Biometrika (1983), **70**, 3, *pp.* 537–52
Printed in Great Britain

Graphical and recursive models for contingency tables

BY NANNY WERMUTH
Psychologisches Institut, Johannes Gutenberg-Universität, Mainz, Federal Republic of Germany

AND STEFFEN L. LAURITZEN
Institute of Electronic Systems, Aalborg University Centre, Aalborg, Denmark

SUMMARY

We discuss two classes of models for contingency tables, graphical and recursive models, both of which arise from restrictions that are expressible as conditional independencies of variable pairs. The first of these is a subclass of hierarchical log linear models. Each of its models can be represented by an undirected graph. In the second class each model corresponds to a particular kind of a directed graph instead and can be characterized by a nontrivial factorization of the joint distribution in terms of response variables. We derive decomposable or multiplicative models as the intersecting class. This result has useful consequences for exploratory types of analysis as well as for the model interpretation: we can give an aid for detecting well-fitting decomposable models in a transformation of the observed contingency table and each decomposable model may be interpreted with the help of an undirected or directed graph.

Some key words: Collapsibility; Conditional independence; Data reduction; Decomposable model; Directed graph; Hierarchical model; Log linear model; Maximum likelihood estimate; Multiplicative model; Path analysis; Reducible zero-pattern; Undirected graph; Zero partial association; Zero partial dependence.

1. INTRODUCTION

Log linear models for contingency tables as defined by Birch (1963) have received considerable attention during the last 20 years as documented in the books by Haberman (1974), Bishop, Fienberg & Holland (1975), Andersen (1980) and Plackett (1981). The most appealing features of a hierarchical log linear model are: that it has a set of minimal sufficient statistics, which is a set of proper marginal contingency tables; and that each of the jointly sufficient tables matches exactly the corresponding table derived from the maximum likelihood estimate of the joint table. Variable sets corresponding to tables listed in the set of minimal sufficient statistics are the important ones in each given model, since their observed marginal tables contain the relevant information for the joint distribution of all variables. A well-fitting hierarchical log linear model then provides the researcher with a guide to classifying certain variable subsets as less important because knowledge of their observed tables is unnecessary for a good approximation of the observed table of all variables by the estimated table, and because their observed tables can be closely reproduced from the estimated joint table. It is in this sense that data reduction is well achieved by all hierarchical log linear models.

Nevertheless, for most hierarchical log linear models it remains true that their meaning for the investigated subject matter is difficult to grasp and to communicate. One reason is that within these models all variables are treated alike. The models offer no natural way of specifying one variable as being dependent or a response variable, another variable as being independent, a treatment or background variable, or yet another being both in relation to different subsets of variables.

In contrast, the models for contingency tables proposed by Goodman (1973) as a kind of modified path analysis permit such a view of the variables. But Goodman's models have a disadvantage as far as their application to path analysis is concerned. There is no one-to-one correspondence between such models and directed graphs, which have been at the heart of path analysis as proposed by the geneticist Wright (1923, 1934) for quantitative variables. The graph is to represent a system of dependencies or, in the terminology of Wright, a system of causal relations. For quantitative variables directed graphs may represent systems of linear recursive equations with independent errors. These systems have been studied in econometrics, and they are used nowadays as a basis for path analysis. By defining recursive models of dependencies for qualitative variables, we specify a subclass of the models considered by Goodman. Each recursive model is shown to have a unique representation as a particular kind of a directed graph.

One of the purposes of path analysis is to find out whether the hypothesized system of dependencies is compatible with the observations. To this Goodman's (1973) results on a larger class of models apply. But Goodman's models as well as their subclass of recursive models do not, in general, lead to good data reductions in the sense described above. Our aim is to specify decomposable models as the intersecting class of hierarchical log linear with recursive models. They combine advantages and avoid disadvantages tied to both classes. Furthermore, we want to show how to check, with the help of a simple well-known transformation of the observed contingency table, whether a particular decomposable model is likely to fit the table or not.

To this end, we first give an introductory discussion for only three variables; secondly we describe for t qualitative variables, denoted by the indices $r \in \{1, \ldots, t\}$, a subclass of hierarchical log linear models in which each member has a one-to-one correspondence to an undirected graph. This property of the class, derived by Darroch, Lauritzen & Speed (1980), was the reason to call its members graphical models. They had previously been studied by Andersen (1974), and by Wermuth (1976a, b) under the name of zero partial association models. The models arise from conditional independence restrictions for variable pairs (i,j) of the following kind: a pair is to be conditionally independent given all other $t-2$ variables. In the notation used by Dawid (1979), we can write these restrictions as $i \perp\!\!\!\perp j \mid \{1, \ldots, t\} \setminus \{i,j\}$. Thirdly we specify a subclass of Goodman's path analysis models, the recursive models. Each of these has a one-to-one correspondence to a particular kind of a directed graph. We assume that the first $k < t$ variables have been ordered so that variable $i \in \{1, \ldots, k\}$ may be considered to be a response variable with respect to some or all variables $j \in \{i+1, \ldots, t\}$, but it is not thought of as a response to any one of the variables $h \in \{1, \ldots, i-1\}$. This accounts for the recursiveness in the system of variables. The different models stem from conditional independencies for variable pairs (i,j) given all other variables which may influence the response variable i: $i \perp\!\!\!\perp j \mid \{i+1, \ldots, t\} \setminus \{j\}$, referred to as zero partial dependence in the following. Models within each of the two classes are distinguished by distinct sets of restricted variable pairs, subsets of the $\frac{1}{2}t(t-1)$ pairs, $(1,2), (1,3), \ldots, (t, t-1)$. Two models taken one from each class may have an identical set I of restricted variable pairs. In general, two such models are distinct since the zero restrictions involve associations for graphical models and dependencies for recursive models. But, fourthly we show that a set I characterizes a model belonging to both classes if and only if it is a reducible set defined as follows: a set

$I \subseteq \{(i,j) \mid 1 \leqslant i < j \leqslant t\}$ is reducible if for each $(i,j) \in I$ and all $h < i$ either $(h,i) \in I$ or $(h,j) \in I$ or both.

For convenience, we speak of a reducible zero-pattern in partial associations and partial dependencies whenever the set I of a model is reducible. This name had been chosen by Wermuth (1980) in the context of a multivariate normal distribution, where a partial association is captured by a concentration and a partial dependence by a regression coefficient: for reducible zero-patterns in the concentrations of a t-dimensional normal distribution the dimension may be reduced or collapsed for all r less than t over the variables 1 to $r-1$ with the result that the set of zero concentrations in the marginal distribution of variables $r, \ldots, t$ is identical to the set of zero concentrations for variables $r, \ldots, t$ in the joint distribution of all t variables.

Since it is known that variables in a graphical model can be renumbered to imply a reducible set of zero partial associations if and only if it is a decomposable model (Wermuth, 1980), we have with this result derived decomposable log linear models as the intersecting class of graphical and recursive models. The reducible zero-pattern shows up in the maximum likelihood estimates of the usual log linear parameters, too. Therefore this result is shown to be useful for exploratory stages of data analysis. Finally, we illustrate the two different approaches to defining a structure in a contingency table on a set of data.

2. The three dimensional table

2·1. *Notation*

Let the three variables be denoted by 1, 2 and 3 and their corresponding categories or levels by $i = 1, \ldots, I; j = 1, \ldots, J$ and $k = 1, \ldots, K$; let $p_{ijk} > 0$ be the probability that one of n given observations belongs to cell (i,j,k) of the corresponding $I \times J \times K$ contingency table, n_{ijk} be the observed cell count. We denote summing over an index by a dot, so that for instance $p_{..k} = \Sigma_{i,j} p_{ijk}$ is the probability that an observation belongs to category k of variable 3, and for instance $p_{j|k} = p_{.jk}/p_{..k}$ is the conditional probability that an observation belongs to category j of variable 2 given this observation is known to be in category k of variable 3. For the log linear parameters we adopt the usual symmetry constraints as for instance $\lambda^{(123)}_{.jk} = \lambda^{(123)}_{i.k} = \lambda^{(123)}_{ij.} = 0$, in order to obtain exactly $IJK-1$ independent parameters. Log linear parameters for the marginal probabilities $p_{.jk}$ and $p_{..k}$ are distinguished from those for p_{ijk} as follows:

$$\log p_{ijk} = \lambda_0 + \lambda^{(1)}_i + \lambda^{(2)}_j + \lambda^{(3)}_k + \lambda^{(12)}_{ij} + \lambda^{(13)}_{ik} + \lambda^{(23)}_{jk} + \lambda^{(123)}_{ijk},$$

$$\log p_{.jk} = \lambda^{(1*)}_0 + \lambda^{(1*2)}_j + \lambda^{(1*3)}_k + \lambda^{(1*23)}_{jk}, \quad \log p_{..k} = \lambda^{(1*2*)}_0 + \lambda^{(1*2*3)}_k$$

2·2. *No restrictions*

A model with no other restriction than $p_{...} = 1$ and the symmetry constraints on the log linear parameters is called saturated or unrestricted. Different equivalent parameterizations for it result depending on how many variables are considered to be responses. After adopting by convention a numbering of the variables such that a response may depend only on variables denoted by larger indices, there are three distinct systems of interest: those with no, one or two response variables.

For each response variable dependencies are expressed with the help of conditional probabilities. Recursive systems of dependencies can then be viewed as applications of the factorization property of probability functions as stated, for instance, by Wilks (1962, p. 66). The two factorizations,

$$p_{ijk} = p_{i|jk}\, p_{.jk}, \quad p_{ijk} = p_{i|jk}\, p_{j|k}\, p_{..k},$$

give the recursive system with only one response, variable 1, and with two responses, variables 1 and 2, respectively.

The graphical representation of the unrestricted model contains three points and three lines connecting each pair of points. Each point stands for a variable. With no variable as a response all lines are drawn without arrows or, as is conventional in path analysis, as two-headed arrows; see Fig. 1a. Each line represents the partial associations of the corresponding variable pair given the third variable.

For the recursive system with only one response variable two single-headed arrows point at variable 1 each denoting dependencies of variable 1 in its conditional distributions given variables 2 and 3. Pair (2, 3) is still connected by a simple line or by a line with two-headed arrows as in Fig. 1b, but in contrast to Fig. 1a the line stands for the marginal association of this variable pair.

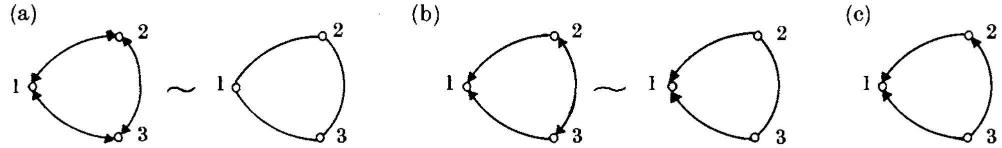

Fig. 1. Graphs of the unrestricted model (a) with no response variable, (b) with one response variable and (c) with two response variables.

In the graph for the recursive system with two responses, Fig. 1c, the arrows point from 2 and 3 to 1 as in Fig. 1b, the arrow from 3 to 2 denotes dependencies of variable 2 in the conditional distributions given variable 3.

In order to relate the partial and marginal association of variable pair (2, 3) we equate parameters of the unrestricted model with no response and with one response. We obtain (2·1) from

$$\log p_{ijk} = \log p_{i|jk} + \log p_{.jk}$$

after substituting log linear parameters for p_{ijk} and $p_{.jk}$ and computing

$$\log p_{ijk} - \Sigma_j \log p_{ijk}/J - \Sigma_k \log p_{ijk}/K + \Sigma_{j,k} \log p_{ijk}/(JK).$$

Equation (2·2) is derived similarly by using the log linear parameterizations for the unrestricted models with one and with two response variables:

$$\lambda_{jk}^{(23)} + \lambda_{ijk}^{(123)} = \{\log p_{i|jk} - \Sigma_j \log p_{i|jk}/J - \Sigma_k \log p_{i|jk}/K + \Sigma_{j,k} \log p_{i|jk}/(JK)\} + \lambda_{jk}^{(1*23)}, \tag{2·1}$$

$$\lambda_{jk}^{(1*23)} = \log p_{j|k} - \Sigma_j \log p_{j|k}/J - \Sigma_k \log p_{j|k}/K + \Sigma_{j,k} \log p_{j|k}/(JK). \tag{2·2}$$

The left-hand side of (2·1) measures the partial association of pair (2, 3) in terms of log linear parameters for p_{ijk}. The left-hand side of (2·2) measures the dependence of pair (2, 3) in terms of log linear parameters for $p_{.jk}$. The particular way in which the two are linked to each other by conditional probabilities in (2·1) provides us with the basic ingredients to develop models with reducible zero-patterns in the restrictions as a special class.

2·3. *Restricted graphical and recursive models*

Restricted graphical models for three variables are defined by zero partial associations. Each given graphical model can be characterized by a list $I^A \subseteq \{(1,2),(1,3),(2,3)\}$ such that for each $(r,s) \in I^A$ we have ZPA(r,s) meaning $r \perp\!\!\!\perp s \mid \{1,2,3\} \backslash \{r,s\}$. For instance ZPA$(1,2)$ denotes $1 \perp\!\!\!\perp 2 \mid 3$, the conditional independence of pair $(1,2)$ given variable 3. This may equivalently be expressed as a restriction on (a) p_{ijk}, (b) $p_{i|jk}$, on (c) the undirected graph, or on (d) the λ parameters for $\log p_{ijk}$ as follows:

(a) $p_{ijk} = p_{i.k}\, p_{.jk}/p_{..k}$,
(b) $p_{i|jk} = p_{i|k}$,
(c) a missing line for pair $(1,2)$,
(d) $\lambda_{ij}^{(12)} = \lambda_{ijk}^{(123)} = 0$.

Similarly, restricted recursive models result from imposing zero partial dependencies on a corresponding recursive system. Each given recursive model can be characterized by the number of response variables and a list I^D. For a system in only one response variable $I^D \subseteq \{(1,2),(1,3)\}$ and for a system in two response variables $I^D \subseteq \{(1,2),(1,3),(2,3)\}$, such that the possible restrictions are $1 \perp\!\!\!\perp 2 \mid 3$, $1 \perp\!\!\!\perp 3 \mid 2$ or in addition $2 \perp\!\!\!\perp 3$. Zero partial associations and dependencies coincide for variable pairs involving variable 1, they are distinct for pair $(2,3)$. Thus ZPD$(2,3)$ means $2 \perp\!\!\!\perp 3$, independence of variables 2 and 3. This can equivalently be expressed as a restriction on (a) $p_{.jk}$, on (b) $p_{j|k}$, on (c) the directed graph of a recursive system in two responses, or (d) on the λ-parameters for $\log p_{.jk}$ as follows:

(a) $p_{.jk} = p_{.j.}\, p_{..k}$,
(b) $p_{j|k} = p_{.j.}$,
(c) a missing arrow for pair $(2,3)$,
(d) $\lambda_{jk}^{(1*23)} = 0$.

As a consequence, graphical models can differ from recursive ones in the case of three variables only, if the restrictions involve pair $(2,3)$. Of the four such subsets of $\{(1,2),(1,3),(2,3)\}$ the sets

$$I = \{(1,2),(2,3)\}, \quad I = \{(1,3),(2,3)\}, \quad I = \{(1,2),\,(1,3),\,(2,3)\}$$

are reducible, while only the set $I = \{(2,3)\}$ is nonreducible. As equation (2·1) shows, the log linear model with $I^A = \{(2,3)\}$ does not imply $I^D = \{(2,3)\}$, nor does $I^D = \{(2,3)\}$ imply $I^A = \{(2,3)\}$. Birch (1963) discussed the model with $I^D = \{(2,3)\}$ as being distinct from a log linear model for all three variables and derived the maximum likelihood estimates $\hat{p}_{ijk}$ for p_{ijk} as

$$\hat{p}_{ijk} = (n_{ijk}\, n_{.j.}\, n_{..k})/(n_{.jk}\, n_{...}^2).$$

This tells us first that no data reduction can be achieved with this model since all observations n_{ijk} are needed to compute the estimate $\hat{p}_{ijk}$, and secondly that the sufficient table n_{ijk} does not match the corresponding estimated table $n_{...}\hat{p}_{ijk}$ unless by mere accident in a sample $n_{.jk} = n_{.j.}\, n_{..k}/n_{...}$.

On the other hand, ZPA$(2,3)$ and ZPD$(2,3)$ are both satisfied for the three reducible zero patterns. To see this we take as example $I = \{(1,2),(2,3)\}$ and show first that $I^A = I$ implies ZPD $(2,3)$. If we take $I^A = \{(1,2),\,(2,3)\}$ this means $1 \perp\!\!\!\perp 2 \mid 3$ and $2 \perp\!\!\!\perp 3 \mid 1$. We obtain

$$p_{i.k}\,p_{.jk}/p_{..k} = p_{ij.}\,p_{i.k}/p_{i..}, \quad p_{.jk}\,p_{i..} = p_{ij.}\,p_{..k}, \quad p_{.jk} = p_{.j.}\,p_{..k},$$

and hence $2 \perp\!\!\!\perp 3$, and ZPD $(2,3)$ follows. Secondly, we show that $I^D = I$ implies ZPA $(2,3)$. If we take $I^D = \{(1,2),(2,3)\}$ this means $1 \perp\!\!\!\perp 2 \mid 3$ and $2 \perp\!\!\!\perp 3$ and we get

$$p_{ijk} = p_{i.k}\,p_{.j.}, \quad p_{.j.} = p_{ij.}/p_{i..}, \quad p_{ijk} = p_{i.k}\,p_{ij.}/p_{i..}$$

and hence $2 \perp\!\!\!\perp 3 \mid 1$ and ZPA $(2,3)$ follows.

Similar arguments can be used for the other two reducible zero-patterns. By the use of (2·1) and (2·2) our main result can be obtained more directly: for each reducible I, I^A and I^D denote equivalent models, and so if $(r,s) \in I$ then ZPA (r,s) and ZPD (r,s) are both satisfied. This result is to be derived for more than three variables in §6.

3. The t-dimensional table

3·1. *Notation*

Variables for a t-dimensional contingency table are again denoted by indices $j \in \{1,\ldots,t\}$, their categories or levels by $l_j = 1,\ldots,L_j$; with $l(A)$ denoting for any $A \subseteq \{1,\ldots,t\}$ the subvector of $(l_1,\ldots,l_t)$ containing all l_i with $i \in A$, we write marginal probabilities as $p_{l(A)}$. For $A = \{1,2\}$ for instance, $p_{l(A)} = p_{l_1,l_2} = \Sigma_{l_3,\ldots,l_t} p_{l_1,\ldots,l_t}$. To simplify notation we write sometimes $p(A)$ for $p_{l(A)}$ and $\Sigma_r p(A)$ for $\Sigma_{l_r} p_{l(A)}$. As for three variables we consider log linear parameters with the usual symmetry constraints: λ-terms for $p(1,\ldots,t)$ and $\lambda^{(1*,\ldots,i-1*)}$ terms for $p(i,\ldots,t)$. Before we look at partial associations and partial dependencies in terms of log linear parameters, we identify them within graphs for the unrestricted case.

3·2. *Undirected and directed graphs*

We consider the following kind of graphs: graphs consisting of t points and at most one connecting line for each pair of points. Each point represents a variable and each connection between two points either an association or a dependence.

A graph with t points is called complete, if it has exactly $\frac{1}{2}t(t-1)$ connecting lines and incomplete, if it has fewer lines. If all connecting lines have no arrows or all are two-headed arrows, the graph is undirected, if at least one connecting line is a one-headed arrow, the graph is called directed. A complete undirected graph corresponds to an unrestricted log linear model for the t variables with a connecting line between points i and j representing the partial association of variables i and j given all the remaining $t-2$ variables. A restricted graphical model is then represented by an incomplete undirected graph.

A subgraph of $r < t$ points is obtained by deleting all other $t-r$ points as well as all connecting lines to these points. Thus, a subgraph of r points can have at most $\frac{1}{2}r(r-1)$ connecting lines, in which case it is complete.

A complete directed graph has a one-to-one correspondence to an unrestricted recursive model with k response variables, if the following two statements are satisfied: (I) there are $k < t$ points at which one-headed arrows are directed and they can be ordered such that $t-i$ arrows beginning at points $i+1$ to t point at each $i \in \{1,\ldots,k\}$ so that the directions of the arrows are induced by a recursive ordering of the dependent variables; and (II) the subgraph of the $t-k$ remaining points is undirected so that this undirected part of the graph contains the background or exogeneous variables. A single-headed arrow denotes partial dependence of variable i on variable j given variables

$\{i+1, \ldots, t\} \setminus \{j\}$; a connecting line of pair (i,j) in the subgraph of the last $t-k$ points represents the partial association of this pair in the marginal distribution of the last $t-k$ variables: of variables i and j given the variables $\{k+1, \ldots, t\} \setminus \{i,j\}$. A restricted recursive model is then represented by a graph obtained from the complete graph of a recursive system by removing some of the single-headed arrows.

3·3. *Partial association and partial dependence*

The relevant log linear parameters for the partial association of variable pair (r, s) are the two-factor term and all higher-order λ-terms involving r and s. In order to relate them to the relevant parameters for the partial dependence of pair (r, s) we observe that

$$p(1, \ldots, t) = \left\{ \prod_{i=1}^{j} p(i \mid i+1, \ldots, t) \right\} p(j+1, \ldots, t);$$

by equating the corresponding parameters we get analogous to (2·1), (2·2) that, for $r < s$,

$$\begin{aligned} \lambda_{l_r, l_s}^{(r,s)} + \mathscr{H}_{rs}(\lambda) = \sum_{i=1}^{j} \{ & \log p(i \mid i+1, \ldots, t) - \Sigma_r \log p(i \mid i+1, \ldots, t)/L_r \\ & - \Sigma_s \log p(i \mid i+1, \ldots, t)/L_s + \Sigma_{r,s} \log p(i \mid i+1, \ldots, t)/(L_r L_s) \} \\ & + \lambda_{l_r, l_s}^{(1^*, \ldots, j^*, r, s)} + \mathscr{H}_{rs}(\lambda^{(1^*, \ldots, j^*)}), \end{aligned} \quad (3·1)$$

$$\begin{aligned} \lambda_{l_r, l_s}^{(1^*, \ldots, r-1^*, r, s)} + \mathscr{H}_{rs}^{(1^*, \ldots, r-1^*)} = & \log p(r \mid r+1, \ldots, t) - \Sigma_r \log p(r \mid r+1, \ldots, t)/L_r \\ & - \Sigma_s \log p(r \mid r+1, \ldots, t)/L_s \\ & + \Sigma_{r,s} \log p(r \mid r+1, \ldots, t)/(L_r L_s), \end{aligned} \quad (3·2)$$

where, for example, $\mathscr{H}_{rs}(\lambda^{(1^*, \ldots, j^*)})$ denotes higher order $\lambda^{(1^*, \ldots, j^*)}$ terms involving r and s. For a proof only summation and subtraction are involved, but the proof is omitted because the notation becomes cumbersome. First, the equations say that partial associations are linked to partial dependencies, measured by marginal partial associations, through conditional probabilities. Secondly, the equations help to show that ZPA (r, s) coincides with ZPD (r, s) for all reducible zero-patterns of restrictions on partial associations or on partial dependencies, and thirdly to show that any given well-fitting recursive model having a reducible zero-pattern in the dependencies has to show up as a reducible zero-pattern in the log linear λ-parameters of the saturated model.

4. The class of graphical models

A member of the class of graphical models as defined by Darroch *et al.* (1980) can be thought of as characterized by a set $I^A \subseteq \{(i,j)\} \mid 1 \leqslant i < j \leqslant t\}$ such that $r \perp\!\!\!\perp s \mid \{1, \ldots, t\} \setminus \{r, s\}$ for each $(r, s) \in I^A$. A variable number is only a label of a variable, the numbers do not induce an order for the variables. First, we list equivalent formulations of the simplest type of restricted graphical models, the one with $I^A = \{(r, s)\}$.

Proposition 1. *For the model with* $I^A = \{(r, s)\}$ *the following statements are equivalent:*

(i) ZPA (r, s);

(ii) $r \perp\!\!\!\perp s \mid \{1, \ldots, t\} \setminus \{r, s\}$;

(iii) $p(1, \ldots, t) = \Sigma_r p(1, \ldots, t) \Sigma_s p(1, \ldots, t) / \Sigma_{r,s} p(1, \ldots, t)$;

(iv) $p(r \mid 1, \ldots, r-1, r+1, \ldots, t) = p(r \mid 1, \ldots, r-1, r+1, \ldots, s-1, s+1, \ldots, t)$;

(v) *the undirected graph has* $\frac{1}{2}t(t-1)$ *minus one connecting lines, the one for* (r, s) *being missing;*

(vi) *two observed marginal tables form the set of minimal sufficient statistics, one contains variables* $\{1, \ldots, t\} \setminus \{r\}$, *the other contains variables* $\{1, \ldots, t\} \setminus \{s\}$;

(vii) $\lambda_{i_r, i_s}^{(r,s)}$ *and all higher-order* λ*-terms involving* r *and* s *are zero.*

All these are either familiar results from the theory of log linear models or immediate consequences of definitions given in § 3.

In general, the list I^A of zero partial associations can equivalently be viewed in three ways: the list of all missing two-factor and higher-order λ-terms; the list of all missing lines in the undirected graph; and as complementary to the list of all connecting lines in this graph. As has been noted by Darroch *et al.* (1980) an equivalent formulation for the list of all connecting lines is the set $\{N\}_T = \{N_1, \ldots, N_T\}$ of maximal complete subsets in the graph. A subset of points in the graph is called maximal complete, if the subgraph of these points is a complete graph and, if by including one more point an incomplete subgraph results. The equivalence to I^A is then defined by $(r, s) \notin I^A$ if and only if there exists an $N_i \in \{N\}_T$ such that $\{r, s\} \subseteq N_i$. The set $\{N\}_T$ is known as the generating class of the corresponding log linear model; a list of its elements separated by dashes has been used as a short-cut model notation by Wermuth (1976a). It can be viewed as the list of minimal sufficient statistics of the model, if the elements of each $N_i \in \{N\}_T$ are interpreted as the variables in the observed marginal table that is obtained from the t-dimensional one by summing over all variables not listed in N_i.

For the interpretation of a graphical model in terms of conditional independencies the following result by Darroch *et al.* (1980) is helpful.

PROPOSITION 2. *If in an undirected graph two disjoint subsets of points* A *and* B *are separated by a subset* D, *in the sense that all paths from* A *to* B *go through* D, *then the variables in* A *are conditionally independent from those in* B *given the variables in* D.

In the example in Fig. 2 there are four maximal complete subsets in the graph, which jointly form the generating class of the model. The set of pairs with missing lines is I^A, the list of zero partial associations. Additional independencies can be read off the graph with the help of Proposition 2; for instance, $1 \perp\!\!\!\perp 2 \mid (4,5,6)$, $1 \perp\!\!\!\perp 2 \mid (4,5)$, $1 \perp\!\!\!\perp 2 \mid 4$, $3 \perp\!\!\!\perp 6 \mid 4$, $(5,6) \perp\!\!\!\perp (1,2,3) \mid 4$. Probably the simplest interpretation of this model can be stated as $1 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3 \perp\!\!\!\perp (5,6) \mid 4$, which says that given variable 4, the variables 1, 2, 3 and the joint variable (5, 6) are mutually independent.

5. THE CLASS OF RECURSIVE MODELS

The class of recursive models has been defined for quantitative variables by Wold (1954). In our terminology a recursive model for a contingency table is fully characterized by: (a) a recursive system in $k < t$ response variables, that is an ordering of the responses, such that variable $i \leqslant k$ may depend only on variables $j \in \{i+1, \ldots, t\}$; and (b) a set $I^D \subseteq \{(i, j) \mid 1 \leqslant i < j \leqslant t \text{ and } i \leqslant k\}$ such that $r \perp\!\!\!\perp s \mid \{r+1, \ldots, t\} \setminus \{s\}$ for each $(r, s) \in I^D$. Here the numbers $1, \ldots, k$ induce a recursive ordering.

As for graphical models we first look at the case of only one restriction.

Fig. 2

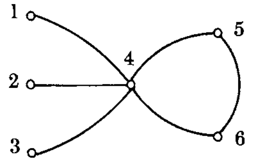

Fig. 3

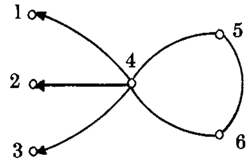

Fig. 2. The graphical model with generating class $\{\{1,4\},\{2,4\},\{3,4\},\{4,5,6\}\}$

Fig. 3. The recursive model with $p(1,\ldots,6) = p(1\mid 4)\,p(2\mid 4)\,p(3\mid 4)\,p(4,5,6)$.

PROPOSITION 3. *For the model with* $I^D = \{(r,s)\}$ *the following statements are equivalent:*

(i) ZPD (r,s);

(ii) $r \perp\!\!\!\perp s \mid \{r+1,\ldots,t\}\backslash\{s\}$;

(iii) $p(r,\ldots,t) = \Sigma_r\, p(r,\ldots,t)\, \Sigma_s\, p(r,\ldots,t)/\Sigma_{r,s}\, p(r,\ldots,t)$;

(iv) $p(r\mid r+1,\ldots,t) = p(r\mid r+1,\ldots,s-1,s+1,\ldots,t)$;

(v) $\lambda_{l_r,l_s}^{(1^*,\ldots,r-1^*,r,s)}$ *and all higher-order* $\lambda^{(1^*,\ldots,r-1^*)}$*-terms involving* r *and* s *are zero;*

(vi) *the directed graph has* $\frac{1}{2}t(t-1)-1$ *connecting lines, the arrow pointing from* s *to* r *being missing,*

and the next statement is implied:

(vii) $\lambda^{(1^*,\ldots,r-1^*)}$*-terms equal* $\lambda^{(1^*,\ldots,r^*)}$*-terms, whenever they involve* s.

The equivalencies of (ii), (iii), (iv) and (vi) to (i) result directly from the definitions, the one in (v) following from (3·2). The result (vii) can be derived similarly to (3·1) and it states that given ZPD (r,s) all $\lambda^{(1^*,\ldots,r-1^*)}$ terms involving variable s are collapsible over variable r; which is to say that they are identical for two marginal contingency tables, one with variables r to t, the other with only variables $r+1$ to t.

In general, the list I^D of zero partial dependencies can equivalently be seen as implying for each $(r,s) \in I^D$: (I) a zero partial association of this variable pair in the marginal distribution of variables r to t, (II) a missing single-headed arrow in reference to the graph of a recursive system, and (III) as defining for each response variable $i \in \{1,\ldots,k\}$ the subset of the variables $i+1$ to t on which it actually depends, $A_i = \{j \mid j > i \text{ and } (i,j) \notin I^D\}$, and a complementary subset of variables $B_i = \{j \mid j > i \text{ and } (i,j) \in I^D\}$. For the interpretation of any given recursive model the following result is useful.

PROPOSITION 4. *In the recursive model specified by* I^D *a response variable* $i \in \{1,\ldots,k\}$ *is conditionally independent of the variables in* B_i *given the variables in* A_i.

Proposition 4 follows from repeated applications of the simple result that ZPD (i,j) and ZPD (i,r) taken together, imply

$$p(i\mid i+1,\ldots,t) = p(i\mid i+1,\ldots,j-1,j+1,\ldots,r-1,r+1,\ldots,t).$$

The effect of this result on the joint probability is a nontrivial factorization: for each given set $\{A_1,\ldots,A_k\}$ of a recursive system we get

$$p(1,\ldots,t) = \left\{\prod_{i=1}^{k} p(i\mid A_i)\right\} p(k+1,\ldots,t). \tag{5·1}$$

Thus, each recursive model can be viewed as a sequence of k hypotheses, each of which is expressible as $p(i\mid A_i \cup B_i) = p(i\mid A_i)$ or equivalently as: $\{N\}_2 = \{\{i\}\cup A_i, A_i \cup B_i\}$ is the generating class for the table with variables $i, \ldots, t$. There are situations, in which these hypotheses concerning k contingency tables coincide with a hypothesis for the t-dimensional table only with the generating class of graphical model.

For the example displayed in Fig. 3 Proposition 4 tells that $1 \perp\!\!\!\perp (2,3,5,6) \mid 4$, $2 \perp\!\!\!\perp (3,5,6) \mid 4$ and $3 \perp\!\!\!\perp (5,6) \mid 4$. The corresponding factorization of the joint probability (5·1) can be written as

$$p(1, \ldots, 6) = p(1,4)\,p(2,4)\,p(3,4)\,p(4,5,6)/p(4)^3,$$

so that for this model the probabilities in the numerator of the factorization concern exactly those marginal tables listed as a generating class for the graphical model in Fig. 2. A necessary and sufficient condition for the equivalence of the two types of models follows in §6.

6. The intersecting class of models

We wish to characterize the class of models in which each zero partial association coincides with a zero partial dependence. The models in this class combine advantages of graphical and recursive models in the following sense: each model represents, like all graphical models, a condensed description for the contingency table, a good data reduction. Further each model gives, like all recursive models, a simple factorization of the joint distribution and of the maximum likelihood estimates.

PROPOSITION 5. *A recursive model for a t-dimensional contingency table given by* $I^D = I$ *is equivalent to the graphical model with* $I^A = I$, *if and only if* I *is reducible.*

The definition of a reducible set has been given at the end of §1. Proposition 5 is analogous to a result proven by Wermuth (1980) for normally distributed variables. An equivalent graph theoretic formulation is contained in an unpublished paper by H. T. Kiiveri, T. P. Speed and J. B. Carlin. Our proof in the Appendix uses the parameterization of graphical models and of recursive systems presented in the previous sections.

In order to explore in more detail the meaning of a reducible zero-pattern we give several equivalent interpretations of the lack of a reducible zero-pattern in a graphical model and a recursive system in Propositions 6 and 7, respectively.

PROPOSITION 6. *For graphical models in contingency tables the following statements are equivalent:*

(i) *the variables cannot be ordered so that* I^A, *the list of zero partial associations, is reducible;*
(ii) *the graphical model for the unordered variables is not a decomposable or multiplicative model;*
(iii) *the undirected graph contains a subset of* $r \geq 4$ *points having a subgraph with* r *connecting lines such that each starting point is reached again with* r *lines, such as in Fig.* 4;
(iv) *the marginal tables corresponding to the generating class* $\{N\}_T$ *cannot be combined multiplicatively to form a factorization of the joint distribution;*
(v) *the maximum likelihood estimates of the cell counts have to be obtained iteratively from the observed marginal tables corresponding to* $\{N\}_T$.

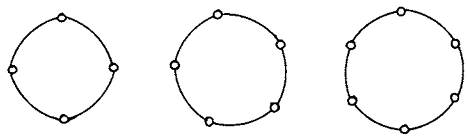

Fig. 4. Characteristic subgraphs of nondecomposable models.

These statements just form a summary of results scattered in the statistical literature. Proofs of the equivalence of (ii) and (iv) and of (ii) and (v) are due to Goodman (1970) and Haberman (1974, p. 170), of (ii) and (iii) to Lauritzen, Speed & Vijayan (1983), and of (i) and (ii) to Wermuth (1980).

For models having subgraphs as in Fig. 4 no labelling of the variables exists which gives a reducible zero-pattern of restrictions. In contrast, the nonreducible zero-pattern in Fig. 5,

$$I^A = \{(1,2),(1,4),(1,5),(1,6),(2,4),(2,5),(2,6),(4,5),(4,6)\},$$

can be removed by renumbering the variables, for instance by exchanging the roles of 3 and 4. This gives the model in Fig. 2 with a reducible list I^A.

Fig. 5

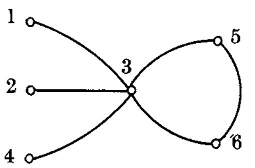

Fig. 6

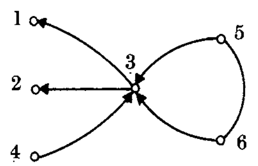

Fig. 5. The graphical model with generating class $\{\{1,3\},\{2,3\},\{3,4\},\{3,5,6\}\}$.
Fig. 6. The recursive model with $p(1,\ldots,6) = p(1\,|\,3)\,p(2\,|\,3)\,p(3\,|\,4,5,6)\,p(4)\,p(5,6)$.

Proposition 7. *For recursive systems in a contingency table with $k < t$ responses i the following statements are equivalent:*

(i) *the list I^D of zero partial dependencies is not reducible;*
(ii) *there is a response variable i having two variables r and s in its set of influencing variables A_i such that $(r,s) \in I^D$;*
(iii) *the directed graph contains three points i, r, s such that two arrows point at i, one from r, the other from s, but r and s are not connected;*
(iv) *there is a response i for which the marginal table of the variables A_i derived from the maximum likelihood estimate for the joint table deviates, in general, from the corresponding observed contingency table;*
(v) *the set obtained from $\{\{1\} \cup A_1, \ldots, \{k\} \cup A_k, \{k+1,\ldots,t\}\}$ after deleting all subsets is not the set $\{N\}_T$ of maximal complete subsets in the graph.*

While (ii) and (iii) are simple reformulations of (i), the last statements (iv), (v) are a consequence of recursive models being nonhierarchical models if the pattern of restrictions is not reducible.

As an example for a recursive model having a nonreducible zero-pattern of restrictions we take the model displayed in Fig. 6. The list of missing arrows I^D in Fig. 6 as compared to a complete recursive system in four responses is identical to the list of missing lines in Fig. 5, but since it is a nonreducible set, the two models are not equivalent. This may be seen more directly by looking at the first variable, for which the reducibility criterion breaks down, at variable 4. Proposition 2 shows for Fig. 5 that $4 \perp\!\!\!\perp (5,6)\,|\,3$ while the complete independence $4 \perp\!\!\!\perp (5,6)$ is implied by Fig. 6.

7. An example

The following four-dimensional contingency table is taken from a prospective study on determinants of early retirement; see Michaelis *et al.* (1980). In this study two or three extensive questionnaires were answered within a time span of two to five years. For the purpose of demonstrating differences between graphical and recursive models we selected four variables and restricted the analysis to men aged 62 to 64 at the time the first questionnaire was answered. The four variables are given in Table 1. We assume

Table 1. *Example for demonstrating differences between graphical and recursive models*

Variable	Categories
1	$i = 1$: retired at age 65 or later
	$i = 2$: retired before age 65 years
2	$j = 1$: no bilious or liver complaints
	$j = 2$: bilious or liver complaints
3	$k = 1$: white collar worker
	$k = 2$: blue collar worker
4	$l = 1$: satisfied with conditions at work
	$l = 2$: dissatisfied with conditions at work

that these variables form a recursive system in two response variables: variable 1 may possibly depend on variables 2, 3, 4, and variable 2 can be thought of as a response to variables 3 and 4, but because of a time lag not to variable 1.

Within this framework we formulate and test two hypotheses: H_1, there is no direct dependence of the occurrence of bilious or liver complaints on the professional status, and H_2, there is no direct dependence of early retirement on the professional status.

Hypothesis H_1 can equivalently be stated as $p_{ijkl} = p_{i|jkl}\, p_{j|l}\, p_{kl}$ or as $I^D = \{(2,3)\}$, or a missing single-headed arrow pointing from 3 to 2 in the directed graph of the recursive system, Fig. 7(a), or as $\{N\}_2 = \{\{2,4\},\{3,4\}\}$ is the generating class for the marginal table of variables 2, 3 and 4.

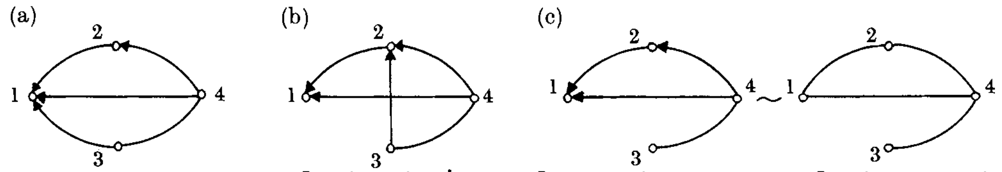

Fig. 7. The model (a) with $I^D = \{(2,3)\}$, (b) with $I^D = \{(1,3)\}$, and (c) with $I^D = \{(1,3),(2,3)\}$.

The second hypothesis H_2 can be expressed as $p_{ijkl} = p_{i|jl}\, p_{j|kl}\, p_{kl}$, Fig. 7b. Since I^D is reducible this hypothesis can equivalently be written as $I^A = \{(1,3)\}$ or with the help of the generating class as $\{N\}_2 = \{\{1,2,4\},\{2,3,4\}\}$, with the model notation as 124/234 or with an undirected graph having connecting lines for pairs $(1,2)$, $(1,4)$, $(2,3)$, $(2,4)$, $(3,4)$.

If the two hypotheses are combined the recursive system is again equivalent to a graphical model as shown in Fig. 7c. The combined hypothesis is expressible as $p_{ijkl} = p_{i|jl}\, p_{j|l}\, p_{kl}$, as $I^D = I^A = \{(1,3),(2,3)\}$ or as model 124/34. If this model fits well, variable 3 can be deleted from the recursive system since variables 1 and 2 jointly can then be regarded as conditionally independent of variable 3 given variable 4.

The usual likelihood ratio test statistics have value 0·99 on 2 degrees of freedom for the first hypothesis and 13·53 on 4 degrees of freedom for the second. Since the latter corresponds to a fractile value of $p = 0{\cdot}009$ this hypothesis, as well as model 124/34, is judged to be incompatible with the observations.

At this stage it is appropriate to switch to an exploratory type of analysis. We know from (3·2) that a zero partial dependence will show up in log linear parameters of a corresponding marginal contingency table. Therefore, we compute the estimated $\lambda^{(1*)}$ parameters with a simple paper and pencil method. We use Yates's (1937) algorithm (Good, 1958; Winer, 1971, p. 629) on the log cell counts. Table 2 shows a numbering for each cell, the cell counts $n_{.jkl}$, the logarithms of the cell counts, and the estimated parameters for the unrestricted model 234. The small three-factor interaction $\lambda_{111}^{(1*234)}$ together with the two small two-factor interactions $\lambda_{11}^{(1*23)}$ and $\lambda_{11}^{(1*24)}$ suggest that there is no direct dependence of variable 2 on variables 3 and 4.

Table 2. *Yates's algorithm applied to* $\log n_{.jkl}$

Cells *jkl*	$n_{.jkl}$	$\log n_{.jkl}$	Stages of Yates's algorithm 1	2	3	Estimates for model 234 Parameters	Values
111	984	6·8916	10·1497	19·9185	33·4182	$\lambda_0^{(1*)}$	4·18
211	26	3·2581	9·7688	13·4997	25·0218	$\lambda_1^{(1*2)}$	1·88
121	920	6·8244	5·3565	7·5135	−2·4085	$\lambda_1^{(1*3)}$	−0·30
221	12	2·9444	8·1432	7·5083	0·1858	$\lambda_{11}^{(1*23)}$	0·02
112	106	4·6634	3·6335	0·3809	6·4188	$\lambda_1^{(1*4)}$	0·80
212	2	0·6931	3·8000	−2·7867	0·0052	$\lambda_{11}^{(1*24)}$	0·00
122	344	5·8406	3·9703	−0·2465	3·1676	$\lambda_{11}^{(1*34)}$	0·40
222	10	2·3026	3·5380	0·4323	−0·6788	$\lambda_{111}^{(1*234)}$	−0·08

Similarly, variable 1 shows up as being weakly related to variable 2 in the estimated λ parameters of the saturated model for all four variables. These impressions are supported by the results of a model search procedure (Wermuth, 1976b, 1980; Wermuth, Wehner & Gönner, 1976) which showed a good fit for model 134/2 or $I^A = \{(1,2),(2,3),(2,4)\}$ but for no other model with more restrictions. Model 134/2 demonstrates that variable 2 is completely independent of variables 1, 3, 4. It corresponds to the following hypothesis on the recursive system: $p_{ijkl} = p_{i|kl}\,p_j\,p_{kl}$. This says that there is no direct dependence of early retirement on bilious or liver complaints and that there is no direct dependence of bilious or liver complaints on the professional status and on the satisfaction with conditions at work. In Table 3 we display the observed and estimated cell counts for this model.

Table 3. *Observed and estimated cell counts for model* 134/2

Cells *iklj*	Counts n_{ijkl}	Estimates $\hat{m}_{ijkl}$	Cells *iklj*	Counts n_{ijkl}	Estimates $\hat{m}_{ijkl}$
1111	833	834·8	1112	22	20·2
2111	151	151·3	2112	4	3·7
1211	731	725·5	1212	12	17·5
2211	189	191·4	2212	7	4·6
1121	81	81·0	1122	2	2·0
2121	25	24·4	2122	0	0·6
1221	244	246·1	1222	8	5·9
2221	100	99·6	2222	2	2·4

To summarize, recursive models form a framework to study dependency structures with one response variable or with several recursively ordered responses while with graphical models the interdependency structure of several unordered variables may be investigated. We characterized the equivalence of the two types of models by a reducible zero-pattern of restrictions on the dependencies or on the associations. Reducibility means that stepwise collapsing of a contingency table in a given order over the responses retains the pattern of restrictions for the remaining variable pairs. Gains are obtained for both approaches to analysing a contingency table by this characterization of the intersecting class of decomposable models. These concern first the interpretation of the models, secondly properties of the set of minimal sufficient statistics and of the corresponding maximum likelihood estimates, and thirdly the interpretation of the log linear parameters for the saturated model. Extensions and modifications of the results for models containing both qualitative and quantitative variables shall be treated in a forthcoming paper.

This research was supported in part by the Danish Natural Science Research Council. We thank John Darroch and Hanns-Georg Leimer for stimulating discussions, a referee and Inge Leimer for their thoughtful comments.

APPENDIX

Proof of Proposition 5

First let us give a slightly more mathematical formulation of the statement in Proposition 5. A reducible I is a necessary and sufficient condition for the following:

$$\text{for all } (i,j) \in I, i \perp\!\!\!\perp j \mid \{i+1,\ldots,t\}\backslash\{j\} \text{ if and only if,}$$
$$\text{for all } (i,j) \in I, i \perp\!\!\!\perp j \mid \{1,\ldots,t\}\backslash\{i,j\}. \qquad (\text{A}\cdot 1)$$

We first consider the situation in which I is not reducible and show that in this case the two statements in (A·1) are not equivalent. For this purpose it is sufficient to find a joint distribution of all t variables satisfying the first of the two statements in (A·1) but not the second. For $t = 3$ we did so in § 2. Using this example we can find a distribution for $t > 3$ as follows.

If I is assumed to be not reducible there exist h, u, v with $(u,v) \in I$, $(h,u) \notin I$, $(h,v) \notin I$. For these we choose the marginal probabilities $p_{l_h, l_u, l_v} > 0$ such that $u \perp\!\!\!\perp v$ but not $u \perp\!\!\!\perp v \mid h$, as in the example of § 2. A simple joint distribution of all t variables is then obtained by requiring equiprobability for the remaining $t-3$ variables: $p_{l_1,\ldots,l_t} = K p_{l_h,l_u,l_v}$, with $\Sigma p_{l_h,l_u,l_v} = 1$, $K = (\Pi_{k \neq h,u,v} L_k)^{-1}$ and L_k as the number of categories of variable k. In this distribution we have $i \perp\!\!\!\perp j \mid \{i+1,\ldots,t\}\backslash\{j\}$ for all $(i,j) \in I$ so that the first statement of (A·1) is satisfied. But, at the same time $u \perp\!\!\!\perp v \mid \{1,\ldots,t\}\backslash\{u,v\}$ does not hold.

We now consider the case where I is reducible. We have to show

$$I \text{ reducible implies (A}\cdot 1). \qquad (\text{A}\cdot 2)$$

This is trivial for $i = 1$, since the two statements of (A·1) coincide for this case. For $i \geqslant 2$ this is done by induction on the number of variables t. For $t = 3$ the statement was demonstrated in § 2. Suppose that the above statement is known to be true for $t = n$ and assume $t = n+1$. If,

$$\text{for all } (i,j) \in I, i \perp\!\!\!\perp j \mid \{i+1,\ldots,t\}\backslash\{j\} \qquad (\text{A}\cdot 3)$$

then the induction assumption gives,

$$\text{for all } (i,j) \in I, i \perp\!\!\!\perp \text{j} \mid \{2,\ldots,t\}\backslash\{i,j\} \qquad (\text{A}\cdot 4)$$

since I reducible implies $\tilde{I}$ reducible, where $\tilde{I} = \{(i,j) \in I;\ i \geqslant 2\}$.

If we factorize the joint probability of all t variables as

$$p_{l_1,\ldots,l_t} = p_{l_1 \mid l_2,\ldots,l_t} p_{l_2,\ldots,l_t}, \qquad (\text{A}\cdot 5)$$

statement (A·4) says that the second term for all $(i,j) \in I$ is a product of a function not depending on l_i and one not depending on l_j. If $(i,j) \in I$ and I is reducible, the first term does either not depend on i or not on j, since either $(1,i) \in I$ or $(1,j) \in I$, which means that the joint probability is a product of a function not depending on i and one not depending on j. This then implies,

$$\text{for all } (i,j) \in I, i \perp\!\!\!\perp j \mid \{1,\ldots,t\} \setminus \{i,j\}, \tag{A·6}$$

which shows half of the biimplication in (A·1).

Now assume (A·6). Then, using (A·5) we get

$$p_{l_2,\ldots,l_t} = p_{l_1,\ldots,l_t} / p_{l_1 | l_2,\ldots,l_t}. \tag{A·7}$$

If $(i, j) \in I$, equation (A·6) gives that the denominator is a product of a function not depending on i and one not depending on j. The reducibility implies that the denominator either does not depend on i or not on j. Thus the marginal probability factorizes and we have,

$$\text{for all } (i,j) \in \tilde{I}, i \perp\!\!\!\perp j \mid \{2,\ldots,t\} \setminus \{i,j\}$$

by the induction assumption; this in turn gives,

$$\text{for all } (i,j) \in \tilde{I}, i \perp\!\!\!\perp j \mid \{i+1,\ldots,t\} \setminus \{j\};$$

but for $i = 1$ the first and the second part of (A·1) coincide such that we have,

$$\text{for all } (i,j) \in I, i \perp\!\!\!\perp j \mid \{i+1,\ldots,t\} \setminus \{j\}$$

and (A·1) is demonstrated. Thus (A·2) follows from induction and the proof of the proposition is complete.

References

ANDERSEN, A. H. (1974). Multidimensional contingency tables. *Scand. J. Statist.* **1**, 115–27.

ANDERSEN, E. B. (1980). *Discrete Statistical Models with Social Science Applications.* Amsterdam: North-Holland.

BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc.* B **25**, 220–3.

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

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

DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc.* B **41**, 1–31.

GOOD, I. J. (1958). The interaction algorithm and practical Fourier analysis. *J. R. Statist. Soc.* B **20**, 361–72.

GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: Interaction among multiple classifications. *J. Am. Statist. Assoc.* **65**, 226–56.

GOODMAN, L. A. (1973). The analysis of 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.

LAURITZEN, S. L., SPEED, T. P. & VIJAYAN, K. (1983). Decomposable graphs and hypergraphs. *J. Aust. Math. Soc.* To appear.

MICHAELIS, J., SCHEIDT, E., BUDENZ, M. & KOLLER, S. (1980). Prospective study of disability and early retirement. In *MEDINFO 80*, Proceedings of the Third World Conference on Medical Informatics, Eds D. A. B. Lindberg and S. Kahaira, pp. 956–60. Amsterdam: North-Holland.

PLACKETT, R. L. (1981). *The Analysis of Categorical Data*, 2nd edition. London: Griffin.

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

WERMUTH, N. (1976b). Model search among multiplicative models. *Biometrics* **32**, 253–63.

WERMUTH, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Am. Statist. Assoc.* **75**, 963–72.

WERMUTH, N. WEHNER, T. & GÖNNER, H. (1976). Finding condensed descriptions for multidimensional data. *Computer Programs Biomed.* **6**, 23–38.

WILKS, S. S. (1962). *Mathematical Statistics*, 2nd edition. New York: Wiley.

WINER, B. J. (1971). *Statistical Principles in Experimental Designs*, 2nd edition. New York: McGraw-Hill.

WOLD, H. O. A. (1954). Causality and econometrics. *Econometrica* **28**, 443–63.

WRIGHT, S. (1923). The theory of path coefficients: A reply to Niles' criticism. *Genetics* **8**, 239–55.
WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Statist.* **5**, 161–215.
YATES, F. (1937). *The Design and Analysis of Factorial Experiments*. Harpenden: Imperial Bureau of Soil Science.

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