Biometrika (1985), **72**, 1, *pp.* 133–44
Printed in Great Britain

Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints

By ALEXANDER SHAPIRO
Department of Mathematics and Applied Mathematics, University of South Africa, Pretoria 0001, South Africa

SUMMARY

The analysis of moment structural models has become an important tool of investigation in behavioural, educational and economic studies. The chi-squared large-sample test is routinely employed to assess the goodness of fit of the model considered. However, in order to invoke the standard asymptotic distribution theory certain regularity conditions have to be met. Here we consider the case where the population value of the parameter vector is a boundary point of the feasible region. We show that in this case the asymptotic distribution of test statistic is a mixture of chi-squared distributions. The problem of finding the corresponding weights is discussed.

Some key words: Asymptotic normality; Chi-bar-squared statistic; Factor analysis; Heywood case; Moment structure; Test statistic.

1. INTRODUCTION

The analysis of moment structural models has become an important tool of investigation in behavioural, educational and economic studies (Jöreskog, 1970, 1981; Anderson, 1973; Browne, 1974, 1982; Shapiro, 1983; Dijkstra, 1983 and references therein). In such a model the elements of the first two moments, for example, are regarded as functions of a $q\times 1$ parameter vector θ, which belongs to a specified parameter space Θ. That is the elements of the mean vector and the covariance matrix Ω. forming the $m\times 1$ vector $\xi = (\xi_1, ..., \xi_m)'$, are considered as functions of θ,

$$\xi_i = g_i(\theta) \quad (i = 1, ..., m). \tag{1·1}$$

For instance, the well-known factor analysis model assumes that the $p\times p$ matrix Ω is representable in the form

$$\Omega = \Lambda\Lambda' + \Psi, \tag{1·2}$$

where Λ is the $p\times k$ matrix of factor loadings and Ψ is the diagonal matrix of the residual variances (Lawley & Maxwell, 1971). Here the covariance structural model is given by (1·2) with the parameter vector θ composed of the elements of Λ and the diagonal elements of Ψ and ξ composed of $\frac{1}{2}p(p+1)$ nonduplicated elements of Ω.

The model is said to hold if there exists a vector $\theta_0 \in \Theta$ such that

$$\xi_0 = g(\theta_0), \tag{1·3}$$

where ξ_0 is the population value of ξ and $g(\theta) = (g_1(\theta), ..., g_m(\theta))'$. Given a sample estimate $\hat{x}$ of ξ_0 one fits the model (1·1) by minimizing the discrepancy between $\hat{x}$ and $\xi = g(\theta)$, using a real-valued discrepancy function $F(x, \xi)$ of two vector variables $x, \xi \in \mathbb{R}^m$

(Browne, 1982). Then the validity of the model is tested by means of the minimum discrepancy function test statistic $n\hat{F}$, where n is the sample size and $\hat{F}$ is the minimal value of the discrepancy function

$$\hat{F} = \min_{\theta \in \Theta} F\{\hat{x}, g(\theta)\}. \tag{1·4}$$

It is known that for an appropriate choice of F and under certain regularity conditions and the null hypothesis (1·3) the asymptotic distribution of $n\hat{F}$ is central chi-squared with $m-q$ degrees of freedom (Jöreskog, 1970; Browne, 1974, 1984; Shapiro, 1983). This result is routinely employed to assess the goodness of fit of the model.

One of the regularity conditions to be met here is that θ_0 must be an interior point of the permissible set Θ. However, it often happens in practice that the minimization procedure (1·4) leads to a solution which lies on the boundary of Θ. Then in some cases it is reasonable to believe that the population value θ_0 is a boundary point or at least is sufficiently close to the boundary of Θ. If such a situation occurs the standard theory is not applicable and the asymptotic distribution of $n\hat{F}$ need not be chi-squared. For example in the factor analysis model (1·2), the matrix Ψ is the covariance matrix of error components and hence its diagonal elements must be nonnegative. It can be seen from Table 8 of Jöreskog (1967, p. 474) that boundary solutions, called Heywood cases, are quite frequent in the factor analysis model. Geweke & Singleton (1980) found in their Monte Carlo study of the factor analysis model, that the standard asymptotic theory can be misleading in the boundary case where some of the diagonal elements of Ψ_0 are zero. Such situations happen quite often in practical applications; see the second example of Andersen in the discussion of Jöreskog (1981).

The present paper is concerned with the asymptotic distribution of the test statistic $n\hat{F}$ when θ_0 is a boundary point of Θ. We extend a result of Chernoff (1954) about the asymptotic distribution of likelihood ratio statistics to the minimum discrepancy function test statistics. It will be shown that for correctly specified discrepancy function, $n\hat{F}$ is asymptotically distributed as a mixture of chi-squared distributions. Afterwards we address the problem of finding the corresponding weights. The development presented here is related to works of Kudô (1963), Nüesch (1966) and Kudô & Choi (1975) on a multivariate analogue of the one-sided test.

2. Asymptotic distribution of the minimum discrepancy function test statistics

In this section we study the asymptotic distribution of the test statistic $n\hat{F}$. Throughout this paper we suppose that the functions $g_i(\theta)$ $(i = 1, \ldots, m)$ are continuously differentiable and that the model (1·1) holds. Let Ξ denote the image of the mapping $g(\theta) = \{g_1(\theta), \ldots, g_m(\theta)\}'$,

$$\Xi = \{\xi\colon \xi = g(\theta),\ \theta \in \Theta\}.$$

For the time being we shall be dealing with ξ and Ξ rather than θ and Θ. Note that $\hat{F}$ can be expressed as

$$\hat{F} = \min_{\xi \in \Xi} F(\hat{x}, \xi). \tag{2·1}$$

We suppose that the discrepancy function satisfies the following conditions (Browne, 1982, p. 81).

Condition 1. $F(x,\xi) \geqslant 0$ for all x, ξ.

Condition 2. $F(x,\xi) = 0$ if and only if $x = \xi$.

Condition 3. F is twice continuously differentiable in x and ξ.

The discrepancy functions currently in use in the analysis of covariance structures, maximum likelihood and generalized least squares, both satisfy Conditions 1–3. For other examples of discrepancy functions satisfying Conditions 1–3, see Swain (1975) and Browne (1984). We denote by $2V_0$ the Hessian matrix $\partial^2 F/\partial\xi\,\partial\xi'$ of F at the point (ξ_0, ξ_0). Note that Conditions 1 and 2 imply that function $F(\xi_0, .)$ attains its minimal value of zero at the point ξ_0 and hence the matrix V_0 is nonnegative-definite. It can be shown that if Conditions 1–3 hold, then the second-order Taylor approximation of F at (ξ_0, ξ_0) is given by $(x-\xi)'V_0(x-\xi)$ and consequently

$$\partial^2 F/\partial x\,\partial x' = 2V_0, \quad \partial^2 F/\partial x\,\partial\xi' = -2V_0.$$

For a given $x, \xi^*(x)$ denotes a minimizer of $F(x,.)$ over Ξ and $F_{\min}(x)$ denotes the corresponding minimum

$$F_{\min}(x) = \inf_{\xi\in\Xi} F(x,\xi) = \inf_{\theta\in\Theta} F(x, g(\theta)). \tag{2·2}$$

Because $\hat{F} = F_{\min}(\hat{x})$, the asymptotic behaviour of the test statistic $n\hat{F}$ is closely related to analytical properties of the function $F_{\min}(x)$.

Note that ξ_0 is the unique minimizer of $F(\xi_0, .)$ and hence $\xi^*(\xi_0) = \xi_0$. In order to ensure the continuity of $\xi^*(x)$, and of $F_{\min}(x)$, at ξ_0 we need to impose the following condition.

Condition 4. There exist positive constants M and ε such that $F(x,\xi) \geqslant \varepsilon$ whenever $x, \xi \in \Xi$ and $\| x-\xi \| \geqslant M$.

Condition 4 prevents $\xi^*(x)$ from going to infinity as x approaches ξ_0. Together with Conditions 1–3 this implies that the minimizer $\xi^*(x)$ tends to ξ_0 as x tends to ξ_0. For more details and a discussion, see Shapiro (1984a). Usually $F(x,\xi) \to \infty$ as $\| x-\xi \| \to \infty$, which implies Condition 4, so that Condition 4 is a mild restriction on the discrepancy function. Of course if the set Ξ is bounded, then Condition 4 holds automatically.

Definition 2·1. We say that the set Ξ is approximated at ξ_0 by a cone $\mathscr{C}$, called the approximating cone, if

$$\inf_{\zeta\in\mathscr{C}} \| (\xi-\xi_0)-\zeta \| = o(\| \xi-\xi_0 \|) \quad (\xi \in \Xi), \quad \inf_{\xi\in\Xi} \| (\xi-\xi_0)-\zeta \| = o(\| \zeta \|) \quad (\zeta \in \mathscr{C}).$$

Given the approximating cone $\mathscr{C}$ we would like to approximate $F_{\min}(x)$ by a simpler function, namely $F^*(x-\xi_0)$, where

$$F^*(v) = \inf_{\zeta\in\mathscr{C}} (v-\zeta)'V_0(v-\zeta). \tag{2·3}$$

To do so we shall need the additional condition.

Condition 5. The matrix V_0 is nonsingular.

For a discussion of the following technical result, see Shapiro (1984b).

LEMMA 2·1. *Suppose that Conditions* 1–5 *hold. Then*

$$F_{\min}(x) - F^*(x-\xi_0) = o(\| x-\xi_0 \|^2). \tag{2·4}$$

From now on we suppose that the asymptotic distribution of $n^{\frac{1}{2}}(\hat{x}-\xi_0)$ is multivariate normal with zero mean and covariance matrix Γ. Usually this assumption can be justified by an application of the Central Limit Theorem. Then Lemma 2·1 implies the following result.

LEMMA 2·2. *Suppose that Ξ is approximated at ξ_0 by a cone $\mathscr{C}$ and that Conditions 1–5 hold. Then the asymptotic distribution of the test statistic $n\hat{F}$ is the same as the distribution of*

$$\inf_{\zeta \in \mathscr{C}} (z-\zeta)' V_0 (z-\zeta), \tag{2·5}$$

where z is a multivariate normal random variable with zero mean and covariance matrix Γ.

Proof. It follows from (2·4) that

$$n\hat{F} - nF^*(\hat{x}-\xi_0) = o(\| n^{\frac{1}{2}}(\hat{x}-\xi_0) \|^2)$$

and consequently

$$n\hat{F} = nF^*(\hat{x}-\xi_0) + o_p(1), \tag{2·6}$$

where $o_p(1)$ converges in probability to zero as $n \to \infty$. From the definition of $F^*(.)$ we have

$$nF^*(\hat{x}-\xi_0) = \inf_{\zeta \in \mathscr{C}} \{n^{\frac{1}{2}}(\hat{x}-\xi_0) - n^{\frac{1}{2}}\zeta\}' V_0 \{n^{\frac{1}{2}}(\hat{x}-\xi_0) - n^{\frac{1}{2}}\zeta\}. \tag{2·7}$$

Since $\mathscr{C}$ is a cone, $n^{\frac{1}{2}}\zeta$ in the right-hand side of (2·7) can be replaced by ζ. Together with (2·6) this implies (2·5). □

We may remark that there is a certain resemblance between Lemma 2·2 and the result of Chernoff (1954, Theorem 1) about the asymptotic distribution of likelihood ratio statistics. This resemblance becomes even more evident in the next theorem.

Now let us return to the mapping $g(\theta)$ and the parameter space Θ. The following regularity assumptions will be imposed.

Assumption 1. The set Θ is compact.

Assumption 2. The parameter vector θ is identified at θ_0, that is $g(\theta^*) = g(\theta_0)$ and $\theta^* \in \Theta$ implies that $\theta^* = \theta_0$.

Assumption 3. The mapping $g(\theta)$ is defined in a neighbourhood of θ_0 and the $m \times q$ Jacobian matrix $\Delta = \partial g/\partial\theta'$ at θ_0 is of full column rank q.

Assumption 4. The covariance matrix Γ is nonsingular and $V_0 = \Gamma^{-1}$.

Assumption 1 is needed to guarantee that the minimum discrepancy function estimator $\hat{\theta}$ of θ_0 is consistent. In many practical situations Assumption 1 does not hold. For instance in the factor analysis model (1·2) the parameter space Θ is unbounded and hence is not compact. Fortunately Assumption 1 can be replaced by the following condition of boundedness (Shapiro, 1984a).

Assumption 1*. There does not exist an unbounded sequence $\{\theta_k\}$ in Θ such that

$$g(\theta_k) \to g(\theta_0).$$

It can be easily verified that for the factor analysis model the boundedness condition holds.

Assumptions 1–4, and especially 4, are the usual regularity assumptions ensuring that the test statistic $n\hat{F}$ is asymptotically chi-squared if θ_0 is an interior point of Θ. Here we consider a more general case where θ_0 is possibly a boundary point of Θ. We assume that Θ is approximated at θ_0 by a convex cone $\mathscr{K}$. Then Lemma 2·2 together with some matrix algebra implies the following result.

THEOREM 2·1. *Suppose that Θ is approximated at θ_0 by a convex cone $\mathscr{K}$ and that Conditions 1–5 and Assumptions 1–4 hold. Then $n\hat{F}$ is asymptotically distributed as $\chi^2_\nu + \bar{\chi}^2$, where χ^2_ν and $\bar{\chi}^2$ are independent random variables, χ^2_ν has chi-squared distribution with $\nu = m-q$ degrees of freedom and*

$$\bar{\chi}^2 = \inf_{\eta \in \mathscr{K}} (y-\eta)'U(y-\eta), \tag{2·8}$$

such that y is a $N(0, U^{-1})$ random variable with $U = \Delta'\Gamma^{-1}\Delta$.

Proof. Assumptions 1–3 imply that the set Ξ is approximated at ξ_0 by the convex cone $\mathscr{C}$,

$$\mathscr{C} = \{\zeta\colon \zeta = \Delta\eta,\ \eta \in \mathscr{K}\}.$$

Let Φ be an $m \times (m-q)$ matrix of rank $m-q$ such that $\Phi'\Delta = 0$. Denote by $\mathscr{L}$ the linear space generated by the column vectors of Δ. Let γ^* be the solution of

$$\min_{\gamma \in \mathscr{L}} (z-\gamma)'V_0(z-\gamma).$$

Then, since $\mathscr{C} \subseteq \mathscr{L}$, the infimum (2·5) is equal to

$$(z-\gamma^*)'V_0(z-\gamma^*) + \inf_{\zeta \in \mathscr{C}} (\gamma^*-\zeta)'V_0(\gamma^*-\zeta). \tag{2·9}$$

It can be calculated that $\gamma^* = \Delta(\Delta'V_0\Delta)^{-1}\Delta'V_0 z$, and the first term in the sum (2·9) is $z'\Phi(\Phi'V_0^{-1}\Phi)^{-1}\Phi'z$. Let us define

$$y = (\Delta'V_0\Delta)^{-1}\Delta'V_0 z, \quad v = (\Phi'V_0^{-1}\Phi)^{-\frac{1}{2}}\Phi'z,$$

so that the sum (2·9) becomes

$$v'v + \inf_{\eta \in \mathscr{K}} (y-\eta)'\Delta'V_0\Delta(y-\eta). \tag{2·10}$$

Since z is $N(0, \Gamma)$ and $V_0 = \Gamma^{-1}$ we have that y is $N(0, U^{-1})$ and v is $N(0, I)$. Moreover, y and v are uncorrelated because $\Phi'\Delta = 0$, and hence are independent. This together with (2·10) completes the proof. □

Let $\tilde{\theta}$ be the unrestricted minimum discrepancy function estimator of θ_0, that is $\tilde{\theta}$ is the unconstrained minimizer of $F(\hat{x}, g(.))$ in a neighbourhood of θ_0. Then the standard asymptotic theory implies that U^{-1} is the asymptotic covariance matrix of $n^{\frac{1}{2}}(\tilde{\theta}-\theta_0)$ (Browne, 1974; Shapiro, 1983). In the case of maximum likelihood, U is Fisher's information matrix (Chernoff, 1954).

If θ_0 is an interior point of Θ, then Θ is approximated at θ_0 by the whole space $\mathbb{R}^q$ and consequently the variable $\bar{\chi}^2$ is identically zero. In this case we obtain the usual result that the asymptotic distribution of $n\hat{F}$ is chi-squared with $m-q$ degrees of freedom.

Let the parameter space Θ be defined by equality and inequality constraints

$$\Theta = \{\theta \in \mathbb{R}^q\colon h_i(\theta) = 0, i = 1, \ldots, l;\ h_i(\theta) \leqslant 0, i = l+1, \ldots, k\}, \tag{2·11}$$

where h_i $(i = 1, \ldots, k)$ are continuously differentiable functions. It is shown in the theory of nonlinear programming that, if a constraint qualification holds, then Θ is approximated at θ_0 by the polyhedral convex cone

$$\mathscr{K} = \{\theta: a_i'\theta = 0, i = 1, \ldots, l; \; a_i'\theta \leqslant 0, \; i \in \mathscr{I}(\theta_0)\}, \tag{2·12}$$

where $a_i = \partial h_i/\partial\theta$ $(i = 1, \ldots, k)$ is the gradient vector of h_i at θ_0 and the index set

$$\mathscr{I}(\theta_0) = \{i: h_i(\theta_0) = 0, i = l+1, \ldots, k\}$$

corresponds to the inequality constraints active at θ_0. For extensive discussion of various constraint qualifications the reader is referred to Bazaraa, Goode & Shetty (1972). For example, one may use the convenient constraint qualification proposed by Mangasarian & Fromovitz (1967); namely (i) there exists a vector b such that $a_i'b = 0$ for $i = 1, \ldots, l$, $a_i'b < 0$ for $i \in \mathscr{I}(\theta_0)$, and (ii) the gradient vectors $a_1, \ldots, a_l$ are linearly independent.

Finally we may remark that if only equality constraints are present in the definition of Θ then $\mathscr{K}$ becomes a $q-l$ dimensional linear space and $\bar{\chi}^2$ has chi-squared distribution with l degrees of freedom. Consequently in this case $n\hat{F}$ is asymptotically chi-squared with $m-q+l$ degrees of freedom.

3. The distribution of $\bar{\chi}^2$

In this section we study the random variable $\bar{\chi}^2$ defined by (2·8). The distribution of $\bar{\chi}^2$ is related to the one-sided testing problem in multivariate analysis and has been studied by several authors (Bartholomew, 1961; Kudô, 1963; Nüesch, 1966; Kudô & Choi, 1975). Kudô (1963) showed that if $\mathscr{K}$ is the nonnegative orthant, then $\bar{\chi}^2$ is distributed as a mixture of chi-squared distributions; see also Nüesch (1966). This result has been generalized by Kudô & Choi (1975), to deal with the case where $\mathscr{K}$ is defined by a number of linear inequalities satisfying a certain 'rank assumption'. Here we propose a simple proof of this result for any convex cone $\mathscr{K}$. For that purpose we shall use some geometrical properties of polyhedral convex cones.

The reader may note that particular cases of $\bar{\chi}^2$ statistic, called chi-bar-squared statistic, also appear in isotonic regression (Barlow et al., 1972, Ch. 1) and in testing order restrictions on multinomial parameters (Robertson, 1978; Dykstra & Robertson, 1982).

THEOREM 3·1. *The random variable $\bar{\chi}^2$ is distributed as a mixture of chi-squared distributions, namely*

$$\operatorname{pr}(\bar{\chi}^2 \geqslant c^2) = \sum_{i=0}^{q} w_i \operatorname{pr}(\chi_i^2 \geqslant c^2), \tag{3·1}$$

where χ_i^2 is a chi-squared random variable with i degrees of freedom, $\chi_0^2 \equiv 0$, and $w_i = w_i(U, \mathscr{K})$ are nonnegative weights such that $w_0 + \ldots + w_q = 1$.

Proof. To save calculations and simplify notations we prove Theorem 3·1 for a particular case when U is the identity matrix I. The general case can be easily reduced to this particular case by a linear transformation.

For a given y the infimum in (2·8) is attained at a unique point, denoted by $\pi(y)$, and represents the squared distance from y to $\mathscr{K}$. Let $\mathscr{K}$ be a polyhedral closed convex cone. We denote by $\mathscr{F}(\mathscr{K})$ the set of all faces of $\mathscr{K}$ and by $\mathscr{K}^0$ the polar cone

$$\mathscr{K}^0 = \{\beta \in \mathbb{R}^q: \beta'\eta \leqslant 0 \text{ for all } \eta \in \mathscr{K}\}.$$

For the definition and basic properties of faces and polar cones the reader is referred to Stoer & Witzgall (1970, Ch. 2). Note that $\mathscr{K}^0$ is also a polyhedral convex cone and $(\mathscr{K}^0)^0 = \mathscr{K}$.

To a face $\phi \in \mathscr{F}(\mathscr{K})$ corresponds the polar face $\phi^* \in \mathscr{F}(\mathscr{K}^0)$ such that the linear spaces generated by ϕ and ϕ^* are the orthogonal complements to each other. We denote by P_ϕ and P_{ϕ^*} the symmetric idempotent matrices giving the orthogonal projections onto the spaces generated by ϕ and ϕ^* respectively. Note that $P_{\phi^*} = I - P_\phi$. We also define $\mathscr{S}_\phi = \{v \in \mathbb{R}^q\colon \pi(v) \in \phi\}$. We have that $\pi(v) \in \phi$ if and only if $P_\phi v \in \mathscr{K}$ and $P_{\phi^*} v \in \mathscr{K}^0$. Since $(\mathscr{K}^0)^0 = \mathscr{K}$, the set $\mathscr{S}_\phi$ is defined by the inequalities

$$a' P_\phi v \leqslant 0 \quad (a \in \mathscr{K}^0), \quad b' P_{\phi^*} v \leqslant 0 \quad (b \in \mathscr{K}). \tag{3·2}$$

Because $\mathscr{K}$ and $\mathscr{K}^0$ are polyhedral, in (3·2) it is sufficient to take a finite number of vectors a and b generating the cones $\mathscr{K}^0$ and $\mathscr{K}$ respectively.

Now we shall need the following result, which in various forms has been used by several authors (Kudô, 1963, p. 414; Perlman, 1969, p. 552; Barlow et al., 1972, p. 128).

LEMMA 3·1. *Suppose that y is $N(0, I)$, Q is an idempotent symmetric matrix of rank r and $\mathscr{A}$ is a cone defined by*

$$\mathscr{A} = \{v \in \mathbb{R}^q\colon d_i' v \leqslant 0,\ i = 1, \ldots, k\},$$

where d_i $(i = 1, \ldots, k)$ are vectors such that either $Qd_i = 0$ or $Qd_i = d_i$. Then the conditional distribution of $y'Qy$, given $y \in \mathscr{A}$, is that of a chi-squared random variable with r degrees of freedom.

Proof of Theorem 3·1 (continued). Consider $\phi \in \mathscr{F}(\mathscr{K})$ and $y \in \mathscr{S}_\phi$. Then $\pi(y) = P_\phi y$ and $\bar{\chi}^2 = y'(I - P_\phi) y = y' P_{\phi^*} y$. The set $\mathscr{S}_\phi$ is defined by linear inequalities and hence is a convex polyhedral cone. Moreover, the corresponding inequalities (3·2) satisfy the conditions of Lemma 3·1 with respect to the matrix $Q = P_{\phi^*}$. Therefore the conditional distribution of $\bar{\chi}^2$, given $y \in \mathscr{S}_\phi$, is chi-squared with $r = \operatorname{rank} P_{\phi^*} = q - \operatorname{rank} P_\phi = q - \dim \phi$ degrees of freedom. Then noting that the sets $\mathscr{S}_\phi$, $\phi \in \mathscr{F}(\mathscr{K})$, are almost disjoint and cover $\mathbb{R}^q$ and applying Bayes's formula we obtain that $\bar{\chi}^2$ is distributed as a mixture of chi-squared distributions.

Now let $\mathscr{K}$ be a convex cone and consider a sequence $\{\mathscr{K}_s\}$ of polyhedral convex cones converging to $\mathscr{K}$. Let $w_i^{(s)}$ $(i = 0, 1, \ldots, q)$ be the weights corresponding to the cone $\mathscr{K}_s$. Since the weights belong to the bounded interval $[0, 1]$ we can suppose, by passing to a subsequence if necessary, that $w_i^{(s)} \to w_i$ as $s \to \infty$ $(i = 0, \ldots, q)$. Then we obtain that $\bar{\chi}^2$ is distributed as a mixture of chi-squared distributions with the weights w_i. Finally we note that the weights corresponding to a given cone $\mathscr{K}$ are unique since the density functions corresponding to chi-squared distributions with different degrees of freedom are linearly independent. □

The weights w_i depend on U and $\mathscr{K}$ and seem to appear in various problems of one-sided testing in multivariate analysis (Shorack, 1967; Perlman, 1969, § 7). The question of their calculation is not a simple one and will be discussed in § 4.

4. The weights corresponding to $\bar{\chi}^2$

In this section we address the problem of finding the weights $w_i = w_i(U, \mathscr{K})$ for some matrix U and convex cone $\mathscr{K}$. If the cone $\mathscr{K}$ is polyhedral, then it follows from the proof of Theorem 3·1, that $w_i = \text{pr}\{\pi(y)$ belongs to a $(q-i)$-dimensional face of $\mathscr{K}\}$, that is

$$w_i(U, \mathscr{K}) = \sum \text{pr}\,(y \in \mathscr{S}_\phi), \tag{4·1}$$

where the sum is over $\phi \in \mathscr{F}(\mathscr{K})$, $\dim \phi = q-i$.

First we consider the case $U = I$. The general case can be reduced to this specific case by a linear transformation and will be discussed later. The inequalities (3·2) are orthogonal to each other. Therefore we have

$$\text{pr}\,(y \in \mathscr{S}_\phi) = \text{pr}\,(y \in \Pi_\phi)\,\text{pr}\,(y \in \Pi_{\phi^*}), \tag{4·2}$$

where Π_ϕ and Π_{ϕ^*} are the sets $\{v\colon P_\phi v \in \phi\}$ and $\{v\colon P_{\phi^*} v \in \phi^*\}$ respectively. For the null-dimensional face $\phi = \{0\}$, the set Π_ϕ is the whole space $\mathbb{R}^q$ and hence the corresponding probability $\text{pr}\,(y \in \Pi_\phi)$ is one. The polar face of $\{0\}$ is $\mathscr{K}^0$ and thus

$$w_q(I, \mathscr{K}) = \text{pr}\,(y \in \mathscr{K}^0). \tag{4·3}$$

For a one-dimensional face ϕ, the set Π_ϕ is a half-space and hence $\text{pr}\,(y \in \Pi_\phi)$ is $\frac{1}{2}$. Consequently

$$w_{q-1}(I, \mathscr{K}) = \tfrac{1}{2}\Sigma\,\text{pr}\,(y \in \Pi_{\phi^*}), \tag{4·4}$$

where the summation is taken over all faces ϕ^* of $\mathscr{K}^0$ of dimension $q-1$. For a two-dimensional face ϕ, $\text{pr}\,(y \in \Pi_\phi)$ is $\frac{1}{2}\alpha(\phi)/\pi$, where $\alpha(\phi)$ denotes the plane angle of ϕ. Then

$$w_{q-2}(I, \mathscr{K}) = \tfrac{1}{2}\pi^{-1}\,\Sigma\,\alpha(\phi)\,\text{pr}\,(y \in \Pi_{\phi^*}), \tag{4·5}$$

where the summation is taken over all faces ϕ^* of $\mathscr{K}^0$ of dimension $q-2$ or, equivalently, over all faces ϕ of $\mathscr{K}$ of dimension 2. And so on, until

$$w_0(I, \mathscr{K}) = \text{pr}\,(y \in \mathscr{K}). \tag{4·6}$$

One may note that there exists a certain symmetry between the weights corresponding to polar cones. It can be seen that the set $\mathscr{S}_\phi$ with respect to the cone $\mathscr{K}$ is the same as the set $\mathscr{S}_{\phi^*}$ with respect to the cone $\mathscr{K}^0$. Consequently $w_{q-i}(I, \mathscr{K}) = w_i(I, \mathscr{K}^0)$ for $i = 0, \ldots, q$.

In the three-dimensional space $\mathbb{R}^3$, formulae (4·3)–(4·6) have a clear geometric interpretation. Note that in $\mathbb{R}^3$ the probability $\text{pr}\,(y \in \mathscr{K})$ is equal to $\frac{1}{4}\beta(\mathscr{K})/\pi$, where $\beta(\mathscr{K})$ denotes the solid angle of $\mathscr{K}$. Consequently the weights w_0, w_1, w_2, w_3 are given by $\frac{1}{4}\beta(\mathscr{K})/\pi$, $\frac{1}{4}\alpha(\mathscr{K})/\pi$, $\frac{1}{4}\alpha(\mathscr{K}^0)/\pi$ and $\frac{1}{4}\beta(\mathscr{K}^0)/\pi$ respectively, where $\alpha(\mathscr{K})$ denotes the plane angle corresponding to the surface of $\mathscr{K}$. Although these formulae have been established for polyhedral cones, they hold, by passing to a limit if necessary, for arbitrary convex cones. For example consider a circular cone in $\mathbb{R}^3$ with half-angle γ. Then the weights w_0, w_1, w_2, w_3 are equal to $\frac{1}{2}(1-\cos\gamma)$, $\frac{1}{2}\sin\gamma$, $\frac{1}{2}\cos\gamma$ and $\frac{1}{2}(1-\sin\gamma)$ respectively (Wynn, 1975, p. 396).

It often happens that the cone $\mathscr{K}$ is given by the nonnegative orthant $\mathbb{R}^q_+ = (x\colon x \geqslant 0)$, or can be reduced to this case by a linear transformation, while the matrix U is arbitrary. For $\mathscr{K} = \mathbb{R}^q_+$ the weights $w_i = w_{i,q}(U)$ can be calculated in a closed form, at least for small values of q. We give below a list of corresponding formulae for $q \leqslant 4$, which will suffice for many practical situations; see examples and discussion of § 5.

For $q = 1$ we clearly have

$$w_{0,1}(U) = w_{1,1}(U) = \tfrac{1}{2}. \tag{4·7}$$

Let us denote by ρ_{ij} the (i, j)th element of the correlation matrix of y,

$$(\text{diag}\,U^{-1})^{-\frac{1}{2}}\,U^{-1}(\text{diag}\,U^{-1})^{-\frac{1}{2}}. \tag{4·8}$$

Then for $q = 2$ and 3, Kudô (1963, pp. 414–5), building on work of Kendall (1941) and David (1953), proposed formulae which in our case can be written as follows:

$$\begin{gathered} w_{0,2}(U) = \tfrac{1}{2}\pi^{-1}(\pi - \cos^{-1}\rho_{12}), \quad w_{1,2}(U) = \tfrac{1}{2}, \quad w_{2,2}(U) = \tfrac{1}{2}\pi^{-1}\cos^{-1}\rho_{12}; \\ w_{0,3}(U) = \tfrac{1}{4}\pi^{-1}(2\pi - \cos^{-1}\rho_{12} - \cos^{-1}\rho_{13} - \cos^{-1}\rho_{23}), \\ w_{1,3}(U) = \tfrac{1}{4}\pi^{-1}(3\pi - \cos^{-1}\rho_{12.3} - \cos^{-1}\rho_{13.2} - \cos^{-1}\rho_{23.1}), \\ w_{2,3}(U) = \tfrac{1}{2} - w_{0,3}(U), \quad w_{3,3}(U) = \tfrac{1}{2} - w_{1,3}(U), \end{gathered} \tag{4·9}$$

where $\rho_{ij.k} = (\rho_{ij} - \rho_{ik}\rho_{jk})(1-\rho_{ik}^2)^{-\frac{1}{2}}(1-\rho_{jk}^2)^{-\frac{1}{2}}$ is the conditional correlation between y_i and y_j given y_k.

For $q = 4$ the weight $w_{0,4}(U)$ is equal to the probability that all coordinates of y are nonnegative. Unfortunately there does not seem to exist a simple method for calculation of this probability. One may use a formula proposed by Kendall (1941) and summarized by Kudô (1963, formulae (3·45), (3·46)).

The rest of the weights are given by

$$\begin{gathered} w_{1,4}(U) = \tfrac{1}{8}\pi^{-1}\Big(8\pi - \sum_{i>j;\, i,j \neq k} \cos^{-1}\rho_{ij.k}\Big), \\ w_{2,4}(U) = \tfrac{1}{4}\pi^{-2} \sum_{i>j,\, k>l;\, k,l \neq i,j} (\cos^{-1}\rho_{ij})(\pi - \cos^{-1}\rho_{kl.ij}), \\ w_{3,4}(U) = \tfrac{1}{8}\pi^{-1}\Big(-4\pi + \sum_{i>j;\, i,j \neq k} \cos^{-1}\rho_{ij.k}\Big), \quad w_{4,4}(U) = \tfrac{1}{2} - w_{0,4}(U) - w_{2,4}(U), \end{gathered} \tag{4·10}$$

where $\rho_{kl.ij}$ denotes the conditional correlation between y_k and y_l given y_i and y_j. We remark that first two formulae in (4·10) essentially coincide with the corresponding ones of Kudô (1963, formula (3·47)), while the expression for $w_{3,4}$ is different. This expression is obtained from geometrical considerations and shows that $w_{1,4} + w_{3,4} = \frac{1}{2}$. It seems that this is a general property, i.e. weights on even places, as well as weights on odd places, always sum up to $\frac{1}{2}$.

Finally, if the cone $\mathscr{K}$ is contained in a $(q-l)$-dimensional linear space, then it has no faces of dimensionality greater than $q-l$. Consequently in this case the first l weights $w_0, \ldots, w_{l-1}$ vanish. Typically such a situation occurs when l equality constraints are present in the definition of the parameter space Θ. On the other hand it may happen that $\mathscr{K}$ contains a linear space. Then $\mathscr{K}$ is the direct sum of its linearity space $\mathscr{L}$ and a pointed cone $\bar{\mathscr{K}}$, where $\bar{\mathscr{K}} = \mathscr{L}^{\perp} \cap \mathscr{K}$ and $\mathscr{L}^{\perp}$ denotes the orthogonal complement of $\mathscr{L}$ (Stoer & Witzgall, 1970, Theorem 2·10·5). In this case $\mathscr{K}^0$ belongs to the linear space $\mathscr{L}^{\perp}$, and hence the weights $w_{q+1-d}, \ldots\ldots, w_q$, where $d = \dim \mathscr{L}$, vanish. The remaining weights $w_0, \ldots, w_{q-d}$ can be calculated as the weights corresponding to the cone $\bar{\mathscr{K}}$ in the linear space $\mathscr{L}^{\perp}$. Note that if Θ is defined as in (2·11) by inequality constraints only, then the linearity space $\mathscr{L}$ is $\mathscr{L} = \{\theta \colon a_i'\theta = 0,\ i \in \mathscr{I}(\theta_0)\}$, with $\dim \mathscr{L} > 0$ if the number of inequality constraints active at θ_0 is less than q.

5. Discussion and applications

The results of §§ 2 and 3 show that the asymptotic distribution of the test statistic $n\hat{F}$ is a mixture of chi-squared distributions. More specifically, let the parameter space Θ be defined by equality and inequality constraints as in (2·11) and suppose that precisely s inequality constraints are active at θ_0, so that the index set $\mathscr{I}(\theta_0)$ contains s elements.

Then under suitable regularity conditions and for $l+s \leqslant q$

$$\lim_{n\to\infty} \text{pr}(n\hat{F} \geqslant c^2) = \sum_{i=l}^{l+s} w_i \,\text{pr}(\bar{\chi}^2_{m-q+i} \geqslant c^2). \tag{5·1}$$

Formula (5·1) indicates that neither of the chi-squared distributions with $v = m-q+l$ and with $v+s$ degrees of freedom is the correct asymptotic distribution of $n\hat{F}$. The true asymptotic distribution rather lies between these two extreme cases.

Chi-squared distributions are well known so there is no problem in calculating the probabilities in the right-hand side of (5·1). Hence once the weights w_i are found the corresponding upper tail probabilities can be easily evaluated.

Let us consider as an example the factor analysis model (1·2). As we have mentioned the diagonal elements of Ψ must be nonnegative. Also $\frac{1}{2}k(k-1)$ equality constraints on the elements of Λ are needed to identify the model (Lawley & Maxwell, 1971). We suppose that the underlying distribution is multivariate normal and that the regularity conditions are met. Assumption 4 is ensured, for example, if one selects F to be the maximum likelihood or a 'best' generalized least-squares discrepancy function (Browne, 1974). If all diagonal elements of the matrix Ψ_0 are positive, then $\bar{\chi}^2$ is identically zero and we obtain the well-known result that $n\hat{F}$ has asymptotic chi-squared distribution with

$$v = \tfrac{1}{2}\{(p-k)^2-(p+k)\} \tag{5·2}$$

degrees of freedom (Lawley & Maxwell, 1971, p. 36).

Now let s diagonal elements of Ψ_0 be zero, so that s inequality constraints become active at Ψ_0. Then we have that

$$\lim_{n\to\infty} \text{pr}(n\hat{F} \geqslant c^2) = \sum_{i=0}^{s} w_{i,s} \,\text{pr}(\chi^2_{v+i} \geqslant c^2), \tag{5·3}$$

where the number v is given by (5·2). Expression (5·3) explains an experimental effect observed by Geweke & Singleton (1980) that '... when one or more specific factors are zero the factor model is rejected too often when the log-likelihood ratio statistic is used'.

The corresponding weights $w_{i,s}$ can be calculated as follows. The asymptotic covariance matrix of the unrestricted minimum discrepancy function estimators of the diagonal elements of Ψ_0 is equal to the inverse of the matrix $U = \frac{1}{2}W^{(2)}$, where

$$W = \Omega_0^{-1} - \Omega_0^{-1}\Lambda_0(\Lambda_0'\Omega_0^{-1}\Lambda_0)^{-1}\Lambda_0\Omega_0^{-1}$$

and $W^{(2)}$ denotes the matrix whose elements are the squares of the corresponding elements of W. Equivalently W is given by $W = E(E'\Omega_0 E)^{-1}E'$, where E is an orthogonal complement of Λ_0 (Lawley & Maxwell, 1971; Shapiro, 1983, p. 76). Denote by ρ_{ij} the (i,j)th element of the correlation matrix (4·8). Also suppose without loss of generality that the first s diagonal elements of Ψ_0 are zero. Then for $s \leqslant 4$ the weights $w_{i,s} = w_{i,s}(U)$ are given by formulae (4·7), (4·9), (4·10) respectively.

Consider an example analysed by Lawley & Maxwell (1971, pp. 44–5, Data 2) with $n = 810$, $p = 10$ and $k = 4$. Here $v = 11$ and the test statistic $n\hat{F}$ has value of 18·5. In this example the maximum likelihood approach leads to a boundary solution, Heywood case, with $s = 1$. Assuming that the population matrix Ψ_0 has one zero diagonal element we obtain, with $w_{0,1} = w_{1,1} = \frac{1}{2}$,

$$\text{pr}(n\hat{F} \geqslant 18{\cdot}5) \simeq \tfrac{1}{2}\text{pr}(\chi^2_{11} \geqslant 18{\cdot}5) + \tfrac{1}{2}\text{pr}(\chi^2_{12} \geqslant 18{\cdot}5) = 0{\cdot}085.$$

Another example, where developed techniques may prove to be useful, occurs in equating tests of ability using the single factor analysis model (Rock, 1982). There the factor loadings should be required to be nonnegative (Gleser, 1982, p. 264). Consequently if some of the population values of factor loadings are zero, or near zero, then the asymptotic distribution of $n\hat{F}$ is a mixture of chi-squared distributions. Once the covariance matrix U is calculated the corresponding weights can be found by formulae (4·9) and (4·10) as in the previous example.

References

Andersen, E. B. (1981). Discussion of paper by K. G. Jöreskog. *Scand. J. Statist.* **8**, 83–4.

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1**, 135–41.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. New York: Wiley.

Bartholomew, D. J. (1961). Ordered tests in the analysis of variance. *Biometrika* **48**, 325–32.

Bazaraa, M. S., Goode, J. J. & Shetty, C. M. (1972). Constraint qualification revisited. *Manag. Sci.* **28**, 567–73.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *S. Afr. Statist. J.* **8**, 1–24.

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.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573–8.

David, F. N. (1953). A note on the evaluation of the multivariate normal integral. *Biometrika* **40**, 458–9.

Dijkstra, T. (1983). Some comments on maximum likelihood and partial least squares methods. *J. Econometrics* **22**, 67–90.

Dykstra, R. L. & Robertson, T. (1982). Order restricted statistical tests on multinomial and Poisson parameters: the starshaped restriction. *Ann. Statist.* **10**, 1246–52.

Geweke, J. F. & Singleton, K. J. (1980). Interpreting the likelihood ratio statistic in factor models when sample size is small. *J. Am. Statist. Assoc.* **75**, 133–7.

Gleser, L. J. (1982). Discussion of 'Equating using the confirmatory factor analysis'. In *Test Equating*, Ed. P. W. Holland and D. B. Rubin, pp. 259–70. Princeton, New Jersey: Academic Press.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32**, 443–82.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika* **57**, 239–51.

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

Kendall, M. G. (1941). Proof of relations connected with the tetrachoric series and its generalizations. *Biometrika* **32**, 196–8.

Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–18.

Kudô, A. & Choi, J. R. (1975). A generalized multivariate analogue of the one sided test. *Mem. Fac. Sci.*, Kyushu Univ., A., **29**, 303–28.

Lawley, D. N. & Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method*. New York: American Elsevier.

Mangasarian, O. L. & Fromovitz, S. (1967). The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Applic.* **17**, 37–47.

Nüesch, P. E. (1966). On the problem of testing location in multivariate populations for restricted alternatives. *Ann. Math. Statist.* **37**, 113–9.

Perlman, M. D. (1969). One sided problems in multivariate analysis. *Ann. Math. Statist.* **40**, 549–67.

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

Rock, D. A. (1982). Equating using the confirmatory factor analysis model. In *Test Equating*, Ed. P. W. Holland and D. B. Rubin, pp. 247–57. Princeton, New Jersey: Academic Press.

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *S. Afr. Statist. J.* **17**, 33–81.

Shapiro, A. (1984a). A note on the consistency of estimators in the analysis of moment structures. *Br. J. Math. Statist. Psychol.*, **37**, 84–8.

Shapiro, A. (1984b). Sensitivity analysis and second-order optimality conditions in minimax optimization. In *Abstracts of Workshop on "Nondifferentiable Optimization"*, Ed. V. F. Demyanov, pp. 153–7. Laxenburg, Austria: I.I.A.S.A.

SHORACK, G. R. (1967). Testing against ordered alternatives in model I analysis of variance; normal theory and nonparameteric. *Ann. Math. Statist.* **38**, 1740–53.

STOER, J. & WITZGALL, C. (1970). *Convexity and Optimization in Finite Dimensions*, **I**. Berlin: Springer-Verlag.

SWAIN, A. J. (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika* **40**, 315–35.

WYNN, H. P. (1975). Integrals for one-sided confidence bounds: A general result. *Biometrika* **62**, 393–6.

[*Received November* 1983. *Revised October* 1984]