Søren Asmussen David Edwards

Collapsibility and Response Variables in Contingency Tables

Preprint
August
1982

7

Institute of Mathematical Statistics
University of Copenhagen

Søren Asmussen and David Edwards*

COLLAPSIBILITY AND RESPONSE VARIABLES IN CONTINGENCY TABLES

Preprint 1982 No. 7

INSTITUTE OF MATHEMATICAL STATISTIC
UNIVERSITY OF COPENHAGEN

August 1982

*RECKU, University of Copenhagen.

Collapsibility and Response Variables in Contingency Tables.

By SØREN ASMUSSEN and DAVID EDWARDS

Institute of Mathematical Statistics, University of Copenhagen,
Universitetsparken 5, DK-2100 Denmark
and
RECKU, University of Copenhagen, Vermundsgade 5, DK-2100 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 models; Interaction graph; Log-linear models; Response variables.

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. 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. It is common practice, however, to ignore the distinction between response and explanatory (background) variables. Simple examples (one of which is given in section 3) suffice to show the pitfalls of this approach: first, that inappropriate models may be used, and second that natural and relevant models that are not log-linear may be overlooked. This article characterises 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 (the complement of a in Γ). We identify a hierarchical log-linear model L (i.e., 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:

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

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

We assume the distribution of the table is multinomial:

$$Pr[N=\{n(\mathbf{i})\}]=\{n!/\Pi_i n(\mathbf{i})!\}\Pi_i p(\mathbf{i})^{n(\mathbf{i})}$$

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) $\hat{p}(\mathbf{i}_c) = n(\mathbf{i}_c)/n \ \forall$ 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 Figure 1.

insert Figure 1 about here

One may interpret the interaction graph in the following way (Darroch, Lauritzen and 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.

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 $\bar{a}$. 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 $\alpha \sim \beta$ iff 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 (ibid.) that all decomposable (direct) models are graphical.

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 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 $\mathbf{i}_a$ under L_a as say $p_a(\mathbf{i}_a)$.

Definition

L is *collapsible* onto a if one of the two following equivalent properties hold:

(i) for all $p = p(\mathbf{i}) \in L$ it holds that $p(\mathbf{i}_a) \in L_a$,

(ii) for all $\mathbf{i}_a$, $\hat{p}(\mathbf{i}_a) = \hat{p}_a(\mathbf{i}_a)$.

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

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

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

Note in connection with (i) that always $L_a \subseteq \{p(\mathbf{i}_a) : p \in L\}$.

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

$$\hat{p}(\mathbf{i}) = \hat{p}_a(\mathbf{i}_a)\hat{p}_b(\mathbf{i}_b)/(n(\mathbf{i}_{a\cap b})/n)$$

Proof

(Haberman, 1974, or Lauritzen, 1979).

Corollary 2.2

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

Proof (Lauritzen, 1979)

a and $b = a^c$ 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

Sufficiency: let b be a connected component of a^c. Then the stated condition is easily seen to hold for L_{b^c} as well and collapsibility follows by applying Corollary 2.2 to each connected component in turn.

Necessity: suppose that for some connected component b of a^c, $\partial b \nsubseteq c$ for all generators c of L. Write $b = \{z_1, \ldots z_p\}$. For each factor in ∂b choose an adjacent factor in b and write accordingly $\partial b = \{y_{rs}\}$ where y_{rs} is adjacent to z_r, $1 \leqslant r \leqslant p$, $s = 1 \ldots s_r$. Define for z_i, z_j adjacent

$$\lambda^{z_i z_j} = \begin{cases} 0 & z_i = z_j = 1 \text{ or } z_i = z_j = 2 \\ -\infty & \text{otherwise} \end{cases}$$

and

$$\lambda^{z_r y_{rs}} = \begin{cases} \theta & z_r = y_{rs} = 1 \\ 0 & \text{otherwise} \end{cases}$$

If u,v...;k,l... are the factors in a^c-b, a-∂b respectively, define

$$p_\theta(z_r,y_{rs};u,v...;k,l...) = c(\theta)\exp(\Sigma_{r,s}\lambda^{z_r y_{rs}} + \Sigma_{i,j}\lambda^{z_i z_j})$$

where $c(\theta)$ is a normalizing constant. Then the marginal distribution of the factors in a is readily seen to be of the form

$$p_\theta(y_{rs};k,l...) = \Sigma_{z_r,u,v...}p_\theta(z_r,y_{rs};u,v...;k,l...)$$

$$= d(\theta)[1+\exp\{\theta\,\#(r,s:y_{rs}=1)\}] \qquad (1)$$

Now let m be the number of factors in ∂b and suppose that (i) holds. Then the m-factor interaction between the y_{rs} vanishes, and hence the cross-product ratio between 11...1 and 22...2 is unity, ie.

$$\Pi_{y\in S_1}p_\theta(y;k,l...) = \Pi_{y\in S_2}p_\theta(y;k,l...) \qquad (2)$$

where $S_1 = \{y:y_{rs}=1 \text{ or } 2,\ \Sigma_{r,s}y_{rs} \text{ even}\}$, and $S_2 = \{y:y_{rs}=1 \text{ or } 2,\ \Sigma_{r,s}y_{rs} \text{ odd}\}$.

Combining (1) and (2), $d(\theta)$ cancels and with $\xi = e^\theta$ we obtain

$$\Pi_{k\leqslant m,k \text{ even}}(1+\xi^k)^{\binom{m}{k}} = \Pi_{k\leqslant m,k \text{ odd}}(1+\xi^k)^{\binom{m}{k}}$$

which cannot hold for all ξ (e.g. the constant term is 1 on the r.h.s. and 2 on the l.h.s). Hence (i) fails.

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 subsets 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.

examples

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.

Let $a=(A,B,C)$, $b=(D,E)$ and $L=[AB][BC][AC][ABD][BCE]$. 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

$$\hat{p}(\mathbf{i})=\hat{p}_a(\mathbf{i}_a)\Pi_b(\hat{p}_{\bar{b}}(\mathbf{i}_{\bar{b}})/(n(\mathbf{i}_{\partial b})/n))$$

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

Proof

Apply Theorem 2.1 to each connected subset in turn.

Expressed loosely, collapsibility onto a subset a means that inference concerning the factors in a 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 differ in terms involving the variables in a (but not included in the boundary of a connected component of a^c) only. Then the likelihood ratio test statistic for testing L_1 against L_2 is

$$\begin{aligned} G^2 &= 2\Sigma_i n(\mathbf{i})\ln(\hat{m}^1(\mathbf{i})/\hat{m}^2(\mathbf{i})) \\ &= 2\Sigma_i n(\mathbf{i})\ln(\hat{p}^1(\mathbf{i})/\hat{p}^2(\mathbf{i})) \\ &= 2\Sigma_i\, n(\mathbf{i}_a)\ln(\hat{p}_a^1(\mathbf{i}_a)/\hat{p}_a^2(\mathbf{i}_a)) \end{aligned}$$

from Corollary 2.4, i.e. the test can be performed in the marginal table N_a. The same applies to the Pearson 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. Moreover we note that the tests do not depend on the model terms involving variables in a^c. This gives rise to a property analogous to orthogonality in normal linear models, as illustrated in the following example.

Consider the model L = [ABC][BCD] (see Figure 2).

insert Figure 2 about here

Clearly, L is collapsible onto (A,B,C) and (B,C,D), as are all hierarchical submodels of L that contain the B.C interaction. Thus the sets of terms S_1 = (ABC,AB,AC,A) and S_2 = (BCD,BD,BC,D) are such that the tests for whether a term in the one set is zero are the same irrespective of which terms in the other set are present in the model. In this way model selection can be put on a more secure foundation. Strategies for model selection ought to take this property into account since the number of tests which it is necessary to perform can be greatly reduced.

The definition of collapsibility given here is apparently stated for the first time in Lauritzen (1979) in the form (ii). Tolver Jensen (1978) has some related discussion from the point of view of hypothesis testing. Whittemore (1978) has defined collapsibility in terms of log-linear parameters, but the definitions are not directly comparable. In our opinion the definition given here is more generally useful: the log-linear parametrisation is a mathematical convenience and little intrinsic interest is attached to the parameters themselves.

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 individuals 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:

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

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:

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

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.

This illustrates that ignoring the distinction between response and explanatory variables has, as mentioned above, 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), ch. 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 factorised into a product of the marginal density of a and the conditional density of b given a:

$$p^J(\mathbf{i}) = p^M(\mathbf{i}_a)p^C(\mathbf{i}_b|\mathbf{i}_a) \tag{4}$$

The class of 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. 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

$$\hat{m}^J(\mathbf{i}) = \hat{m}^M(\mathbf{i}_a)(\hat{m}^C(\mathbf{i})/n(\mathbf{i}_a))$$

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:

$$\begin{aligned} G^2_J &= 2\Sigma n(\mathbf{i})\ln\{\hat{m}^J(\mathbf{i})/n(\mathbf{i})\} \\ &= 2\Sigma n(\mathbf{i})\ln[\hat{m}^M(\mathbf{i}_a)\hat{m}^C(\mathbf{i})/\{n(\mathbf{i}_a)n(\mathbf{i})\}] \\ &= 2\Sigma n(\mathbf{i}_a)\ln\{\hat{m}^M(\mathbf{i}_a)/n(\mathbf{i}_a)\} + 2\Sigma n(\mathbf{i})\ln\{\hat{m}^C(\mathbf{i})/n(\mathbf{i})\} \\ &= G^2_M + G^2_C \end{aligned}$$

The corresponding degrees of freedom are similarly additive.

The present section addresses the question of when the joint model J is log-linear. This is important for several reasons. Firstly, it enables us to characterise 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 $\boldsymbol{L}$ be the set of hierarchical log-linear models for N, $\boldsymbol{M}_a$ be the set of hierarchical log-linear models for the marginal table N_a, $\boldsymbol{C}_a$ the set of conditional models (ie. containing all interactions between the factors in a), and $\boldsymbol{J}_a$ the set of joint density models generated from $\boldsymbol{M}_a$ and $\boldsymbol{C}_a$.

Theorem 3.1

For $L \in \boldsymbol{L}$, $L \in \boldsymbol{J}_a$ if and only if L is collapsible onto a. In that case $M = L_a$ and $C = [a] \cup L_{\bar{a}}$.

Proof

If $L \in \boldsymbol{J}_a$ then clearly $\hat{p}^L(\mathbf{i}_a) = \hat{p}^M(\mathbf{i}_a) \in M$ so that collapsibility will follow from $M \subseteq L_a$. That this is indeed the case can be seen, e.g., by taking $c = p^C(\mathbf{i}_b | \mathbf{i}_a)$ independent of $\mathbf{i}_a$, $\mathbf{i}_b$ in (4). Then $cp^M(\mathbf{i}_a) \in L$ for all $p^M(\mathbf{i}_a) \in M$ 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 $C = [a] \cup L_{\bar{a}}$. Then if $b_1 \ldots b_v$ are the connected components of a^c, it is easily seen that

$$\hat{p}^M(i_a) = \hat{p}_a(i_a) = \hat{p}(i_a), \text{ and}$$

$$\hat{p}^C(i_{a^c}|i_a) = \Pi_{k=1}^{v}\hat{p}^C(i_{b_k}|i_{\partial a^c})$$
$$= \Pi_{k=1}^{v}\hat{p}_b^L(i_{b_k})/\{n(i_{\partial b_k})/n\}.$$

Forming the joint model J = (M,C) and using Corollary 2.4 shows that $\hat{p}^J = \hat{p}^L$. Hence $L = J \in \mathcal{J}_a$.

Theorem 3.2

For J = (M,C) $\in \mathcal{J}_a$, J $\in \mathcal{L}$ 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 $L = M \cup C_{\overline{a^c}}$.

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 joint model (M',C') given by $M' = L_a$, $C' = [a] \cup L_{\overline{a^c}}$. 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...v.

Conversely if the stated condition holds, then the log-linear model $L = M \cup C_{\overline{a^c}}$ is collapsible and by Corollary 2.4

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

so that J = L is log-linear.

example

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, [AB], and the boundary of E, [BC] are contained in generators of M. By construction J = [AC][ABD][BCE].

The conditions we have obtained for log-linear models to be appropriate as response variable models can in part be interpreted in terms of conditional independence. To see this, suppose for a given model L that L_a is graphical. If for some connected component b of a^c, ∂b is not in a generator of L (or L_a), it follows that ∂b is not complete in the interaction graph of L. Thus for two explanatory variables, e_1 and e_2 say, a set of response variables R and a (possibly empty) set of explanatory variables E, we obtain that e_1 and e_2 are conditionally independent given R and E. Thus essentially all models that imply that two explanatory variables are conditionally independent given a (set of) response variables (and possibly other explanatory variables) are excluded as inappropriate. This reinforces and clarifies our intuition.

In addition, when L_a is not graphical, other models are excluded. For example L = [AB][BC][AC][AD][BD][CD], where A,B and C are explanatory and D a response variable, is excluded, but this does not have a conditional independence interpretation.

We note that three approaches can be adopted to the analysis of tables with response variables.

Firstly, one can simply condition on the explanatory variables. This approach is appropriate 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 appropriate log-linear models 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), where a sequence of sets $a, b_1, \ldots b_k$ is given. Here the variables in the set b_r (r = 1...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.

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

$$p^J = p^C(d_0) \Pi_{i=1}^k p^C(d_i | d_{i-1})$$

where $C_0, C_1 \ldots C_k$ are log-linear models defined on the appropriate marginal tables. The theorems of the previous section can easily be extended giving

1) $L \in \mathcal{L}$ is a response variable model iff it it is collapsible onto d_i (i = 0...k-1).

2) A response variable model $J = (C_0, C_1 \ldots C_k)$ is log-linear iff the boundary of each connected component of b_r under C_r is contained in a generator of C_{r-1}, for r = 1...k.

4. S-SUFFICIENCY

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

$$p_{\theta,\eta}(x) = p_\theta(t) p_\eta(x|t) \qquad (5)$$

where the parameters θ (of the marginal distribution of T) and η (of the conditional distribution of T given X) are variation independent. Then T is called S-sufficient for θ, cf. Barndorff-Nielsen (1978). 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 $p(\mathbf{i}_a)$, or equivalently if and only if $p(\mathbf{i})$ factorises as

$$p(\mathbf{i}) = p_\theta(\mathbf{i}_a)p_\eta(\mathbf{i}_{a^c}|\mathbf{i}_a) \qquad (6)$$

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. 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

$$\Pi_{k=1}^{n} p_\theta(\mathbf{i}_a(k))\Pi_{k=1}^{n} p_\eta(\mathbf{i}_{a^c}(k)|\mathbf{i}_a(k))$$

Since N_a is sufficient for M, the first factor can be factorised 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 a^c irrelevant, we can find η such that $p_\eta(\mathbf{i}_{a^c}|\mathbf{i}_a) = c$ independent of $\mathbf{i}_{a^c}$, $\mathbf{i}_a$. Thus for any θ, $p(\mathbf{i}) = cp_\theta(\mathbf{i}_a) \in L$ by variation independence. But this implies $p_\theta(\mathbf{i}_a) \in L_a$, i.e. (i) holds.

ACKNOWLEDGEMENTS

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

REFERENCES

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

BIRCH,M.W. (1963).Maximum likelihood in three-way contingency tables. *J.R.Statist.Soc.* **B**,25,220-233.

DARROCH, J.N., LAURITZEN, S.L. and SPEED, T.P. (1980). Markov Fields and log-linear interactions models for contingency tables. *Ann. Statist.*,**8**,522-539.

FIENBERG, S.E. (1980). **The Analysis of cross-classified data.** MIT Press.

FRASER, D.A.S (1956). Sufficient statistics with nuisance parameters. *Ann.Math.Statist.*, **27**, 838-842.

GOODMAN, L.A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika*, **60**,1,179-192.

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

LAURITZEN, S.L. (1979). **Lectures on Contingency Tables.** Institute of Mathematical Statistics, University of Copenhagen. (revision to appear 1982 at Ålborg University Press).

TOLVER JENSEN, S. (1978). **Flersidede kontingenstabeller**. Institute of Mathematical Statistics, University of Copenhagen.

WHITTEMORE, A.S. (1978). Collapsibility of Multidimensional Contingency Tables. *J.R.Statist.Soc.* B,**40**,328-340.

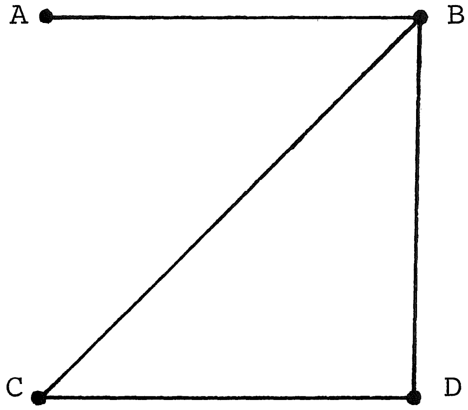

Fig. 1. The interaction graph of [AB][BCD]

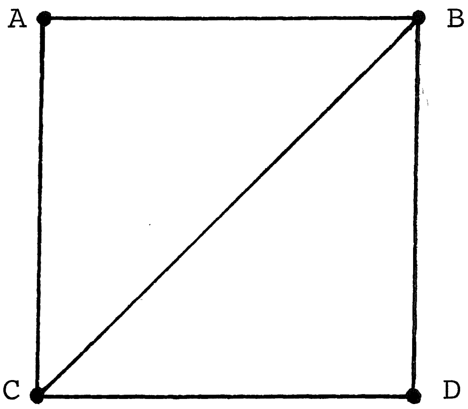

Fig. 2. The interaction graph of [ABC][BCD]

PREPRINTS 1981

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Johansen, Søren: Asymptotic Inference in Random Coefficient Regression Models.

No. 2 Asmussen, Søren: On the Role of a Certain Eigenvalue in Estimating the Growth Rate of a Branching Process.

No. 3 Lauritzen, Steffen.: Time Series Analysis in 1880. A Discussion of Contributions made by T.N. Thiele.

No. 4 Asmussen, Søren: Conditioned Limit Theorems Relating a Random Walk to its Associate, with Applications to Risk Reserve Processes and the GI/G/1 Queue.

No. 5 Johansen, Søren: The Statistical Analysis of a Markov Branching Process.

No. 6 Asmussen, Søren: Time - Dependent Approximations in some Queueing Systems with Imbedded Markov Chains Related to Random Walks.

No. 7 Skovgaard, Ib M.: A Second - Order Investigation of Asymptotic Ancillarity.

No. 8 Rootzén, Holger: The Rate of Extremes of Stationary Normal Sequences.

No. 9 Skovgaard, Ib M.: Large Deviation Approximations for Maximum Likelihood Estimators.

No. 10 Jensen, Ulla Funck: A Stochastic Projection Model with Implications for Multistate Demography and Manpower Analysis.

No. 11 Johansen, Søren: An Extension of Cox's Regression Model.

No. 12 Skovgaard, Ib M.: Edgeworth Expansions in Statistics.

PREPRINTS 1982

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Holmgaard, Simon and Yu, Song Yu: Gaussian Markov Random Fields Applied to Image Segmentation.

No. 2 Andersson, Steen A., Brøns, Hans K. and Jensen, Søren Tolver: Distribution of Eigenvalues in Multivariate Statistical Analysis.

No. 3 Tjur, Tue: Variance Component Models in Orthogonal Designs.

No. 4 Jacobsen, Martin: Maximum-Likelihood Estimation in the Multiplicative Intensity Model.

No. 5 Leadbetter, M.R.: Extremes and Local Dependence in Stationary Sequences.

No. 6 Henningsen, Inge and Liestøl, Knut: A Model of Neurons with Pacemaker Behaviour Recieving Strong Synaptic Input.

No. 7 Asmussen, Søren and Edwards, David: Collapsibility and Response Variables in Contingency Tables.