Collapsibility and Response Variables in Contingency Tables

Soren Asmussen; David Edwards

Biometrika, Vol. 70, No. 3 (Dec., 1983), 567-578.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198312%2970%3A3%3C567%3ACARVIC%3E2.0.CO%3B2-U

Biometrika is currently published by Biometrika Trust.

Biometrika (1983), **70**, 3, *pp.* 567–78
Printed in Great Britain

Collapsibility and response variables in contingency tables

BY SØREN ASMUSSEN
Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, Denmark

AND DAVID EDWARDS
Regional Computing Centre, University of Copenhagen, Copenhagen, Denmark

SUMMARY

Various definitions of the collapsibility of a hierarchical log linear model for a multidimensional contingency table are considered and shown to be equivalent. Necessary and sufficient conditions for collapsibility are found in terms of the generating class. It is shown that log linear models are appropriate for tables with response and explanatory variables if and only if they are collapsible onto the explanatory variables.

Some key words: Collapsibility; Contingency table; Graphical model; Interaction graph; Log linear model; Response variable; S-sufficiency.

1. INTRODUCTION AND PRELIMINARIES

Two topics in the field of hierarchical log linear models for multidimensional contingency tables, collapsibility and response variable models, are considered and shown to be closely related.

Some models have the property that relations between a set of the classifying factors may be studied by examination of the table of marginal totals formed by summing over the remaining factors. Such models are said to be collapsible onto the given set of factors. Collapsibility has important consequences for hypothesis testing and model selection, and can be useful in data reduction. We consider various definitions of collapsibility and show their equivalence. Furthermore, necessary and sufficient conditions for collapsibility are found in terms of the generating class.

Many tables analysed in practice involve response variables. Simple examples, one of which is given in § 3, suffice to show the importance of distinguishing between response and explanatory variables: first, that inappropriate models may be avoided, and second that natural and relevant models that are not log linear may be considered. This paper characterizes appropriate and inappropriate log linear models for tables with response variables and some alternative approaches for the analysis of such tables are briefly considered.

We consider a multidimensional contingency table N based on a set of classifying factors Γ. For a given subset a of Γ we are interested in the table of marginal totals N_a, that is to say the table of cell counts summed over the remaining factors a^c, that is the complement of a in Γ. We identify a hierarchical log linear model L, that is the set of probabilities $p \in L$, with its generating class, whose elements, generators, are given in square brackets: thus for example the model $[AB][BCD]$ for a 4-way table corresponds in the usual notation to

$$\log m_{ijkl} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} + \lambda_{kl}^{CD} + \lambda_{jl}^{BD} + \lambda_{jkl}^{BCD}.$$

We denote an arbitrary cell in N as i and the corresponding marginal cell as i_a. We denote the number of objects in cell i as $n(i)$ and the number of objects in the marginal cell i_a as $n(i_a)$. We are interested in the probabilities $p(i)$ of an object falling in cell i, and the corresponding marginal probabilities $p(i_a)$ formed by summing $p(i)$ over a^c. Similarly $m(i)$ denotes the expected number of objects in cell i and $m(i_a)$ the corresponding marginal quantity.

We assume the distribution of the table is multinomial, that is

$$\text{pr}\,[N = \{n(i)\}] = \{n!/\Pi_i n(i)!\}\,\Pi_i p(i)^{n(i)},$$

where n is the total number of objects. It is well known that the maximum likelihood estimate $\hat{p}$ of $p \in L$ is given as the unique solution to the system of equations:

(i) $\hat{p} \in L$,

(ii) $\hat{p}(i_c) = n(i_c)/n$ for all generators c of L.

For a given log linear model L we define the interaction graph of L as the undirected graph whose vertices correspond to the classifying factors in Γ and whose edges are given by the 2-factor interactions present in the model. See for example Fig. 1a. One may interpret the interaction graph in the following way (Darroch, Lauritzen & Speed, 1980); if two disjoint subsets of vertices a_1 and a_2 are separated by a subset a_3 in the sense that all paths from a_1 to a_2 go through a_3, then the variables in a_1 are conditionally independent of those in a_2 given the variables in a_3.

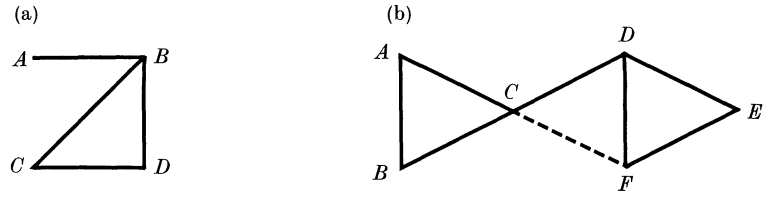

Fig. 1a. The interaction graph of $[AB][BCD]$ and $[AB][BC][CD][BD]$.

Fig. 1b. The interaction graphs of $[ABC][CD][DEF]$ and $[ABC][CDF][DEF]$.

We use the notation $a \perp b \mid c$ to denote the conditional independence of a and b given c.

We say that two vertices in a graph are adjacent if there is an edge between them and we define the boundary of a subset a of Γ, written ∂a, as those vertices that are not in a but are adjacent to some vertex in a. The closure of a is defined as the union of a and its boundary and is denoted $cl(a)$. A set a is called complete if all possible edges between the vertices of a are present in the graph.

We can define an equivalence relation on the graph as $\alpha \sim \beta$ if and only if there is a path connecting α and β. The subgraphs induced by the equivalence relation are termed the connected components of the graph.

Clearly, many different log linear models may have the same interaction graph, as long as they contain the same 2-factor interactions. Models with the maximal permissible higher-order interactions corresponding to a given graph are termed graphical models: it is shown by Darroch *et al.* (1980) that all decomposable models are graphical. More specifically, decomposable models are graphical models whose graphs are triangulated, i.e. contain no cycle of length greater than 3 without a chord.

Graphical models are in many ways analogous to covariance selection models, as shown by, for example, Wermuth (1976). Presumably much of the discussion of the

present paper has counterparts within that framework as well, but we have not looked into this.

2. Collapsibility

For a given hierarchical log linear model L defined on N we define its restriction L_a on N_a in the following way: the generating class of L_a is formed by deleting all occurrences of factors in a^c, the complement of a in Γ, in the generating class of L, and then removing unnecessary elements. Thus if $a = (A, B, C)$ and $L = [AB][BCD][AD]$, then $L_a = [AB][BC][A] = [AB][BC]$.

Write the probability of cell i_a under L_a as say $p_a(i_a)$.

Definition. L is collapsible onto a if one of the two following equivalent properties hold:
(i) for all $p = p(i) \in L$, we have that $p(i_a) \in L_a$;
(ii) for all i_a, $\hat{p}(i_a) = \hat{p}_a(i_a)$.

A further characterization in terms of S-sufficiency and motivation is given in §4, while some discussion of the concept is given at the end of the present section. As justification, we give here only a proof of the equivalence of the criteria. To prove that (ii) implies (i), note that if $p \in L$ is the true probability measure, then as n tends to ∞, $\hat{p}$ tends to p and hence, since all $p(i_a) > 0$,

$$p(i_a) = \lim \hat{p}(i_a) = \lim \hat{p}_a(i_a) \in L_a.$$

To prove (i) implies (ii), note that if $c \subseteq a$ is contained in a generator, then $\hat{p}(i_c) = n(i_c)/n$. But in conjunction with $\hat{p}_a \in L_a$ these are the equations determining $\hat{p}_a$. Thus $\hat{p}(i_a) \in L_a$ implies $\hat{p}_a(i_a) = \hat{p}(i_a)$.

It is easy to see that L is always collapsible onto a if a is contained in a generator. Further simple examples can be found in Theorem 2·1 and Corollary 2·2 below, and the property is completely characterized in Theorem 2·3. The simplest example of non-collapsibility is given by $L = [AB][BC]$ and $a = (A, C)$.

Note in connexion with (i) that always $L_a \subseteq \{p(i_a)\colon p \in L\}$.

Definition. Two subsets a and b form a decomposition of Γ relative to a hierarchical log linear model L if $a \cup b = \Gamma$, a and b are separated by $a \cap b$, and $a \cap b \subseteq c$ for some generator c of L.

Theorem 2·1. *If a and b form a decomposition of Γ relative to L then*

$$\hat{p}(i) = \hat{p}_a(i_a)\,\hat{p}_b(i_b)/\{n(i_{a\cap b})/n\}, \quad \hat{p}(i_a) = \hat{p}_a(i_a).$$

Proof. The first formula is given by Haberman (1974, p. 166 ff.) and Lauritzen (1982). The second formula follows by summing over $b \setminus a$ and noting that $\hat{p}_b(i_{a\cap b}) = n(i_{a\cap b})/n$ since $a \cap b$ is contained in a generator of L_b.

Corollary 2·2. *If $\partial(a^c) \subseteq c$ for some generator c of L, then L is collapsible onto a.*

Proof (Lauritzen, 1982). The sets a and $b = cl(a^c)$ form a decomposition of Γ relative to L exactly when $\partial(a^c) \subseteq c$ for some generator c of L.

Theorem 2·3. *A hierarchical log linear model L is collapsible onto a if and only if the boundary of every connected component of a^c is contained in a generator of L.*

Proof. For sufficiency, let $b_1, \ldots, b_p$ be the connected components of a^c. By Corollary 2·2, L is collapsible onto $b = b_1^c$. Clearly the boundary of b_2 is contained in a generator of L_b, so that, again citing Corollary 2·2, L_b is collapsible onto $(b_1 \cup b_2)^c$. Thus L is collapsible onto $(b_1 \cup b_2)^c$ too. Continuing in this fashion, we obtain that L is collapsible onto a as required.

For necessity, we show that if the stated condition does not hold we can construct a $p \in L$ such that (i) does not hold. We consider first an illustrative particular example, and then proceed to the general case.

Let $L = [AD][BC][CD]$ and $a = (A, B)$. Then the boundary of (C, D) is (A, B) and since this set is not contained in a generator of L, the stated condition does not hold.

We construct a $p \in L$ as follows. Let

$$p_{ijkl} = c(\theta) \exp(\lambda_{il}^{AD} + \lambda_{jk}^{BC} + \lambda_{kl}^{CD}),$$

where

$$\lambda_{kl}^{CD} = \begin{cases} 0 & (k = l = 1 \text{ or } k = l = 2), \\ -\infty & \text{otherwise}, \end{cases}$$

$$\lambda_{il}^{AD} = \begin{cases} \theta & (i = l = 1), \\ 0 & \text{otherwise}, \end{cases} \qquad \lambda_{jk}^{BC} = \begin{cases} \theta & (j = k = 1), \\ 0 & \text{otherwise}, \end{cases}$$

and $c(\theta)$ is a normalizing constant. Then

$$\begin{aligned} p_{ij..} &= c(\theta) \Sigma_{kl} \exp(\lambda_{il}^{AD} + \lambda_{jk}^{BC} + \lambda_{kl}^{CD}) \\ &= c(\theta) \sum_{k=1}^{2} \exp(\lambda_{ik}^{AD} + \lambda_{jk}^{BC}) \\ &= c(\theta)(1 + \exp[\theta\{\delta(i, 1) + \delta(j, 1)\}]), \end{aligned}$$

where δ is the Kronecker delta.

If $p_{ij..} \in L_a = [A][B]$, then the cross-product ratio between the cells $(1, 1)$ and $(2, 2)$ is unity, that is $p_{11..}p_{22..} = p_{12..}p_{21..}$. Writing $\zeta = e^{\theta}$, we obtain $(1 + \zeta^2)(1 + \zeta^0) = (1 + \zeta)^2$. But this cannot be true for all ζ. Hence $p_{ij..}$ is not in L_a and L is not collapsible onto a.

We now consider the general case. There exists a connected component b of a^c such that ∂b is not contained in c for any generator c of L. Write $b = \{Z_1, \ldots, Z_p\}$, $\partial b = \{Y_1, \ldots, Y_q\}$ and $a = \{Y_1, \ldots, Y_r\}$ say, where $r \geqslant q$. For each factor Y_j in ∂b choose an adjacent factor $Z_{(j)}$ in b. Define for Z_j, Z_k adjacent

$$\lambda_{jk}(z_j, z_k) = \begin{cases} 0 & (z_j = z_k = 1 \text{ or } z_j = z_k = 2), \\ -\infty & \text{otherwise}, \end{cases}$$

and, for $j = 1, \ldots, q$,

$$\mu_j(y_j, z_{(j)}) = \begin{cases} \theta & (y_j = z_{(j)} = 1), \\ 0 & \text{otherwise}. \end{cases}$$

Then define

$$p(i) = c(\theta) \exp\left\{ \sum_{j=1}^{q} \mu_j(y_j, z_{(j)}) + \Sigma_{j,k} \lambda_{jk}(z_j, z_k) \right\}.$$

Clearly $p \in L$, and since b is connected we obtain

$$p(i_a) = Mc(\theta) \sum_{k=1}^{2} \exp\left\{\sum_{j=1}^{q} \mu_j(y_j, k)\right\},$$

where M is the number of cells in N_d, where $d = a^c \backslash b$. Hence

$$p(i_a) = Mc(\theta)\left(1+\exp\left[\theta\left\{\sum_{i=1}^{q} \delta(y_i, 1)\right\}\right]\right). \tag{1}$$

Now suppose that $p(i_a) \in L_a$. Then the q-factor interaction between $Y_1, \ldots, Y_q$ vanishes, and hence the cross-product ratio between cells $j_1 = \{(i_\alpha)_{\alpha \in a}: i_\alpha = 1 \text{ for all } \alpha\}$ and $j_2 = \{(i_\alpha)_{\alpha \in a}: i_\alpha = 2, \alpha \in \partial b;\ i_\alpha = 1 \text{ otherwise}\}$ is unity, i.e.

$$\Pi p(i_a) = \Pi p(i_a), \tag{2}$$

where the products on the left-hand side and right-hand side are respectively over

$$S_1 = \left\{i_a: y_s = 1 \text{ or } 2, s = 1, \ldots, q;\ y_s = 1, s = q+1, \ldots, r;\ \sum_{s=1}^{q} y_s \text{ even}\right\},$$

and S_2 the corresponding set for odd Σy_s. Inserting (1) in (2) and writing $\zeta = e^\theta$ we obtain

$$\prod_{k \leq q, k \text{ even}} (1+\zeta^k)^{\psi(k)} = \prod_{k \leq q, k \text{ odd}} (1+\zeta^k)^{\psi(k)},$$

where $\psi(k) = q!\{k!(q-k)!\}^{-1}$. This cannot hold for all ζ, since, for example, the constant term is 2 on the left-hand side but 1 on the right-hand side. Hence L is not collapsible onto a.

Before proceeding further, we give some remarks and examples to illustrate the theorem. Note first that the connected components describe the maximal partitioning of a^c into subjects which are conditionally independent given the factors in a. Also, if L is graphical, the condition simply means that the boundary of every connected component is complete.

Example 1. Let $\Gamma = (A, B, C, D)$, $a = (A, B, C)$ and

$$L = [AB][BC][AC][AD][BD][CD].$$

Then the boundary of $a^c = (D)$ is (A, B, C) and since the term ABC is not contained in a generator of L, L is not collapsible onto a.

Example 2. Let $a = (A, B, C)$, $b = (D, E)$ and $L = [AC][ABD][BCE]$. Then the components of b are not connected, the boundary of D is (A, B), the boundary of E is (B, C), and since $[AB]$ and $[BC]$ are contained in generators of L, L is collapsible onto a.

COROLLARY 2·4. *If L is collapsible onto a, then*

$$\hat{p}(i) = \hat{p}_a(i_a)\, \Pi_b[\hat{p}_{cl(b)}(i_{cl(b)})/\{n(i_{\partial b})/n\}],$$

where the product is over connected components b of a^c.

Proof. Apply Theorem 2·1 to each connected component in turn.

COROLLARY 2·5. *A graphical model is collapsible onto a if and only if*

$$a_1, a_2 \subseteq a, \quad a_1 \perp a_2 \mid s \text{ implies } a_1 \perp a_2 \mid s \cap a.$$

Proof. We use here the easily proved fact that $a_1 \perp a_2 \mid s$ if and only if s separates a_1 and a_2. Suppose first that L is collapsible onto a, and that the stated condition does not hold, so that there exists a path from a_1 to a_2 which intersects s but not $s \cap a$. For each connected component b_i say of a^c the path intersects, we can replace the segment in b_i by an edge in ∂b, e_i say, since ∂b_i is complete. The edges $\{e_i\}$ must intersect s at some vertex, for otherwise we have constructed a path from a_1 to a_2 that does not intersect s. Thus the original path intersects $s \cap a$, contrary to assumption.

Conversely, if L is not collapsible onto a, there exist a connected component b of a^c and nonadjacent vertices $\alpha, \beta \in \partial b$. Hence $\alpha \perp \beta \mid \Gamma - \alpha - \beta$ is true but $\alpha \perp \beta \mid a - \alpha - \beta$ is not.

We note also that the condition stated in Corollary 2·5 is necessary, but not sufficient, for the collapsibility of hierarchical, nongraphical models.

Expressed loosely, collapsibility onto a subset a means that inference concerning the factors in a not contained in a boundary of a connected component of a^c can be performed in the marginal table N_a. Suppose for example that two models $L_1 \subseteq L_2$ both are collapsible onto a and that they only differ in terms involving variables in a. Then L_b for $b = cl(a^c)$ is the same in both models and so the likelihood ratio test statistic for testing L_1 against L_2 is

$$\begin{aligned} -2 \log Q &= 2 \Sigma_i n(i) \log \{\hat{m}^2(i)/\hat{m}^1(i)\} \\ &= 2 \Sigma_i n(i) \log \{\hat{p}^2(i)/\hat{p}^1(i)\} \\ &= 2 \Sigma n(i_a) \log \{\hat{p}_a^2(i_a)/\hat{p}_a^1(i_a)\} \end{aligned}$$

from Corollary 2·4, i.e. the test can be performed in the marginal table N_a. The same applies to Pearson's test for goodness-of-fit. Since the marginal table always has larger cell counts than the whole table, this enables asymptotic results to be cited with more confidence.

Example. Let $L_1 = [ABC][CD][DEF]$ and $L_2 = [ABC][CDF][DEF]$ and consider the test for $L_1 \subseteq L_2$; see Fig. 1b. Here L_1 and L_2 differ with respect to the presence of the terms λ^{CDF} and λ^{CF}; L_2 is collapsible onto (C, F) but L_1 is not. However, both L_1 and L_2 are collapsible onto $a = (C, D, F)$, so that the test can be performed in the marginal table N_a, that is as a test of $C \perp F \mid D$.

As suggested by a referee, collapsibility can be linked to the idea of invariance of models when some variables are unobserved, as the following example shows. Consider two possible models for a five-way table, $L_1 = [ABC][DE]$ and $L_2 = [ABC][CDE]$, and suppose the factor C were unobserved. If L_1 is the true model, then from collapsibility we can see that $L_1' = [AB][DE]$ holds for the observed factors. Inferences from this model, for example that $A, B \perp D, E$, are valid for the complete, unobserved table.

Under L_2, however, the table of totals over C would not simplify, since L_2 is not collapsible onto (A, B, D, E). Latent class analysis could perhaps be attempted. Clearly it makes no sense to fit latent class models that are collapsible onto the observed variables.

Bishop, Fienberg & Holland (1975, Chapter 2) and Whittemore (1978) have defined collapsibility in a somewhat different spirit by stressing the log linear parameters. The definition studied in the present paper was apparently first stated in the 1979 edition of

Lauritzen (1982) in the form (ii), though Jensen (1978) has some related discussion within the framework of hypothesis testing. One reason for adhering to this point of view is the opinion that the log linear parameters are mainly a mathematical convenience of little intrinsic interest and that the important feature of the model is rather the specification of which interactions are present or not.

3. Response variables

As a simple example, suppose that we have a 3-way table of counts of individuals, where S denotes sex, R denotes race and A attitude to some question of topical interest, and where we suppose that the response A depends on both the individual's sex and race. If one performs a conventional analysis by choosing the log linear model with the best fit, regardless of its interpretation, one may accept the model

$$\log(m_{ijk}^{SRA}) = \lambda + \lambda_i^S + \lambda_j^R + \lambda_k^A + \lambda_{ik}^{SA} + \lambda_{jk}^{RA}.$$

However this model asserts that sex and race are conditionally independent given attitude, which is absurd. A more appropriate model is that sex and race are marginally independent. Birch (1963) considered this model: it has explicit maximum likelihood estimates given by

$$\hat{m}_{ijk} = (n_{i..}\, n_{.j.}/n_{...})\,(n_{ijk}/n_{ij.}) \tag{3}$$

but is not log linear in the three variables. A closely related model also discussed by Birch (1963) specifies in addition to marginal independence that there is no 3-factor interaction between the three variables. This has maximum likelihood estimates given by

$$\hat{m}_{ijk} = (n_{i..}\, n_{.j.}/n_{...})\,(m_{ijk}^{*}/n_{ij.}),$$

where m_{ijk}^{*} are the fitted values obtained by fitting the model of no 3-factor interaction to the whole table. Neither is this model log linear.

Thus ignoring the distinction between response and explanatory variables has two dangers: first that inappropriate models may be used, and second that natural and relevant models that are not log linear may be overlooked.

The class of appropriate models was defined by Goodman (1973); see also Fienberg (1980, Chapter 7). To define the class, let a be the set of explanatory variables, and b the set of response variables. The joint density of (a, b) can be factorized into a product of the marginal density of a and the conditional density of b given a:

$$p^J(i) = p^M(i_a)\, p^C(i_b \mid i_a). \tag{4}$$

The class of response variable models is then defined by specifying a log linear model M for the marginal density of a, and a log linear model C for the conditional density of b given a. In practice we can fit M in the ordinary way to the table of marginal totals N_a. Here C is fitted as a log linear model for the whole table: since we are conditioning on a we must include all interactions between the variables in a.

The fitted values for the final joint model J are then obtained as

$$\hat{m}^J(i) = \hat{m}^M(i_a)\,\{\hat{m}^C(i)/n(i_a)\}.$$

For example, the model whose fitted values are given in (3) has $M = [S][R]$ and $C = [SRA]$.

Inference concerning the marginal model and conditional model can be performed separately: useful here is the additivity of the residual deviances, which can easily be

obtained:

$$\begin{aligned} -2\log Q_J &= 2\Sigma n(i)\log\{n(i)/m^J(i)\} \\ &= 2\Sigma n(i)\log[\{n(i_a)\,n(i)\}/\{\hat{m}^M(i_a)\,m^C(i)\}] \\ &= 2\Sigma n(i_a)\log\{n(i_a)/\hat{m}^M(i_a)\}+2\Sigma n(i)\log\{n(i)/\hat{m}^C(i)\} \\ &= -2\log Q_M-2\log Q_C. \end{aligned}$$

The corresponding degrees of freedom are similarly additive.

As we have seen, not all log linear models are response variable models and not all response variable models are log linear. The next two theorems characterize the intersection of these classes. Theorem 3·1 gives conditions for a log linear model to be a response variable model, and Theorem 3·2 gives conditions for a response variable model to be log linear. These results are important for several reasons. First, they enable us to characterize appropriate and inappropriate joint log linear models for contingency tables with response variables. Secondly, having fitted a marginal and a conditional model to a table, it is useful to know when these can be combined to form a log linear model, since these are more familiar and allow a better data reduction to sufficient marginal tables. Thirdly, we can formulate model selection strategies based initially on joint log linear models that may be more convenient to carry out in practice.

Fix now a and let $\mathscr{L}$ be the set of hierarchical log linear models for N, M_a be the set of hierarchical log linear models for the marginal table N_a, C_a the set of conditional models, i.e. containing all interactions between the factors in a, and J_a the set of response variable models generated from M_a and C_a.

THEOREM 3·1. *For $L \in \mathscr{L}$, $L \in J_a$ if and only if L is collapsible onto a. In that case the log linear model M for the marginal density of a is given by $M = L_a$ and the log linear model C for the conditional density of a^c given a is given by $C = [a] \cup L_b$, where $b = cl(a^c)$.*

For interpretation of the formula for C see the example below.

Proof. If $L \in J_a$ then clearly $\hat{p}^L(i_a) = \hat{p}^M(i_a) \in M$ so that collapsibility will follow from $M \subseteq L_a$. That this is indeed the case can be seen, for example, by taking $c = p^C(i_b \mid i_a)$ independent of i_a, i_b in (4). Then $cp^M(i_a) \in L$ for all $p^M(i_a) \in M$ so that L must include at least the interactions in M and this implies immediately that $M \subseteq L_a$.

Conversely, suppose that L is collapsible onto a and define $M = L_a$ and $C = [a] \cup L_b$ where $b = cl(a^c)$. Then if $b_1, \ldots, b_v$ are the connected components of a^c, it is easily seen that $\hat{p}^M(i_a) = \hat{p}_a(i_a) = \hat{p}(i_a)$,

$$\hat{p}^C(i_{a^c} \mid i_a) = \prod_{k=1}^{v} \hat{p}^C(i_{b_k} \mid i_{\partial a^c}) = \prod_{k=1}^{v} \hat{p}_{cl(b_k)}\{i_{cl(b_k)}\}/\{n(i_{\partial b_k})/n\}.$$

Forming the joint model $J = (M, C)$ and using Corollary 2·4 shows that $\hat{p}^J = \hat{p}$. Hence $L = J \in J_a$.

Example 3. Let $L = [ABC][BD][CDE]$ and $a = (B, C, D)$. Then, by Theorem 2·4, L is collapsible onto a and so by Theorem 3·1 it coincides with a response variable model $J = (M, C)$, where $M = [BC][BD][CD]$ and

$$C = [BCD] \cup \{[ABC][BD][CDE]\} = [ABC][BCD][CDE].$$

Example 4. Let $L = [AB][BC][AC][AD][BD][CD]$ and $a = (A, B, C)$. By Theorem 2·4, L is not collapsible onto a and hence does not coincide with a response variable model.

THEOREM 3·2. *For $J = (M, C) \in J_a, J \in \mathscr{L}$ if and only if the boundary of every connected component of a^c in C is contained in a generator of M. In that case $L = M \cup C_b$, where $b = cl(a^c)$.*

Proof. Suppose that $J = (M, C)$ coincides with the log linear model L. Then L is collapsible and by Theorem 3·1 it follows that J coincides with the response variable model (M', C') given by $M' = L_a$, $C' = [a] \cup L_b$, where $b = cl(a^c)$. Since (M, C) are uniquely determined, $M' = M$ and $C' = C$. Thus the connected components of a^c relative to C and L are the same and it is clear by collapsibility that $\partial b_k \subseteq c_k$ for generators c_k of M $(k = 1, \ldots, v)$.

Conversely if the stated condition holds, then the log linear model $L = M \cup C_b$ is collapsible and, by Corollary 2·4,

$$\hat{p}^L(i) = \hat{p}^M(i_a)\hat{p}^C(i_{a^c} \mid i_a) = \hat{p}^J(i)$$

so that $J = L$ is log linear.

Example 5. For $\Gamma = (A, B, C, D, E)$, let $a = (A, B, C)$, $M = [AB][BC][AC]$ and $C = [ABC][ABD][BCE]$. Then $J = (M, C)$ is hierarchical log linear since the boundary of D, (A, B), and the boundary of E, (B, C), are contained in generators of M. The generating class of J is $[AC][ABD][BCE]$.

The conditions we have obtained for log linear models to be appropriate for tables with response variables can in part be interpreted in terms of conditional independence. Corollary 2·5 stipulates that conditional independencies between the explanatory variables must hold in the marginal distribution of the explanatory variables. Thus for example if A, B are explanatory and C a response variable the model $L = [AC][BC]$ is inappropriate since it implies that $A \perp B \mid C$ but not that $A \perp B$; $L = [A][BC]$, on the other hand, is appropriate since it implies both $A \perp B \mid C$ and $A \perp B$.

We note that, dependent, for example, on the sampling scheme, three approaches can be adopted to the analysis of tables with response variables.

First, one can simply condition on the explanatory variables. This is a suitable approach when there is no interest in the mutual dependencies exhibited by the explanatory variables. It may be relevant, for example, when the explanatory variables are demographic, and better demographic information is available from other sources.

Secondly, one can fit marginal and conditional models as described above. When a final model has been selected, Theorem 3·2 can be cited to determine whether it is log linear.

Thirdly, and this may be the more convenient approach in practice, one may choose to remain within the class of log linear models that are collapsible onto the explanatory variables as long as possible. When a 'best' model has been chosen, it may be examined to see which marginal independence relations are not testable in the joint framework, and these may be tested in the marginal table.

We finally mention that the results are easily extended to the models discussed by Goodman (1973) and Fienberg (1980, Chapter 7), where a sequence of sets $a, b_1, \ldots, b_k$ is given. Here the variables in the set b_r $(r = 1, \ldots, k)$ are responses to the variables in the

sets $a, \ldots, b_{r-1}$, and themselves explanatory with regard to $b_{r+1}, \ldots, b_k$. This framework may for example be appropriate when the variables are measured in time sequence, and we exclude the possibility that a variable can depend on another variable measured at a subsequent point of time.

If we define the sets $d_0 = a$, $d_i = b_i \cup d_{i-1}$ $(i = 1, \ldots, k)$, the class of models is defined by the equation

$$p^J = p^{C_0}(d_0) \prod_{i=1}^{k} p^{C_i}(d_i \mid d_{i-1}),$$

where $C_0, C_1, \ldots, C_k$ are log linear models defined on the appropriate marginal tables. We denote this class of models causal chain models, shown in Fig. 2. The theorems of the previous section can easily be extended giving Theorem 3·3.

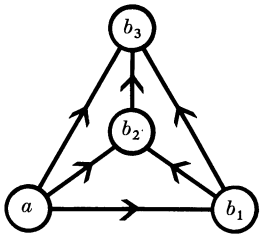

Fig. 2. A causal chain model.

THEOREM 3·3. (a) *Model $L \in \mathscr{L}$ is a causal chain model if and only if it is collapsible onto d_i $(i = 0, \ldots, k-1)$.*
(b) *A causal chain model $J = (C_0, C_1, \ldots, C_k)$ is log linear if and only if the boundary of each connected component of b_r under C_r is contained in a generator of C_{r-1} $(r = 1, \ldots, k)$.*

A subclass of causal chain models is obtained when the sets $b_1, \ldots, b_k$ all consist of single variables, C_0 is the saturated model on N_a, and $C_1, \ldots, C_k$ are graphical. These models have been termed recursive models (Wermuth & Lauritzen, 1983).

Theorem 3·3 can be applied to give an interesting characterization of decomposable models in terms of recursive models. Fulkerson & Gross (1965) proved that an undirected graph is triangulated if and only if there exists an ordering $\sigma = (v_1, \ldots, v_k)$ of the vertices such that each set $X_i = \{v_j \in \delta v_i : j > i\}$ is complete; see also Golumbic (1980, Chapter 4). We thus obtain, using Theorem 3·3, Corollary 3·4.

COROLLARY 3·4. *A graphical model is decomposable if and only if it is recursive.*

This was obtained by Wermuth & Lauritzen (1983) by other means.

4. S-SUFFICIENCY

Let $T = T(X)$ be statistics and suppose that the density of X factorizes as

$$p_{\theta,\eta}(x) = p_\theta(t)\, p_\eta(x \mid t), \tag{5}$$

where the parameters θ, of the marginal distribution of T, and η, of the conditional distribution of X given T, are variation independent. Then T is called S-sufficient for θ

(Barndorff-Nielsen, 1978, p. 49). The notion was introduced by Fraser (1956) for describing '. . . sufficiency for the parameter of interest'.

THEOREM 4·1. *A hierarchical log linear model L is collapsible onto a if and only if, for any n, the marginal table N_a is S-sufficient for $p(i_a)$, or equivalently if and only if $p(i)$ factorizes as*

$$p(i) = p_\theta(i_a)\, p_\eta(i_{a^c} \mid i_a), \tag{6}$$

where θ, η are variation independent.

Proof. We first note that (6) is equivalent to S-sufficiency for $n = 1$. Suppose first that L is collapsible onto a. Then, by Theorem 3·1, $L = (M, C)$ and (4) shows immediately that (6) holds. Thus S-sufficiency of N_a for $n > 1$ follows now by elementary properties of sufficiency and conditioning along the following lines. Write $N = I(1) + \ldots + I(n)$, where $I(k)$ is the table for individual k and similarly $N_a = I_a(1) + \ldots + I_a(n)$. By (6), the joint density of the $I(k)$ is

$$\prod_{k=1}^{n} p_\theta\{i_a(k)\} \prod_{k=1}^{n} p_\eta\{i_{a^c}(k) \mid i_a(k)\}.$$

Since N_a is sufficient for M, the first factor can be factorized according to Neyman's criterion and the S-sufficiency of N_a, based on observation of the $I(k)$, follows easily. To obtain the conclusion based on observation of N, appeal once more to Neyman's criterion and the sufficiency of N for L. We omit the details.

Suppose next that (6) holds. Since L always includes p with the factors in $b = a^c$ irrelevant, we can find η such that $p_\eta(i_b \mid i_a) = c$ independent of i_b, i_a. Thus for any $\theta, p(i) = c p_\theta(i_a) \in L$ by variation independence. But this implies $p_\theta(i_a) \in L_a$, that is L is collapsible onto a.

We would like to thank Steffen Lauritzen for stimulating discussions, and Søren Tolver Jensen, Nanny Wermuth and the referee for useful comments which helped clarify the final formulation of the paper.

REFERENCES

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* New York: Wiley.

BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc.* B **25**, 220–33.

BISHOP, Y. M. M., FIENBERG, S. E. & HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge: Massachusetts Institute of Technology Press.

DARROCH, J. N., LAURITZEN, S. L. & SPEED, T. P. (1980). Markov fields and log linear interactions models for contingency tables. *Ann. Statist.* **8**, 522–39.

FIENBERG, S. E. (1980). *The Analysis of Cross-classified Data.* Cambridge: Massachusetts Institute of Technology Press.

FRASER, D. A. S. (1956). Sufficient statistics with nuisance parameters. *Ann. Math. Statist.* **27**, 838–42.

FULKERSON, D. R. & GROSS, O. A. (1965). Incidence matrices and interval graphs. *Pac. J. Math.* **15**, 835–55.

GOLUMBIC, M. C. (1980). *Algorithmic Graph Theory and Perfect Graphs.* New York: Academic Press.

GOODMAN, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika* **60**, 179–92.

HABERMAN, S. J. (1974). *The Analysis of Frequency Data.* University of Chicago Press.

JENSEN, S. T. (1978). *Flersidede Kontigenstabeller.* Copenhagen: Institute of Mathematical Statistics.

LAURITZEN, S. L. (1982). *Lectures on Contingency Tables*, 2nd edition. University of Aalborg Press, Denmark.

WERMUTH, N. (1976). Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics* **32**, 95–108.

WERMUTH, N. & LAURITZEN, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70**, 537–52.

WHITTEMORE, A. S. (1978). Collapsibility of multidimensional contingency tables. *J. R. Statist. Soc.* B **40**, 328–40.

[*Received August* 1982. *Revised March* 1983]