The Annals of Statistics
1981, Vol. 9, No. 6, 1178–1186

TESTS FOR INDEPENDENCE IN TWO-WAY CONTINGENCY TABLES BASED ON CANONICAL CORRELATION AND ON LINEAR-BY-LINEAR INTERACTION[1]

BY SHELBY J. HABERMAN

University of Chicago

Tests for independence of rows and columns in an $r \times s$ contingency table are developed from canonical correlation analysis and from models of linear-by-linear interaction. The resulting test statistics are asymptotically equivalent under the null hypothesis. They are consistent and asymptotically unbiased. Approximate critical values are available from existing tables. The proposed tests are most appropriate when the matrix of joint probabilities is well approximated by a matrix of rank 2. Against some alternatives which may arise in such tables, the proposed statistics have greater asymptotic power than conventional chi-square tests of independence.

1. Introduction. Both the canonical correlation analysis of Fisher (1939) and the models of linear-by-linear interaction of Andersen (1980) and Goodman (1979) can be employed to test for independence of the row and column variables in an r by s contingency table. As shown in this paper, the resulting test statistics from these approaches are all asymptoticaly equivalent under the null hypothesis, and approximate critical values for these statistics may be obtained from Pearson and Hartley (1972). These test statistics all have the same asymptotic power, and they share the properties of asymptotic unbiasedness and consistency with the more familiar Pearson and likelihood-ratio chi-square tests for independence. As is evident from results of Sections 4, 5, and 6, the proposed test statistics have neither uniformly greater nor uniformly smaller asymptotic power than conventional chi-square tests. Based on a result of Section 1, the new statistics appear most useful when the r by s matrix of all probabilities is well approximated by a matrix of rank 2. Thus the new statistics work best if the data are well approximated by a latent-class model with two latent classes or a model of linear-by-linear interaction.

2. The canonical correlation test of independence. Consider an $r \times s$ contingency table with frequencies n_{ij}, $1 \le i \le r$, $1 \le j \le s$, $r \ge 2$, $s \ge 2$. Let the $n_{i\cdot} = \sum_{j=1}^{s} n_{ij}$, $1 \le i \le r$, be the row marginal totals, and let the $n_{\cdot j} = \sum_{j=1}^{r} n_{ij}$, $1 \le j \le s$, be the column marginal totals. Assume that the n_{ij} have a multinomial distribution with sample size N and with cell probabilities $p_{ij} > 0$, $1 \le i \le r$, $1 \le j \le s$. Let $f_{ij} = N^{-1}n_{ij}$, $f_{i\cdot} = N^{-1}n_{i\cdot}$, and $f_{\cdot j} = N^{-1}n_{\cdot j}$ denote the relative frequencies corresponding respectively to n_{ij}, $n_{i\cdot}$, and $n_{\cdot j}$. The canonical correlation test is a test of the independence hypothesis

$$p_{ij} = p_{i\cdot}p_{\cdot j}, \qquad 1 \le i \le r, \qquad 1 \le j \le s, \tag{2.1}$$

where the $p_{i\cdot} = \sum_{j=1}^{s} p_{ij}$, $1 \le i \le r$, are row marginal probabilities and the $p_{\cdot j} = \sum_{i=1}^{r} p_{ij}$, $1 \le j \le s$, are column marginal probabilities. The test is appropriate when row and column categories are ordered but no specific scores are known to be associated with the categories. To perform the test, assign scores x_i, $1 \le i \le r$, to the row categories and scores y_j, $1 \le j$

Received August 16, 1980; revised February 20, 1981.

[1] Support for this research was provided in part by National Science Foundation Grant No. SOC76-80389.

AMS 1970 *subject classifications.* Primary 62F05; secondary 62H20.

Key words and phrases. Likelihood ratio, Pearson chi-square, Wishart distribution, eigenvalue, asymptotic unbiasedness, consistency, asymptotic power, maximum likelihood.

$\leq s$, to the column categories so that the correlation

$$R_1 = \sum_{i=1}^{r} \sum_{j=1}^{s} f_{ij} x_i y_j \tag{2.2}$$

of scores of row and column variables is maximized subject to the constraints that the mean row score $\sum_{i=1}^{r} f_{i.} x_i$ and the mean column score $\sum_{j=1}^{s} f_{.j} y_j$ are both 0 and the respective observed variances $\sum_{i=1}^{r} f_{i.} x_i^2$ and $\sum_{j=1}^{s} f_{.j} y_j^2$ of the row and column scores are both 1.

Computation of R_1 is straightforward. Let $g_{ij} = f_{i.} f_{.j}$ be the maximum likelihood estimate of p_{ij} under (2.1), and let $e_{ij} = f_{ij}/g_{ij}^{1/2}$. By Kendall and Stuart (1973, page 591), R_1^2 is the second largest eigenvalue of the $r \times r$ matrix C with elements

$$c_{ii'} = \sum_{j=1}^{s} e_{ij} e_{i'j}, \qquad 1 \leq i \leq r, \qquad 1 \leq i' \leq r.$$

Since C always has an eigenvalue of 1 and an eigenvector $\mathbf{x}$ with coordinates $f_{i.}^{1/2}$, $1 \leq i \leq r$, it follows from Wilkinson (1965, page 585) that R_1^2 is the largest eigenvalue of the $r \times r$ matrix B with elements

$$\begin{aligned} b_{ii'} &= \sum_{j=1}^{s} d_{ij} d_{i'j} \\ &= \sum_{j=1}^{s} e_{ij} e_{i'j} - f_{i.} f_{i'.}, \qquad 1 \leq i \leq r, \qquad 1 \leq i' \leq r, \end{aligned}$$

where $d_{ij} = (f_{ij} - g_{ij})/g_{ij}^{1/2}$.

The correlation R_1 is the test statistic in the canonical correlation test. Conditional on the marginal totals $n_{i.}$, $1 \leq i \leq r$, and $n_{.j}$, $1 \leq j \leq s$, an exact significance level of R_1 can be obtained in principle; however, in practice an approximate significance level is more likely to be used. Let $W(r-1, s-1)$ be the $r-1$ by $r-1$ central Wishart matrix with $s-1$ degrees of freedom with elements

$$w_{ii'}(r-1, s-1) = \sum_{j=1}^{s-1} v_{ij} v_{i'j}, \qquad 1 \leq i \leq r-1, \qquad 1 \leq i' \leq r-1,$$

such that the v_{ij}, $1 \leq i \leq r-1$, $1 \leq j \leq s-1$, are independent $N(0, 1)$ random variables. Let $F(r-1, s-1)$ be the maximum eigenvalue of $W(r-1, s-1)$, and for $0 < \alpha < 1$, let $\lambda(r-1, s-1, \alpha)$ be the upper α point of $F(r-1, s-1)$. Let $\to_{\mathscr{D}}$ denote convergence in distribution. As shown by Corsten (1976) and O'Neill (1978a) if the independence model (2.1) holds as $N \to \infty$, then $NR_1^2 \to_{\mathscr{D}} F(r-1, s-1)$. Thus R_1 has an approximate significance level less than α if $NR_1^2 > \lambda(r-1, s-1, \alpha)$. For values of $\lambda(r-1, s-1, \alpha)$ for α equal to 0.05 or 0.01, see Table 51 of Pearson and Hartley (1972). Note that if $r = 2$, then $F(r-1, s-1)$ has a chi-square distribution with $s-1$ degrees of freedom, while $F(r-1, s-1)$ has a chi-square distribution with $r-1$ degrees of freedom if $s = 2$. In other cases, $F(r-1, s-1)$ does not have a chi-square distribution.

3. The model of linear-by-linear interaction. Alternate tests of independence may be based on the model of linear-by-linear interaction of Andersen (1980) and Goodman (1979). In this model, it may be assumed that $\mathbf{p}$ is in the closure $\bar{Q}$ of Q, where Q consists of vectors $\mathbf{p} = \{p_{ij}: 1 \leq i \leq r, 1 \leq j \leq s\}$ such that $\sum_{i=1}^{r} \sum_{j=1}^{s} p_{ij} = 1$ and

$$p_{ij} = p_{i.} p_{.j} \exp(\alpha + \beta_i + \gamma_j + \psi \mu_i \nu_j), \qquad 1 \leq i \leq r, \qquad 1 \leq j \leq s, \tag{3.1}$$

for some α, β_i, γ_j, ψ, μ_i, and ν_j such that

$$\sum_{i=1}^{r} p_{i.} \mu_i^2 = \sum_{j=1}^{s} p_{.j} \nu_j^2 = 1$$

and

$$\sum_{i=1}^{r} p_{i.} \mu_i = \sum_{j=1}^{s} p_{.j} \nu_j = \sum_{i=1}^{r} p_{i.} \beta_j = \sum_{j=1}^{s} p_{.j} \gamma_j = 0.$$

(An alternative parameterization in which $p_{i.} p_{.j}$ is omitted from (3.1) is equivalent if all p_{ij} are positive.) Let $\hat{\mathbf{p}} = \{\hat{p}_{ij}: 1 \leq i \leq r, 1 \leq j \leq s\}$ be a vector of maximum likelihood estimates (MLEs) of $\mathbf{p} = \{p_{ij}: 1 \leq i \leq r, 1 \leq j \leq s\}$ under the model $\mathbf{p} \in \bar{Q}$. To test the independence model against the alternative $\mathbf{p} \in \bar{Q}$, one may use the Pearson chi-square statistic

$$X^2 = N \sum_{i=1}^{r} \sum_{j=1}^{s} (\hat{p}_{ij} - g_{ij})^2/g_{ij}$$

or the likelihood-ratio chi-square statistic

$$L^2 = 2 \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij} \log\left(\frac{\hat{p}_{ij}}{g_{ij}}\right).$$

In Section 6, the following result is shown. If the independence model (2.1) holds, then X^2 and L^2 are both asymptotically equivalent to NR_1^2, so that as $N \to \infty$, $X^2 \to_{\mathscr{D}} F(r-1, s-1)$ and $L^2 \to_{\mathscr{D}} F(r-1, s-1)$. The approximate significance level of X^2 (or L^2) is less than α if X^2 (or L^2) exceeds $\lambda(r-1, s-1, \alpha)$. Note that the chi-square statistics X^2 and L^2 do not have approximate chi-square distributions when $r > 2$ and $s > 2$. The basic cause of this result is that ψ, μ_i, and ν_j are uniquely determined, except for sign, if (2.1) does not hold, but μ_i and ν_j are undetermined if (2.1) holds.

4. Comparison with conventional chi-square tests. Most tests for independence rely on the Pearson chi-square statistic

$$X_I^2 = N \sum_{i=1}^{r} \sum_{j=1}^{s} (f_{ij} - g_{ij})^2/g_{ij}$$

or the likelihood-ratio chi-square statistic

$$L_I^2 = 2 \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij} \log(f_{ij}/g_{ij}).$$

It is of interest to compare the properties of these test statistics with those of NR_1^2, X^2, and L^2.

To begin, comparisons are only of real interest if $r > 2$ and $s > 2$, for $X_I^2 = NR_1^2 = X^2$ and $L_I^2 = L^2$ when $r = 2$ or $s = 2$. Proofs are elementary. If $k = \min(r-1, s-1)$ and R_j^2 is the jth largest eigenvalue of B, then

$$X_I^2 = N \sum_{j=1}^{k} R_j^2$$

(Kendall and Stuart, 1973, page 594). Thus $X_I^2 = NR_1^2$ when $k = 1$. If $r = 2$ or $s = 2$, then the model $\mathbf{p} \in \bar{Q}$ imposes no restrictions on $\mathbf{p}$. Thus $\hat{p}_{ij} = f_{ij}$, $X^2 = X_I^2$, and $L^2 = L_I^2$. Given these observations, it is appropriate to assume for the remainder of the section that $r > 2$ and $s > 2$.

It is well known that X_I^2 and L_I^2 provide consistent and asymptotically unbiased tests of independence. As shown in Sections 5 and 6, these properties are shared by NR_1^2, X^2, and L^2. Asymptotic powers of the five statistics are not identical, and no one statistic uniformly dominates any other statistic in this regard.

To examine this situation, define

$$a_{ij} = (p_{ij} - p_{i\cdot}p_{\cdot j})/(p_{i\cdot}p_{\cdot j})^{1/2}, \qquad 1 \le i \le r, \qquad 1 \le j \le s,$$

to be a normalized deviation of p_{ij} from the value $p_{i\cdot}p_{\cdot j}$ found under independence, and let

$$h_{ii'} = \sum_{j=1}^{s} a_{ij}a_{i'j}, \qquad 1 \le i \le r, \qquad 1 \le i' \le r.$$

Let H be the $r \times r$ matrix with elements $h_{ii'}$, let ρ_j^2, $1 \le j \le k$, be the jth largest eigenvalue of H, and let

$$\phi^2 = \operatorname{tr} H = \sum_{i=1}^{r} \sum_{j=1}^{s} a_{ij}^2 = \sum_{j=1}^{k} \rho_j^2$$

be the coefficient of mean square contingency. Note that $\rho_j^2 = \phi^2 = 0$, $1 \le j \le k$, if and only if (2.1) holds.

It is well known that the asymptotic power of X_I^2 or L_I^2 at level α is $C(N\phi^2, \alpha) = P\{\chi_\nu^2(N\phi^2) > \chi_{\nu,\alpha}^2\}$, where $\nu = (r-1)(s-1)$, $\chi_{\nu,\alpha}^2$ is the upper-α point of a chi-square distribution with ν degrees of freedom, and $\chi^2(N\phi^2)$ is a noncentral chi-square random variable with noncentrality parameter $N\phi^2$ and with ν degrees of freedom.

In contrast, the asymptotic power of NR_1^2, X^2, or L^2 at level α is $D(N^{1/2}\boldsymbol{\rho}, \alpha) = P\{F'(r-1, s-1, N^{1/2}\boldsymbol{\rho}) > \lambda(r-1, s-1, \alpha)\}$. Here for $\mathbf{z} = \{z_j: 1 \le j \le k\}$ and $z_j = 0, j > k$, $F'(r-1, s-1, \mathbf{z})$ is the largest eigenvalue of the noncentral Wishart matrix

$$W'(r-1, s-1, \mathbf{z}) = \{w'_{ii'}(r-1, s-1, \mathbf{z}): 1 \le i \le r,\ 1 \le i' \le r\}$$

such that

$$w'_{ii'}(r-1, s-1, \mathbf{z}) = \textstyle\sum_{j=1}^{s-1}(v_{ij} + \delta_{ij}z_j)(v_{i'j} + \delta_{i'j}z_j), \qquad 1 \le i \le r, \qquad 1 \le i' \le r,$$

and δ denotes the Kronecker δ. As in the case of the central Wishart, the v_{ij} are independent $N(0, 1)$ random variables. These results are proven in Sections 5 and 6.

Given that tr $W'(r-1, s-1, N^{1/2}\boldsymbol{\rho})$ has a $\chi^2_\nu(N\phi^2)$ distribution, comparison of NR_1^2, X^2, or L^2 to X_I^2 or L_I^2 corresponds to comparison in multivariate analysis of the largest root criterion to the trace criterion when the underlying covariance matrix is known. By Perlman and Olkin (1980), $C(0, \alpha) = D(\mathbf{0}, \alpha) = \alpha$, $C(N\phi^2, \alpha)$ is strictly increasing in $N\phi^2$, and $D(N^{1/2}\boldsymbol{\rho}, \alpha)$ is strictly increasing in the $N^{1/2}\rho_j$, $1 \le j \le k$. Thus all tests under study are asymptotically unbiased.

By John (1971), $C(N\phi^2, \alpha) > D(N^{1/2}\boldsymbol{\rho}, \alpha)$ for all $N\phi^2$ sufficiently small; i.e., the conventional chi-square tests dominate the tests based on NR_1^2, X^2, or L^2 if the deviation from independence is sufficiently small. On the other hand, $C(N\phi^2, \alpha) < D(N^{1/2}\boldsymbol{\rho}, \alpha)$ if $\rho_j = 0, j > 1$, and if $N^{1/2}\rho_1$ is sufficiently large. For a proof see Section 7.

The condition $N^{1/2}\rho_j = 0$, $2 \le j \le k$, holds if the matrix of p_{ij}, $1 \le i \le r$, $1 \le j \le s$, has rank two, as is the case if the p_{ij} are consistent with a latent-class model with two latent classes (Good, 1965, and Gilula, 1979). If (3.1) holds and $N^{1/2}\psi \to \delta > 0$ as $N \to \infty$, then it is easily verified that $N^{1/2}\rho_j \to 0$, $2 \le j \le k$, and $N^{1/2}\rho_1 \to \delta$. Thus the canonical correlation tests and the X^2 and L^2 tests are likely to be most satisfactory if the p_{ij} have rank 2 or if (3.1) holds.

5. Asymptotic properties of the canonical correlation statistic. In this section, the p_{ij} are assumed to depend implicitly on the sample size N. In this way, asymptotic power can be considered. The basic results on NR_1^2 are provided by Theorem 1.

THEOREM 1. *Assume that* $p_{ij} \to p^*_{ij}$, $1 \le i \le r$, $1 \le j \le s$. *Let* $0 < \alpha < 1$. *Then*

$$P\{NR_1^2 > \lambda(r-1, s-1, \alpha)\} - P\{F'(r-1, s-1, N^{1/2}\boldsymbol{\rho}) > \lambda(\mathrm{r}-1, \mathrm{s}-1, \alpha)\} \to 0. \tag{5.1}$$

PROOF. *Case* 1. Assume $p^*_{ij} = p^*_{i\cdot}p^*_{\cdot j}$ and

$$N^{1/2}a_{ij} \to \lambda_{ij}, \qquad 1 \le i \le r, \quad 1 \le j \le s. \tag{5.2}$$

Define the $r \times s$ matrices $\Lambda = \{\lambda_{ij}: 1 \le i \le r, 1 \le j \le s\}$ and $D = \{d_{ij}: 1 \le i \le r, 1 \le j \le s\}$. Then $N^{1/2}D$ is the matrix of standardized residuals in the independence model (Haberman, 1973). Let $T = \{t_{ii'}: 1 \le i \le r, 1 \le i' \le r\}$ be the orthogonal projection on R^r of rank $r-1$ with elements

$$t_{ii'} = \delta_{ii'} - p_{i\cdot}^{1/2}\, p_{i'\cdot}^{1/2}, \qquad 1 \le i \le r, \quad 1 \le i' \le r,$$

and let $U = \{u_{jj'}: 1 \le j \le s, 1 \le j' \le s\}$ be the orthogonal projection on R^s of rank $s-1$ with elements

$$u_{jj'} = \delta_{jj'} - p_{\cdot j}^{1/2}\, p_{\cdot j'}^{1/2}, \qquad 1 \le j \le s, \quad 1 \le j' \le s.$$

Here δ is the Kronecker function. Let $T \otimes U$ be the Kronecker product of T and U. By Haberman (1974, pages 104, 138, 140, 212), $N^{1/2}D$ converges in distribution to the $r \times s$ random matrix Z with distribution $N(\Lambda, T \otimes U)$. If asterisks are used to denote transposes, so that D^* is the transpose of D and Z^* is the transpose of Z, then $NDD^* \to_{\mathscr{D}} ZZ^*$. Since NR_1^2 is a continuous function of NDD^* (Wilkinson, 1965, pages 72–77), NR_1^2 converges in distribution to the largest eigenvalue of ZZ^*. Let $V = \{v_{ij}: 1 \le i \le r-1, 1 \le j \le s-1\}$. Let $\boldsymbol{\sigma} = \{p_{i\cdot}^{1/2}: 1 \le i \le r\}$ and $\boldsymbol{\tau} = \{p_{\cdot j}^{1/2}: 1 \le j \le s\}$, so that $\Lambda^*\boldsymbol{\sigma} = \mathbf{0}$ and $\Lambda\boldsymbol{\tau} = \mathbf{0}$. There exists an $r \times (r-1)$ matrix Γ, an $s \times (s-1)$ matrix Δ, and an $(r-1) \times (s-1)$ matrix Ω such that

$\Lambda = \Gamma\Omega\Delta^*$, $\Gamma\Gamma^* = T$, $\Delta\Delta^* = U$, $\Gamma^*\Gamma = I_{r-1}$, the $(r-1) \times (r-1)$ identity matrix, $\Delta^*\Delta = I_{s-1}$, and $\Omega = \{\delta_{ij}\xi_j : 1 \le i \le r-1, 1 \le j \le s-1\}$ for a nonincreasing sequence ξ_j, $1 \le j \le s-1$, such that $\xi_j = 0$ for any $j > k$. By Eaton (1970), Z is distributed as $\Gamma(V + \Omega)\Delta^*$ and ZZ^* is distributed as $\Gamma(V + \Omega)(V + \Omega)^*\Gamma^*$. Since $\Gamma^*\Gamma = I_{r-1}$, the largest eigenvalue of ZZ^* is also the largest eigenvalue of $(V + \Omega)(V + \Omega)^*$. This latter eigenvalue has a $F'(r-1, s-1, \boldsymbol{\xi})$ distribution if $\boldsymbol{\xi} = \{\xi_j : 1 \le j \le k\}$. Thus $NR_1^2 \to_{\mathscr{D}} F'(r-1, s-1, \boldsymbol{\xi})$. Since $N^{1/2}\boldsymbol{\rho} \to \boldsymbol{\xi}$, $F'(r-1, s-1, N^{1/2}\boldsymbol{\rho}) \to F'(r-1, s-1, \boldsymbol{\xi})$ and (5.1) holds in this case.

Case 2. Let $N\phi^2$, $N \ge 1$, be bounded above. Then every subsequence of the $\{p_{ij} : 1 \le i \le r, 1 \le j \le s\}$ contains a subsequence such that (5.2), and therefore (5.1), holds. Thus (5.1) holds if $N\phi^2$ is bounded above.

Case 3. Let $N\phi^2 \to \infty$. Since $\phi^2 = \sum_{j=1}^{k} \rho_j^2$, $N\rho_1^2 \to \infty$. Since

$$F'(r-1, s-1, N^{1/2}\boldsymbol{\rho}) \geqslant \sum_{j=1}^{k-1} (v_{1j} + N^{1/2}\rho_1\delta_{ij})^2,$$

it follows that $P\{F'(r-1, s-1, N^{1/2}\boldsymbol{\rho}) > \lambda(r-1, s-1, \alpha)\} \to 1$ as $N \to \infty$. Similarly, let $\boldsymbol{\mu} = \{\mu_i : 1 \le i \le r\}$ be an eigenvector of H corresponding to the eigenvalue ρ_1^2 such that $\sum_{i=1}^{r} \mu_i^2 = 1$. Then

$$NR_1^2 \ge \sum_{j=1}^{s} (\sum_{i=1}^{r} \mu_i d_{ij})^2$$

and $d_{ij} - a_{ij} \to_p 0$, where $\to_p$ denotes convergence in probability. Since

$$\sum_{j=1}^{s} (\sum_{i=1}^{r} \mu_i a_{ij})^2 = \rho_1^2$$

it follows that $NR_1^2 \to_p \infty$. Thus $P\{NR_1^2 > \lambda(r-1, s-1, \alpha)\} \to 1$ as $N \to \infty$, and (5.1) holds. Given these three cases, it follows that (5.1) holds whenever $p_{ij} \to p_{ij}^*$, $1 \le i \le r$, $1 \le j \le s$. □

The case of constant p_{ij} implies that $NR_1^2 \to_{\mathscr{D}} F(r-1, s-1)$ if (2.1) holds and that $P\{NR_1^2 > \lambda(r-1, s-1, \alpha)\} \to 1$ if (2.1) does not hold. Thus the theorem implies consistency of the canonical correlation test and implies that approximate significance levels of the test may be found by reference to the distribution of $F(r-1, s-1)$.

6. Asymptotic properties of X^2 and L^2. Results for X^2 and L^2 closely resemble those for NR_1^2. To avoid technical difficulties, it is assumed in this section that $\hat{\mathbf{p}}$ is a function of the counts n_{ij}. Since the likelihood kernel

$$L(\mathbf{q}) = \prod_{i=1}^{r} \prod_{j=1}^{s} q_{ij}^{n_{ij}} \tag{6.1}$$

is continuous for $\mathbf{q}$ in the simplex S of $\mathbf{q} = \{q_{ij} : 1 \le i \le r, 1 \le j \le s\}$ such that $q_{ij} \ge 0$, $1 \le i \le r$, $1 \le j \le s$, and $\sum_{i=1}^{r} \sum_{j=1}^{s} q_{ij} = 1$ and since $\bar{Q}$ is closed, some $\hat{\mathbf{p}}$ always exists. By Goodman (1979), $\hat{p}_{i\cdot} = \sum_{j=1}^{s} \hat{p}_{ij} = f_{i\cdot}$ and $\hat{p}_{\cdot j} = \sum_{i+1}^{r} \hat{p}_{ij} = f_{\cdot j}$. Since $L(\mathbf{q}) = 0 < L(\mathbf{g})$ if $q_{ij} = 0$ and $n_{ij} > 0$ for some i and j, $\hat{p}_{ij} > 0$ if $n_{ij} > 0$. Thus X^2 and L^2 are finite if the conventions $0/0 = 0$ and $0 \log 0 = 0$ are used.

Corresponding to Theorem 1, one has

THEOREM 2. *Assume that $p_{ij} \to p_{ij}^* > 0$, $1 \le i \le r$, $1 \le j \le s$. Then for $0 < \alpha < 1$,*

$$P\{X^2 > \lambda(r-1, s-1, \alpha\} - P\{F'(r-1, s-1, N^{1/2}\boldsymbol{\rho}) > \lambda(r-1, s-1, \alpha)\} \to 0 \tag{6.2}$$

and

$$P\{L^2 > \lambda(r-1, s-1, \alpha)\} - P\{F'(r-1, s-1, N^{1/2}\boldsymbol{\rho}) > \lambda(r-1, s-1, \alpha)\} \to 0. \tag{6.3}$$

If $N\phi^2$ is bounded above as $N \to \infty$, then $X^2 - L^2$, $X^2 - NR_1^2$, and $L^2 - NR_1^2$ all converge in probability to 0.

PROOF. *Case 1.* Assume (5.2) and (2.1). Let

$$K(\mathbf{q}) = \sum_{i=1}^{r} \sum_{j=1}^{s} f_{ij} \log(f_{ij}/q_{ij}) = N^{-1} \log\{L(\mathbf{f})/L(\mathbf{q})\} \tag{6.4}$$

and

$$J(\mathbf{q}) = \sum_{i=1}^{r} \sum_{j=1}^{s} (f_{ij} - q_{ij})^2/g_{ij} \tag{6.5}$$

for $\mathbf{q} \in S$. Thus

$$L^2 = 2N\{K(\mathbf{g}) - K(\hat{\mathbf{p}})\}. \tag{6.6}$$

The function L^2 is readily approximated by $N\{J(\mathbf{g}) - J(\hat{\mathbf{p}})\}$. To verify this claim, observe that

$$0 \le K(\hat{p}) \le K(\mathbf{g}) \tag{6.7}$$

and that $2NK(\mathbf{g})$ is well known to converge in distribution to $\chi^2_\nu(\mu)$, where $\mu = \sum_{i=1}^{r} \sum_{j=1}^{s} \lambda_{ij}^2$. The estimate g_{ij} is well known to be the maximum likelihood estimate of p_{ij} under the independence model, and it is well known that $f_{ij} \to_p p_{ij}^*$. The argument of Rao (1975, page 356) implies that $g_{ij} \to_p p_{ij}^*$ and $p_{ij} \to_p p_{ij}^*$.

By Rao (1973, page 356), if $q_{ij} > 0$ whenever $f_{ij} > 0$, then

$$K(\mathbf{q}) = \tfrac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{s} f_{ij}(f_{ij} - q_{ij})^2/z_{ij}^2 \tag{6.8}$$

for some z_{ij} between q_{ij} and f_{ij}. Thus

$$2K(\mathbf{g}) = J(\mathbf{g})(1 + \epsilon) \tag{6.9}$$

and

$$2K(\hat{\mathbf{p}}) = J(\hat{\mathbf{p}})(1 + \epsilon'), \tag{6.10}$$

where $\epsilon \to_p 0$ and $\epsilon' \to_p 0$. It follows that

$$L^2 - N\{J(\mathbf{g}) - J(\hat{\mathbf{p}})\} \to_p 0. \tag{6.11}$$

An approximation for $J(\hat{\mathbf{p}})$ can then be obtained. Let $t_{ij} = \log(\hat{p}_{ij}/g_{ij})$, so that $-\infty \le t_{ij} < \infty$. If all $\hat{p}_{ij}$ are positive, then $t_{ij} = \hat{\alpha} + \hat{\beta}_i + \hat{\gamma}_j + \hat{\psi}\hat{\mu}_i\hat{\nu}_j$, where

$$\sum_{i=1}^{r} f_{i\cdot}\, \hat{\beta}_i = \sum_{j=1}^{s} f_{\cdot j}\hat{\gamma}_j = \sum_{i=1}^{r} f_{i\cdot}\hat{\mu}_i = \sum_{j=1}^{s} f_{\cdot j}\hat{\nu}_j = 0$$

and

$$\sum_{i=1}^{r} f_{i\cdot}\hat{\mu}_1^2 = \sum_{j=1}^{s} f_{\cdot j}\hat{\nu}_j^2 = 1.$$

Let $w_{ij} = g_{ij}(1 + t_{ij})$. One may then approximate $J(\hat{\mathbf{p}})$ by $J(\mathbf{w})$.

To verify the approximation using $J(\mathbf{w})$, note that

$$J(\hat{\mathbf{p}}) = J(\mathbf{w}) - 2G_1 + G_2,$$

where

$$G_1 = \sum_{i=1}^{r} \sum_{j=1}^{s} (f_{ij} - w_{ij})(\hat{p}_{ij} - w_{ij})/g_{ij}$$

and

$$G_2 = \sum_{i=1}^{r} \sum_{j=1}^{s} (\hat{p}_{ij} - w_{ij})^2/g_{ij}$$

By Schwarz's inequality, $G_1 = e_1\{J(\mathbf{w})G_2\}^{1/2}$, where $|e_1| \le 1$. By the quadratic formula,

$$\{J(\mathbf{w})\}^{1/2} = e_1 G_2^{1/2} + s_1\{e_1^2 G_2 + J(\hat{\mathbf{p}}) - G_2\}^{1/2}, \tag{6.12}$$

where s_1 is 1 or -1.

To use (6.12) to achieve the desired approximation, it must be shown that $NG_2 \to_p 0$. To do so, let

$$o(x) = e^x - 1 - x,$$

so that $|o(x)| \le x^2$ for $|x|$ sufficiently small. Then

$$G_2 = \sum_{i=1}^r \sum_{j=1}^s g_{ij}\{(o(t_{ij}))\}^2.$$

By Taylor's theorem,

$$X^2/N = \sum_{i=1}^r \sum_{j=1}^s (\hat{p}_{ij} - g_{ij})^2/g_{ij} = \sum_{i=1}^r \sum_{j=1}^s g_{ij}\{t_{ij}\exp(t^*_{ij})\}^2 \tag{6.13}$$

for some t^*_{ij} between 0 and t_{ij}. Since $t_{ij} \to_p 0$, $G_2 = e_2 X^2/N$, where $e_2 \to_p 0$. By Schwarz's inequality,

$$\begin{aligned} X^2/N &= \sum_{i=1}^r \sum_{j=1}^s (f_{ij} - g_{ij} - f_{ij} + \hat{p}_{ij})^2/g_{ij} \\ &= J(\mathbf{g}) + J(\hat{\mathbf{p}}) - 2\sum_{i=1}^r \sum_{j=1}^s (f_{ij} - g_{ij})(f_{ij} - \hat{p}_{ij})/g_{ij} \\ &\le J(\mathbf{g}) + J(\hat{\mathbf{p}}) + 2\{J(\mathbf{g})J(\hat{\mathbf{p}})\}^{1/2}. \end{aligned}$$

Since $NJ(\mathbf{g})$ is the Pearson chi-square statistic for testing independence, $NJ(\mathbf{g}) \to_{\mathscr{D}} \chi^2_\nu(\mu)$. Given that $2NK(\mathbf{g}) \to_{\mathscr{D}} \chi^2_\nu(\mu)$, $K(\hat{\mathbf{p}}) \le K(\mathbf{g})$, and (6.10) holds, it follows that $NG_2 \to_p 0$. Squaring (6.12) shows that

$$N\{J(\mathbf{w}) - J(\hat{\mathbf{p}})\} \to_p 0. \tag{6.14}$$

Thus

$$L^2 - N\{J(g) - J(\mathbf{w})\} \to_p 0. \tag{6.15}$$

For any $\epsilon > 0$, as $N \to \infty$

$$P(L^2 - NR_1^2 > \epsilon) \to 0. \tag{6.16}$$

This claim follows since whenever all $\hat{p}_{ij}$ are positive,

$$J(g) - J(\mathbf{w}) = h^2 - (\hat{\psi} - h)^2 - \hat{\alpha}^2 - \sum f_{i\cdot}\hat{\beta}_i^2 - \sum f_{\cdot j}\hat{\gamma}_j^2, \tag{6.17}$$

where

$$h = \sum_{i=1}^r \sum_{j=1}^s (f_{ij} - g_{ij})\hat{\mu}_i\hat{\nu}_j = \sum_{i=1}^r \sum_{j=1}^s f_{ij}(\hat{\mu}_i - a)(\hat{\nu}_j - b)$$

for $a = \sum_{i=1}^r f_{i\cdot}\hat{\mu}_i$ and $b = \sum_{j=1}^s f_{\cdot j}\hat{\nu}_j$. Since

$$\sum_{i=1}^r f_{i\cdot}(\hat{\mu}_i - a) = \sum_{j=1}^s f_{\cdot j}(\hat{\nu}_j - b) = 0,$$

$$\sum_{i=1}^r f_{i\cdot}(\hat{\mu}_i - a)^2 = 1 - a^2,$$

and

$$\sum_{j=1}^s f_{\cdot j}(\hat{\nu}_j - b)^z = 1 - b^2, \qquad |h| \le R_1.$$

Thus (6.16) follows.

On the other hand, as $N \to \infty$,

$$P(L^2 - NR_1^2 < -\epsilon) \to 0. \tag{6.18}$$

To verify this claim, let $\mathbf{p}^+ \in \bar{Q}$ satisfy

$$p^+_{ij} = c^{-1}g_{ij}\exp(R_1 x_i y_j).$$

Here x_i and y_i are defined as in Section 2 so that (2.2) holds, $\sum_{i=1}^r f_{i\cdot}x_i = \sum_{j=1}^s f_{\cdot j}y_j = 0$, and $\sum_{i=1}^r f_{i\cdot}x_i^2 = \sum_{j=1}^s f_{\cdot j}y_j^2 = 1$. One has

$$c = \sum_{i=1}^r \sum_{j=1}^s g_{ij}\exp(R_1 x_i y_j) = 1 + \sum_{i=1}^r \sum_{j=1}^s o(R_1 x_i y_j).$$

Clearly

$$L^2 = 2N\{K(\mathbf{g}) - K(\hat{\mathbf{p}})\} \ge 2N\{K(\mathbf{g}) - K(\mathbf{p}^+)\}. \tag{6.19}$$

Since $NR_1^2 \to_{\mathcal{D}} F(r-1, s-1, \xi)$, where ξ is defined as in the proof of Theorem 1, it readily follows that

(6.20) $$N\{2K(\mathbf{p}^+) - \sum_{i=1}^r \sum_{j=1}^s (f_{ij} - g_{ij} - g_{ij}R_1x_iy_j)^2/g_{ij}\} \to_p 0.$$

Since, as in Householder and Young (1938),

$$\sum_{i=1}^r \sum_{j=1}^s (f_{ij} - g_{ij} - g_{ij}R_1x_iy_j)^2/g_{ij} = J(\mathbf{g}) - R_1^2,$$

(6.9), (6.19), and (6.20) imply (6.17). Thus $L^2 - NR_1^2 \to_p 0$ and (6.2) holds.

In the case of X^2, if $\hat{p}_{ij} > 0$, then

(6.21) $$\sum_{i=1}^r \sum_{j=1}^s g_{ij}t_{ij}^2 = \hat{\psi}^2 + \hat{\alpha}^2 + \sum f_{i\cdot}\,\hat{\beta}_i^2 + \sum f_{\cdot j}\,\hat{\gamma}_j^2.$$

The inequality $h^2 \le R_1^2$, (6.2), (6.13), (6.15), and (6.17) imply that $N\hat{\psi}^2 - NR_1^2 \to_p 0$, $X^2 - N\hat{\psi}^2 \to_p 0$, and $X^2 - NR_1^2 \to_p 0$. Thus (6.2) holds, and $X^2 - L^2 \to_p 0$.

Case 2. Let $N\phi^2$ be bounded above as $N \to \infty$. Then the argument proceeds as in Case 2 of the proof of Theorem 1.

Case 3. Let $N\phi^2 \to \infty$. Then $NR_1^2 \to_p \infty$. Let $\beta \le 1$ be a function of N such that $\beta R_1 \to_p 0$ and $\beta NR_1^2 \to_p \infty$. Let x_i and y_j be defined as in Case 1, and let

$$p'_{ij} = c^{-1}g_{ij}\exp(\beta R_1 x_i y_j),$$

where

$$c = \sum_{i=1}^r \sum_{j=1}^s g_{ij}\exp(\beta R_1 x_i y_j).$$

Then $\mathbf{p}' = \{p'_{ij} : 1 \le i \le r, 1 \le j \le s\} \in \bar{Q}$ and as in (6.19),

$$L^2 \ge 2N\{K(\mathbf{g}) - K(\mathbf{p}')\} = \beta NR_1^2 - N\log c.$$

Since $(\beta^2R_1^2)^{-1}\log c \to_p \frac{1}{2}$, $L^2 \to_p \infty$. As in Case 3 of the proof of Theorem 1, it follows that (6.3) holds. Since $L^2 \to_p \infty$ and since $\sum_{i=1}^r \sum_{j=1}^s x_iy_if_{ij} = \sum_{i=1}^r \sum_{j=1}^s x_iy_j\hat{p}_{ij}$ (Goodman, 1979),

$$L^2 = 2N\sum_{i=1}^r \sum_{j=1}^s \hat{p}_{ij}\log(\hat{p}_{ij}/g_{ij}) \to_p \infty.$$

Using a relationship analogous to (6.8), it easily follows that $X^2 \to_p \infty$, so that (6.3) holds. □

7. Asymptotic power comparison for $\rho_j = 0$, $2 \le j \le k$. To show that $C(N\phi^2, \alpha) < D(N^{1/2}\boldsymbol{\rho}, \alpha)$ for $N^{1/2}\rho_1 = N^{1/2}\phi$ sufficiently large in this case, one proceeds by analogy. Let X_{ij}, $1 \le i \le r-1$, $1 \le j \le s-1$, be independent normal observations with respective means μ_{ij} and variances 1. One may argue as in Ghosh (1964) and Stein (1956). The joint density $f(\cdot, \mu)$ of the X_{ij} relative to Lebesgue measure has the exponential family form,

$$f(\mathbf{x}, \mu) = (2\pi)^{-(1/2)(r-1)(s-1)}\exp(-\tfrac{1}{2}\sum_{i=1}^{r-1}\sum_{j=1}^{s-1}x_{ij}^2 - \tfrac{1}{2}\sum_{j=1}^{r-1}\sum_{i=1}^{s-1}\mu_{ij}^2 + \sum_{i=1}^{r-1}\sum_{j=1}^{s-1}x_{ij}\mu_{ij})$$

A possible unbiased test of size α of the hypothesis $\mu_{ij} = 0$, $1 \le i \le r-1$, $1 \le j \le s-1$, rejects only if $(\sum_{i=1}^{r-1}\sum_{j=1}^{s-1}x_{ij}a_ib_j)^2 \ge \lambda(r-1, s-1, \alpha)$ for some $\mathbf{a} = \{a_i : 1 \le i \le r-1\}$ and $\mathbf{b} = \{b_j : 1 \le j \le s-1\}$ such that $\sum_{i=1}^{r-1} a_i^2 = \sum_{j=1}^{s-1} b_j^2 = 1$. This test is well known to have power $D(\mathbf{c}, \alpha)$, where $\mathbf{c} = \{c_j : 1 \le j \le k\}$ and c_j is the jth largest eigenvalue of the $(r-1) \times (r-1)$ matrix with elements

$$\sum_{j=1}^{s-1}\mu_{ij}\mu_{i'j}, \qquad 1 \le i \le r-1, \quad 1 \le i' \le r-1.$$

An alternate unbiased test of size α rejects if and only if $\sum_{i=1}^{r-1}\sum_{j=1}^{s-1}X_{ij}^2 \ge \chi_\nu^2$. This latter test has power $C(d, \alpha)$, where $d = \sum_{j=1}^k c_j^2$. Let $\mu_{ij} = N^{1/2}\phi\epsilon_i\eta_j$, where $\sum_{i=1}^{r-1}\epsilon_i^2 = \sum_{j=1}^{s-1}\eta_j^2 = 1$. As in Stein (1956), it follows that for sufficiently large $N^{1/2}\phi$,

$$C(d, \alpha) = C(N\phi^2, \alpha) < D(\mathbf{c}, \alpha),$$

where $c_1 = N^{1/2}\phi$ and $c_j = 0$, $2 \le j \le k$.

REFERENCES

ANDERSEN, E. B. (1980). *Discrete Statistical Models with Social Science Applications.* North-Holland, Amsterdam.

BIRCH, M. W. (1964). A new proof of the Pearson-Fisher theorem. *Ann. Math. Statist.* **35** 817–824.

CORSTEN, L. C. A. (1976). Matrix approximation, a key to application of multivariate methods. *Proc. 9th Internat. Conf.* **1** 61–77. Biometric Society, Raleigh, N.C.

EATON, M. L. (1970). Gauss-Markov estimation for multivariate linear models: a coordinate free approach. *Ann. Math. Statist.* **41** 528–538.

FISHER, R. A. (1940). The precision of discriminant functions. *Ann. Eugen. Lond.* **10** 422–429.

GHOSH, M. N. (1964). On the admissibility of some test of Manova. *Ann. Math. Statist.* **41** 789–794.

GILULA, Z. (1979). Singular value decomposition of probability matrices: probabilistic aspects of latent dichotomous variables. *Biometrika* **66** 339–344.

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press, Cambridge, Mass.

GOODMAN, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74** 537–552.

GOODMAN, L. A. (1980). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. Technical Report No. 101, Department of Statistics, University of Chicago.

HABERMAN, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics* **29** 205–220.

HOUSEHOLDER, A. S. AND YOUNG, G. (1938). Matrix approximation and latent roots. *Amer. Math. Monthly* **45** 165–171.

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127.

KENDALL, M. G. AND STUART, A. (1973). *The Advanced Theory of Statistics* **2**., 3rd. ed. Hafner, New York.

O'NEILL, M. E. (1978a). Distribution expansions for canonical correlations from contingency tables. *J. Roy. Statist. Soc. B* **40** 303–312.

O'NEILL, M. E. (1978b). Asymptotic distributions of the canonical correlations from contingency tables. *Austral. J. Statist.* 75–82.

PEARSON, E. S. AND HARTLEY, H. O. (1972). *Biometrika Tables for Statisticians* **2**. Cambridge University Press.

PERLMAN, M. D. AND OLKIN, I. (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems. *Ann. Statist.* **8** 1326–1341.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd. ed. Wiley, New York.

STEIN, C. (1956). The admissibility of Hotelling's T^2-test. *Ann. Math. Statist.* **27** 616–623.

WILKINSON, J. H. (1965). *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637