymbol{\mu}$; see, for example, Johnson & Kotz (1970) for a description of noncentral beta distributions.

In the situation above the distance $\|\mathbf{y} - P(\mathbf{y}, C)\|$ from $\mathbf{y}$ to C is greater than or equal to the distance $\|\mathbf{y} - P(\mathbf{y}, L)\|$, which is greater than zero with probability one. Therefore in this case it is possible to define the statistic

$$\bar{F} = \frac{\bar{\boldsymbol{\eta}}'\mathbf{U}^{-1}\bar{\boldsymbol{\eta}}}{(\mathbf{y} - \bar{\boldsymbol{\eta}})'\mathbf{U}^{-1}(\mathbf{y} - \bar{\boldsymbol{\eta}})}. \tag{7.5}$$

It follows from (7.2) that $\bar{E}^2 = \bar{F}/(1 + \bar{F})$ and hence the tests based on statistics $\bar{E}^2$ and $\bar{F}$ are equivalent. In the definition of the $\bar{F}$ statistic we do not make an adjustment for the degrees of freedom which are random variables in the present situation (Barlow et al., 1972, p. 122). However, the $\bar{F}$ statistic is less useful than $\bar{E}^2$ since if the cone C has a nonempty interior, then with a positive probability the denominator in ratio (7.5) is zero in which case $\bar{F}$ is not defined. Similarly to (7.4) it can be shown that $\bar{F}$ is distributed as a mixture of distributions corresponding to the ratio of chi-squared variables; that is

$$\Pr\{\bar{F} \geqslant c\} = \sum_{i=0}^{k} w_i \Pr\{\chi_i^2/\chi_{m-i}^2(\delta) \geqslant c\}.$$

Let us discuss some examples. Consider a sequence $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of $k \times 1$ independent $N(\boldsymbol{\tau}, \boldsymbol{\Sigma})$ variables representing a data from a sample of size n. Suppose that it is required to test the null hypothesis $\boldsymbol{\tau} = \mathbf{0}$ against the alternative $\boldsymbol{\tau} \in K$, where K is a convex cone. It is assumed that $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Lambda}$ where the matrix $\boldsymbol{\Lambda}$ is known but the scalar σ^2 is unknown. Then the likelihood ratio test is based on the statistic

$$\bar{E}^2 = \frac{n\hat{\boldsymbol{\tau}}'\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{\tau}}}{\sum \mathbf{y}_i'\boldsymbol{\Lambda}^{-1}\mathbf{y}_i}, \tag{7.6}$$

where the sum is over $i = 1, \ldots, n$, and where $\hat{\boldsymbol{\tau}} = P(\bar{\mathbf{y}}, K)$ and $\bar{\mathbf{y}}$ is the sample mean (Barlow et al., 1972, p. 178). In our framework this statistic can be derived as follows. Let $\mathbf{y}$ be the $nk \times 1$ vector formed by stacking the column vectors $\mathbf{y}_i$, $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')'$. We have that $\mathbf{y}$ is $N(\mathbf{0}, \sigma^2\boldsymbol{\Phi})$, where $\boldsymbol{\Phi} = \operatorname{diag}(\boldsymbol{\Phi}_{ii})$ is block-diagonal with $\boldsymbol{\Phi}_{ii} = \boldsymbol{\Lambda}$ $(i = 1, \ldots, n)$. Consider the $nk \times k$ matrix $\boldsymbol{\Delta} = [\mathbf{I}_k, \ldots, \mathbf{I}_k]'$ and the cone $C = \boldsymbol{\Delta}K$. Then the projection $P(\mathbf{y}, C)$, with respect to the matrix $\boldsymbol{\Phi}^{-1}$, is given by $\boldsymbol{\Delta}\hat{\boldsymbol{\tau}}$. Consequently the $\bar{E}^2$ statistic in (7.6) can be written

$$\bar{E}^2 = \frac{\hat{\boldsymbol{\tau}}'\boldsymbol{\Delta}'\boldsymbol{\Phi}^{-1}\boldsymbol{\Delta}\hat{\boldsymbol{\tau}}}{\mathbf{y}'\boldsymbol{\Phi}^{-1}\mathbf{y}}.$$

Therefore under the null hypothesis this statistic is distributed as a mixture of (central) beta distributions

$$\Pr\{\bar{E}^2 \geq c\} = \sum_{i=0}^{k} w_i \Pr\{\beta_{i/2,(nk-i)/2} \geq c\}, \tag{7.7}$$

with the weights $w_i = w_i(k, \boldsymbol{\Lambda}, K)$ (Barlow et al., 1972, Th. 4.2).

This result can be extended to the noncentral case as follows. Let $\mathbf{y}_i$ be independent normal variables $N(\boldsymbol{\tau}_i, \sigma^2\boldsymbol{\Lambda})$ and suppose that the mean vector $\bar{\boldsymbol{\tau}}$ of $\bar{\mathbf{y}}$ is *null*. Then $\bar{E}^2$ has a distribution which is a mixture of noncentral beta with the same shape parameters as in (7.7), zero noncentrality parameter and

$$\delta = \sum_{i=1}^{n} \sigma^{-2}\boldsymbol{\tau}_i'\boldsymbol{\Lambda}^{-1}\boldsymbol{\tau}_i.$$

The corresponding weights remain the same as in (7.7). Similar results can be obtained for the $\bar{F}$ statistic.

Now consider the linear model (4.1). As earlier we assume $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Lambda}$ such that $\boldsymbol{\Lambda}$ is known but σ^2 is unknown. Suppose that it is required to test the null hypothesis (4.2) against the alternative (4.3). Consider the quadratic function

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Lambda}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{7.8}$$

and let $\bar{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ be the minimizers of $Q(\boldsymbol{\beta})$ subject to $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ and $\mathbf{R}_1\boldsymbol{\beta} \geq \mathbf{r}_1$, respectively. The likelihood ratio test rejects the null hypothesis for large values of

$$\bar{E}^2 = \frac{Q(\bar{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})}{Q(\bar{\boldsymbol{\beta}})} \tag{7.9}$$

(Barlow et al., 1972, p. 180). This statistic is the ratio of the difference (4.4) to its first term. The numerator in the ratio (7.9) can be reduced to the form (4.6) with $\boldsymbol{\Omega} = (\mathbf{X}'\boldsymbol{\Lambda}^{-1}\mathbf{X})^{-1}$. Also the ratio will not be changed if $\mathbf{z}$ is replaced by its projection onto the space $M^{\perp}$. Therefore under the null hypothesis $\bar{E}^2$ is distributed as a mixture of beta distributions. Since $M^{\perp}$ had dimension p, the degrees of freedom of chi-squared variables which appear in the denominator of (7.9) are i and $(N - k) + (p - i)$. Consequently

$$\Pr\{\bar{E}^2 \geq c\} = \sum_{i=p-q}^{p} w_i \Pr\{\beta_{i/2,(N-k+p-i)/2} \geq c\}, \tag{7.10}$$

where the weights $w_i = w_i(k, \boldsymbol{\Omega}, C^*)$ are given by (5.10).

Suppose now that we wish to test the null hypothesis (4.7) against the unrestricted alternative. Similarly to (7.9) we consider the test statistic

$$\bar{E}^2 = \frac{Q(\hat{\boldsymbol{\beta}}) - Q(\tilde{\boldsymbol{\beta}})}{Q(\hat{\boldsymbol{\beta}})}, \tag{7.11}$$

where $\tilde{\boldsymbol{\beta}}$ is the unrestricted minimizer and $\hat{\boldsymbol{\beta}}$ is the minimizer of $Q(\boldsymbol{\beta})$ subject to constraints (4.7). As in the case of known covariance matrix the least favorable distribution of $\bar{E}^2$ occurs when all inequality constraints are active.

The numerator of the ratio (7.11) can be represented in the form (4.8) and the denominator as the sum of numerator and $Q(\tilde{\boldsymbol{\beta}})$. Recall that $Q(\tilde{\boldsymbol{\beta}})$ is chi-squared with $N-k$ degrees of freedom and is independent of $Q(\hat{\boldsymbol{\beta}})-Q(\tilde{\boldsymbol{\beta}})$. Therefore we obtain that the least favorable null distribution of $\bar{E}^2$, defined in (7.11), is given by

$$\Pr\{\bar{E}^2 \geqslant c\} = \sum_{i=0}^{k} w_i \Pr\{\beta_{i/2,(N-k)/2} \geqslant c\}, \tag{7.12}$$

where $w_i = w_i(k, \boldsymbol{\Omega}, K^0)$ or, alternatively, $w_i = w_{k-i}(k, \boldsymbol{\Omega}, K)$ $(i = 0, \ldots, k)$.

In a similar way $\bar{F}$ statistics can be introduced. For example the denominator of ratio (7.11) can be replaced by the unrestricted minimum $Q(\tilde{\boldsymbol{\beta}})$. Then the obtained ratio statistic has a least favorable distribution which is a mixture of distributions associated with variables χ_i^2/χ_{N-k}^2. The corresponding weights w_i will be the same as in (7.12).

8 Concluding remarks

It was mentioned in §§ 4 and 7 that in the case of testing inequality constraints against an unrestricted alternative the least favorable distribution occurs when all inequality constraints of the null hypothesis are active. This is not necessarily so if we have two nested hypotheses *both* of which involve inequality constraints. A discussion of an interesting example of that type where the least favorable configuration occurs at 'infinity' is given by Warrack & Robertson (1984). In the analysis of structural models the situation is complicated further by the fact that the asymptotic covariance matrix $\boldsymbol{\Gamma}$ usually depends on the population value of the parameter vector.

Under alternative hypotheses the $\bar{\chi}^2$ and $\bar{E}^2$ statistics in general are not mixtures of (noncentral) chi-squared and beta distributions, respectively. This makes an exact calculation of the corresponding power functions, and for that matter a comparison of competing test statistics, quite a difficult problem. Available results are limited in scope and only some rather simple examples have been analysed analytically (Barlow et al., 1972, § 3.4; Pincus, 1975). Monte Carlo experiments support an intuitive conjecture that the power function increases with the restrictiveness of the alternative hypothesis (Barlow et al., 1972, p. 158). In an extreme case when the alternative region is reduced to a ray, the corresponding chi-bar-squared test is most powerful as follows from the Neyman–Pearson fundamental lemma.

The problem of testing the multivariate normal mean $\boldsymbol{\tau} = \mathbf{0}$ against the alternative $\boldsymbol{\tau} \in K$, when the covariance matrix $\boldsymbol{\Sigma}$ is completely unknown, has been considered by Perlman (1969). His results, which are not reproducible in our framework, can be summarized as follows. Suppose that $n > k$ and let

$$\mathbf{S} = (n-1)^{-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$$

be the sample covariance matrix giving an estimator of $\boldsymbol{\Sigma}$. Then the likelihood ratio test rejects the null hypothesis for large values of

$$U = \frac{\hat{\boldsymbol{\tau}}_*' \mathbf{S}^{-1} \hat{\boldsymbol{\tau}}_*}{1 - n^{-1} + (\bar{\mathbf{y}} - \hat{\boldsymbol{\tau}}_*)' \mathbf{S}^{-1} (\bar{\mathbf{y}} - \hat{\boldsymbol{\tau}}_*)}, \tag{8.1}$$

where $\hat{\boldsymbol{\tau}}_*$ is given by the projection $\hat{\boldsymbol{\tau}}_* = P(\bar{\mathbf{y}}, K)$ with respect to the matrix $\mathbf{S}^{-1}$. The null distribution of this statistic is

$$\Pr\{U \geq c\} = \sum_{i=0}^{k} w_i \Pr\{\chi_i^2/\chi_{n-k}^2 \geq c\}, \tag{8.2}$$

where $w_i = w_i(k, \boldsymbol{\Sigma}, K)$. Note that the null distribution of U depends on the unknown matrix $\boldsymbol{\Sigma}$ through the corresponding weights w_i.

The distributional result (8.2) is exact. As n tends to infinity $\mathbf{S}$ and $\bar{\mathbf{y}}$ converge in probability to $\boldsymbol{\Sigma}$ and $\boldsymbol{\tau}$, respectively. Therefore under the null hypothesis $\boldsymbol{\tau} = \mathbf{0}$, nU is asymptotically equivalent to the $\bar{\chi}^2$ statistic $n\hat{\boldsymbol{\tau}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\tau}}$. Note that by the law of large numbers χ_{n-k}^2/n tends in probability to one and hence asymptotically (8.2) coincides with the corresponding result for the $\bar{\chi}^2$ statistic.

References

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. & Brunk, H.D. (1972). *Statistical Inference under Order Restrictions. The Theory and Application of Isotonic Regression.* New York: Wiley.

Bartholomew, D.J. (1959). A test of homogeneity for ordered alternatives. *Biometrika* **46,** 36–48.

Bartholomew, D.J. (1961). Ordered tests in the analysis of variance. *Biometrika* **48,** 325–332.

Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass: MIT Press.

Bohrer, R. & Chow, W. (1978). Weights for one-sided multivariate inference. *Appl. Statist.* **27,** 100–104.

Browne, M.W. (1982). Covariance structures. In *Topics in Applied Multivariate Analysis,* Ed. D.M. Hawkins, pp. 72–141. Cambridge University Press.

Browne, M.W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *Br. J. Math. Statist. Psychol.* **37,** 62–83.

Chacko, V.J. (1966). Modified chi-square test for ordered alternatives. *Sankhyā* B **28,** 185–190.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25,** 573–578.

Dykstra, R.L. (1983). An algorithm for restricted least squares regression. *J. Am. Statist. Assoc.* **78,** 837–842.

Dykstra, R.L. & Robertson, T. (1982a). Order restricted statistical tests on multinomial and Poisson parameters: the starshaped restriction. *Ann. Statist.* **10,** 1246–1252.

Dykstra, R.L. & Robertson, T. (1982b). An algorithm for isotonic regression for two or more independent variables. *Ann. Statist.* **10,** 708–716.

Farebrother, R.W. (1986). Testing linear inequality constraints in the standard linear model. *Comm. Statist.* A **15,** 7–31.

Feder, P.I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. *Ann. Math. Statist.* **39,** 2044–2055.

Gouriéroux, C., Holly, A. & Monfort, A. (1982). Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* **50,** 63–80.

Johnson, N.L. & Kotz, S. (1970). *Continuous Univariate Distributions,* **2.** Boston: Houghton-Mifflin.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32,** 443–482.

Jöreskog, K.G. (1981). Analysis of covariance structures (with discussion). *Scan. J. Statist.* **8,** 65–92.

Kendall, M.G. & Stuart, A. (1979). *The Advanced Theory of Statistics,* **2,** 4th ed. London: Griffin.

Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50,** 403–418.

Kudô, A. & Choi, J.R. (1975). A generalized multivariate analogue of the one sided test. *Mem. Fac. Sci.,* Kyushu Univ. A, **29,** 303–328.

Kudô, A. & Yao, J.S. (1982). An approximate formula for $p(l, k)$ used in order restricted inferences, and the analysis of freezing dates of lake Mendota. *Comm. Statist.* A **11,** 607–614.

Lehmann, E.L. (1959). *Testing Statistical Hypothesis.* New York: Wiley.

Mitchell, B.F., Demyanov, V.F. & Malozemov, V.N. (1974). Finding the point of a polyhedron closest to the origin. *SIAM J. Control* **12,** 19–26.

Nüesch, P.E. (1964). Multivariate tests of location for restricted alternatives. Doctoral dissertation, Swiss Federal Institute of Technology, Zürich.

Nüesch, P.E. (1966). On the problem of testing location in multivariate populations for restricted alternatives. *Ann. Math. Statist.* **37,** 113–119.

Perlman, M.D. (1969). One-sided testing problems in multivariate analysis. *Ann. Math. Statist.* **40,** 549–567.

Pincus, R. (1975). Testing linear hypotheses under restricted alternatives. *Math. Oper. u. Statist.* **6,** 733–751.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications.* New York: Wiley.

Rao, C.R. & Mitra, S.K. (1971). *Generalized Inverse of Matrices and its Applications.* New York: Wiley.

Robertson, T. (1978). Testing for and against an order restriction on multinomial parameters. *J. Am. Statist. Assoc.* **73,** 197–202.

Robertson, T. & Wegman, E.J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6,** 485–505.

Robertson, T. & Wright, F.T. (1982). Testing for ordered alternatives with increased precision in one of the samples. *Biometrika* **69,** 579–586.

Robertson, T. & Wright, F.T. (1983). On approximation of the level probabilities and associated distributions in order restricted inference. *Biometrika* **70,** 597–606.

Shapiro, A. (1985a). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72,** 133–144.

Shapiro, A. (1985b). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *S. Afr. Statist. J.* **19,** 73–81.

Shapiro, A. (1987). On differentiability of metric projections in $\mathbb{R}^n$, 1: Boundary case. *Proc. Am. Math. Soc.* **99,** 123–128.

Shorack, G.R. (1967). Testing against ordered alternatives in model I analysis of variance; normal theory and nonparametric. *Ann. Math. Statist.* **38,** 1740–1753.

Silvey, S.D. (1970). *Statistical Inference.* Baltimore: Penguin.

Siskind, V. (1976). Approximate probability integrals and critical values for Bartholomew's test for ordered means. *Biometrika* **63,** 647–654.

Stoer, J. & Witzgall, C. (1970). *Convexity and Optimization in Finite Dimensions,* **2.** Berlin: Springer-Verlag.

Stroud, T.W. (1972). Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistic. *Ann. Math. Statist.* **43,** 447–454.

Swain, A.J. (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika* **40,** 315–335.

Warrack, G. & Robertson, T. (1984). A likelihood ratio test regarding two nested but oblique order-restricted hypotheses. *J. Am. Statist. Assoc.* **79,** 881–886.

Wolfe, P. (1976). Finding the nearest point in a polytope. *Math. Program.* **11,** 128–149.

Résumé

Dans cet article on examine une théorie de distribution de fonctions des observations utilisées dans des tests pour des problèmes divers dans l'analyse à plusieurs variables avec des contraintes d'inégalité. On présente un point de vue uni basé sur les caractéristiques géométriques de cônes convexes. Les tests $\bar{\chi}^2$ et $\bar{E}^2$ sont introduits. On discute leurs applications à des problèmes concernant des tests d'hypothèse.

[*Received July* 1985, *revised July* 1987]