Exact Inference for Contingency Tables with Ordered Categories
Author(s): Alan Agresti, Cyrus R. Mehta and Nitin R. Patel
Source: *Journal of the American Statistical Association*, Vol. 85, No. 410 (Jun., 1990), pp. 453-458
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2289783
Accessed: 20-11-2024 18:41 UTC

Exact Inference for Contingency Tables With Ordered Categories

ALAN AGRESTI, CYRUS R. MEHTA, and NITIN R. PATEL*

This article proposes an efficient numerical algorithm for small-sample exact inferences in contingency tables having ordinal classifications. The inferences, which apply conditional on the observed marginal totals, also provide an exact analysis for the log-linear model of linear-by-linear association for cell probabilities. An exact test of independence has a one-sided P value equal to the null probability that model-based maximum likelihood estimates of odds ratios are at least as large as the observed estimates. The conditional nonnull distribution yields confidence intervals for odds ratios having a linear-by-linear structure. The computations utilize an extension of the network algorithm proposed by Mehta and Patel (1983).

KEY WORDS: Linear-by-linear association; Log-linear models; Multiple hypergeometric distribution; Network algorithm; Odds ratio; Rank tests; Score tests; Uniform association.

1. INTRODUCTION

This article presents exact small-sample inferences for cross-classifications of ordinal variables. Let $\{n_{ij}\}$ denote cell counts in an $r \times c$ contingency table, and let $\{m_{ij}\}$ denote their expected values. Denote the row variable by X and the column variable by Y. We assume a standard Poisson or multinomial sampling model for $\{n_{ij}\}$. Let $n = \sum\sum n_{ij}$.

To eliminate nuisance parameters and permit exact analysis, the usual approach conditions on marginal totals and uses the resulting hypergeometric distribution. For ordinal classifications, powerful tests usually result from basing the P value on some ordinal measure of distance from independence. Agresti and Wackerly (1977) used the difference between the numbers of concordant and discordant pairs. Patefield (1982) suggested alternative ordinal statistics, and he and Agresti, Wackerly, and Boyett (1979) implemented sampling schemes to estimate the P value for cases in which calculations are excessive.

This article presents an alternative analysis that applies to a popular log-linear model for doubly-ordered tables, the *linear-by-linear association* model. For fixed scores $u_1 < \cdots < u_r$ for the rows and $v_1 < \cdots < v_c$ for the columns, this model is

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \beta u_i v_j. \qquad (1.1)$$

Let $\alpha_{ij} = (m_{ij}m_{rc})/(m_{ic}m_{rj})$. Model (1.1) has odds ratio structure

$$\log \alpha_{ij} = \beta(u_r - u_i)(v_c - v_j), \quad i = 1, \ldots, r-1, \quad j = 1, \ldots, c-1.$$

For equal-interval scores, the model implies uniformity in the values of *local odds ratios* $\{\theta_{ij} = m_{ij}m_{i+1,j+1}/m_{i,j+1}m_{i+1,j}\}$. A wealth of applications of the model have appeared in the past decade, many of them summarized by Agresti (1990, chap. 8).

Minimal sufficient statistics for Model (1.1) are $\{n_{i+}\}$, $\{n_{+j}\}$, and $T = \sum\sum u_i v_j n_{ij}$. For fixed marginal counts, the maximum likelihood (ML) estimate $\hat{\beta}$ of β is a strictly monotone function of T (Gilula, Krieger, and Ritov 1988). This suggests that we base exact inference for β on the distribution of T, which itself is a monotone function of the correlation between X and Y for the given scores. Yates (1948) proposed a large-sample test using T as an alternative to Pearson's chi-squared test, and Mantel (1963) used it in tests of partial association. Such tests concentrate the search for association on a single degree of freedom, often resulting in greatly increased power compared to the chi-squared test.

Our test has a P value equal to the null conditional probability that the ML estimator of β is at least as large as the observed value. We also construct confidence intervals for β, and hence for odds ratios, based on the nonnull conditional distribution. Computations use an extension of the network algorithm of Mehta and Patel (1983).

2. EXACT PERMUTATION DISTRIBUTION OF ORDINAL ODDS RATIOS

Conditional on the marginal totals, Cornfield (1956) showed that the distribution of $\{n_{ij}\}$ is proportional to

$$\prod_{i=1}^{r-1}\prod_{j=1}^{c-1} \alpha_{ij}^{n_{ij}} \Big/ \prod_{i=1}^{r}\prod_{j=1}^{c} n_{ij}!. \qquad (2.1)$$

For Model (1.1), this simplifies to $e^{\beta T}/(\prod\prod n_{ij}!)$. The conditional distribution of T is therefore

$$\Pr(T = t \mid \{n_{i+}\}, \{n_{+j}\}; \beta) = C_t e^{\beta t}/(\sum C_u e^{\beta u}), \qquad (2.2)$$

where C_t is the sum of $(\prod\prod n_{ij}!)^{-1}$ for all tables with the given marginal distributions having $T = t$.

* Alan Agresti is Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Cyrus R. Mehta is Associate Professor, Department of Biostatistics, Harvard University School of Public Health, Boston, MA 02115, and President, Cytel Software Corporation, Cambridge, MA 02139. Nitin R. Patel is Visiting Professor, Department of Biostatistics, Harvard University School of Public Health, Boston, MA 02115. Mehta's research was partially supported by National Institutes of Health Grant GM38971 and National Cancer Institute Grants CP95627 and CA33019. The research was partially completed while Agresti was Visiting Professor at the Harvard University School of Public Health, Department of Biostatistics and the Boston University Mathematics Department. The authors thank Pralay Senchaudhuri for computing help and a referee for helpful comments.

Journal of the American Statistical Association
June 1990, Vol. 85, No. 410, Theory and Methods

Now, let $\{x_{ij}\}$ denote a generic set of cell counts, abbreviated as $\mathbf{x}$, and let t denote the value of T for the observed data $\{n_{ij}\}$. Let $S = \{\mathbf{x} : x_{i+} = n_{i+}, x_{+j} = n_{+j}, \text{ all } i \text{ and } j\}$ and $S_t = \{\mathbf{x} : \mathbf{x} \text{ in } S \text{ and } T \geq t\}$. To test H_0: $\beta = \beta_0$ against H_1: $\beta > \beta_0$, the exact conditional P value is

$$\Pr(T \geq t \mid \{n_{i+}\}, \{n_{+j}\}; \beta_0) = \frac{\sum_{S_t}\left[e^{\beta_0 T}\Big/(\prod\prod x_{ij}!)\right]}{\sum_{S}\left[e^{\beta_0 T}\Big/(\prod\prod x_{ij}!)\right]} = \frac{\sum_{T\geq t} C_T e^{\beta_0 T}}{\sum_T C_T e^{\beta_0 T}}. \quad (2.3)$$

For H_1: $\beta \neq \beta_0$, there are several possible ways to form P values. We recommend letting $P = \Pr[|T - E(T)| \geq |t - E(T)|]$, where $E(T)$ is the expected value of T when $\beta = \beta_0$.

Since this test uses category orderings, it can give results very different from tests that ignore the ordering. For an extreme case, consider the test of independence ($\beta = 0$) against H_1: $\beta \neq 0$ for an $r \times r$ table having $n_{ii} = 1$ for all i and $n_{ij} = 0$ for all $i \neq j$. This ordinal test has a P value equal to $(2/r!)$, whereas an exact test using a nominal-scale statistic such as the Pearson chi-squared statistic has $P = 1.0$.

We construct a confidence interval for β by inverting the test. An interval having confidence coefficient at least $1 - \gamma$ is (β_-, β_+), where $\sum_{T\geq t} P(T \mid \{n_{i+}\}, \{n_{+j}\}, \beta_-) = \gamma/2$ and $\sum_{T\leq t} P(T \mid \{n_{i+}\}, \{n_{+j}\}, \beta_+) = \gamma/2$. These sums are monotone functions of β_- and β_+, by the same argument given by Mehta, Patel, and Tsiatis (1984). The confidence interval translates to ones for odds ratios. For instance, when we assume (1.1) with unit-spaced scores, $[\exp(\beta_-), \exp(\beta_+)]$ is a confidence interval for all $(r - 1)(c - 1)$ local odds ratios.

The ML estimate of β for multinomial or Poisson sampling treats at most one margin of the table as fixed. Given both margins, the *conditional* ML estimate maximizes (2.2). It satisfies $t = E(T)$, where $E(T)$ refers to distribution (2.2). From this expression, it is simple to show that the conditional ML estimate is also a monotone function of T, given the marginal totals. Thus, for H_1: $\beta > 0$, the P value is also the probability that the conditional ML estimator is at least as large as observed.

3. SPECIAL CASES OF THE TEST OF INDEPENDENCE

A wide variety of exact tests of independence are special cases of our test, for $\beta_0 = 0$. For the 2×2 case, our test is Fisher's exact test. For the $2 \times c$ case, Graubard and Korn (1987) tabulated 14 tests based on T. For $r \times c$ tables with fixed scores, the test is an exact test for the Pearson correlation. For $r \times c$ tables, the ridit or midrank scores are fixed for the set S, and we have an exact test for the Spearman correlation. For the $2 \times c$ case, this test is equivalent to an exact Wilcoxon test for ordered categorical data (Mehta et al. 1984).

Suppose Y is an ordinal response, and X is an explanatory variable for which levels have a natural set of scores $\{u_i\}$. Let $\Pr(Y_i \leq j)$ denote the probability that a randomly selected subject at level i of X makes response in category j or below of Y. For the cumulative logit model

$$\log[\Pr(Y_i \leq j)/\Pr(Y_i > j)] = \alpha_j - \beta u_i,$$

T with midrank scores for $\{v_j\}$ is the efficient score for testing H_0: $\beta = 0$ (McCullagh 1980). The statistic T with arbitrary ordered scores is the efficient score for testing $\beta = 0$ in Model (1.1). Thus for these models the tests of independence presented in this article are locally most powerful against one-sided alternatives.

Suppose we want to test independence against the more general alternative that all local log odds ratios are nonnegative, and at least one is positive. Distribution (2.1) of $\{n_{ij}\}$ can be expressed in terms of $\{\theta_{ij}\}$ by noting that

$$\prod_{i=1}^{r-1}\prod_{j=1}^{c-1} \alpha_{ij}^{n_{ij}} = \prod_{i=1}^{r-1}\prod_{j=1}^{c-1} \theta_{ij}^{s_{ij}},$$

where $s_{ij} = \sum_{a\leq i}\sum_{b\leq j} n_{ab}$. Hirotsu (1982) showed that statistics that have a convex acceptance region and are monotone increasing in each s_{ij} (given the marginal totals) yield efficient score tests. Cohen and Sackrowitz (1988) showed that such tests form a complete class, being the set of exact, unbiased, and admissible tests. A simple way to construct a test in this class is to let the test statistic be some positive linear combination of $\{s_{ij}\}$ and form the critical region from large values of the statistic. For strictly ordered scores, T can be expressed as such a positive linear combination, so tests discussed here fall in this complete class.

4. NETWORK ALGORITHM

We use a network algorithm to compute, recursively, the coefficients $\{C_t\}$ that specify the conditional null distribution of T when $\beta = 0$. These coefficients can, in turn, yield the distribution of T for any β, through (2.2). Inference about β then proceeds in accordance with (2.3).

4.1 Network Representation for S

It is convenient to represent the reference set S as a network of *nodes* and *arcs*. The network is constructed in $c + 1$ stages, labeled $0, 1, \ldots, c$. Each stage consists of a set of nodes. Arcs emanate from each node at any stage k, and each arc is connected to a distinct node at stage $k + 1$. The nodes at stage k have form $(k, \mathbf{w}_k)$, where

$$\mathbf{w}_k = \begin{pmatrix} w_{1k} \\ w_{2k} \\ w_{rk} \end{pmatrix} = \begin{pmatrix} x_{11} + x_{12} + \cdots + x_{1k} \\ x_{21} + x_{22} + \cdots + x_{2k} \\ x_{r1} + x_{r2} + \cdots + x_{rk} \end{pmatrix}$$

represents one possible partial sum for the first k columns of the $r \times c$ tables in S, and $\mathbf{w}_0 = \mathbf{0}$. There are as many nodes at stage k as there are possible partial sums for the first k columns of the tables in S.

The network begins with the single node $(0, \mathbf{0})$ at stage 0. It is constructed recursively by specifying all successor nodes that are connected by arcs to each node $(k, \mathbf{w}_k)$ at stage k. These successor nodes have form $(k + 1, \mathbf{w}_{k+1})$

and satisfy, for $i = 1, \ldots, r$,

$$\max\left[w_{ik}, n_{i+} - \sum_{h=k+2}^{c} n_{+h}\right] \le w_{i,k+1} \le \min\left[n_{i+}, n_{+k} + w_{ik} - \sum_{h=1}^{i-1}(w_{h,k+1} - w_{hk})\right]. \quad (4.1)$$

When we apply (4.1) successively at stages $0, 1, \ldots, c - 1$, we end up with a single terminal node $(c, \mathbf{w}_c)$ at stage c, and the network is fully constructed.

For $0 \le j < k \le c$, a *path* connecting two nodes $(j, \mathbf{w}_j)$ and $(k, \mathbf{w}_k)$ is defined as a sequence of arcs of form $(j, \mathbf{w}_j) \to (j + 1, \mathbf{w}_{j+1}) \to \cdots \to (k, \mathbf{w}_k)$. A path through the entire network is a sequence of arcs $(0, \mathbf{0}) \to (1, \mathbf{w}_1) \to \cdots \to (c, \mathbf{w}_c)$ connecting the initial node to the terminal node. It represents a distinct $r \times c$ contingency table in S, with entries $(\mathbf{w}_{k+1} - \mathbf{w}_k)$ in column $k + 1$. Each table $\mathbf{x}$ in S corresponds to a single path through the network. This unique path can also be denoted by $\mathbf{x}$, and the network is equivalent to the reference set S.

It is necessary to order all paths through the network in accordance with the test statistic T, and to assign a hypergeometric probability to each one. To this end we define, for an arc connecting two nodes $(k, \mathbf{w}_k) \to (k + 1, \mathbf{w}_{k+1})$, the *rank length*

$$a_k(\mathbf{w}_k, \mathbf{w}_{k+1}) = v_{k+1} \sum_{i=1}^{r} u_i(w_{i,k+1} - w_{ik}),$$

and the *probability length*

$$b_k(\mathbf{w}_k, \mathbf{w}_{k+1}) = \left[\prod_{i=1}^{r}(w_{i,k+1} - w_{ik})!\right]^{-1}.$$

Let the rank length of any path be the sum of rank lengths of the arcs constituting the path, and let its probability length be the product of probability lengths of those same arcs. Then the rank length of any path $\mathbf{x}$ in S equals the test statistic $T = \sum\sum u_i v_j x_{ij}$, and its probability length equals the (unnormalized) hypergeometric probability $[\prod\prod x_{ij}!]^{-1}$. The null distribution of T is the same as the distribution of the rank lengths of all paths through the network. The tail area $\Pr(T \ge t)$ is proportional to the sum of the probability lengths of all paths through the network whose rank lengths are at least t.

The next two sections exploit the network representation of S to generate efficiently the null distribution of T, and its tail area.

4.2 Generating the Distribution of T

We have constructed the network such that the distribution of T is equivalent to the distribution of the rank lengths of paths through the network. This distribution is obtained by stagewise recursion. Consider each node $(k, \mathbf{w}_k)$ at stage k. Suppose the set of all paths connecting $(0, \mathbf{0})$ to $(k, \mathbf{w}_k)$ yields $q(k, \mathbf{w}_k)$ distinct rank lengths, denoted by $\{t_{hk}; h = 1, 2, \ldots, q(k, \mathbf{w}_k)\}$. Let C_{hk} be the sum of probability lengths of all paths from $(0, \mathbf{0})$ to $(k, \mathbf{w}_k)$ that have the same t_{hk}. Thus, the set of *records* $\Omega(k, \mathbf{w}_k) = \{(t_{hk}, C_{hk}); h = 1, \ldots, q(k, \mathbf{w}_k)\}$ is the probability distribution of rank lengths of all paths from $(0, \mathbf{0})$ to $(k, \mathbf{w}_k)$, up to a normalizing constant. If $\Omega(k, \mathbf{w}_k)$ is known for each node at stage k, the following recursive procedure generates $\Omega(k + 1, \mathbf{w}_{k+1})$ for each node at stage $k + 1$.

Step 1. Select a record t_{hk} in $\Omega(k, \mathbf{w}_k)$.

Step 2. Transmit a copy of this record to each successor node $(k + 1, \mathbf{w}_{k+1})$, where the successors are identified by (4.1).

Step 3. At each successor node $(k + 1, \mathbf{w}_{k+1})$, transform the transmitted record to (t^*, C^*), where $t^* = t_{hk} + a_k(\mathbf{w}_k, \mathbf{w}_{k+1})$ and $C^* = C_{hk} b_k(\mathbf{w}_k, \mathbf{w}_{k+1})$.

Step 4. Insert (t^*, C^*) into $\Omega(k + 1, \mathbf{w}_{k+1})$ as follows. (a) If there already exists a record $(t_{h,k+1}, C_{h,k+1})$ in $\Omega(k + 1, \mathbf{w}_{k+1})$ such that $t^* = t_{h,k+1}$, then merge the two records by replacing $(t_{h,k+1}, C_{h,k+1})$ with $(t_{h,k+1}, C^* + C_{h,k+1})$ in $\Omega(k + 1, \mathbf{w}_{k+1})$. (b) If no record currently in $\Omega(k + 1, \mathbf{w}_{k+1})$ has rank length equal to t^*, then augment $\Omega(k + 1, \mathbf{w}_{k+1})$ by adding (t^*, C^*) to it as a new record.

The technique of hashing with a linear probe (Sedgewick 1983, p. 201) is used to search for matches and either merge or augment records in $\Omega(k + 1, \mathbf{w}_{k+1})$. This ensures an optimum trade-off between efficient use of available memory and fast search.

Step 5. Return to Step 1 and select another record from $\Omega(k, \mathbf{w}_k)$. Repeat Steps 2–5 with it. When every record in $\Omega(k, \mathbf{w}_k)$ has been processed, select another node at stage k and repeat Steps 1–5 with it. When all nodes at stage k have been exhausted, increment k by 1 and repeat Steps 1–5 for the new stage k.

Starting with $\Omega(0, \mathbf{0}) = (0, 1)$ and moving through stages $0, 1, \ldots, c - 1$ in accordance with these five steps, we end up with $\Omega(c, \mathbf{w}_c)$ as the null distribution of T, up to a normalizing constant. To keep notation consistent with that introduced in Section 3, we hereafter drop the subscripts representing stage c and denote the null distribution of T by the set of records $\Omega(c, \mathbf{w}_c) = \{(t, C_t); t = t_1, \ldots, t_N\}$. To compute the distribution of T when $\beta \ne 0$, we multiply each coefficient C_t by $\exp(\beta t)$, and divide by the normalizing constant $\sum_u C_u \exp(\beta u)$. When $\beta = 0$, the normalizing constant is the sum of rank lengths of all paths through the network, which equals $n!/[\prod\prod (n_{i+}!)(n_{+j}!)]$ (Mehta and Patel 1983).

Step 1 relates to the efficiency of this algorithm. The more often records merge with other records, the fewer records remain for processing at later stages. We refer to the merging of two records as *clubbing*. The extent of clubbing depends on the choice of scores $\{u_i\}$ and $\{v_j\}$. Greater clubbing results in fewer possible values for T. For *any* table and sets of scores, nrc is a crude upper bound for the number of possible values of T; $u_i v_j$ can assume rc values, and the sum of this cross-product can take at most nrc values for the n subjects in the sample. Maximum clubbing occurs for equal-interval scores, such as $\{u_i = i\}$ and $\{v_j = j\}$. For this case, suppose $r = c$ and $n_{i+} = n_{+j} = n/r$ for all i and j. It follows from theorems in the next section that the maximum value of T occurs when $\{n_{ii} = n/r\}$ and the minimum occurs when $\{n_{i,r-i+1} = n/r\}$. Thus,

T is bounded below by $(n/r)\sum[i(n - r + i)] = n[(r + 1)^2/2 - (r + 1)(2r + 1)/6]$ and bounded above by $(n/r)\sum i^2 = n(r + 1)(2r + 1)/6$. For these scores, T can assume each integer value between the bounds. Thus, the number of possible values for T equals $n(r^2 - 1)/6 + 1$.

4.3 Computing the Tail Area of the Distribution of T

Often we are not interested in the entire distribution of T, but merely need to compute a P value, such as $\Pr(T \geq t)$. It is then possible to improve the network algorithm considerably by conducting a *shortest path* and a *longest path* test on each record (t^*, C^*) prior to its insertion into $\Omega(k + 1, \mathbf{w}_{k+1})$. If the record passes either test, it is dropped from further processing, thereby reducing the remaining work.

Define $SP(k + 1, \mathbf{w}_{k+1})$ as the shortest rank length and $LP(k + 1, \mathbf{w}_{k+1})$ as the longest rank length among all paths from node $(k + 1, \mathbf{w}_{k+1})$ to the terminal node $(c, \mathbf{w}_c)$. Also define $TP(k + 1, \mathbf{w}_{k+1})$ as the sum of the probability lengths of all paths from node $(k + 1, \mathbf{w}_{k+1})$ to node $(c, \mathbf{w}_c)$. By a slight modification of the method in the appendix of Mehta and Patel (1983), we can show that

$$TP(k + 1, \mathbf{w}_{k+1}) = \frac{\left(n - \sum_{j=1}^{k+1} n_{+j}\right)!}{\prod_{j=k+2}^{c} n_{+j}! \prod_{i=1}^{r} (w_{ic} - w_{i,k+1})!}.$$

Now, perform the longest path test on record (t^*, C^*). If

$$t^* + LP(k + 1, \mathbf{w}_{k+1}) < t, \tag{4.2}$$

then no completion of the record can exceed t. Therefore, this record cannot contribute to the P value, and can be dropped from $\Omega(k + 1, \mathbf{w}_{k+1})$.

If (4.2) is not satisfied, perform the shortest path test on (t^*, C^*). If

$$t^* + SP(k + 1, \mathbf{w}_{k+1}) \geq t, \tag{4.3}$$

then every completion of the current record must exceed t and thereby contribute to the P value. Of all of these completions of the current record, the total contribution to the P value can be shown to equal

$$C^* \times TP(k + 1, \mathbf{w}_{k+1})(n!)^{-1} \prod_{i=1}^{r} n_{i+}! \prod_{j=1}^{c} n_{+j}!. \tag{4.4}$$

The record can be dropped after adding its contribution (4.4) to the desired tail probability. If neither Test (4.2) nor Test (4.3) is satisfied, we insert the record (t^*, C) into $\Omega(k + 1, \mathbf{w}_{k+1})$ and proceed with the algorithm described in Section 4.2.

It remains only to obtain $LP(k + 1, \mathbf{w}_{k+1})$ and $SP(k + 1, \mathbf{w}_{k+1})$. The following two theorems show how these quantities can be rapidly computed from any node of the network.

Theorem 1. Let $u_1 < u_2 < \cdots < u_r$ and $v_1 < v_2 < \cdots < v_c$ be fixed. To maximize $\sum\sum u_i v_j x_{ij}$ subject to $x_{ij} \geq 0$ and integer, $x_{i+} = n_{i+}$, $x_{+j} = n_{+j}$ for $i = 1, \ldots, r$ and $j = 1, \ldots, c$, an optimal solution must satisfy $x_{11} = \min(n_{1+}, n_{+1})$.

Theorem 2. Let $u_1 < u_2 < \cdots < u_r$ and $v_1 < v_2 < \cdots < v_c$ be fixed. To minimize $\sum\sum u_i v_j x_{ij}$ subject to $x_{ij} \geq 0$ and integer, $x_{i+} = n_{i+}$, $x_{+j} = n_{+j}$ for $i = 1, \ldots, r$ and $j = 1, \ldots, c$, an optimal solution must satisfy $x_{1c} = \min(n_{1+}, n_{+c})$.

The proofs are given in the Appendix. Theorem 1 can be used repeatedly to determine the unique optimal solution $\{x_{ij}^*\}$ that maximizes $\sum\sum u_i v_j x_{ij}$. The flow chart in Figure 1 shows how. By applying the flow chart at each node $(k, \mathbf{w}_k)$, we can easily compute $LP(k, \mathbf{w}_k)$. In a similar manner, Theorem 2 is used to compute $SP(k, \mathbf{w}_k)$.

5. COMMENTS

For large samples with Model (1.1), we can test independence using the statistic $z = \hat{\beta}/\mathrm{ASE}$, obtaining the estimated asymptotic standard error (ASE) from the inverse information matrix. Alternatively, we could use a likelihood-ratio test, or a statistic based on the efficient score,

$$z = \left[T - \sum\sum u_i v_j (n_{i+} n_{+j}/n)\right] \div \left\{\left[\sum u_i^2 n_{i+} - \left(\sum u_i n_{i+}\right)^2/n\right] \times \left[\sum v_j^2 n_{+j} - \left(\sum v_j n_{+j}\right)^2/n\right]\right\}^{1/2}.$$

The exact test given in this article should be used when the sample may be too small to use such tests. Guidelines for validity of these large-sample tests are not yet well developed, though the tests apply for smaller samples and sparser tables than tests of fit (Agresti and Yang 1987). We note that ML estimates of parameters do not exist when T assumes its maximum or minimum possible value for the given margins. In that case, our methodology still yields an exact P value and produces a confidence interval for β of form (β_-, ∞) or $(-\infty, \beta_+)$.

For fixed sample size n, the number of tables in S increases rapidly as r and c increase or as the row and column totals become more homogeneous. Gail and Mantel (1977) and Good (1976) gave approximations for the cardinality of S, and Agresti and Wackerly (1977) gave upper bounds for several table dimensions and sample sizes. Using a FORTRAN 77 program on a high-speed computer to implement our algorithm, we can handle tables of moderate dimensions for which asymptotics may not apply. In fact, we can use a personal computer when the order of the cardinality of S is no more than about 10^6 or 10^7. The first two columns of Table 1 illustrate the maximum central processing unit (CPU) time the algorithm needed to compute the entire distribution of T, using MS-DOS with an IBM PS 2 (model 70) personal computer, running at 20 megahertz with a math coprocessor. For each r, c, and n these maxima occur for tables having homogeneous marginal totals. The times represent the maximum CPU time needed to construct an exact confidence interval for β, or to calculate a moderate-sized P value. Results are reported for both equally spaced row and column scores and the

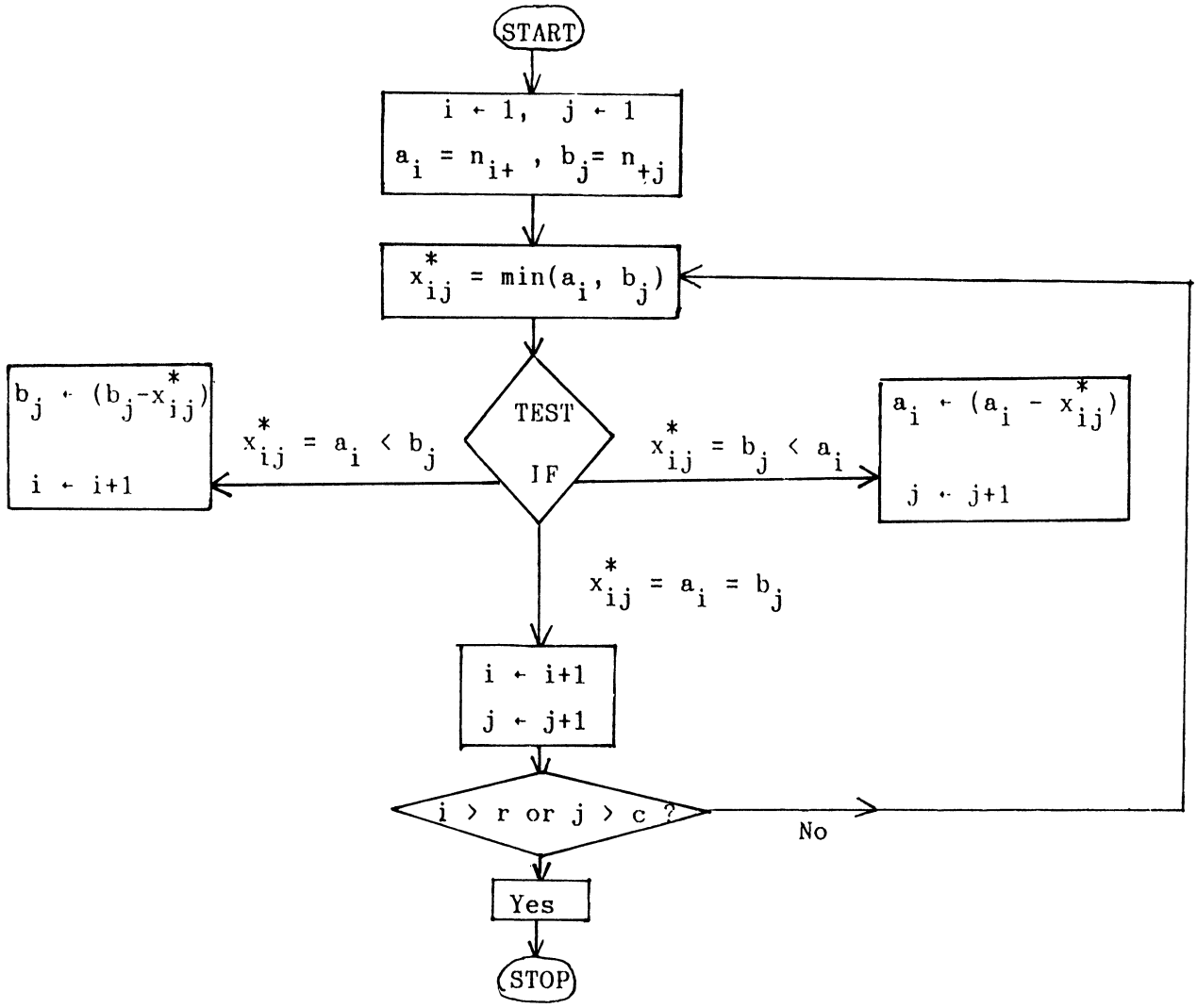

Figure 1. Flow Chart for Determining $\{x^*_{ij}\}$ That Maximize T.

unequal spacings $\{u_i = 2^{i-1}\}$ and $\{v_j = 3^{j-1}\}$. For times not reported, the memory limitation of 640K under MS-DOS was insufficient. These cases can be handled using computers with larger memory, however.

The CPU times are faster when we only need a P value, and that P value is small, since the algorithm then exploits the longest and shortest path tests. To illustrate, Table 1 also reports CPU times for calculating a P value of about .05, when the sample counts have a uniform association pattern and homogeneous marginal totals. The table also reports CPU times when T takes its maximum value, again for the worst case of homogeneous marginal totals. The CPU times are much faster when the marginal counts are highly nonuniform. For instance, a 5 × 5 table with marginal totals of (96, 1, 1, 1, 1) took five seconds. For such highly unbalanced cases, one is especially wary of using asymptotic approximations. For tables in which the algorithm is infeasible, one could use an importance-sampling algorithm (such as in Mehta, Patel, and Senchaudhuri 1988) to estimate precisely the exact P value.

Table 1. Maximum CPU Time (seconds) to Compute the Distribution of T or a P Value on an IBM PS 2 (model 70) Personal Computer

		Distribution		P = .05		Extreme P value	
		Score spacing		Score spacing		Score spacing	
Dimension	Sample size	Equal	Unequal	Equal	Unequal	Equal	Unequal
3 × 3	10	2	2	1	1	1	1
	25	5	6	3	4	1	1
	40	16	18	13	15	1	1
4 × 4	10	5	6	2	3	1	1
	25	190	340	97	248	1	1
	40	15,730	—	1,430	4,779	3	3
5 × 5	10	17	41	3	5	2	1
	25	3,290	—	834	3,925	3	3
	40	—	—	—	—	6	5

NOTE: A dash represents infeasibility under MS-DOS because of its 640K memory limitation.

Our procedures assume Model (1.1) is correct, but they would normally be applied in situations where sparseness makes model checking difficult. Regardless of whether the model holds, the P value based on T is a valid way to summarize a type of evidence against the hypothesis of independence. When the association contains a strong monotone component, the exact test using T is usually more powerful than exact tests that ignore the ordering. Confidence intervals for odds ratios no longer apply when the model does not hold, however, and the extent of de-

parture from (1.1) determines their validity as approximations. The smoothing induced by the model does imply that for small samples, model-based conditional ML estimates of odds ratios tend to be better than empirical estimates corresponding to the fit of the saturated model.

One can construct an exact goodness-of-fit test of the model by also conditioning on T. Given $\{n_{it}\}$, $\{n_{tj}\}$, and T and given the fitted values for Model (1.1) that these sufficient statistics determine, one could evaluate a goodness-of-fit statistic (such as the Pearson chi-squared statistic) for all tables having those sufficient statistic values. The P value for testing the model is then the conditional probability that the value of the goodness-of-fit statistic is at least as large as the observed one.

Some results in this article extend naturally to multiway tables with ordinal classifications. We are currently generalizing the network algorithm to handle exact analyses for multiway tables with ordinal variables.

Upon receipt of a floppy diskette, we will be pleased to supply an executable code for performing the exact two-way analyses described in this article on IBM-PC-compatible machines. A more complete implementation of the algorithm will be available subsequently in the StatXact package distributed by Cytel Software Corporation.

APPENDIX: PROOF OF THEOREMS

Theorem 1. Any feasible solution $\{x_{ij}\}$ must have $x_{11} \leq \min(n_{1+}, n_{+1})$. Suppose $x_{11} < \min(n_{1+}, n_{+1})$. Since $x_{1+} = n_{1+}$, there exists $k \neq 1$ such that $x_{1k} \geq 1$. Also, since $x_{+1} = n_{+1}$, there exists $h \neq 1$ such that $x_{h1} \geq 1$. Now, consider feasible solution $\{x'_{ij}\}$ with $x'_{11} = x_{11} + 1$, $x'_{1k} = x_{1k} - 1$, $x'_{h1} = x_{h1} - 1$, $x'_{hk} = x_{hk} + 1$, and $x'_{ij} = x_{ij}$ for all other (i, j). Now,

$$\sum\sum u_i v_j x'_{ij} = \sum\sum u_i v_j x_{ij} + (u_h - u_1)(v_k - v_1).$$

Since $(u_h - u_1)(v_k - v_1) > 0$, $\sum\sum u_i v_j x'_{ij} > \sum\sum u_i v_j x_{ij}$, and thus $\{x_{ij}\}$ cannot be an optimal solution.

We have established that $x_{11} = \min(n_{1+}, n_{+1})$ is a necessary condition for an optimal solution. Now there is only a finite number of nonnegative integer arrays $\{x_{ij}\}$ satisfying $x_{i+} = n_{i+}$ and $x_{+j} = n_{+j}$ all i and j. One or more of them must maximize $\sum\sum u_i v_j x_{ij}$, thereby guaranteeing that an optimal solution exists.

The proof of Theorem 2 is similar to that of Theorem 1.

[*Received April 1989. Revised October 1989.*]

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis*, New York: John Wiley.

Agresti, A., and Wackerly, D. (1977), "Some Exact Conditional Tests of Independence for $R \times C$ Cross-Classification Tables," *Psychometrika*, 42, 111–125.

Agresti, A., Wackerly, D., and Boyett, J. (1979), "Exact Conditional Tests for Cross-Classifications: Approximation of Attained Significance Levels," *Psychometrika*, 44, 75–83.

Agresti, A., and Yang, M. (1987), "An Empirical Investigation of Some Effects of Sparseness in Contingency Tables," *Computational Statistics & Data Analysis*, 5, 9–21.

Cohen, A., and Sackrowitz, H. B. (1988), "Tests for Independence in Contingency Tables With Ordered Categories," unpublished manuscript.

Cornfield, J. (1956), "A Statistical Problem Arising From Retrospective Studies," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 4), ed. J. Neyman, Berkeley: University of California Press, pp. 135–148.

Gail, M., and Mantel, N. (1977), "Counting the Number of $r \times c$ Contingency Tables With Fixed Margins," *Journal of the American Statistical Association*, 72, 859–862.

Gilula, Z., Krieger, A., and Ritov, Y. (1988), "Ordinal Association in Contingency Tables: Some Interpretive Aspects," *Journal of the American Statistical Association*, 83, 540–545.

Good, I. J. (1976), "On the Application of Symmetric Dirichlet Distributions and Their Mixtures to Contingency Tables," *The Annals of Statistics*, 4, 1159–1189.

Graubard, B. I., and Korn, E. L. (1987), "Choice of Column Scores for Testing Independence in Ordered $2 \times K$ Contingency Tables," *Biometrics*, 43, 471–476.

Hirotsu, C. (1982), "Use of Cumulative Efficient Scores for Testing Ordered Alternatives in Discrete Models," *Biometrika*, 69, 567–577.

Mantel, N. (1963), "Chi-Square Tests With One Degree of Freedom: Extensions of the Mantel–Haenszel Procedure," *Journal of the American Statistical Association*, 58, 690–700.

McCullagh, P. (1980), "Regression Models for Ordinal Data" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 42, 109–142.

Mehta, C. R., and Patel, N. R. (1983), "A Network Algorithm for Performing Fisher's Exact Test in $r \times c$ Contingency Tables," *Journal of the American Statistical Association*, 78, 427–434.

Mehta, C. R., Patel, N. R., and Senchaudhuri, P. (1988), "Importance Sampling for Estimating Exact Probabilities in Permutational Inference," *Journal of the American Statistical Association*, 83, 999–1005.

Mehta, C. R., Patel, N. R., and Tsiatis, A. A. (1984), "Exact Significance Testing to Establish Treatment Equivalence With Ordered Categorical Data," *Biometrics*, 40, 819–825.

Patefield, W. M. (1982), "Exact Tests for Trends in Ordered Contingency Tables," *Applied Statistics*, 31, 32–43.

Sedgewick, R. (1983), *Algorithms*, Reading, MA: Addison-Wesley.

Yates, F. (1948), "The Analysis of Contingency Tables With Grouping Based on Quantitive Characters," *Biometrika*, 35, 176–181.