On a Separation Theorem for Generalized Eigenvalues and a Problem in the Analysis of Sample Surveys

Alastair J. Scott
Department of Mathematics and Statistics
The University of Auckland
Private Bag
Auckland, New Zealand

and

George P. H. Styan*
Department of Mathematics and Statistics
McGill University
805 ouest, rue Sherbrooke
Montréal, Québec, Canada H3A 2K6

Submitted by C. R. Rao

ABSTRACT

We obtain usable bounds for the asymptotic percentage points of chi-squared tests of fit for log-linear models fitted to contingency tables estimated from survey data, by applying some new separation inequalities for the generalized eigenvalues of a matrix $X'AX$ with respect to a matrix $X'BX$, when both the matrices $A$ and $B$ are nonnegative definite. We also present some historical remarks on the Poincaré separation theorem for eigenvalues from which our new inequalities are shown to follow.

1. INTRODUCTION AND SOME PRELIMINARY RESULTS

Our main purpose in this paper is to obtain bounds (Theorem 3) for the asymptotic percentage points of chi-squared tests of fit for log-linear models fitted to tables of counts estimated from survey data. These bounds are obtained by applying some new separation inequalities (Theorem 2) for the

*On sabbatical leave 1984/1985 at The University of Auckland.
generalized eigenvalues of a matrix $X'AX$ with respect to a matrix $X'BX$, when both the matrices $A$ and $B$ are assumed to be nonnegative definite, symmetric and real; we suppose, in addition, that the column space of $A$ is contained in the column space of $B$. We also present some historical remarks on the Poincaré separation theorem (Theorem 1) for eigenvalues from which our new inequalities are shown to follow.

We begin with some preliminary results.

**Lemma 1.** Let the matrices $A$ and $B$ both have $n$ columns. If any one of the following three conditions holds, then all three hold:

1. $\mathcal{R}(A) \subset \mathcal{R}(B)$,
2. $\mathcal{N}(B) \subset \mathcal{N}(A)$,
3. $A(I - B^{-}B) = 0$,

where $\mathcal{R}(\cdot)$ denotes row space, $\mathcal{N}(\cdot)$ null space, and $B^{-}$ is any generalized inverse of $B$ satisfying $BB^{-}B = B$. If (3) holds for a particular generalized inverse $B^{-}$, then (3) holds for every generalized inverse $B^{-}$.

We omit the proof which is straightforward and given in our technical report [34, Lemma 1].

We will define the scalar $\lambda$ to be a **generalized eigenvalue** of the matrix $A$ with respect to the matrix $B$ whenever

$$Ax = \lambda Bx$$

for some nonnull vector $x$ that does not belong to both $\mathcal{N}(A)$ and $\mathcal{N}(B)$.

As the scalar $\mu$ varies over the whole real line, the matrix $A - \mu B$ is called a **matrix pencil**—see e.g. Gantmacher [14, Chapter 12 (Chapter 2 in the Interscience edition)]—and the **pencil rank** $p$ may be defined as the order of the largest minor that does not vanish identically in $\mu$. There are then $p$ generalized eigenvalues $\lambda = \mu$ that satisfy (4), and following Mitra and C. R. Rao [27, Section 4] we will call these **proper**. We will augment these $p$ proper generalized eigenvalues with $n - p$ zeros (called **improper** generalized eigenvalues), so that there are in all $n$ generalized eigenvalues of the $n \times n$ matrix $A$ with respect to the $n \times n$ matrix $B$. We note that $n - p$ is equal to the dimension of the intersection of the null spaces $\mathcal{N}(A)$ and $\mathcal{N}(B)$, so that $p = n - \text{dim}[\mathcal{N}(A) \cap \mathcal{N}(B)]$. 
From Theorems 4.1 and 4.2 in Mitra and C. R. Rao [27] we obtain the following:

**Lemma 2.** Let the matrices $A$ and $B$ both be real, symmetric, and $n \times n$, and suppose that the matrix $B$ is nonnegative definite and that

$$H = I - B^{-1}B.$$  \hspace{1cm} (5)

Let

$$\text{rank}(H'AH) = \text{rank}(AH).$$  \hspace{1cm} (6)

Then the eigenvalues of $B^{-1}[A - AH(H'AH)^{-1}H'A]$ are all real and do not depend on the choices of generalized inverses, and

(a) are all real and do not depend on the choices of generalized inverses, and

(b) are precisely the generalized eigenvalues of $A$ with respect to $B$.

Moreover

$$\text{In}\left\{B^{-1}[A - AH(H'AH)^{-1}H'A]\right\} = \text{In}\left[A - AH(H'AH)^{-1}H'A\right],$$  \hspace{1cm} (7)

where the inertia $\text{In}$ is defined by the ordered triple $\{\pi, \eta, \nu\}$, in which $\pi$ is the number of positive eigenvalues, $\eta$ is the number of negative eigenvalues, and $\nu$ the number of zero eigenvalues (or the nullity).

The condition (6) may also be written as

$$\text{rank}\begin{pmatrix} A & B \\ B & 0 \end{pmatrix} = \text{rank}(A,B) + \text{rank}(B)$$  \hspace{1cm} (8)

using, e.g., Theorem 19 of [24]. The partitioned matrix in (8) is a variation of the "fundamental bordered matrix of linear estimation" extensively considered by C. R. Rao [29, 2nd ed., Section 4i.1]; our Equation (8) is Equation (4i.1.21) in [29, 2nd ed., p. 296]. A sufficient condition for Equation (8), or for Equation (6), to hold is that the matrix $A$ is nonnegative definite.

The inertia formula (7) may be established by choosing $B^{-1} = L^2$, where $L$ is positive definite and symmetric; the result then follows using Sylvester’s law of inertia [26, p. 377].

**Lemma 3.** Let the matrices $A$ and $B$ both be real, symmetric, and $n \times n$, with $\mathcal{C}(A) \subset \mathcal{C}(B)$, where $\mathcal{C}(\cdot)$ denotes column space. Suppose, further-
more, that \( B \) is nonnegative definite, and let \( X \) be an \( n \times k \) real matrix. Then

(i) it follows that

\[
\mathcal{C}(X'AX) \subset \mathcal{C}(X'BX);
\]  

(ii) the eigenvalues of \((X'BX)^{-} X'AX\)

(a) are all real and do not depend on the choice of generalized inverse, and

(b) are precisely the generalized eigenvalues of \( X'AX \) with respect to \( X'BX \);

and

(iii) furthermore

\[
\text{In}
\left[
(X'BX)^{-} X'AX
\right] = \text{In}(X'AX).
\]  

where the inertia \( \text{In} \) is defined as in Lemma 2.

Proof. (i): From Lemma 1 it suffices to show that

\[
X'AX(X'BX)^{-}X'BX = X'AX.
\]  

Since \( B \) is nonnegative definite, it follows that \( \text{rank}(X'BX) = \text{rank}(BX) \), and so using the rank cancellation rule [24, Theorem 2] we find that

\[
X'BX(X'BX)^{-}X'BX = X'BX \Rightarrow BX(X'BX)^{-}X'BX = BX.
\]  

Premultiplying (12) by \( X'AB^{-} \) yields (11), since \( AB^{-}B - A \leftrightarrow \mathcal{C}(A) \subset \mathcal{C}(B) \) from Lemma 1. This proves (i).

(ii) and (iii): Since \( \mathcal{C}(X'AX) \subset \mathcal{C}(X'BX) \), it follows from Lemma 1 that \( \mathcal{M}(X'BX) \subset \mathcal{M}(X'AX) \) and so

\[
\text{rank}(H'X'AXH) = \text{rank}(X'AXH) = 0,
\]  

where \( H = I - (X'BX)^{-}X'BX \). Thus (6) holds (with \( A \) replaced by \( X'AX \)) and so (ii) and (iii) follow at once from Lemma 2 (with \( B \) replaced by \( X'BX \)).
2. TWO EIGENVALUE-SEPARATION THEOREMS

We now present the so-called Poincaré separation theorem (Theorem 1) for eigenvalues. Our version is based on that given by Mäkeläinen in [23, Theorem 4.1; Corollary 4.2.2] and by C. R. Rao in [31, Theorem 2.11], which are the only references that we have found where the characterizations for equality on the left and on the right of (14) are established. See also [30] for some closely related results.

**Theorem 1.** Let $A$ be a real symmetric $n \times n$ matrix, and let $F_1$ be a real $n \times p$ matrix such that $F_1'F_1 = I_p$. Then

$$
\chi_{n-p+i}(A) \leq \chi_i(F_1'AF_1) \leq \chi_i(A), \quad i = 1, \ldots, p,
$$

(14)

where $\chi_i(\cdot)$ denotes the $i$th largest eigenvalue.

Equality holds on the left of (14) simultaneously for all $i = 1, \ldots, p$ if and only if there exists a real $n \times p$ matrix $P_0$ such that

$$
P_0'P_0 = I_p, \quad AP_0 = P_0A_0, \quad \text{and} \quad \mathcal{C}(P_0) = \mathcal{C}(F_1),
$$

(15)

where $A_0$ is a $p \times p$ diagonal matrix containing the $p$ smallest eigenvalues of $A$.

Equality holds on the right of (14) simultaneously for all $i = 1, \ldots, p$ if and only if there exists a real $n \times p$ matrix $P_1$ such that

$$
P_1'P_1 = I_p, \quad AP_1 = P_1A_1, \quad \text{and} \quad \mathcal{C}(P_1) = \mathcal{C}(F_1),
$$

(16)

where $A_1$ is a $p \times p$ diagonal matrix containing the $p$ largest eigenvalues of $A$.

The inequalities (14) have been named after [Jules] Henri Poincaré (1854–1912) in view of his 1890 paper [28, pp. 259–260 (pp. 78–79 in the Œuvres version)] where he (apparently only?) obtained the right-hand side of (14) (and apparently only?) for \( i = p \).

The first complete treatment, however, of the separation inequalities (14) for eigenvalues appears to be in 1922 by Richard Courant (1888–1972), who considered [8] the vibration frequencies (Schwingungszahlen) of an oscillating mechanical system restricted by linear constraints. See also Courant and Hilbert [9, 1st ed., p. 17; 2nd ed., p. 28; 10, p. 331 and Julia [21, pp. 199–200] for other early treatments of (14).

The first application of the separation inequalities (14) to statistics is by Durbin and Watson [11, pp. 415, (177)], who in 1950 obtained bounds for the eigenvalues of the matrix product \( AB \), where \( A \) is nonnegative definite and \( B \) is symmetric idempotent [of rank \( p \) say, and so can be expressed as \( F_i F_i' \) (with \( F_i' F_i = I_p \)), whence \( \chi_i(AB) = \chi_i(F_i' F_i') \) for \( i = 1, \ldots, p \) when \( A \) is nonnegative definite]. The well-known Durbin-Watson bounds test for serial correlation in regression analysis is based on these separation inequalities. Extensions to the eigenvalues of \( AB \), where \( A \) is symmetric but not necessarily nonnegative definite and \( B \) is symmetric and nonnegative definite but not necessarily idempotent, were obtained in considerable detail by Mäkeläinen [23] in 1970.

**Theorem 2.** Let the matrices \( A \) and \( B \) both be nonnegative definite, real, symmetric, and \( n \times n \), with \( \mathcal{C}(A) \subset \mathcal{C}(B) \), where \( \mathcal{C}(\cdot) \) denotes column space, and let \( X \) be an \( n \times k \) real matrix, with

\[
b = \text{rank}(B) \quad \text{and} \quad r = \text{rank}(BX).
\]

Then

\[
\chi_{b-r+i}(B \ A) \leq \chi_i\left[(X'BX)^{-1}X'AX\right] \leq \chi_i(B^{-1}A), \quad i = 1, \ldots, r,
\]

where \( \chi_i(\cdot) \) denotes the \( i \)th largest eigenvalue. In (18) any choices of generalized inverses \( B^{-1} \) and \( (X'BX)^{-1} \) may be made.

Equality holds on the left of (18) simultaneously for all \( i = 1, \ldots, r \) if and only if there exists a real \( n \times r \) matrix \( Q_0 \) such that

\[
Q_0BQ_0 = I_r, \quad AQ_0 = BQ_0A_0, \quad \text{and} \quad \mathcal{C}(BQ_0) = \mathcal{C}(BX),
\]

where

\[
\chi_i(X) = \chi_i(AX).
\]
where $\Lambda_0$ is an $r \times r$ diagonal matrix containing the $r$ smallest proper generalized eigenvalues of $A$ with respect to $B$.

Equality holds on the right of (18) simultaneously for all $i = 1, \ldots, r$ if and only if there exists a real $r \times r$ matrix $Q_1$ such that

$$Q_1^* B Q_1 = I_r, \quad A Q_1 = B Q_1 \Lambda_1, \quad \text{and} \quad \mathcal{C}(B Q_1) = \mathcal{C}(B X),$$

where $\Lambda_1$ is an $r \times r$ diagonal matrix containing the $r$ largest (proper) generalized eigenvalues of $A$ with respect to $B$.

Proof. Since the $n \times n$ matrix $B$ is nonnegative definite with rank equal to $b$, there exists an $n \times b$ matrix $G$, say, such that

$$B = GG',$$

where

$$b = \text{rank}(G) = \text{rank}(B).$$

Since $\mathcal{C}(A) \subset \mathcal{C}(B)$, it follows from Lemma 1 that we may write

$$A = GG^{-} A(GG^{-})' = GEG',$$

say, where the $b \times b$ matrix

$$E = G^{-} A(G^{-})'$$

and $G^{-}$ is any choice of generalized inverse of $G$. Then

$$(X'BX)^{-} X'AX = (X'CG'X)^{-} X'GEG'X$$

$$= Q_r D_r^{-1} P_r' E P_r D_r Q_r',$$

where $G'X = P_r D_r Q_r'$ is a singular value decomposition with $P_r$ and $Q_r$ of full column rank $r$ and $D_r$ nonsingular, diagonal, and $r \times r$. In (25) we have chosen $(X'BX)^{-} = (X'BX)^{+} = Q_r D_r^{-2} Q_r$, the Moore-Penrose generalized inverse of $X'BX$.

Then by moving $D_r Q_r'$ in (25) to the front we find that

$$\text{ch}_i[(X'BX)^{-} X'AX] = \text{ch}_i(P_r' E P_r), \quad i = 1, \ldots, r,$$
since $E$ is nonnegative definite; cf. (24). The eigenvalues of the $b \times b$ matrix $E = G'A(G')'$ are all also eigenvalues of the $n \times n$ matrix

$$\begin{align*}
(G^-)'G^-A &= (GG')^-A = B^-A, \\
&= (G^-)'G^-A = (GG')^-A = B^-A,
\end{align*}$$

(27)

since $G^-G = I_b$. Thus (18) follows at once from (14) in Theorem 1.

Using (15) and (26), we see that equality holds on the left of (18) if and only if there exists a real $b \times r$ matrix $P_0$, say, such that

$$P_0'P_0 = I_r,$$

(28)

$$EP_0 = P_0\Lambda_0,$$

(29)

and

$$\varphi(P_0) = \varphi(P_r),$$

(30)

where $\Lambda_0$ is an $r \times r$ diagonal matrix containing the $r$ smallest eigenvalues of the $b \times b$ matrix $E$. Let

$$Q_0 = (G^-)'P_0.$$  

(31)

Then

$$AQ_0 = A(G^-)'P_0 = GG^-A(G^-)'P_0 = GEP_0 = GP_0\Lambda_0,$$

(32)

using (24) and (29). Since $G$ has full column rank equal to $b$, it follows that $G^-G = I_b$ and so

$$G'Q_0 = P_0;$$

(33)

thus $BQ_0 = GG'Q_0 = GP_0$, and we see that (29) is equivalent to the second equation in (19). The third equation in (19) is equivalent to (30), since

$$\varphi(P_0) = \varphi(P_r) \iff \varphi(G'Q_0) = \varphi(G'X) \iff \varphi(BQ_0) = \varphi(BX).$$

(34)

Moreover, using (33) we may write (28) as $I_r = P_0'P_0 = Q_0'GG'Q_0 = Q_0'BQ_0$, which is the first equation in (19). The eigenvalues of the $b \times b$ matrix
E = G^−A(G^−)' are also eigenvalues of the n × n matrix B^−A [cf. (27)]; the b eigenvalues of B^−A that are not necessarily zero are the \textit{proper} generalized eigenvalues of A with respect to B, and the r smallest of these are also the r smallest of the eigenvalues of E.

We have therefore characterized equality on the left of (18). The condition (20) characterizing equality on the right of (18) is proved similarly.

When the matrix product BX in (18) has full column rank

\[ r = \text{rank}(BX) = k. \]  

then from (26) we see that the \( k = r \) eigenvalues of \((X'BX)^-X'AX\) are precisely the \( r \) eigenvalues of \( P_r'E_P \), and so (26) holds for all symmetric E (nonnegative definite, indefinite, or nonpositive definite); it follows, therefore, that when BX has full column rank, then Theorem 2 remains valid for all symmetric matrices A, not necessarily nonnegative definite.

When BX has less than full rank, so that

\[ r = \text{rank}(BX) < k \]  

then (18) does not necessarily hold when A is not nonnegative definite. For example let \( n = 3, k = 2, A = -I_3, B = I_3 \), and

\[ X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \]  

so that \( r = 1 \). Then \( B^{-A} = B^{-1}A = -I_3 \), while

\[ (X'BX)^-X'AX = -(X'X)^-X'X = \begin{pmatrix} -1 & 0 \\ x & 0 \end{pmatrix} \]  

for some scalar \( x \). Hence

\[ 0 = \text{ch}_1[(X'BX)^-X'AX] > \text{ch}_1(B^−A) = -1, \]  

contradicting the right inequality in (18) with \( i = r = 1 \).

The special case of Theorem 2 when \( r = k = 1 \) has been considered by McDonald, Torii, and Nishisato [25], who also characterized all the stationary values of \((X'BX)^-X'AX = x'Ax/x'Bx\), where the \( n \times 1 \) vector \( x = X \).
3. AN APPLICATION TO THE ANALYSIS OF SAMPLE SURVEYS

Methods using log-linear models for analyzing multiway tables of categorical data arising from sample surveys have been developed extensively in the last two decades. Very good accounts of the general theory and methodology are given by Haberman [16] and by Bishop, Fienberg, and Holland [4].

Let \( \pi = \{ \pi_t \} \) denote the \( T \times 1 \) vector of population cell proportions when the cells of the multiway table are ordered in some way. We assume that \( \pi_t > 0 \) for all \( t = 1, \ldots, T \). Then a log-linear model for \( \pi \) takes the form

\[
\mu = u(\theta)e + X\theta,
\]

where \( \mu \) is the \( T \times 1 \) vector of log-proportions with \( t \)th component \( \mu_t = \log \pi_t \), \( (t = 1, \ldots, T) \), \( e \) denotes the \( T \times 1 \) vector of ones, \( X \) is a \( T \times p \) matrix of known constants with full rank \( p \leq T - 1 \) and with \( X'e = 0 \), \( \theta \) is a \( p \times 1 \) vector of unknown constants, and \( u(\theta) \) is a normalizing (scalar) constant chosen so that \( \sum_{t} \pi_t = 1 \). In the simplest case, a random sample of \( n \) observations is drawn (with replacement) from the population and the results classified according to the cells of the table. Let \( \hat{\pi} \) denote the resulting \( T \times 1 \) vector of observed proportions. It can be shown [16] that the maximum likelihood estimator of \( \pi \) is the unique solution \( \hat{\pi} \) to

\[
X'\hat{\pi} = X'\hat{\theta}
\]

satisfying [cf. (40)]

\[
\hat{\mu} = \hat{u}(\theta)e + X\theta
\]

for some \( \theta \) with \( \hat{u}(\theta) \) chosen so that \( \sum_{t} \pi_t = 1 \). For some special models, \( \hat{\pi} \) can be found explicitly, and very efficient algorithms are available [13] for calculating the solution \( \hat{\pi} \) to (41) and (42) in general.

A major aim in the analysis is to obtain as parsimonious a representation for \( \hat{\pi} \) as possible. Thus we are led to consider nested models of the form

\[
\mu = u_1(\theta_1)e + X_1\theta_1
\]

with

\[
\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \quad \text{and} \quad X = (X_1, X_2),
\]
A SEPARATION THEOREM FOR EIGENVALUES

where $\theta_1$ is $p_1 \times 1$, $\theta_2$ is $p_2 \times 1$, and correspondingly $X_1$ is $T \times p_1$ and $X_2$ is $T \times p_2$ ($p_1 + p_2 = p$). In other words, we are interested in testing the hypothesis $H_0: \theta_2 = 0$ in (40). There are two standard tests, one based on the Pearson chi-squared statistic

$$\chi^2 = n \sum_t \frac{\left( \hat{\pi}_t - \hat{\pi}_t^* \right)^2}{\hat{\pi}_t},$$

(44)

where $\hat{\pi}_t^*$ is the maximum likelihood estimate of $\pi$ under the restricted model (43), and the other one the likelihood ratio statistic

$$\mathcal{G}^2 = 2n \sum_t \hat{\pi}_t \log \left( \frac{\hat{\pi}_t}{\hat{\pi}_t^*} \right)$$

$$= 2n \sum_t \hat{\pi}_t \log \left( \frac{\hat{\pi}_t}{\hat{\pi}_t^*} \right) - 2n \sum_t \hat{\pi}_t \log \left( \frac{\hat{\pi}_t}{\pi_t^*} \right).$$

(45)

The statistics $\chi^2$ and $\mathcal{G}^2$ are asymptotically equivalent under mild regularity conditions [4, Section 14.8.1], and both have asymptotic central $\chi^2$ distributions with $p_2$ degrees of freedom under the null hypothesis $H_0$.

These methods have proved extremely fruitful in making sense of complex interrelationships (see the paper [19] by Imrey, Koch, and Stokes for an extensive bibliography of successful applications), and researchers in many disciplines, particularly in the social and health sciences, have tended to use the same methods (without modification!) to analyze data from more complex sample surveys; see e.g. [5, 7]. All large-scale surveys, however, involve some kind of stratification and multistage sampling where clusters of linked units are drawn together. This means that the assumption of complete independence underlying the standard theory is very far from true, and there has been considerable interest over the last few years in the effect of these violations; see e.g. [1, 12, 18, 32, 33].

Let $\hat{\pi}_1$ denote the $T \times 1$ vector of estimated proportions obtained from a sample of $n$ units, which are now no longer all drawn independently. This estimator $\hat{\pi}_1$ may be extremely complicated, involving ratio estimation and incorporating survey design weights, for example. We assume that $\hat{\pi}_1$ is a consistent estimator of $\pi$ and that a central limit theorem [32] is available for the specified combination of design and estimator so that $\sqrt{n}(\hat{\pi}_1 - \pi)$ converges in distribution to a $T$-variate normal random vector with mean vector $0$ and covariance matrix $A$, say. We note that $Ae = 0$ (since $e^T\pi = 1$), so that $A$ is singular (positive semidefinite). In the case of independent sampling $A$
reduces to the multinomial covariance matrix \( \text{diag}(\pi) - \pi \pi' = B \), say. It is impossible to write down the likelihood function for a general survey design, so it is standard practice to use a pseudo maximum likelihood estimate, \( \hat{\pi}_1 \) say, obtained by using \( \hat{\beta}_1 \) instead of \( \hat{\beta} \) in (41). The consistency of \( \hat{\beta}_1 \) guarantees that of \( \hat{\pi}_1 \), and it can be shown [20, 33] that \( \sqrt{n}(\hat{\pi}_1 - \pi) \) is asymptotically \( T \)-variate normal with mean vector \( 0 \) and covariance matrix

\[
BX(X'BX)^{-1}X'AX(X'BX)^{-1}X'B
\]

under the regularity conditions as given in [4, Theorem 14.8.1]. We note that \( X'BX \) is nonsingular, since the null space of \( B \) is spanned by \( e \) and \( X'e = 0 \). Asymptotically valid alternatives to the statistics \( X^2 \) and \( g^2 \) for testing the adequacy of the submodel (41) can be constructed if a consistent estimator of the covariance matrix \( A \) is available [22]. Few published tables, however, give such matrices, since the space required would be prohibitive for a table of any size. Partly for this reason (but perhaps more because package programs to implement the standard theory are so widely available and easy to use), many practitioners use the multinomial-based \( X^2 \) or \( g^2 \) tests given by (44) and (45). The asymptotic distribution of \( X^2 \) and \( g^2 \) for a complex design has been investigated by J. N. K. Rao and Scott [33]; they show that \( X^2 \) and \( g^2 \) are again asymptotically equivalent and are asymptotically distributed as

\[
\sum_{i=1}^{p_2} \delta_i Z_i^2
\]

under the model (42), where \( Z_1, \ldots, Z_{p_2} \) are independent \( N(0,1) \) random variables,

\[
\delta_i = \text{ch}_i(\tilde{X}^2_i B \tilde{X}^2_2)^{-1} \tilde{X}^2_i A \tilde{X}^2_2,
\]

and

\[
\tilde{X}^2_2 = X_2 - X_1(X_1'BX_1)^{-1}X_1'BX_2.
\]

We note that \( \tilde{X}^2_2 B \tilde{X}^2_2 \) is nonsingular, since [24, Theorem 5]

\[
\text{rank}(X_1, X_2) = \text{rank}(X_1) + \text{rank}[(I - X_1 X_1')X_2]
= \text{rank}(X_1) + \text{rank}(\tilde{X}_2),
\]

choosing \( X^-_1 = (X_1'BX_1)^{-1}X_1'B \).
If we could obtain estimates of the eigenvalues $\delta_1, \ldots, \delta_{p_2}$ in (47), then it would be straightforward to find reasonably accurate percentage points for the true asymptotic distribution of $X^2$ (or $G^2$) using, for example, the approximations in Solomon and Stephens [35]. We note that the $\delta_i$'s depend on the hypothesis being tested as well as on the true covariance matrix, and it would require complete specification of $A$ to allow computation of the $\delta_i$'s for every possible choice of $X_1$ and $X_2$. Publication of an estimate of the full covariance matrix is simply not feasible for a table of any substantial size; no survey organization in the world currently publishes such estimates for anything beyond a $2 \times 2$ table. Consequently a great deal of effort in recent years has gone into attempts to find reasonable approximations or bounds for the $\delta_i$'s based on partial information about the cell covariances; see, e.g., Bedrick [2] and Gross [15]. The main result of this paper is that we can use Theorem 2 to obtain bounds on the asymptotic distribution of $X^2$ (or $G^2$) which do not depend on $X_1$ or $X_2$.

**Lemma 4.** Let the $T \times T$ matrices $A$ and $B$ be defined as in the discussion leading to (46). Then the eigenvalues of $B - A$ do not depend on the choice of generalized inverse $B^-$, and if $\lambda_i$ is the $i$th largest such eigenvalue, it follows that

$$\lambda_{T - 1 - p_2 + 1} \leq \delta_i \leq \lambda_i, \quad i = 1, \ldots, p_2,$$

where $\delta_i$ is as defined by (48).

**Proof.** From the discussion leading to (46) we note that $\mathcal{N}(B) = \mathcal{C}(e) \subset \mathcal{N}(A)$, and so by Lemma 1 and symmetry it follows that $\mathcal{C}(A) \subset \mathcal{C}(B)$, and hence by Lemma 3(ii) the eigenvalues of $B - A$ do not depend on the choice of generalized inverse $B^-$. From (50) we note that $\text{rank}(BX_2) = p_2$, and since $\mathcal{N}(B) = \mathcal{C}(e)$, we see that (51) follows at once from (18).

A particularly convenient choice for the generalized inverse $B^-$ of the matrix $B = \text{diag}(\pi) - \pi \pi'$ in Lemma 4 is $[\text{diag}(\pi)]^{-1}$.

**Theorem 3.** Let the percentage points $c_L$, $c_U$, and $c_a$ be defined by

$$P \left( \sum_{i=1}^{p_2} \lambda_{T - 1 - p_2 + i} Z_i^2 > c_L \right) = P \left( \sum_{i=1}^{p_2} \lambda_i Z_i^2 > c_U \right) = P \left( \sum_{i=1}^{p_2} \delta_i Z_i^2 > c_a \right) = \alpha,$$

as defined by (52).
where $\lambda_i$ is the $i$th largest eigenvalue of $[\text{diag}(\pi)]^{-1}A$, the $\delta_i$ are the eigenvalues defined by (48), and $Z_1, \ldots, Z_{p_2}$ are independent $N(0,1)$ random variables. Then

$$c_L \leq c_\alpha \leq c_U. \quad (53)$$

Of course, we still require estimates of the eigenvalues of $[\text{diag}(\pi)]^{-1}A$ to use the bounds (53). For this, however, we need to specify only $T - 1$ numbers rather than the $\frac{1}{2}T(T + 1)$ necessary for the full covariance matrix $A$. This should be feasible at least for important surveys, the results of which are likely to be widely analyzed.

We are indebted to an anonymous referee for drawing our attention to the paper by Mitra and C. R. Rao [27] and for pointing out the factorization (23); our proofs of Lemma 3 and Theorem 2 have thereby been substantially shortened over those presented in the original version [34] of this paper.

The first author wishes to thank Professor J. N. K. Rao for arranging a series of visits to Carleton University and for many stimulating discussions there on the analysis of categorical survey data. These visits were supported by the Natural Sciences and Engineering Research Council of Canada.

The second author is most grateful to Professor C. Radhakrishna Rao for his invitation to the Indian Statistical Institute in New Delhi during February and March 1976, and for the very helpful discussions during this visit concerning the Poincaré separation theorem and its history.

The research of the second author was supported in part by the Natural Sciences and Engineering Research Council of Canada, by the Social Sciences and Humanities Research Council of Canada, and by the Gouvernement du Québec, Programme de formation de chercheurs et d'action concertée, while he was on sabbatical leave 1984–85 in the Department of Mathematics and Statistics of The University of Auckland, Auckland, New Zealand.

REFERENCES

A SEPARATION THEOREM FOR EIGENVALUES 223


Received 16 January 1985; revised 13 May 1985