Improved Exact Inference About Conditional Association in Three-Way Contingency Tables

Author(s): Donguk Kim and Alan Agresti

Source: *Journal of the American Statistical Association*, Vol. 90, No. 430 (Jun., 1995), pp. 632-639
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2291076
Accessed: 20-11-2024 18:42 UTC

Improved Exact Inference About Conditional Association in Three-Way Contingency Tables

Donguk KIM and Alan AGRESTI*

We propose modified exact inferential methods for contingency tables. Ordinary "exact" inference is conservative, because of the discreteness. For estimating a common odds ratio in several 2×2 tables, two modifications of the ordinary "exact" confidence interval maintain at least a fixed confidence level but tend to be much narrower. One approach inverts results of a test with a modified P value utilizing the test statistic and table probabilities. The second approach inverts one two-sided test rather than two one-sided tests. This approach is much less conservative when the true odds ratio is relatively small or large. We also generalize results of Cohen and Sackrowitz and relate modified P values to construction of exact, unbiased, and admissible tests for an ordinal alternative to conditional independence.

KEY WORDS: Confidence interval; Fisher's exact test; Linear-by-linear association; Mid P value; Odds ratio; P value.

1. INTRODUCTION

The title of this article may seem like a contradiction. How can one improve on a procedure that is "exact?" The improvement refers to decreasing the conservativeness that occurs due to discreteness. For instance, we present a $2 \times 2 \times 5$ table for which the ordinary 95% confidence interval for an assumed common odds ratio is (1.1, 531.5). The discreteness implies that .95 is a lower bound for the actual confidence coefficient. A modified confidence interval that we propose also guarantees at least 95% confidence but takes the much shorter range (2.1, 67.3). Our approach is applicable for any contingency table, but we illustrate the arguments for inferences about conditional associations in three-way contingency tables.

For three-way $I \times J \times K$ tables, let X, Y, and Z denote the row, column, and layer classifications. Let $\mathbf{N} = \{n_{ijk}\}$ denote the cell counts, with expected frequencies $\{m_{ijk}\}$. The data can follow any of the standard sampling models, such as multinomial over the entire table, or independent multinomial within each level of Z or each combination of levels of X and Z. The log-linear model of no three-factor interaction is

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}. \quad (1)$$

Conditional independence of X and Y, given Z, is the special case $\{\lambda_{ij}^{XY} = 0\}$. Exact conditional inference about $\{\lambda_{ij}^{XY}\}$ utilizes the distribution of $\{n_{ij+}\}$, the sufficient statistics for these parameters, conditional on the sufficient statistics $\{n_{i+k}\}$ and $\{n_{+jk}\}$ for the remaining parameters ("+" denotes the sum over the index it replaces).

We present exact inference for this model and also for a narrower one relevant when X or Y are ordinal. Our procedures are adaptations of the ordinary exact conditional methods that are less conservative. Section 2 discusses a modified P value, proposed in a related context by Cohen and Sackrowitz (1992), and shows that its distribution can be much less discrete than that of the ordinary P value. Section 3 proposes a modified "exact" confidence interval for a common odds ratio in several 2×2 tables, based on inverting two one-sided tests using the modified P value. Section 4 presents another "exact" confidence interval, based on inverting a two-sided test. Though these two types of modified intervals are also conservative, they may be much narrower than the ordinary "exact" interval. In particular, the ordinary one is based on two separate one-sided tests. Approaches based on two-sided tests are usually much better, especially when the true odds ratio is relatively small or relatively large.

Section 5 generalizes to three-way tables some results of Cohen and Sackrowitz (1991, 1992) regarding admissibility of tests for two-way tables. The ordinary test of conditional independence for $2 \times 2 \times K$ tables is usually inadmissible. One can use the modified P value to reduce the degree of supplementary randomization while achieving admissibility.

2. A LESS CONSERVATIVE *P*-VALUE

Suppose that one plans to conduct an exact conditional test for categorical data using some preassigned size α, such as .05. It is not usually possible to construct a critical region having that size, because of the discreteness of the distribution. One can artificially achieve the nominal size by performing supplementary randomization in making the decision about whether to reject when a table occurs at the boundary of the critical region (Lehmann 1986, p. 135). In practice, of course, such randomization is unacceptable. For a particular test, let T denote the test statistic and let t_o denote its observed value. When large values of T contradict the null, the usual P value is

$$P = P_{H_0}(T \geq t_o).$$

To make a formal decision about H_0, one rejects if $P \leq \alpha$. The discreteness implies that the actual size is no greater than α, both conditionally and unconditionally.

The exact conditional approach conditions on sufficient statistics for unknown parameters to eliminate them (Agresti 1992). The extra conditioning reduces the set of possible test statistic values, making the distribution more highly dis-

* Donguk Kim is a graduate student and Alan Agresti is Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. This work was partially supported by National Institutes of Health Grant GM 43824. The authors appreciate helpful suggestions and comments from James Booth, Myron Chang, Joseph Gastwirth, Keith O'Rourke, and Ming-Chung Yang. They are grateful to S. Vollset and K. Hirji for supplying a FORTRAN program.

Journal of the American Statistical Association
June 1995, Vol. 90, No. 430, Theory and Methods

crete. Hence tests of nominal size α based on the exact conditional P value can have actual size considerably less than that level (see, for example, Suissa and Shuster 1985). This problem is exacerbated by the overemphasis on testing at sacred levels, such as .05. One can argue for simply reporting the P value and not making comparisons to such arbitrary levels, particularly when data are discrete. But the discreteness also affects interval estimation. If one constructs an "exact" confidence interval with nominal confidence coefficient $1 - \alpha$, then the actual confidence coefficient is *at least* that level and is unknown (Neyman 1935).

To reduce the conservativeness, one can utilize a modified P value based on a less discrete distribution (Cohen and Sackrowitz 1992). The modified P value uses a partition of the sample space that is more refined than using T alone. Within fixed values of T, a secondary partitioning utilizes the null table probability. For a given value of T, tables that are less likely under the null are considered to give greater evidence against the null. Let Γ denote the set of all tables that have the same values for the sufficient statistics fixed by the conditional test. Let $B = \{\mathbf{Z}: \mathbf{Z} \in \Gamma, T = t_o, P(\mathbf{Z}) \le P(\mathbf{N})\}$, where the probabilities are computed under the null. The modified P value is

$$P^* = P_{H_0}(T > t_o) + P_{H_0}(B). \qquad (2)$$

Cohen and Sackrowitz (1992) used this P value for ordinal tests in two-way tables. Streitberg and Roehmel (1990) considered a related idea for nonparametric tests, forming a P value by partitioning all data sets having a primary statistic value according to some secondary statistic that also describes departure from the null.

One can calculate P^* for any test statistic having a discrete distribution, because it satisfies $P_{H_0}(P^* \le \alpha) \le \alpha$ for $0 < \alpha < 1$. The modified P value cannot exceed the usual one, so the test of fixed size based on it is less conservative. When each table with a fixed value of T has the same probability, P^* equals the usual P value. When only one table has each distinct value of T, such as in the one-sided version of Fisher's exact test using $T = n_{11}$, they are identical.

We illustrate the modified P value (2) using Table 1, taken from Mantel (1963). It refers to the effectiveness of immediately injected or $1\frac{1}{2}$ hour-delayed penicillin in protecting rabbits against lethal injection with β-hemolytic streptococci. Assuming a constant odds ratio θ for the five partial tables, we test H_0: $\theta = 1$ against H_a: $\theta > 1$. The test statistic is $T = \sum_k n_{11k}$, given the marginal totals at each penicillin level (Birch 1964). Rejecting the null for large values of T gives a test that is uniformly most powerful (UMP) unbiased of its size (Lehmann 1986, p. 163). For the first and last table, the zero marginal count implies that the conditional distribution of n_{11k} is degenerate, and the table makes no contribution to the test. Therefore, we can conduct the test using the three remaining tables. For these tables, $t_o = 14$, and the four tables with $T \ge 14$ are $\{(n_{111}, n_{112}, n_{113}) = (3, 6, 6), (2, 6, 6), (3, 5, 6), (3, 6, 5)\}$. The ordinary exact P value is $P = P_{H_0}(T \ge 14) = (2 + 9 + 16 + 2)/1452 = .0200$. The modified exact P value is $P^* = (2 + 2)/1452 = .0028$, the null probability for the tables $\{(3, 6, 6), (3, 6, 5)\}$. The second example uses Table 2, the "crying babies" data given by Cox (1970, p. 5), a $2 \times 2 \times 18$ table. Here $P = .045$ and $P^* = .021$.

Table 1. Example 1 for Exact Analyses

Penicillin level	Delay	Response: Cured	Response: Died
1/8	None	0	6
	1-1/2 hour	0	5
1/4	None	3	3
	1-1/2 hour	0	6
1/2	None	6	0
	1-1/2 hour	2	4
1	None	5	1
	1-1/2 hour	6	0
4	None	2	0
	1-1/2 hour	5	0

Source: Mantel (1963).

Table 2. Example 2 for Exact Analyses

Day	Treated: Not crying	Treated: Crying	Control: Not crying	Control: Crying
1	1	0	3	5
2	1	0	2	4
3	1	0	1	4
4	0	1	1	5
5	1	0	4	1
6	1	0	4	5
7	1	0	5	3
8	1	0	4	4
9	1	0	3	2
10	0	1	8	1
11	1	0	5	1
12	1	0	8	1
13	1	0	5	3
14	1	0	4	1
15	1	0	4	2
16	1	0	7	1
17	0	1	4	2
18	1	0	5	3

Source: Cox (1970).

There can be a considerable discrepancy between the behavior of the ordinary and modified "exact" P values, the modified P value having a distribution that can be much less discrete. For Table 1, the total number of possible P values equals 9 for the ordinary P value and 35 for the modified P value. For Table 2, the corresponding numbers are 19 and 13,110. When a test statistic has a continuous distribution, the P value has a uniform(0, 1) null distribution. In the discrete case, the P value is stochastically larger than the uniform. Figure 1 presents the cumulative distribution functions of the ordinary and modified exact P values for null distributions based on the fixed margins of Table 2. The modified cdf is practically indistinguishable from the uniform.

One can summarize the conservativeness of a P value by comparing $E_{H_0}(P \text{ value})$ to the uniform mean of .5. For the

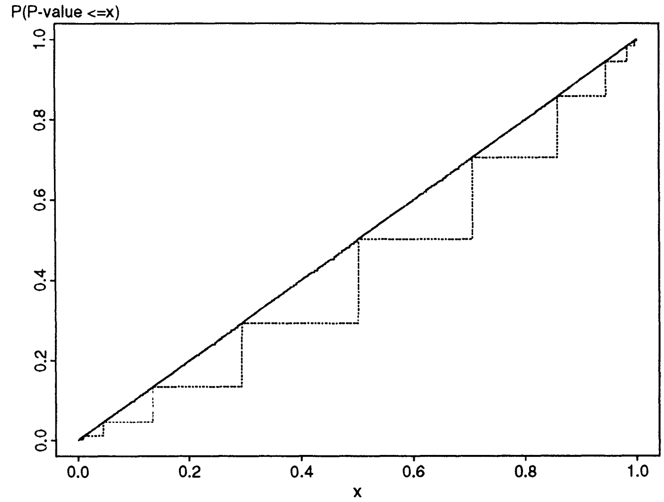

Figure 1. Cumulative Distribution Functions of Exact P Values for the Margins of Table 2: ——, Modified P Value; - - -, Ordinary P Value.

conditional distribution based on the fixed margins of Table 1, $E_{H_0}P = .611$ and $E_{H_0}P^* = .542$. For Table 2, $E_{H_0}P = .576$ and $E_{H_0}P^* = .501$. A P value corresponding to a test using supplementary randomization has the form

$$P_u = P_{H_0}(T > t_o) + UP_{H_0}(T = t_o), \tag{3}$$

where U denotes a uniform(0, 1) random variable. The mid P value (Lancaster 1961), defined by

$$P_{\text{mid}} = P_{H_0}(T > t_o) + (1/2)P_{H_0}(T = t_o),$$

replaces U by its expectation. Both these P values have null expected value of .5. The mid P value and the modified P value attempt to reduce conservativeness without the arbitrariness of supplementary randomization. A disadvantage of the mid P value is that tests or confidence intervals based on it are not "exact," the actual size possibly exceeding the nominal value.

The mid P value assigns weight 1/2 to probabilities of tables comparable to the observed table in the sense that $T = t_o$. For the modified P value (2), the comparable tables are those with $T = t_0$ and $P(\mathbf{Z}) = P(\mathbf{N})$. Thus a mid–$P$ value version of the modified P value (2) is

$$P^*_{\text{mid}} = P^* - \frac{1}{2}P_{H_0}(\{T = t_o, P(\mathbf{Z}) = P(\mathbf{N})\}). \tag{4}$$

The modified mid–P value also has null expected value equal to .5. Note that $(P^* - P^*_{\text{mid}}) \leq (P - P_{\text{mid}})$. For Table 1, $P_{\text{mid}} = .011$ and $P^*_{\text{mid}} = .002$. For Table 2, $P_{\text{mid}} = .028$ and $P^*_{\text{mid}} = .021$.

3. A LESS CONSERVATIVE "EXACT" CONFIDENCE INTERVAL

One can construct an "exact" confidence interval for a parameter by inverting an exact conditional test about that parameter. To construct a narrower interval having actual confidence coefficient closer to the nominal value, one can invert the test based on the modified P value.

We illustrate with estimation of an assumed common odds ratio, θ, in $2 \times 2 \times K$ contingency tables. The conditional probability of any table in the reference set, Γ, is

$$P(\mathbf{N}; \theta) = P(\{n_{11k}\} \mid \{n_{1+k}\}, \{n_{+1k}\}, \{n_{+2k}\}; \theta)$$
$$= \frac{\Pi_k \binom{n_{+1k}}{n_{11k}} \binom{n_{+2k}}{n_{1+k} - n_{11k}} \theta^{n_{11k}}}{\sum_{\mathbf{Z}\in\Gamma} \Pi_k \binom{n_{+1k}}{z_k} \binom{n_{+2k}}{n_{1+k} - z_k} \theta^{z_k}}, \tag{5}$$

where $\{z_1, \ldots, z_K\}$ denote values of $\{n_{111}, \ldots, n_{11K}\}$ for a table in Γ. Let $\Gamma_t = \{\mathbf{Z}: \mathbf{Z} \in \Gamma, \sum_k n_{11k} = t\}$. Ordinary exact confidence limits for the common odds ratio are constructed from the conditional distribution of $T = \sum_k n_{11k}$, which is

$$P(T = t; \theta) = \frac{c_t \theta^t}{\sum_{u=t_{\min}}^{t_{\max}} c_u \theta^u}, \tag{6}$$

where

$$c_t = \sum_{\mathbf{Z}\in\Gamma_t} \Pi_k \binom{n_{+1k}}{z_k} \binom{n_{+2k}}{n_{1+k} - z_k}$$

and where $t_{\min} = \sum_k \max(0, n_{1+k} - n_{+2k})$ and $t_{\max} = \sum_k \min(n_{1+k}, n_{+1k})$. The ordinary interval (Cox 1970; Gart 1970; Mehta, Patel, and Gray 1985; Vollset, Hirji, and Elashoff 1991) is based on inverting two separate one-sided tests. It equals (θ_-, θ_+), where for $t_{\min} \leq t_o \leq t_{\max}$,

$$\sum_{t\geq t_o} P(t; \theta_-) = \alpha/2$$

and

$$\sum_{t\leq t_o} P(t; \theta_+) = \alpha/2. \tag{7}$$

When $t_o = t_{\min}$, the lower endpoint is zero; if $t_o = t_{\max}$, the upper endpoint is ∞.

To obtain a narrower "exact" interval, we invert the one-sided tests using the modified exact P value. Let $B(\theta) = \{\mathbf{Z}: \mathbf{Z} \in \Gamma_{t_o}, P(\mathbf{Z}; \theta) \leq P(\mathbf{N}; \theta)\}$. The modified "exact" confidence limits are found using the functions

$$P^*_1(\theta) = \sum_{t>t_o} P(t; \theta) + P[B(\theta); \theta]$$

and

$$P^*_2(\theta) = \sum_{t<t_o} P(t; \theta) + P[B(\theta); \theta]. \tag{8}$$

The lower limit, θ^*_-, is the smallest θ to satisfy $P^*_1(\theta) \geq \alpha/2$, and the upper limit, θ^*_+, is the largest θ to satisfy $P^*_2(\theta) \geq \alpha/2$. When $P^*_1(\theta)$ and $P^*_2(\theta)$ are strictly monotone, $P^*_1(\theta^*_-) = P^*_2(\theta^*_+) = \alpha/2$.

Appendix A shows that the probability that this interval excludes θ is at most α. It is contained within the ordinary one. Hence the modified confidence interval is "exact," yet it has actual confidence coefficient closer to the nominal value than the ordinary "exact" interval. One can solve for the modified endpoints numerically, using the ordinary endpoints as initial values. When the ordinary and modified P values are identical, the observed table has the largest null probability among tables having $T = t_o$. From (5), one can show that it also has the largest probability among those tables having $T = t_o$ for arbitrary θ. Hence $P(T = t_o; \theta) = P[B(\theta); \theta]$, and the ordinary and modified confidence intervals are identical.

The 95% "exact" interval for a common odds ratio using the ordinary approach is (1.08, 531.51) for Table 1 and (.86, 21.37) for Table 2. The modified "exact" confidence interval is (2.08, 67.35) for Table 1 and (1.04, 14.87) for Table 2. Inferences can be considerably sharper with the modified approach. For Table 1, the ordinary lower bound indicates that the true odds ratio could be quite close to conditional independence. The modified interval suggests that the odds ratio is substantively different from conditional independence.

4. "EXACT" CONFIDENCE INTERVALS BASED ON TWO-SIDED TESTS

Confidence intervals discussed so far are based on inverting two separate one-sided tests each of level $\alpha/2$. An alternative approach forms confidence intervals by inverting a single two-sided test, rather than two one-sided tests. Sterne (1954) used this approach in constructing confidence intervals for a binomial parameter, and Baptista and Pike (1977) used it to construct confidence limits for odds ratios in 2×2 tables. This approach extends directly to common odds ratios in $2 \times 2 \times K$ tables.

For testing a particular value of θ, a two-sided P value is given by

$$P(\theta) = \sum_{\{t: P(t;\theta) \leq P(t_o;\theta)\}} P(t; \theta). \quad (9)$$

When the distribution of T has probabilities increasing in t up to some point and then decreasing after that, this is simply a two-tail probability. (This has happened for all examples we have considered.) The two-sided exact confidence interval consists of all θ for which this two-sided P value equals at least α. Alternatively, one could base the two-sided P value on a nonnull test statistic and construct the confidence interval by inverting that test.

This two-sided approach produces an interval that is usually shorter than the ordinary one based on inverting two separate one-sided tests. Under certain conditions, the two-sided approach is necessarily better, at least for one of the endpoints. For instance, when the upper limit θ_+ of this interval is quite large, the distribution of T often satisfies $P(t; \theta_+) > P(t_o; \theta_+)$ for all $t > t_o$. A special case of this holds

Table 3. Various 95% Confidence Intervals for the Common Odds Ratio

Inverted test	Data set 1	Data set 2
Ordinary one-sided	(1.08, 531.51)	(.86, 21.37)
Modified one-sided	(2.08, 67.35)	(1.04, 14.87)
Ordinary two-sided	(1.29, 261.49)	(.88, 15.92)
Modified two-sided	(1.38, 40.45)	(1.01, 11.14)

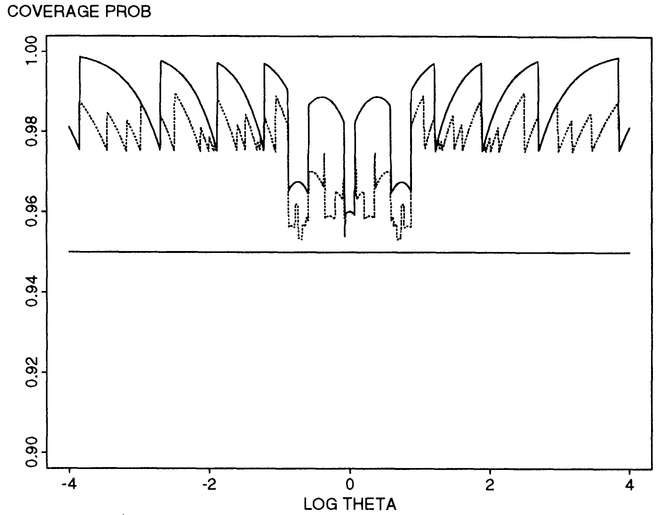

Figure 2. Coverage Probability for Confidence Intervals Based on Inverting One-Sided Tests, for Conditional Distribution Based on Margins of Table 1: - - -, One-Sided Modified P; ——, One-Sided Ordinary P.

when the probabilities are monotone increasing in t, which, from (6), is guaranteed when $\theta_+ > \max_t\{c_{t-1}/c_t\}$. In this case one can show directly that this upper limit θ_+ is the same as the upper limit obtained using the one-sided testing approach with double the error probability. For instance, the upper limit of the 95% interval based on inverting a two-sided test is then the same as the upper limit of the 90% interval for the approach based on inverting two separate one-sided tests. Analogous remarks apply to the lower limit. In such cases, the approach based on two-sided tests has a clear advantage.

We have now considered two modifications of the usual "exact" confidence interval, one based on inverting tests using a modified P value and one based on inverting one two-sided test instead of two one-sided tests. To incorporate both of these modifications simultaneously, we define a modified two-sided P value for testing a particular value of θ as

$$P^*(\theta) = P(\theta) - P(\{\mathbf{Z}: \mathbf{Z} \in \Gamma, P(t; \theta) = P(t_o; \theta), P(\mathbf{Z}; \theta) > P(\mathbf{N}; \theta)\}). \quad (10)$$

The confidence interval consists of the shortest interval containing all values of θ for which $P^*(\theta) \geq \alpha$. Appendix B shows that this confidence interval is "exact." This approach tends to give even narrower intervals than those obtained by inverting the two-sided test with the ordinary P value.

For Tables 1 and 2, Table 3 displays 95% confidence intervals obtained by inverting tests using two separate one-sided ordinary or modified p values and using the ordinary or modified two-sided P values. The confidence interval constructed using the two-sided P value is shorter than the ordinary interval based on two one-sided P values. In fact, for each data set, the upper endpoint for the two-sided–based interval equals the endpoint for the one-sided method for a 90% confidence interval. For each type of interval, the ones based on modified P values are narrower yet.

For the conditional distribution having the fixed marginal counts of Table 1, Figure 2 shows the actual coverage prob-

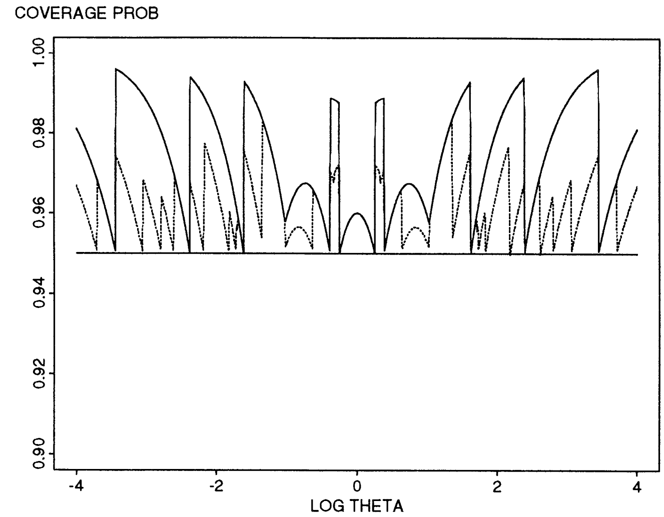

Figure 3. Coverage Probability for Confidence Intervals Based on Inverting Two-Sided Tests, for Conditional Distribution Based on Margins of Table 1: - - -, Two-Sided Modified P; ——, Two-Sided Ordinary P.

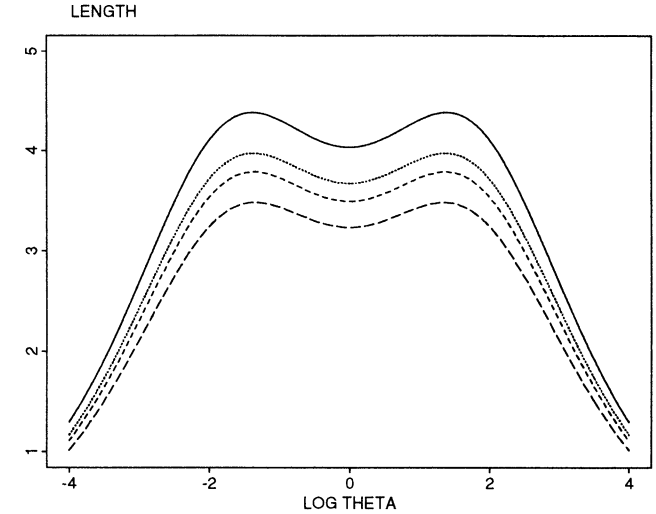

Figure 4. Average Lengths of the Confidence Intervals for Log θ, for Conditional Distribution Based on Margins of Table 1: ——, One-Sided Ordinary P, · · ·, One-Sided Modified P; ---, Two-Sided Ordinary P; - - -, Two-Sided Modified P.

ability as a function of the true log odds ratio, for 95% confidence intervals based on inverting separate one-sided tests using the ordinary or modified P value. There is a clear advantage to using the interval based on the modified P value. For either approach, for sufficiently large θ, all tables with those margins would have lower bound of the interval below θ; for sufficiently small θ, all tables would have upper bound above θ. In such cases the actual probability of coverage of a $100(1-\alpha)\%$ confidence interval has lower bound $1-\alpha/2$. That bound is achieved at values of θ that are potential endpoints of the intervals (Neyman 1935).

Figure 3 is an analogous display for the confidence intervals based on inverting two-sided tests using ordinary or modified P values. Again, there is an advantage to the interval based on the modified P value. A comparison of Figures 2 and 3 shows there is almost always an advantage to using confidence intervals based on inverting two-sided tests. In fact, for the conditional distribution with the margins of Table 1, the interval based on the ordinary two-sided test is always contained within the interval based on two ordinary one-sided tests. Figure 3 shows that for the two-sided approach, for large $|\log \theta|$, the true coverage probability has .95 as a lower bound, rather than .975. This relates to the property mentioned previously, by which an endpoint for the two-sided approach with error probability α can equal one for the one-sided approach with error probability 2α.

Figure 4 compares average lengths of the confidence intervals for log θ under the four approaches, for conditional distributions with the margins of Table 1. These are computed conditionally on T not equaling its minimum or maximum value, since the length is then infinite. Again, we see best results with inverting the two-sided test with modified P value. With the modified approach, the effects of discreteness diminish quickly as the sample size or K increases. To illustrate, Figure 5 shows coverage probabilities for the approaches based on two-sided tests, using the first three and using the first nine of the 18 strata from Table 2. For $K = 9$ strata, the ordinary approach still is quite conservative, but the modified approach is almost uniformly very good.

For each approach, one can construct even narrower intervals (albeit not "exact" ones) by inverting the tests based on corresponding modified mid-P values. Though approximate, confidence intervals based on the ordinary mid-P value have been observed empirically to behave well (Mehta and Walsh 1992).

5. EXACT, UNBIASED, ADMISSIBLE TESTS

Consider now $I \times J \times K$ tables. When X and Y are ordinal, it usually makes sense to test conditional independence against a narrow alternative, corresponding to a monotone trend in the partial association. One would then form a test

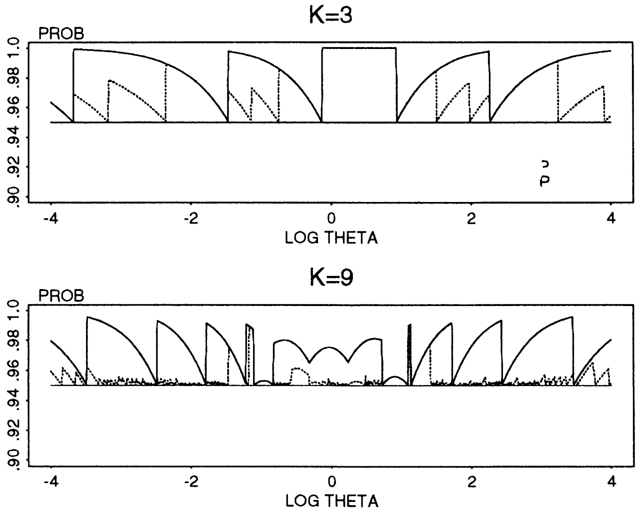

Figure 5. Coverage Probability for Confidence Intervals Based on Inverting Two-Sided Tests, for Conditional Distribution Based on First K Partial Tables of Table 2. · · ·, two-sided modified P; ——, two-sided ordinary P.

statistic using a model that is a special case of (1) and reflects the ordinality. For instance, the model

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \beta u_i v_j + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \quad (11)$$

replaces the general association term λ_{ij}^{XY} by a linear-by-linear term $\beta u_i v_j$, where $\{u_i\}$ and $\{v_j\}$ are monotone scores for levels of X and Y. Conditional independence of X and Y is its special case of $\beta = 0$. For this model, the sufficient statistic for β is $\sum_k [\sum_i \sum_j u_i v_j n_{ijk}]$. The test that orders tables in Γ by this statistic is sensitive to "correlation" alternatives to conditional independence. For $K = 1$, Agresti, Mehta, and Patel (1990) and Cohen and Sackrowitz (1992) presented exact tests of independence for model (11).

This section discusses tests of conditional independence of fixed size α for the ordinal alternative (11). We generalize a result of Cohen and Sackrowitz (1991) regarding exact, unbiased, and admissible tests. As a special case, we obtain results for tests of conditional independence in $2 \times 2 \times K$ tables that relate to use of the modified P value.

Let $T_{ij(k)} = \sum_{m=1}^{j} \sum_{l=1}^{i} n_{lmk}$, $\mathbf{T}_i^{(k)} = (T_{i1(k)}, T_{i2(k)}, \ldots, T_{i(J-1)(k)})$, $i = 1, \ldots, I-1$, and $\mathbf{T} = (\mathbf{T}_1^{(1)}, \ldots, \mathbf{T}_{I-1}^{(1)}, \mathbf{T}_1^{(2)}, \ldots, \mathbf{T}_{I-1}^{(2)}, \ldots, \mathbf{T}_1^{(K)}, \ldots, \mathbf{T}_{I-1}^{(K)})$. Let $\mathbf{s} = (\{n_{i+k}\}, \{n_{+jk}\})$ and let $\theta_{ij(k)} = (m_{ijk} m_{i+1,j+1,k})(m_{i,j+1,k} m_{i+1,j,k})$, $i = 1, \ldots, I-1, j = 1, \ldots, J-1, k = 1, \ldots, K$. Note that $(\mathbf{T}, \mathbf{s})$ is a one-to-one linear transformation from $\mathbf{N}$. A randomized test is characterized by a critical function, φ, satisfying $0 \le \varphi(\mathbf{N}) \le 1$ for all $\mathbf{N}$. For a test $\varphi(\mathbf{N})$ of size α, denote the conditional test as a function of the observed value $\mathbf{t}$ of $\mathbf{T}$, for each fixed $\mathbf{s}$, by $\varphi_{\mathbf{s}}(\mathbf{t})$. The test $\varphi_{\mathbf{s}}(\mathbf{t})$ must have conditional size α; that is, $E_{\theta=1}[\varphi_{\mathbf{s}}(\mathbf{t}) | \mathbf{s}] = \alpha$ for all $\mathbf{s}$.

Suppose that for each $\mathbf{s}$, $\varphi_{\mathbf{s}}(\mathbf{t})$ is monotone nondecreasing in $\mathbf{t}$; that is, when all elements of $\mathbf{t}$ are fixed except for any one, $\varphi_{\mathbf{s}}(\mathbf{t})$ is nondecreasing in that variable. Next, for each fixed $\mathbf{s}$, let $A_{\mathbf{s}}(\mathbf{t}) = \{\mathbf{t}: \varphi_{\mathbf{s}}(\mathbf{t}) < 1\}$ denote the acceptance region of the test, except for possible randomization. A point a in A is called an extreme point if a is not an interior point of any line segment in A. For $K = 1$, Cohen and Sackrowitz (1991) showed that a test $\varphi_{\mathbf{s}}(\mathbf{t})$ that is monotone nondecreasing in $\mathbf{t}$ is conditionally unbiased, and the original test $\varphi(\mathbf{N})$ is unconditionally unbiased. Furthermore, they showed that $\varphi(\mathbf{N})$ is admissible if and only if for each fixed $\mathbf{s}$, $A_{\mathbf{s}}(\mathbf{t})$ is convex and $\varphi_{\mathbf{s}}(\mathbf{t})$ is zero at nonextreme points of $A_{\mathbf{s}}(\mathbf{t})$. Though their proof applies to two-way tables, it can be easily generalized (Kim 1994). The proof for three-way tables is outlined in Appendix C.

We next illustrate exact, unbiased, and admissible tests for testing conditional independence against the alternative H_a: $\beta > 0$ in model (11), using $T = \sum_k [\sum_i \sum_j u_i v_j n_{ijk}]$. We can express T as $T = \sum_k [\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} (u_i - u_{i+1})(v_j - v_{j+1}) T_{ij(k)}] + C$, where C is a constant depending on the scores and the fixed marginal totals. Thus T is monotone in $\{T_{ij(k)}\}$ if the scores satisfy $(u_i - u_{i+1})(v_j - v_{j+1}) > 0$ for $i = 1, \ldots, I-1, j = 1, \ldots, J-1$; that is, if the scores $\{u_i\}$, $\{v_j\}$ are both monotone increasing or both monotone decreasing.

In constructing critical regions, we follow the Cohen and Sackrowitz (1992) approach of ordering the tables for which $T = t_o$ by their conditional null probabilities. This relates naturally to the modified P value (2). Let C_α be a constant, depending on $\mathbf{s}$, such that $P\{T \ge C_\alpha\} \ge \alpha$ and $P\{T > C_\alpha\} = \alpha' < \alpha$. The test rejects if $T > C_\alpha$. When $T = C_\alpha$, tables having smaller probabilities are more contradictory to the null hypothesis, so we reject also for those tables whose probabilities are smallest and total at most $(\alpha - \alpha')$. We allow randomization only at extreme points of a convex acceptance section of the remaining points, so the test is exact, unbiased, and admissible. We denote a test of this form by φ^*. This test is less likely to require randomization than the usual test φ that randomizes whenever $T = C_\alpha$. Also, the modified test is better than φ, because usually that entire set contains nonextreme points, making φ inadmissible.

We utilize the middle three subtables of Table 1 to illustrate. We consider size $\alpha = .05$ tests based on $T = \sum_k n_{11k}$, which results from the scores $u_1 = v_1 = 1$, $u_2 = v_2 = 0$. Now, $P\{T \ge 13\} = .1136 > \alpha$ and $P\{T > 13\} = .0200 < \alpha$, so randomization is required for tables with $T = 13$. The usual .05-size conditional test based on T is

$$\begin{aligned} \varphi &= 1 \quad \text{if } (n_{111}, n_{112}, n_{113}) \\ &\qquad = (3,6,6), (2,6,6), (3,5,6), (3,6,5), \\ &= .3206 \quad \text{if } (n_{111}, n_{112}, n_{113}) \\ &\qquad = (1,6,6), (2,5,6), (3,4,6), \\ &\qquad\quad (2,6,5), (3,5,5), \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Because table (2, 5, 6) is an interior point of the line segment between tables (1, 6, 6) and (3, 4, 6), it is not an extreme point of a convex acceptance region, and this test is inadmissible.

The exact test φ^* that orders the tables by their probabilities is

$$\begin{aligned} \varphi^* &= 1 \quad \text{if } (n_{111}, n_{112}, n_{113}) \\ &\qquad = (3,6,6), (2,6,6), (3,5,6), (3,6,5), \\ &\qquad\quad (1,6,6), (2,6,5), (3,5,5), \\ &= .3200 \quad \text{if } (n_{111}, n_{112}, n_{113}) = (3,4,6), \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

This test randomizes only at an extreme point (3, 4, 6) of its convex acceptance region, so it is unbiased and admissible. Compared to the previous test, it requires randomization for only a single table. The probability that randomization is required is .0207, rather than .0937.

6. CONCLUSIONS

In practice, tests using supplementary randomization to achieve a fixed size are unacceptable. Thus even the tests described in Section 5 that require less randomization than usual are not intended for practical use. But results of that section suggest a way of forming critical regions to achieve actual size closer to a desired value (such as .05) than is possible with the ordinary test.

Using inferences based on modified P values moves us in the direction of the optimal procedures based on supplementary randomization. For instance, the UMP unbiased test of size α (Lehmann 1986, p. 163) corresponds to making a decision based on comparing α to a P value using supplementary randomization, such as (3) for the one-sided case. After generating a uniform random variable, one could invert such a test to obtain a confidence interval, which (unconditionally) would be uniformly most accurate unbiased (Lehmann 1986, p. 217). Our tests and confidence intervals, though not achieving optimality for an arbitrary size α, have the advantage of using properties of the data, rather than being data independent. Our modification, unlike the mid-P value, maintains at least the guaranteed level and uses the data for the P value adjustment. In addition, the modified procedures are less conservative than the usual ones based on the ordinary P value. The improvement can be considerable when K is large but the sample size is not, in which case many tables with different probabilities may have the same test statistic value.

We used the table probability to generate a secondary partitioning of tables having the observed value of T. When T is a score, Wald, or likelihood ratio statistic for a particular alternative, it would not help to form a modified P value by using simultaneously either of the other two statistics. Because these tests all depend only on the sufficient statistics under the alternative, they induce the same partitioning.

For approaches based on ordinary or modified P values, we prefer confidence intervals based on inverting two-sided tests. Even if one prefers not to base inference on modified P values, one can obtain confidence intervals that are often much narrower than the ordinary intervals by inverting results of two-sided tests rather than results of separate one-sided tests, such as is done in existing software (e.g., StatXact 1991).

We have presented our ideas in the context of three-way contingency tables, but they extend to other settings in which exact conditional inference applies. An example is logistic regression modeling. Exact inference in logistic regression is often highly discrete, even degenerate. One can often alleviate this problem somewhat by treating the data as a contingency table and using the modified approach.

The first author has prepared a FORTRAN program, designed for IBM-compatible PC's or UNIX workstations, for computing modified P values and modified confidence intervals for a common odds ratio in $2 \times 2 \times K$ tables. It is available from the first author by sending a formatted 3-1/2 inch floppy disc. This program is an adaptation of one written by Vollset and Hirji (1991) for ordinary exact inference for such tables.

APPENDIX A: EXACTNESS FOR THE MODIFIED CONFIDENCE INTERVAL BASED ON ONE-SIDED TESTS

The lower limit is the smallest value of θ for which $P_1^*(\theta) \geq \alpha/2$. For $\theta < \theta_-^*$, $P_1^*(\theta) < \alpha/2$. It follows that

$$\Pr(\theta_-^* > \theta) \leq \Pr\left(P_1^*(\theta) < \frac{\alpha}{2}\right)$$
$$\leq \Pr\left(P_1^*(\theta) \leq \frac{\alpha}{2}\right) = E \Pr\left(P_1^*(\theta) \leq \frac{\alpha}{2} \middle| \mathbf{s}\right) \leq \frac{\alpha}{2},$$

where $\mathbf{s}$ denotes a possible sufficient marginal configuration, and the last step follows because of discreteness. For the upper limit, by the same arguments we have $\Pr(\theta_+^* < \theta) \leq \alpha/2$. Thus the exclusion probability $\Pr(\theta_+^* < \theta) + \Pr(\theta_-^* > \theta)$ is at most α.

APPENDIX B: EXACTNESS FOR THE MODIFIED CONFIDENCE INTERVAL BASED ON TWO-SIDED TESTS

The lower limit θ_- is the smallest θ satisfying $P^*(\theta) \geq \alpha$, and the upper limit θ_+ is the largest θ satisfying $P^*(\theta) \geq \alpha$. For all values of θ lying outside the closed interval $\theta_- \leq \theta \leq \theta_+$, it follows that $P^*(\theta) < \alpha$. Then

$$\Pr(\theta < \theta_-, \theta > \theta_+) \leq \Pr(P^*(\theta) < \alpha)$$
$$\leq \Pr(P^*(\theta) \leq \alpha) = E \Pr(P^*(\theta) \leq \alpha \mid \mathbf{s}) \leq \alpha.$$

Hence $\Pr(\theta_- \leq \theta \leq \theta_+) \geq 1 - \alpha$.

APPENDIX C: EXACT, UNBIASED, AND ADMISSIBLE TESTS

For the proof of unbiasedness, we need a lemma that generalizes lemma 3.2 of Cohen and Sackrowitz (1991) to three-way tables. The same techniques are applied, with appropriate modifications for three-way tables. We use $\mathbf{T}$ and $\mathbf{s}$ as defined in Section 5. For nondecreasing functions $W(\mathbf{T})$ and $W^*(\mathbf{T})$, one can show that under conditional independence,

$$E\{W(\mathbf{T})W^*(\mathbf{T}) \mid \mathbf{s}\} \geq E\{W(\mathbf{T}) \mid \mathbf{s}\}E\{W^*(\mathbf{T}) \mid \mathbf{s}\}.$$

Let $f_\theta(\mathbf{t} \mid \mathbf{s})$ and $f_0(\mathbf{t} \mid \mathbf{s})$ denote the conditional densities of $(\mathbf{T} \mid \mathbf{s})$ under the alternative and null. We see that $W^*(\mathbf{t}) = f_\theta(\mathbf{t} \mid \mathbf{s})/f_0(\mathbf{t} \mid \mathbf{s})$ is nondecreasing for any $\boldsymbol{\theta}$ in the alternative space, and by the assumption, the test $\varphi_\mathbf{s}(\mathbf{t})$ is also nondecreasing in $\mathbf{t}$. Then for $\boldsymbol{\theta}$ in the alternative space,

$$E_\theta(\varphi_\mathbf{s}(\mathbf{t}) \mid \mathbf{s}) = \sum \varphi_\mathbf{s}(\mathbf{t}) f_\theta(\mathbf{t} \mid \mathbf{s})$$
$$= \sum \varphi_\mathbf{s}(\mathbf{t}) W^*(\mathbf{t}) f_0(\mathbf{t} \mid \mathbf{s})$$
$$\geq \left[\sum \varphi_\mathbf{s}(\mathbf{t}) f_0(\mathbf{t} \mid \mathbf{s})\right]\left[\sum W^*(\mathbf{t}) f_0(\mathbf{t} \mid \mathbf{s})\right], \quad \text{(A.1)}$$
$$= E_0 \varphi_\mathbf{s}(\mathbf{t}) = \alpha. \quad \text{(A.2)}$$

This expression implies conditional unbiasedness of $\varphi_\mathbf{s}(\mathbf{t})$, which in turn implies unbiasedness of the original test $\varphi(\mathbf{N})$.

Next, we outline the proof of admissibility. Using the distribution of $\mathbf{N}$, we test conditional independence against (1). Eaton (1970) gave an essentially complete class for an exponential family. This class consists of tests for which the acceptance region is convex, with possible randomization on the boundary of the acceptance region. Applying his theorem, our tests lie in the complete class of tests. Furthermore, Ledwina (1978, 1984) gave the class of admissible rules for multivariate exponential distributions with discrete support. Ledwina considered the connection between admissible tests based on conditional and joint distributions, and showed that if the conditional test is admissible for every fixed value of nuisance parameters, then the original test is admissible. Admissibility of tests in three-way tables is obtained using the same arguments given there.

[Received October 1993. Revised June 1994.]

REFERENCES

Agresti, A. (1992), "A Survey of Exact Inference for Contingency Tables," *Statistical Science*, 7, 131–177.

Agresti, A., Mehta, C. R., and Patel, N. R. (1990), "Exact Inference for Contingency Tables With Ordered Categories," *Journal of the American Statistical Association,* 85, 453–458.

Baptista, J., and Pike, M. C. (1977), "Exact Two-Sided Confidence Limits for the Odds Ratio in a 2×2 Table," *Journal of the Royal Statistical Society,* Ser. C, 26, 214–220.

Birch, M. W. (1964), "The Detection of Partial Association I: The 2×2 Case," *Journal of the Royal Statistical Society,* Ser. B, 26, 313–324.

Cohen, A., and Sackrowitz, H. B. (1991), "Tests for Independence in Contingency Tables With Ordered Categories," *Journal of Multivariate Analysis,* 36, 56–67.

——— (1992), "An Evaluation of Some Tests of Trend in Contingency Tables," *Journal of the American Statistical Association,* 87, 470–475.

Cox, D. R. (1970), *Analysis of Binary Data,* London: Chapman and Hall.

Eaton, M. L. (1970), "A Complete Class Theorem for Multidimensional One-Sided Alternatives," *Annals of Mathematical Statistics,* 41, 1884–1888.

Gart, J. J. (1970), "Point and Interval Estimation of the Common Odds Ratio in the Combination of 2×2 Tables With Fixed Marginals," *Biometrika,* 57, 471–475.

Kim, D. (1994), "Improved Exact Methods for Statistical Inference in Contingency Tables," unpublished Ph.D. dissertation, University of Florida, Dept. of Statistics.

Lancaster, H. O. (1961), "Significance Tests in Discrete Distributions," *Journal of the American Statistical Association,* 56, 223–234.

Ledwina, T. (1978), "Admissible Tests for Exponential Families With Finite Support," *Statistics,* 9, 105–118.

——— (1984), "A Note on Admissibility of Some Tests of Independence Against 'Positive Dependence' in $R \times C$ Contingency Table," *Statistics,* 15, 565–570.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (2nd ed.), New York: John Wiley.

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.

Mehta, C. R., Patel, N. R., and Gray, R. (1985), "Computing an Exact Confidence Interval for the Common Odds Ratio in Several 2 by 2 Contingency Tables," *Journal of the American Statistical Association,* 80, 969–973.

Mehta, C. R., and Walsh, S. J. (1992), "Comparison of Exact, Mid-p, and Mantel–Haenszel Confidence Intervals for the Common Odds Ratio Across Several 2×2 Contingency Tables," *The American Statistician,* 46, 146–150.

Neyman, J. (1935), "On the Problem of Confidence Limits," *Annals of Mathematical Statistics,* 6, 111–116.

StatXact. (1991), *StatXact: Statistical Software for Exact Nonparametric Inference,* Version 2. Cambridge, MA: Cytel Software.

Sterne, T. E. (1954), "Some Remarks on Confidence or Fiducial Limits," *Biometrika,* 41, 275–278.

Streitberg, B., and Roehmel, J. (1990), "On Tests that are Uniformly More Powerful Than the Wilcoxon–Mann–Whitney Test," *Biometrics,* 46, 481–484.

Suissa, S., and Shuster, J. J. (1985), "Exact Unconditional Sample Sizes for the 2×2 Binomial Trial," *Journal of the Royal Statistical Society,* Ser. A, 148, 317–327.

Vollset, S. E., and Hirji, K. F. (1991), "A Microcomputer Program for Exact and Asymptotic Analysis of Several 2×2 Tables," *Epidemiology,* 2, 217–220.

Vollset, S. E., Hirji, K. F., and Elashoff, R. M. (1991), "Fast Computation of Exact Confidence Limits for the Common Odds Ratio in a Series of 2×2 Tables," *Journal of the American Statistical Association,* 86, 404–409.