SVEND KREINER

Analysis of Multidimensional Contingency Tables by Exact Conditional Tests: Techniques and Strategies

Reprinted from
Scandinavian Journal of Statistics, Vol. 14, No. 2, 1987

Scand J Statist 14: 97–112, 1987

Analysis of Multidimensional Contingency Tables by Exact Conditional Tests: Techniques and Strategies

SVEND KREINER

Danish Institute for Educational Research

ABSTRACT. The tests for zero partial association in a multidimensional contingency table (suggested by Birch (1965) and later used by Wermuth (1976) for a more elaborate model-searching strategy) have gained new importance with the introduction of the class of graphical models. It is shown how these may be performed as exact conditional tests using as test criteria either the ordinary likelihood ratio, the standard χ^2 statistic or any other appropriate statistic (e.g. incorporating ordinal properties of some of the categorical variables). A strategy for model checking based exclusively on exact tests for zero partial association, performed initially in the full table and subsequently in appropriate marginal tables, is presented. An example of application of the strategy is given, which illustrates that the usual approximate p-values are completely unreliable for large sparse tables.

Key words: conditional independence, contingency table, exact conditional test, graphical model, log-linear model, model selection strategy, sparse tables, zero partial association

1. Introduction

This article considers the possibility of performing exact (or approximate) conditional tests of certain hypotheses encountered in connection with the analysis of multidimensional contingency tables.

Throughout the paper, the basic concepts and results for the log-linear models for contingency tables will be assumed known (Andersen, 1974; Bishop *et al.*, 1975; Lauritzen, 1982, provide good introductions to the subject).

Consider an undirected graph, such as Fig. 1, that is a set of vertices and edges. It is shown by Darroch *et al.* (1980) (DLS) that this graph may be identified with a log-linear model in the following way.

A set of vertices of the graph is complete if all possible edges between the vertices in the set are in the graph. A clique is defined as a maximal complete set, i.e. a complete set which cannot be extended to a larger complete set by the addition of more vertices. In Fig. 1 [ABE], [ADE] and [ACD] are cliques, whereas [ABDE] is not complete since the edge BD is not in the graph and therefore [ABDE] is not a clique.

Finally a log-linear model is identified with the graph by identifying the variables as the vertices of the graph and the generators, specifying the minimal set of sufficient marginals, as the set of cliques of the graph.

Models that can be generated in this way are called graphical models, and can be interpreted in terms of conditional independence in the following fashion:

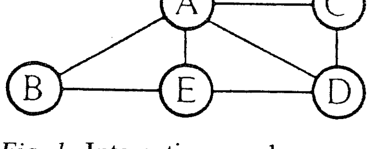

Fig. 1. Interaction graph.

Two variables that are not directly connected in the interaction graph are conditional independent given the rest of the variables of the contingency table. In the example the variables B and C are conditional independent given variables A, D and E. Using the notation of Dawid (1979) this relation may be written: $\{B \perp C | A, D, E\}$.

We thus have three different ways of specifying the same graphical model:

1. By the generators or alternatively the minimal sufficient marginals of the models:

   [ABE][ADE][ACD]

2. By statements concerning independence between pairs of variables given the rest of the variables forming the table:

   $\{B \perp C | ADE\} \cap \{B \perp D | ACE\} \cap \{E \perp C | ABD\}$

3. By the interaction graph.

2. Properties of graphical models

An important property of the class of graphical models for multidimensional contingency tables is that their specification only requires the fundamental statistical concept of conditional independence.

The graphical models may be characterized as the class of models defined by assumptions concerning conditional independence of pairs of variables given the values of the remaining variables, without any further assumptions concerning a more simple (reduced) parametric specification of the probabilities.

It is shown in DLS that it is possible to construct a statistical model from the conditional independence properties alone and that the class of graphical models is a subclass of the general class of hierarchical log-linear models. It also follows from DLS that any non-graphical log-linear model may be regarded as a submodel of a minimal graphical model with some of the higher order interactions vanishing, that is as a model with the same conditional independence properties as the graphical model, but with a simpler parametric structure.

The interaction graph with vertices representing variables and missing edges representing conditional independences is a powerful analytical and practical tool.

It should be emphasized that an interaction graph is more than a mere visualization of the statistical model. It is in itself a mathematical model. It may be analysed using standard graph-theoretical results and several important graph-theoretical concepts may be interpreted in terms of the statistical model. The interaction graph thus facilitates a quick insight in the properties of the statistical model under consideration and may thus be used both for the planning and execution of the statistical analysis and for the interpretation of the results.

The three most important results are:

1. That the generators or the minimal sufficient marginals are represented by the cliques of the interaction graph.
2. That decomposability fully (in the case of a triagulated interaction graph) or partially corresponds to decomposability of the statistical model (Sundberg, 1975).
3. If two disjoint subsets, S_1 and S_2 are separated by a subset S_3 in the sense that all paths from S_1 to S_2 go through S_3, then the variables in S_1 are conditional independent of those in S_2 given the variables in S_3. In the example B and C are not only conditional independent given ADE, but also given AE or AD.

The third property is especially important for the interpretation of the graphical model. It says that the conditional independence is not only valid in the contingency table containing all variables, but also in certain collapsed tables.

Both the theoretical and practical consequences of choosing the smaller class of graphical models as the basis of the statistical analysis are discussed in further detail in Edwards & Kreiner (1983), Edwards & Havranek (1985) and Whittaker (1982).

3. Birch's test for zero partial association

It follows directly from the definition of graphical models as models specified by the conditional independencies, that the problem of finding the "best" graphical model for a given contingency table may be reduced to relatively few questions even when the number of possible graphical models is quite large.

Disregarding the models where one or more variables are irrelevant the total number of graphical models for a seven-dimensional contingency tables is $2^{21} = 2\,097\,152$. The model may be found however by examining at most 21 hypotheses of conditional independence for pairs of variables (or at least for each pair, for which the hypotheses are relevant) given the rest of the variables forming the contingency table.

These hypotheses correspond to the hypotheses of zero partial association (ZPA) discussed by Birch (1965) and later used by Wermuth (1976) for a more elaborate model-searching strategy. The analysis of contingency tables by graphical models may thus be seen to place more emphasis on Birch's test than is usually found in literature on the analysis of contingency tables.

The six ZPA hypotheses which may be tested in a four-dimensional contingency table are shown Table 1.

A convenient aspect of testing ZPA hypotheses in contingency tables of dimensions greater than three is the fact that the conditioning variables may be combined into one categorical variable. For the test of the first hypothesis in Table 1, the variables C and D may thus be combined into one variable, $X = C*D$, with categories defined by all possible combinations of the values of C and D.

Whatever the dimension of the original contingency table, the tests of ZPA hypotheses may therefore always be treated as tests in a three-dimensional table.

Consider then a three-dimensional contingency table formed by variables A, B and X and the ZPA hypothesis, $A \perp B | X$.

The generators for the model under the hypothesis are [AX] and [BX]. It follows from principles of conditioning on ancillary statistics that the hypothesis should be evaluated with

Table 1. *The hypotheses of conditional independency for a four-way table formed by four categorical variables A, B, C and D.*

	Hypothesis	Generators
I	$\{A \perp B \| CD\}$	[ACD][BCD]
II	$\{A \perp C \| BD\}$	[ABD][BCD]
III	$\{A \perp D \| BC\}$	[ABC][BCD]
IV	$\{B \perp C \| AD\}$	[ABD][ACD]
V	$\{B \perp D \| AC\}$	[ABC][ACD]
VI	$\{C \perp D \| AB\}$	[ABC][ABD]

respect to the conditional distribution of the contingency table given the sufficient AX and BX marginals. This and other conditional tests are discussed by Andersen (1974).

The three-dimensional table may be viewed as consisting of several "slices" of two-dimensional AB tables, one for each level of the conditioning X variable, and conditioning with respect to the AX and BX marginals may easily be seen to correspond to conditioning with respect to the marginals in each of the two-dimensional slices. The conditional probability of the three-dimensional table given the two-dimensional AX and BX marginals may therefore be expressed as the product of the conditional probabilities for each "slice" given the corresponding one-dimensional marginals, i.e. a product of generalized hypergeometric probabilities.

The conditional test may be performed according to the following algorithm:

A. Generate all tables with the given AX and BX marginals.
B. Calculate the appropriate test statistic for each table.
C. Calculate the conditional probability of each table.
D. Calculate the attained level of significance (p-value) as the sum of the conditional probabilities for the tables which are at least as favourable to the alternative hypothesis (according to the test statistic) as the observed table.

Note that the conditional approach makes it possible to use test statistics for which no asymptotic results are available, for example tests using ordinal properties of the variables.

When the number of tables to be generated is too large to be economically implemented the following procedure can be used.

A. Generate a random sample of tables according to the conditional distribution of the table given the sufficient marginals. The tables may in the case of the ZPA hypothesis be generated one slice at time because of the factorization property mentioned above. Each slice may be generated using either the method of Agresti *et al.* (1979) and Boyett (1979) or the faster (in the case of moderate or large samples) method of Patefield (1981). The latter method is based on a factorization of the generalized hypergeometric distribution for $r \times c$ tables into a product of ordinary hypergeometric probabilities for 2×2 tables. (This leads to a "one-cell-at-a-time generation of the cells of the slices.)
B. For each generated table the appropriate test statistics are calculated.
C. The attained level of significance is approximated as the number of generated tables which are at least as favourable to the alternative as the observed table divided by total number of tables generated.

The p-values determined in this way are of course not really exact, but may be viewed as *estimates* of the exact p-values. As such they are of course subject to random error.

The estimate is based on binomial sampling and therefore unbiased. The accuracy of the estimate may be evaluated by the confidence intervals of the binomial distribution.

To give a rough survey the 95% confidence intervals for certain sample sizes are given in Table 2.

The primary reason to prefer the "estimated" p-values to the "asymptotic" p-values–both of course being approximations to the true values–are

1. That the estimated p-values are unbiased while the behaviour of the asymptotic p-values are somewhat aberrant especially in the case of the likelihood ratio statistic which tends to reject substantially more often than is expected even for moderate samples (Fienberg, 1980). While several Monte Carlo studies have been reported very little is known about the behaviour of the test statistics in the case of very large sparse tables. The known results suggest that the asymptotic p-values systematically underestimate the p-values. The unbiasedness of the estimated p-values therefore is an attractive feature.

Table 2. *95% confidence limits of estimates of the exact conditional p-values*

	Number of generated tables					
	500		2000		5000	
Estimated p-value	Lower	Higher	Lower	Higher	Lower	Higher
0.500	0.456	0.544	0.478	0.522	0.486	0.514
0.250	0.214	0.290	0.232	0.269	0.238	0.262
0.100	0.077	0.129	0.088	0.114	0.092	0.108
0.050	0.034	0.073	0.041	0.060	0.044	0.056
0.030	0.018	0.049	0.023	0.038	0.025	0.035
0.020	0.011	0.036	0.015	0.027	0.016	0.024
0.010	0.004	0.023	0.006	0.015	0.007	0.013
0.005	0.002	0.016	0.003	0.009	0.003	0.007
0.001	0.000	0.010	0.000	0.004	0.000	0.002

* Calculated according to formula (4.21) of Sachs (1982).

2. That the accuracy of the approximation of the estimated p-values may be controlled by the statistician.

The behaviour of the estimated p-values is in other words known while the behaviour of the asymptotic p-values in the case of very large sparse tables is at best the subject of conjecture.

We illustrate the conditional test with examples from a survey investigation on teachers in the Danish primary school.

The first example presents a small table where the exact conditional distribution of the test statistics may be determined.

In the second example a very large sparse table is analysed. This example is preceded by a discussion on strategical aspects of the analysis of large sparse tables.

4. Example 1

Table 3 presents the distribution of 27 teachers with respect to their evaluation of the pupils in specific classes. This variable is related to the number of pupils in the class and the teacher's evaluation of the possibilities of coping with the individual problems of the pupils.

Table 3. *Distribution of 27 teachers according to evaluation of individual classes, number of pupils in the class and the possibility of coping with the individual problems of the pupils*

(C) Possibility of coping with individual problems	(B) Number of pupils in the class	(A) Frequency of restless pupils in class			
		None	Few	Some/many	Total
Good	12–19	4	1	0	5
	20–30	1	2	0	3
	Total	5	3	0	8
Bad	12–19	1	13	1	15
	20–30	0	2	2	4
	Total	1	15	3	19

It should be emphasized that the 27 teachers only represent a random sample of *ca.* 1% of the teachers included in the study. The purpose of this example is to demonstrate the conditional test for zero partial association including both the exact evaluation of the attained significance and the procedure for approximating the p-value by random sampling of conditional tables. The real problems connected with this kind of survey investigations are far better represented in the following example 2.

The hypothesis to be considered here is the conditional independence of variables A and B given C.

According to the conditional approach this hypothesis may be evaluated with respect to the conditional distribution of appropriate test statistics given the sufficient AC and BC marginals.

Only the standard likelihood ratio statistic will be discussed in this example.

The likelihood-ratio statistic calculated for this hypothesis is 5.719. Because of the one zero in the AC marginal, this value should be evaluated with respect to a χ^2 distribution with 3 and not 4 degrees of freedom. It should be noted, though, that this adjustment of degrees of freedom is based on apparently intuitive arguments (e.g. Bishop *et al.*, 1975, pp. 114–116), and that no formal arguments for this adjustment seem to have been published.

It should also be noted that an automatic use of adjusted degrees of freedom will lead to meaningless results.

Consider for instance the two log-linear models, [AC][BC] and [A][BC].

The degrees of freedom for [AC][BC] against [ABC] should as noted above be adjusted from 4 to 3.

As there are no zeros in the A and BC marginals, the degrees of freedom for [A][BC] against [ABC] should not be corrected. The degrees of freedom for this test may easily be seen to be 6.

The likelihood ratio statistic for this hypothesis may according to standard results be expressed as the sum of the likelihood ratio statistic for the first hypotheses and the likelihood ratio statistic for the "conditional" hypothesis, [A][BC] against [AC][BC]. It follows from the same standard results that the likelihood ratio statistic for [A][BC] against [AC][BC] should be evaluated with respect to the χ^2 distribution with $6-3=3$ degrees of freedom.

It is intuitively clear, that this makes no sense, and it follows directly from results on decomposability by Sundberg (1975) that the "conditional" test is equivalent to the test for marginal independence in the AC table. It should therefore be evaluated with respect to the χ^2 distribution with 2 degrees of freedom.

The point may seem minor, but is nevertheless well worth noting since the adjustment of degrees of freedom with respect to most saturated model are automatically calculated by the standard programmes for contingency tables and used for the calculation of degrees of freedom for hypotheses against non-saturated alternatives.

A total of 32 different tables may be found corresponding to the sufficient margins AC and BC of the table given in Table 3.

For each of these tables the likelihood ratio statistic and the conditional probability was calculated and the p-value determined as the sum of the conditional probabilities for all tables with likelihood ratio observed values equal to or greater than the observed value of 5.719.

At the same time the exact p-value was approximated by estimates from 500, 2000 and 5000 random tables generated according to the conditional distribution given the sufficient marginals.

The results are shown in Table 4 which also shows the asymptotic p-values according the χ^2 distribution with 3 degrees of freedom. It should be noted, that the 32 tables defined by the sufficient marginals leads to only 26 different likelihood ratio values. In three cases the likelihood ratio for different tables differed by less than 0.002, as noted in the table.

Table 4. *Exact conditional distribution of the likelihood ratio statistic for the hypothesis of zero partial association between A and B given the sufficient marginals of table 3*

	p-Values				
		Estimated number of tables generated			
Likelihood ratio	Asymptotic*	500	2000	5000	Exact
0.76	0.8590	1.000	1.000	1.000	1.0000
2.20	0.5330	0.804	0.810	0.810	0.8110
2.49	0.4780	0.618	0.620	0.623	0.6230
3.92	0.2700	0.528	0.526	0.529	0.5280
3.99‡	0.2620	0.434	0.432	0.434	0.4340
4.58†	0.2050	0.348	0.347	0.346	0.3470
5.72‡	0.1260	0.226	0.222	0.218	0.2210
6.01	0.1110	0.182	0.182	0.176	0.1780
6.31	0.0980	0.118	0.120	0.114	0.1150
7.81‡	0.0500	0.084	0.087	0.082	0.0830
8.40	0.0380	0.056	0.056	0.054	0.0540
8.43	0.0380	0.034	0.034	0.032	0.0330
10.15	0.0170	0.026	0.028	0.028	0.0270
11.31	0.0100	0.022	0.023	0.024	0.0240
12.24†	0.0070	0.020	0.018	0.019	0.0180
12.74	0.0050	0.016	0.014	0.014	0.0140
13.97	0.0030	0.008	0.006	0.007	0.0070
14.54†	0.0020	0.006	0.005	0.006	0.0060
15.13	0.0020	0.002	0.001	0.002	0.0030
16.06	0.0010	0	0.001	0.001	0.0010
18.97	< 0.0005	0	0	0	0.0005
19.59	–	–	–	–	< 0.0005
21.32	–	–	–	–	–
22.79	–	–	–	–	–
23.41	–	–	–	–	–
30.14	–	–	–	–	–

* Evaluated with respect to the χ^2 distribution with 3 degrees of freedom.
† Including two tables with slightly different likelihood ratio values.
‡ Two tables with equal likelihood ratio values.

5. A strategy for the analysis of large sparse tables

Several strategies for the analysis of multidimensional contingency tables by graphical models have been suggested (Whittaker, 1982; Edwards & Kreiner, 1983; Edwards & Havranek, 1985). As all strategies are based on likelihood ratio tests with the distributions of the test statistics being approximated by the asymptotic χ^2 distribution, the results of the initial steps of these procedures are problematic, when the contingency table is large and sparse.

To overcome these problems the following strategy for the *initial* analysis is suggested.

1. All hypotheses of zero partial association between pairs of variables (excluding of course pairs of explanatory variables) are performed using "exact" conditional tests.
2. A base model is constructed as the graphical model defined by the conditional independencies accepted in the first step.
3. The base model is analysed with respect to collapsibility and decomposability.
4. The results of the analysis of the base model is checked in collapsed marginal tables and the base model revised according to the results of this analysis.

From this point the analysis may proceed according to the general strategy adhered to (not necessarily restricted to the class of graphical models) with all tests being performed in collapsed marginal tables according to decomposability properties of the revised base model.

6. On collapsibility and the examination of the adequacy of graphical models

The suggestions for checking the adequacy of a graphical model are based on the assumption that a model check should be primarily concerned with the basic properties of the model in question. As the basic properties of the graphical models (both with respect to the statistical analysis and with respect to the interpretation of the results of the analysis) are connected with decomposability and collapsibility it follows that special care should be taken to ensure that these properties are valid.

To facilitate the discussion it will be convenient to introduce the property of w-collapsibility (for weak collapsibility) and certain related concepts in the following way.

Consider a multidimensional contingency table, a hierarchical log-linear model and three distinct subsets of variables, A, B and C, with A containing at least two variables.

We will refer to the problem of estimating and testing with respect to the interactions between the variables of the A-set as the *A-problem*.

The important topic of collapsibility, that is, the property that relations between a set of variables may be studied by analysis of marginal tables, will in this context be discussed with explicit reference to a given A-problem.

Two types of collapsibility will be defined: strong collapsibility corresponding to the kind of collapsibility usually discussed in literature on collapsibility and separability of contingency tables, and weak collapsibility admitting a more limited analysis of the A-problem in the collapsed table than in the strongly collapsible case.

To simplify the discussion of these concepts and their implications for the analysis of the data and interpretation of results, two corollary concepts will be defined: the order of and the separators for the A-problem.

A model is said to be *w-collapsible* over C with respect to the variables in A if conditional independence of the variables in A given B and C is a sufficient condition for the conditional independence of variables in A given only B.

We will call the variables of B the *w-separators* for the A-problem. In the interaction graph of the model, the vertices representing the separators constitute a so-called cut-set cutting all paths between vertices corresponding to the A-variables.

For a given model the contingency table will usually be w-collapsible in several different ways with respect to a given problem.

It will be presumed that the question of how many or how few variables are needed for the conditional independence to be valid is one of the primary conclusions which may be drawn from the analysis. For this reason we find it convenient to define the w-order of the problem concerning the variables in A in the following way:

The *w-order* of the A-problem is defined as the number of variables in a minimal set of w-separators for the A-problem.

In many cases the w-order variables needed for the conditioning may not be uniquely determined because the conditional independence between the A-variables may be implied by the graphical model for several collapsed tables of the same minimal dimension.

The kind of collapsibility usually discussed in connection with contingency tables (Asmussen & Edwards 1983; Lauritzen, 1982; Sundberg, 1975) is more restrictive than the weak collapsibility discussed here. Loosely formulated this kind of collapsibility requires not only that vanishing interactions in the full model vanishes in the collapsed model, but also that the values of non-vanishing interactions between A-variables (in the log-linear parametrization) are the same in the full and the collapsed tables and that the likelihood ratio tests of hypotheses concerning the interactions between the A-variables are the same in the full and the collapsed tables. This kind of collapsibility will be denoted s-collapsibility (strong collapsibility). s-Order and s-separators may be defined analogously to w-order and w-separators.

As s-collapsibility infers w-collapsibility it follows that w-order will be smaller than or equal to corresponding s-order.

When conclusions concerning collapsibility, order and separators are considered important, these conclusions should be carefully checked before the model is accepted. That is, the model should not be accepted just because an analysis of the complete table deems it satisfactory, but only if analyses of collapsed tables confirm the consequences inferred from the model.

The following procedure for the examination of the adequacy of the model is therefore suggested:

1. Determine for all cases of conditional independence the w-order and the w-separators of the conditional independent variables.
2. In the corresponding collapsed tables thus inferred, tests of conditional independence should be performed.
3. The rejectance of the hypotheses of conditional independence should be considered as evidence against the model. Depending on the overall evaluation of the model and the importance of the hypotheses in question the model should be remedied by adding interactions which makes the corresponding results concerning w-order and w-separators invalid. It should be noted, that the inclusion of just one set of interactions may be sufficient to explain the rejection of several different hypotheses of conditional independence even when these hypotheses do not concern the conditional independence between the variables whose interaction is added to the full model. Special care should therefore be taken not to include more interactions than necessary to keep the full model as parsimonious as possible.

It should be noted that examination of the properties derived by s-collapsing in many cases will be unnecessary, when the original strategy adhered to employs partitioned statistics, where each possible model is evaluated with respect to already accepted models. In these cases the likelihood ratio tests performed will in reality be tests in collapsed tables, and the s-collapsed properties will therefore already have been checked. The only problem which may occur is the technical problem of incorrect adjustment of degrees of freedom discussed above.

It is an immediate consequence of the properties of graphical models cited in the second section of this article that w-collapsibility and related properties may be easily determined by a graphical analysis of the interaction graph of the model. It is, as mentioned above, a question of determining the minimal cut-set(s) separating a pair of vertices from the rest of the graph. Algorithms for this problem exist (Kreiner (1987)), but in the case of graphs with relatively few vertices it is easy to determine the cut-sets by visual inspection of the graph. The technique will be illustrated in the following example where the strategy for model checking will be used in connection with the initial analysis of a large sparse table. It should however be stressed, that the strategy may be applied in any phase of the statistical analysis.

7. Example 2

We illustrate the conditional test and the strategy for model checking with an example of data from the survey investigation on teachers in the Danish primary school. The data were collected in 1978 and comprised several different variables from which the following seven were choosen for the present example:

A. **Stress** categorized as low, medium or high degree of stress.
B. **Number of periods absent from work because of sickness** categorized into no periods, 1–3 periods and more than 3 periods.
C. **The degree to which the works seems overwhelming** categorized into four categories: low degree, general high degree, high degree specifically in relation to the educational tasks of the teacher, and high degree specifically in relation to the social work of the teacher.
D. **The degree to which the teacher found the work divided between too many classes, too many subjects and/or too many shifts between classes pr. day** (low, medium or high degree of scatteredness).
E. **Number of problems experienced** (few, medium, many).
F. **Urbanization** classified according to four ordinal categories from low to high degree.
G. **Sex.**

A total of 2 748 teachers were classified in the contingency table formed by the seven variables. The table containing 2 592 cells is far too large to print in this format but have been published in a research report from the Danish Institute for Educational Research (Kreiner, 1984a). Although we have more than one teacher per cell we have a typical example of a large sparse table as the teachers are non-uniformly distributed with respect to all variables.

F and G will be considered explanatory variables and the other response variables. For this reason the FG interaction will be included in all models considered in this analysis.

We notice also that all variables except overwhelmingness (C) are binary or ordinal.

The initial analysis were performed according to the strategy discussed above.

The first step consisted of tests of conditional independence of each pair of factors given the other variables of the table.

Following this step the graphical model defined by the conditional independencies accepted at the first step of the analysis were constructed and the adequacy of this model examined as prescribed by the strategy presented above.

The test statistics used for this analysis were the likelihood ratio test, the standard χ^2 test and–in the case of two ordinal or binary variables–a partial γ coefficient summarizing the values of the Goodman–Kruskal γ's calculated for each slice of the three-dimensional table defined after stacking of the conditioning factors (Davis, 1967).

Exact conditional p-values (one-sided in the case of the partial γ's) were approximated from random samples of 500 tables generated according to the above mentioned algorithm and the hypotheses accepted if all the corresponding p-values were greater than 1%.

Included also in Table 5 are the approximate p-values determined by the asymptotic χ^2 distribution with the degrees of freedom adjusted to compensate for empty estimated cells as suggested by Bishop *et al.* (1975).

The results are shown in Table 5 and Fig. 2.

The hypotheses of conditional independence were rejected in 10 of the 20 cases leading to the graphical model shown in Fig. 2 and with the generators (readily read off the interaction graph):

[ABE][ACE][BDE][BEF][BFG].

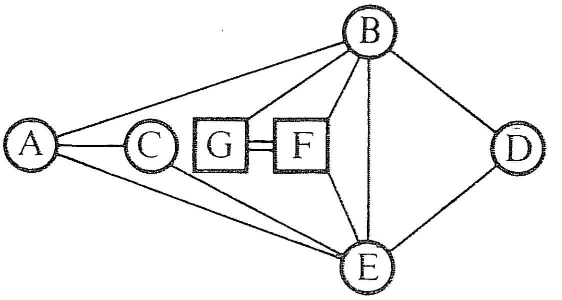

Fig. 2. The interaction graph defined by the conditional independencies accepted in Table 5. Response variables are represented by circles, explanatory variables by squares and interactions fixed in the model by double edges.

Comparison with the confidence limits given in Table 2 shows that the "asymptotic" p-values in most cases fall far below the lower 95% confidence limits of the "estimated" p-values.

Of further interest in Table 5 is the marked discrepancy not only between the exact and "asymptotic" p-values, but also between the asymptotic p-values of the likelihood ratio and the χ^2 test.

Table 5. *Conditional tests of all zero partial association hypotheses for the contingency table classifying teachers according to stress (A), absence (B), overwhelmingness (C), scatteredness (D), problems (E), urbanization (F) and sex (G).* P_e *= exact conditional p-value;* P_a *= asymptotic p-value*

Hypothesis	Degrees of freedom	LR	p-Values P_e^*	P_a	χ^2	p-Values P_e^*	P_a	γ	P_e^*	Comments A† R‡
A⊥B|CDEFG	238	346.1	0.000	0.000	350.0	0.000	0.000	0.396	0.000	R
A⊥C|BDEFG	251	345.5	0.004	0.000	349.9	0.018	0.000	–	–	R
A⊥D|BCEFG	221	269.7	0.314	0.014	255.4	0.280	0.056	0.075	0.162	A†
A⊥E|BCDFG	243	289.8	0.306	0.021	286.2	0.222	0.030	0.277	0.000	R
A⊥F|BCDEG	299	333.3	0.504	0.084	353.5	0.136	0.017	0.061	0.110	A†
A⊥G|BCDEF	140	190.1	0.614	0.003	158.2	0.460	0.139	0.169	0.012	A†
B⊥C|ADEFG	272	342.7	0.442	0.002	290.5	0.640	0.210	–	–	A†
B⊥D|ACEFG	223	320.4	0.102	0.000	277.2	0.104	0.008	0.209	0.000	R
B⊥E|ACDFG	246	335.8	0.184	0.000	309.1	0.078	0.004	0.170	0.000	R
B⊥F|ACDEG	320	440.9	0.046	0.000	432.4	0.004	0.000	0.153	0.000	R
B⊥C|ACDEF	165	268.4	0.042	0.000	228.1	0.018	0.001	0.276	0.000	R
C⊥D|ABEFG	260	325.1	0.374	0.004	299.6	0.318	0.045	–	–	A†
C⊥E|ABDFG	284	396.6	0.004	0.000	367.3	0.060	0.001	–	–	R
C⊥F|ABDEG	369	421.6	0.660	0.030	403.2	0.538	0.107	–	–	A†
C⊥G|ABDEF	178	224.9	0.997	0.010	178.1	0.990	0.483	–	–	A†
D⊥E|ABCFG	241	402.2	0.000	0.000	368.3	0.000	0.000	0.341	0.000	R
D⊥F|ABDEG	291	359.8	0.778	0.004	306.4	0.844	0.256	0.023	0.290	A†
D⊥G|ABCEF	153	208.2	0.710	0.002	171.0	0.700	0.151	0.008	0.441	A†
E⊥F|ACDEG	321	371.4	0.730	0.028	357.5	0.428	0.078	0.115	0.000	R
E⊥G|ABEDF	153	187.6	0.930	0.030	157.4	0.999	0.387	0.008	0.441	A†

* P_e: calculated for 500 randomly generated tables.
† A: acceptance.
‡ R: rejection.

Table 6. *w-Collapsibility properties of the model given in Fig. 2*

Conditional independent variables	w-Order	separators	Marginal conditional independencies
A and D	2	BE	$A \perp D \mid BE$*
A and F	2	BE	$A \perp F \mid BE$*
A and G	2	BE and BF	$A \perp G \mid BE$ and $A \perp G \mid BF$
B and C	2	AE	$B \perp C \mid AE$*
C and D	2	BE and AE	$C \perp D \mid BE$ and $C \perp D \mid AE$
C and F	2	BE and AE	$C \perp F \mid BE$ and $C \perp F \mid AE$
C and G	2	BE, BF and AE	$C \perp G \mid BE$, $C \perp G \mid BF$ and $C \perp G \mid AE$
D and F	2	BE	$D \perp F \mid BE$*
D and G	2	BE and BF	$D \perp G \mid BE$ and $D \perp G \mid BF$
E and G	2	BF	$E \perp G \mid BF$*

* s-Collapsibility.

The second step of the analysis consists of

(a) Derivation of results concerning conditional independence in marginal tables for the base model (Fig. 2) (w- and s-collapsibility).
(b) Treating these properties as hypotheses, the base model may be checked by the corresponding tests of conditional independence in marginal tables, using the same methods as in the first step of the analysis.

Table 6 summarizes the collapsibility properties of all cases of conditionally independent variables. The w-order is in all cases found to be 2. The second step of the analysis will therefore consist of tests in four-dimensional tables.

In five cases the collapsibility corresponded to a proper decomposition of the model, implying not only w-collapsibility but also s-collapsibility.

Consider for instance the case of the conditionally independent variables, A and D. The path from A to D may be cut by the vertices/variables B and E, which therefore constitute the w-separators of the AD problem.

To determine whether the collapsibility over C, F and G corresponds to a proper decomposition of the model it is–according to proposition 3.6 of Lauritzen (1982) or Theorem 2 of Kreiner (1987)–sufficient to consider the boundary of each connected component in the subgraph induced by these variables.

The CFG subgraph contains two connected components, [C] and [FG], with boundaries given by respectively A and E, and B and E. As both these boundaries are complete in Fig. 2, it follows that the graph may be decomposed into two parts containing respectively the A, D, B and E variables and the C, F, G, B and E variables. The implied test of conditional independence of A and D in the collapsed four-dimensional table is therefore equivalent to the test of conditional independence of A and D in the complete table with the model defined by Fig. 2 with the AD interaction included as alternative.

Each of the above mentioned relationships corresponds to a marginal table found by collapsing over three of the original seven variables.

The results of the (marginal) tests of these hypotheses are shown in Table 7.

In four of the 16 cases the hypothesis of marginal conditional independence were clearly rejected indicating the necessity of including certain interactions in the base model.

Which interaction to include is however not necessarily implied from the results of Table 7.

To see this one has to consider the implications of adding only one interaction to the model.

Table 7. *Conditional tests of all zero partial associations accepted in the first step of the analysis (Table 5). All tests are performed in tables collapsed as a consequence of the first step of the analysis*

Hypothesis	Degrees of freedom	p-values			p-Values					Comments A† R‡
		LR	P_e^*	P_a	χ^2	P_e^*	P_a	γ	P_e^*	
A⊥D|BE	32	26.8	0.830	0.730	21.7	0.920	0.920	0.112	0.033	A†
A⊥F|BE	54	47.2	0.830	0.730	47.6	0.710	0.720	−0.061	0.090	A†
A⊥G|BE	17	26.6	0.087	0.064	25.5	0.063	0.084	−0.176	0.000	R
A⊥G|BF	24	31.1	0.240	0.150	29.0	0.220	0.220	−0.092	0.067	A†
B⊥C|AE	48	65.5	0.067	0.047	71.3	0.044	0.016	–	–	A†
C⊥D|AE	45	76.6	0.003	0.002	76.4	0.046	0.002	–	–	R
C⊥D|BE	44	67.7	0.016	0.012	65.1	0.043	0.021	–	–	A†
C⊥F|BE	72	92.0	0.116	0.056	83.3	0.106	0.093	–	–	A†
C⊥F|AE	72	108.6	0.003	0.003	104.5	0.003	0.007	–	–	R
C⊥G|BE	24	29.8	0.310	0.190	26.1	0.350	0.350	–	–	A†
C⊥G|BF	35	48.1	0.160	0.069	40.7	0.210	0.230	–	–	A†
C⊥G|AE	24	33.1	0.210	0.100	28.2	0.230	0.250	–	–	A†
D⊥F|BE	48	56.6	0.300	0.180	52.7	0.300	0.300	0.018	0.330	A†
D⊥G|BE	16	17.2	0.420	0.370	16.5	0.400	0.420	0.099	0.011	A†
D⊥G|BF	24	39.7	0.026	0.023	39.1	0.014	0.027	0.130	0.001	R
E⊥G|BF	24	41.4	0.023	0.015	42.0	0.013	0.013	0.068	0.075	A†

* P_e: calculated for 700 randomly generated tables.
† A: acceptance.
‡ R: rejection.

In the present example evidence was found against the conditional independence of variables A and G, variables C and D, variables C and F, variables D and G. It therefore seems natural to consider one or more of these interactions as candidates for inclusion in the model. It will not however be necessary to include all four interactions.

Consider for instance the model with only the interaction between variables C and F included. This model, [ABE][ACE][BDE][BEF][BFG][CEF], is shown in Fig. 3. It is easily seen, that the inclusion of the CF interaction not only remedies the rejected hypothesis of conditional independence between C and F, but also between variables A and G, which would no longer be conditional independent given B and E (but only given B and F) and variables C and D which, given the model shown in Fig. 3, are conditional independent given B and E, but not given A and E.

The rejection of hypotheses in the second step of the analysis should in fact be interpreted as evidence against the model shown in Fig. 2 as a whole and not necessarily as evidence against hypotheses concerning specific interactions.

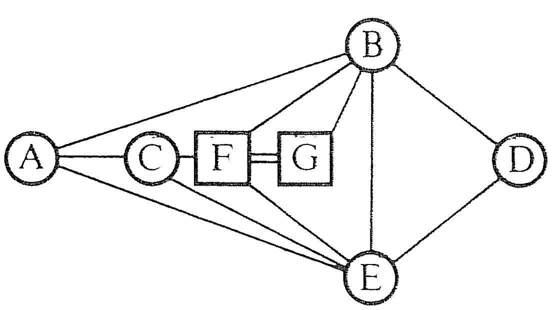

Fig. 3. The interaction graph of the model given in Fig. 2, with the CF interaction included. Response variables are represented by circles, the explanatory variables by squares and interactions fixed in the model by double edges.

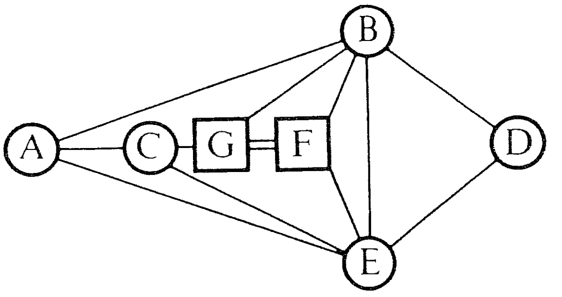

Fig. 4. The interaction graph of the model given in Fig. 2, with the CG interaction included. Response variables are represented by circles, the explanatory variables by squares and interactions fixed in the model by double edges.

This point is further stressed by the fact that in all four cases where the hypotheses of conditional independence in marginal tables were rejected, the hypotheses concerning conditional independence of the same variables in other marginal tables were accepted.

If you are looking for the most parsimonious model which fits the data it seems natural to ask how few and which interactions it is necessary to include in the model to correct these deficiencies.

The answer is that it is sufficient to consider one interaction, namely the one concerning the C and G variables.

Although the hypotheses concerning the conditional independence of C and G were accepted, the inclusion of the CG interaction in the model (Fig. 2) will automatically render the four rejected hypotheses invalid.

As this is the only interaction, the inclusion of which has these consequences, we proceed to accept the model with the CG interaction included as the most parsimonious graphical model which gives a satisfactory description of the contingency table.

The model is shown in Fig. 4. It has generators [ABE][ACE][BDE][BEF][BFG][CG] and is characterized by the conditional independence of the following pairs of variables:

A and D: Stress and scatteredness
A and F: Stress and urbanization
A and G: Stress and sex
B and C: Absence and overwhelmingness
C and D: Overwhelmingness and scatteredness
C and F: Overwhelmingness and Urbanization
D and F: Scatteredness and urbanization
D and G: Scatteredness and sex
E and G: Problems and sex

Subsequent checking of the model (Fig. 4) according to the same strategy as was used for the model (Fig. 2) presents no evidence against the model. Details will not be given here.

8. Discussion

The computer programs developed for the conditional analysis of multidimensional contingency tables have been applied in several different contexts.

The present example was chosen among several alternatives with the following two objectives in mind:

1. It should demonstrate the use of the conditional methods in a situation where the asymptotic methods might be expected to fail.

2. It should demonstrate the applicability of the methods for the analysis of a very large table and thereby demonstrate that the use of these methods is more than an interesting possibility.

The methods are, of course, not inexpensive and the price of computer resources is responsible for the limited number of tables–in most cases 500–generated per test and thus also responsible for a relatively low precision of the approximation of the p-values. The point however is that the inaccuracy of the p-values is quite transparent, and that it is up to the statistician alone to decide how precise the approximation should be (subject only to possible economic constraints). The same cannot be said of the asymptotic p-values, which may be very inaccurate even in cases where the asymptotic approximations could reasonably be expected to be adequate (see for instance some of the cases of tests in four-dimensional tables given in Table 7).

In addition the following conclusions may be drawn from the above analysis:

1. That the partial Goodman–Kruskal gamma in several cases disclose interactions indectable by the likelihood ratio or χ^2 tests.
2. That the asymptotic p-values of the likelihood tests in (almost) all cases are smaller than the asymptotic p-values of the χ^2 test which are again smaller than the exact p-values. The use of asymptotic p-values will thus typically lead to the rejections of too many hypotheses.
3. A marked discrepancy between the asymptotic p-values of the likelihood ratio and χ^2 tests, leading in many cases to markedly different results of the analysis. This discrepancy does not occur in connection with the exact p-values. An observed difference between p-values found from respective likelihood ratio and χ^2 tests calculated for the same hypotheses may in fact be taken as a warning that the asymptotic properties are invalid.
4. That the discrepancy between exact and asymptotic p-values disappears at a much quicker rate for the χ^2 test than for the likelihood ratio test.

Programs

Computer programs written in ALGOL for conditional analysis of contingency tables of dimension up to but not exceeding seven and programs for the graph-theoretical analysis of interaction graphs have been implemented at the Regional Computing Centre at the University of Copenhagen (Kreiner, 1984b and 1987). Copies are available from the author. A PLI version of the programs has been implemented at the County Hospital of Copenhagen.

Acknowledgements

The author wishes to thank David Edwards and the referee for helpful comments and Lone Aalekjaer for developing and testing the algorithms and programs for the graphical analysis.

References

Agresti, Alan, Wackerly, Dennis, & Boyett, J. P. (1979). Exact conditional tests for cross-classifications: approximation of attained significance levels. *Psychometrika* **44,** 75–83.

Andersen, A. H. (1974). Multidimensional contingency tables. *Scand. J. Statist.* **1,** 115–127.

Asmussen, S. & Edwards, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika* **70,** 567–578.

Birch, M. W. (1965). The detection of partial association, II: The general case *J. Roy. Statist. Soc. Ser. B* **27,** 111–124.

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1975). *Discrete multivariate analysis.* MIT Press, Cambridge.

Boyett, James M. (1979). Random R × C tables with given row and column totals. *Appl. Statist.* **29**, 329–332.

Darroch, J. N., Lauritzen, S. L. & Speed, T. P. (1980). Markov fields and log linear interactions model for contingency tables. *Ann. Statist.* **8**, 522–539.

Davis, J. A. (1967). A partial coefficient for Goodman and Kruskal's gamma. *J. Amer. Statist. Assoc.* **62**, 189–193.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* **44**, 1–31.

Edwards, D. G. & Havranek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika* **72**, 339–351.

Edwards, D. G. & Kreiner, S. (1983). The analysis of contingency tables by graphical methods. *Biometrika* **70**, 553–565.

Fienberg, S. (1980). *The analysis of cross-classified categorical data.* MIT Press, Cambridge.

Kreiner, S. (1984a). Analysis of multiple contingency tables by exact conditional tests for zero partial association. Danish Institute of Educational Research, Copenhagen.

Kreiner, S. (1984b). EXABIRCH–a program for exact conditional analysis of multiple contingency tables. Danish Institute of Educational Research, Copenhagen.

Kreiner, S. (1987). On Collapsibility of Multidimensional Contingency Tables. In Mortensen, L.: *Symposium on Applied Statistics 1987.* UNI · C, Aarhus.

Lauritzen, S. L. (1982). *Lectures on contingency tables.* Aalborg University Press.

Patefield, W. M. (1981). An efficient method of generating random R × C tables with given row and column totals. *Appl. Statist.* **30**, 91–7.

Sachs, L. (1982). *Applied statistics. A handbook of techniques.* Springer Verlag, New York.

Sundberg, R. (1975). Some results about decomposable (or Markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests. *Scand. J. Statist.*, **2**, 71–79.

Wermuth, N. (1976). Model search among multiplicative models. *Biometrics* **32**, 253–64.

Whittaker, J. (1982). GLIM syntax and simultaneous test for graphical log linear models. In *GLIM 82: Proceedings of the international conference on generalised linear models.* Springer Verlag, New York.

Williams, D. A. (1976). Improved likelihood ratio tests for complete contingency tables. *Biometrika*, **63**, 33–37.

Received January 1984, in final form October 1986

Svend Kreiner, the Danish Institute for Educational Research, Hermodsgade 28, DK-2200 Copenhagen N Denmark