Ann Inst Stat Math (2009) 61:309–329
DOI 10.1007/s10463-007-0153-1

Log-linear modeling using conditional log-linear structures

P. Vellaisamy · V. Vijay

Received: 12 October 2006 / Revised: 5 March 2007 / Published online: 3 October 2007

Abstract Analysis of large dimensional contingency tables is rather difficult. Fienberg and Kim (1999, *Journal of American Statistical Association*, *94*, 229–239) studied the problem of combining conditional (on single variable) log-linear structures for graphical models to obtain partial information about the full graphical log-linear model. In this paper, we consider the general log-linear models and obtain explicit representation for the log-linear parameters of the full model based on that of conditional structures. As a consequence, we give conditions under which a particular log-linear parameter is present or not in the full model. Some of the main results of Fienberg and Kim follow from our results. The explicit relationships between full model and the conditional structures are also presented. The connections between conditional structures and the layer structures are pointed out. We investigate also the hierarchical nature of the full model, based on conditional structures. Kim (2006, *Computational Statistics and Data Analysis*, *50*, 2044–2064) analyzed graphical log-linear models based on conditional log-linear structures, when a set of variables is conditioned. For this case, we employ the Möbius inversion technique to obtain the interaction parameters of the full log-linear model, and discuss their properties. The hierarchical nature of the full model is also studied based on conditional structures. This result could be effectively used for the model selection also. As applications of our results, we have discussed several typical examples, including a real-life example.

Keywords Categorical data · Conditional log-linear model · Graphical model · Hierarchical model · Interaction factor · Log-linear model · Möbius inversion · Model combining

P. Vellaisamy (✉) · V. Vijay
Department of Mathematics, Indian Institute of Technology, Bombay, Powai, Mumbai 400 076, India
e-mail: pv@math.iitb.ac.in

1 Introduction

A contingency table can be analyzed by a log-linear model which involves interaction factors, also called log-linear parameters. These interaction factors identify the underlying structure of the data given in the contingency table. Even in the case of moderately large dimensional contingency tables, the associated log-linear models involve a large number of interaction factors and hence the analysis becomes difficult. Considering the levels of a categorical variable, say X_n, the full contingency table can also be viewed as a collection of contingency tables (called layers) of dimension less than by unity. Thus, an n-dimensional table, with the data collected on all the n-variables, can also be analyzed based on layer tables of dimension $(n-1)$. Sometimes, the data is collected when one or more variables are fixed. This results in conditional tables of dimension less than n. Observe that the analysis of conditional (say, when a variable is fixed) log-linear structures (models) is rather easier compared to that of the full model. As clearly demonstrated in Fienberg and Kim (1999) and Kim (2006), conditional models play an important role in artificial intelligence and especially in Bayesian networks. Also, Gilula and Haberman (1994) used conditional log-linear models for analyzing panel studies which arise commonly in medical sciences and in the behavioral sciences.

In case of graphical models, Fienberg and Kim (1999) showed that, using these conditional log-linear structures, some amount of partial information about the underlying full log-linear model can be derived. They combined the conditional log-linear structures (conditioned over one variable) to get a class of generators, which contains the generating class of the underlying graphical log-linear model. They showed that if a generator θ is present in all the conditional (when X_n is conditioned) structures, then either θ or $\theta \cup \{n\}$ will be present in the full log-linear model. However, if θ is present in some, but not in all, conditional structures, then $\theta \cup \{n\}$ will be present in the full model. It is important to note that the class of graphical log-linear models, where all the generators are cliques of the graph (see, Whittaker 1990; Lauritzen 1996), is a subclass of hierarchical log-linear models which is again a subclass of general log-linear models.

Observe that log-linear models are widely used in several disciplines, including in survival analysis as accelerated life-time models (Christensen 2000). In this paper, we consider the problem of general log-linear modeling (Agresti 2002; Christensen 1997), based on conditional log-linear structures. Though our focus is combining conditional log-linear structures, the approach can be used for combining log-linear structures for the layer tables as well. We point out the close connections between them as and when necessary. As mentioned earlier, the layer tables and the conditional tables arise mainly from the way the data are collected or modeled. From the log-linear modeling point of view, there is not much difference except the constant term. We first establish an explicit representation for the interaction factors of the full log-linear model, based on that of the conditional log-linear structures, when one variable is conditioned. It is shown that some of the main results of Fienberg and Kim (1999) for the graphical log-linear models follow from our results. We obtain necessary and sufficient conditions, based on conditional interaction parameters, for the full model to be hierarchical. We bring out the relationships between the full log-linear model and the conditional log-linear structures.

Kim (2006) extended the works of Fienberg and Kim (1999), and explored the relationship between the graphical log-linear model for the full table and the conditional tables, when a set of variables is conditioned. In the same spirit, we obtain, using the Möbius inversion technique (Lauritzen 1996), explicit representation for the interaction parameters of the full model, based on the conditional interaction parameters, when a set of variables is conditioned. Using these results, the properties of the full model are analyzed in detail. Also, the hierarchical nature of the full model, using conditional structures, is investigated. Several typical examples, including a real-life example, are discussed as applications of our results. The proofs of the results in this paper are given in the Appendix.

2 Log-linear modeling based on conditional (one variable) structures

2.1 Log-linear models

Consider an n-dimensional contingency table corresponding to n-categorical variables $X_1, \ldots, X_n$. Let the support of X_j be $\bar{I}_j = \{1, \ldots, I_j\}$, $1 \leq j \leq n$, and $V = \{1, 2, \ldots, n\}$ denotes the set of indices of the variables. Let us denote the n-dimensional contingency table by $J_n = I_1 \times \cdots \times I_n$. Also, let $p(i_1, \ldots, i_n) = P(X_1 = i_1, \ldots, X_n = i_n)$, where $i_j \in \bar{I}_j$, denote the probability that an observation $(X_1, \ldots, X_n)$ falls into the cell $(i_1, \ldots, i_n)$, that is, $p(i_1, \ldots, i_n)$ denotes the cell probability. For a subset, say, $Z = \{1, \ldots, k\} \subseteq V$, let $i_Z = (i_1, \ldots, i_k)$, and $\tau_Z(i_Z) = \tau_Z(i_1, \ldots, i_k)$ denote the interaction factor between $X_1, \ldots, X_k$, when $(X_1, \ldots, X_k) = (i_1, \ldots, i_k)$. Now let the saturated log-linear model for the full (n-dimensional) table be (see, Whittemore 1978 or Vellaisamy and Vijay 2007)

$$l(i_1, \ldots, i_n) = \ln p(i_1, \ldots, i_n) = \sum_{Z \subseteq \{1,2,\ldots,n\}} \tau_Z(i_Z), \tag{1}$$

where Z is any subset (including null set) of $\{1, \ldots, n\}$. For instance, the log-linear model for a 3-dimensional table is

$$\begin{aligned} l(i_1, i_2, i_3) = \tau_\phi + \tau_1(i_1) + \tau_2(i_2) + \tau_3(i_3) + \tau_{12}(i_1, i_2) \\ + \tau_{13}(i_1, i_3) + \tau_{23}(i_2, i_3) + \tau_{123}(i_1, i_2, i_3), \end{aligned} \tag{2}$$

where τ_ϕ denotes the overall mean, $\tau_j(i_j)$ denotes the main effect of X_j at level i_j, and $\tau_{12}(i_1, i_2)$ denotes an interaction effect between X_1 and X_2.

Let now $J_{n-1}(i_n) = (I_1 \times \cdots \times I_{n-1})(i_n)$ denote the $(n-1)$-dimensional conditional contingency table, when X_n is conditioned at $i_n \in \bar{I}_n$. Also, let $p(i_1, \ldots, i_{n-1}|i_n)$ be the cell probabilities (conditional probabilities) for the conditional table. Then, the (saturated) log-linear model for the conditional table is

$$l(i_1, \ldots, i_{n-1}|i_n) = \ln p(i_1, \ldots, i_{n-1}|i_n) = \sum_{W \subseteq \{1,2,\ldots,n-1\}} u_W(i_W|i_n), \tag{3}$$

where $u_W(i_W|i_n)$ denotes, for example, a s-factor interaction of the conditional log-linear model if $|W| = s$. For the case $n = 3$, we have

$$l(i_1, i_2|i_3) = u_\phi(i_\phi|i_3) + u_1(i_1|i_3) + u_2(i_2|i_3) + u_{12}(i_1, i_2|i_3),$$

where, for example, $u_{12}(i_1, i_2|i_3)$ denotes an interaction effect between X_1 and X_2, when X_3 is conditioned at the level i_3.

Let, for example, $i_Z = (i_1, \ldots, i_k)$. The τ-terms and u-terms satisfy (see, Bishop et al. 1975)

$$\sum_{i_j} \tau_Z(i_1, \ldots, i_k) = 0, \text{ and } \sum_{i_j} u_Z(i_1, \ldots, i_k|i_n) = 0, \text{ for every } j, \ 1 \le j \le k. \tag{4}$$

Based on conditional log-linear structures, Fienberg and Kim (1999) identified a class of possible graphical models for the full table. We obtain the following result which gives relationships between the τ-terms of the full model and the u-terms of the conditional structures.

Theorem 1 *Consider the log-linear models defined in* (1) *and* (3). *Let* $p(i_n) = P(X_n = i_n)$ *denote the marginal distribution of* X_n. *Then, for every* $i_n \in \bar{I}_n$,

$$\text{(i)} \quad \tau_\phi = \frac{1}{I_n} \sum_{i_n \in \bar{I}_n} \{\ln p(i_n) + u_\phi(i_\phi|i_n)\}, \tag{5}$$

$$\text{(ii)} \quad \tau_n(i_n) = \ln p(i_n) + u_\phi(i_\phi|i_n) - \tau_\phi. \tag{6}$$

Also, for each $W \subseteq \{1, 2, \ldots, n-1\}$, $W \ne \phi$ *and* $i_n \in \bar{I}_n$,

$$\text{(iii)} \quad \tau_W(i_W) = \frac{1}{I_n} \sum_{i_n \in \bar{I}_n} u_W(i_W|i_n), \quad \textit{and} \tag{7}$$

$$\text{(iv)} \quad \tau_{W\cup n}(i_W, i_n) = u_W(i_W|i_n) - \tau_W(i_W). \tag{8}$$

Remark 1 It is clear, from (7) and (8), that all the interaction factors τ_Z, where $Z \ne \phi$ and $Z \ne \{n\}$, can be obtained from the conditional interaction factors. However, the terms τ_ϕ and τ_n, given respectively in (5) and (6), can not be obtained from conditional interaction factors alone, as they require the knowledge of $\ln(p(i_n))$ also. Note also that, from (7) and (8), if $u_W(i_W|i_n)$ does not depend on i_n, then $\tau_{W\cup n} = 0$. Also, conversely, if $\tau_{W\cup n} = 0$, then from (8), $u_W(i_W|i_n) = \tau_W(i_W)$ for all i_n, and so $u_W(i_W|i_n)$ does not depend on i_n.

Thus, we have the following important corollary which shows that the zeroness of τ-terms in the full model can be inferred using the u-terms in the conditional models.

Corollary 1 *Consider the log-linear Models* (1) *and* (3). *Let* W *be any nonempty subset of* $\{1, \cdots, n-1\}$. *Then* $\tau_{W\cup n} = 0$ *if and only if* $u_W(i_W|i_n)$ *is independent of* i_n.

Suppose, for example, $u(i_1, i_3|i_4)$ is independent of i_4 in 3-dimensional conditional models. Then, it follows from Corollary 1 that $\tau_{134}(i_1, i_3, i_4) = 0$ in the 4-dimensional full model.

Observe that an n-dimensional table can also be viewed as a collection of $(n-1)$-dimensional tables (layers) corresponding to the levels of X_n. We state here that the interaction factors corresponding to a particular layer, say $X_n = k$, are the same as that of the $(n-1)$-dimensional conditional table, conditioned on $X_n = k$, except for the constant term. Consider, for example, a 3-dimensional table corresponding to the variables X_1, X_2 and X_3. When $X_3 = 1$, we get a 2-dimensional table (layer) and let the log-linear model for the layer table be given by

$$\ln p(i_1, i_2; 1) = \eta_\phi(1) + \eta_1(i_1; 1) + \eta_2(i_2; 1) + \eta_{12}(i_1, i_2; 1). \tag{9}$$

On the other hand, the log-linear model for the conditional table, when X_3 is conditioned at level 1, is

$$l(i_1, i_2|1) = u_\phi(i_\phi|1) + u_1(i_1|1) + u_2(i_2|1) + u_{12}(i_1, i_2|1), \tag{10}$$

Then, it can be easily seen that

$$\eta_\phi(1) = u_\phi(i_\phi|1) + \ln p(\cdot, \cdot, 1); \quad \eta_1(i_1; 1) = u_1(i_1|1);$$
$$\eta_2(i_2; 1) = u_2(i_2|1); \quad \eta_{12}(i_1, i_2; 1) = u_{12}(i_1, i_2|1),$$

where $p(\cdot, \cdot, 1) = \sum_{i_1, i_2} p(i_1, i_2; 1)$. Thus, the interaction parameters are the same for both the layer table and the conditional table, except for the constant term. Indeed, one can prove the following general result.

Theorem 2 *Let* $\ln p(i_1, \ldots, i_{n-1}; i_n) = \sum_{W \subseteq \{1, \ldots, n-1\}} \eta_W(i_W; i_n)$ *denote the log-linear model for the* $(n-1)$*-dimensional layer table corresponding to level* i_n *of* X_n*. Also, let the conditional log-linear model for* $X_1, \ldots, X_{n-1}$ *be given by (3). Then,*

$$\eta_W(i_W; i_n) = u_W(i_W|i_n) \tag{11}$$

for each nonempty subset W of $\{1, \ldots, n-1\}$*, and*

$$\eta_\phi(i_n) = u_\phi(i_\phi|i_n) + \ln p(i_n). \tag{12}$$

It is now clear from Theorems 1 and 2 that the interaction terms τ_Z can also be obtained from the terms η_W's. That is,

$$\tau_W(i_W) = \frac{1}{I_n} \sum_{i_n \in \bar{I}_n} \eta_W(i_W; i_n); \quad \text{and} \tag{13}$$

$$\tau_{W \cup n}(i_W, i_n) = \eta_W(i_W; i_n) - \tau_W(i_W) \tag{14}$$

for each $W \subseteq \{1, \ldots, n-1\}$. We remark here that the formulas (13) and (14) remain valid when $W = \phi$ also; thus the terms τ_ϕ and τ_n can also be obtained from η-terms.

Table 1 Data corresponding to the conditional levels of C

$C = 1$:

A	B: 1	B: 2
1	30	10
2	20	10

$C = 2$:

A	B: 1	B: 2
1	90	20
2	30	10

$C = 3$:

A	B: 1	B: 2
1	60	10
2	40	10

Next we discuss hierarchical and graphical log-linear models.

The log-linear model, defined in (1), is said to be hierarchical if $\tau_A = 0$ implies $\tau_B = 0$ for all $B \supset A$. It is known that many log-linear models, such as an unsaturated model with all nonzero two-factor interactions, do not have direct MLE's and in that case one must resort to iterative procedure. However, for hierarchical log-linear models the MLE's are relatively easy to compute, since the estimates satisfy certain intuitive marginal constraints. If the model structure of a hierarchical log-linear model is determined by the maximal domain subsets of the interaction factors, then these subsets are called generators and the collection of all the generators is called the generating class of the model. Also, if all the generators are cliques (maximal complete subgraph) of the interaction graph then the model is graphical. For a detailed discussion on hierarchical and graphical models, one may refer to Darroch et al. (1980).

Example 1 Consider the following conditional tables (Table 1)

For each conditional level of C, there are three 2×2 conditional tables. The odds ratios can be seen to be OR(1) = 1.5, OR(2) = 1.5 and OR(3) = 1.5, where OR(i) denotes the odds ratio of the ith conditional table. Also, the interaction factors are given by (see, Bishop et al. 1975, p. 27),

$$u_{12}(1, 1|1) = 1/4 \ln(OR(1)).$$

Thus,

$$u_{12}(1, 1|1) = u_{12}(1, 1|2) = u_{12}(1, 1|3) = 0.1014.$$

From Corollary 1, $\tau_{123}(1, 1, k) = 0$. Also, using (4), we have indeed $\tau_{123}(i, j, k) = 0$ for all (i, j, k). Therefore, the full model does not contain any three factor interaction term. Similarly, it can be checked that all the other u-terms do not satisfy the condition of Corollary 1. Hence, all other τ-terms are present in the full model. Thus, the full model $\{\{A, B\}, \{A, C\}, \{B, C\}\}$ is hierarchical but not graphical.

We next show that Theorems 3 and 4 of Fienberg and Kim (1999), for graphical models, follow from our results. We adopt their notations, for convenience.

(i) Consider the case when a generator θ is common to all the conditional log-linear structures, that is, $u_\theta(i_\theta|i_n) \neq 0$ for all $i_n \in \bar{I}_n$. Note, in this case, either $u_\theta(i_\theta|i_{n_1}) = u_\theta(i_\theta|i_{n_2}) \neq 0$, for all $i_{n_1}, i_{n_2} \in \bar{I}_n$ or $u_\theta(i_\theta|i_{n_1}) \neq u_\theta(i_\theta|i_{n_2})$, for some $i_{n_1}, i_{n_2} \in \bar{I}_n$.
In the first case, when $u_\theta(i_\theta|i_{n_1}) = u_\theta(i_\theta|i_{n_2})$, for all $i_{n_1}, i_{n_2} \in \bar{I}_n$, then by Corollary 1, $\tau_{\theta\cup n} = 0$. Since $u_\theta(i_\theta|i_n) \neq 0$, for all i_n, we have from Eq. (7),

Table 2 Conditional data for each level of C

$C = 1$:

A	B 1	B 2
1	30	10
2	40	20

$C = 2$:

A	B 1	B 2
1	10	20
2	30	40

$\tau_\theta \neq 0$ and therefore θ is a generator for the full model. Note however $\theta \cup n$ is not a generator.

In the second case where $u_\theta(i_\theta|i_{n_1}) \neq u_\theta(i_\theta|i_{n_2})$, for some $i_{n_1}, i_{n_2} \in \bar{I}_n$, we have by Corollary 1, $\tau_{\theta\cup n} \neq 0$. Hence, $\theta \cup n$ is a generator for the full model. Therefore, if θ is present in all conditional log-linear structures, then either θ or $\theta \cup n$ is a generator for the full model. This is essentially Theorem 3 of Fienberg and Kim (1999).

(ii) Consider next the case when a generator θ is not common to all conditional log-linear structures, but is present in at least one of structures. This implies that $u_\theta(i_\theta|i_{n_j}) = 0$, for some $i_{n_j} \in \bar{I}_n$, and $u_\theta(i_\theta|i_{n_l}) \neq 0$, for some $i_{n_l} \in \bar{I}_n$. Again by Corollary 1, $\tau_{\theta\cup n} \neq 0$, and so $\theta \cup n$ is a generator for the full model. This is essentially Theorem 4 of Fienberg and Kim (1999).

2.2 Hierarchical nature of the full model

We discuss here the hierarchical nature of the full and the conditional models. It is possible sometimes that the conditional log-linear models are hierarchical, while the full log-linear models are non-hierarchical. We first discuss an example in this direction.

Example 2 Consider the following conditional data sets (Table 2)

From Table 2, we have $u_{12}(1, 1|1) = 0.1014$ and $u_{12}(1, 1|2) = -0.1014$. Also, it can be easily checked that all other u-terms for both the tables are non-zero and so the conditional models are hierarchical. Also, using (7) and (8), we get $\tau_{12}(1, 1) = 0$, but $\tau_{123}(1, 1, 1) \neq 0$, which implies $\tau_{12} = 0$, but $\tau_{123} \neq 0$. Hence, the full model is non-hierarchical.

The above example motivates the following result.

Theorem 3 *Consider the full and conditional models specified in* (1) *and* (3)*, and assume $\tau_n \neq 0$. Then the full model is hierarchical if and only if for each $W \subseteq \{1, \ldots, n-1\}$, the u-terms satisfy either of the following two conditions:*

(C_1): $u_W(i_W|i_{n_1}) \neq u_W(i_W|i_{n_2})$ *for some* $i_{n_1}, i_{n_2} \in \bar{I}_n \Longrightarrow \sum_{i_n \in \bar{I}_n} u_W(i_W|i_n) \neq 0.$

(C_2): (a) $u_W(i_W|i_n) = u_W(i_W)$, *independent of* $i_n \in \bar{I}_n \Longrightarrow u_{W'}(i_{W'}|i_n) = u_{W'}(i_{W'})$ *is independent of* i_n, *for all* $W' \supset W$.

(b) *Also,* $u_W(i_W) = 0$ *in* $C_2(a) \Longrightarrow u_{W'}(i_{W'}) = 0,$ *for all* $W' \supset W$.

Remark 2 It is not necessary to check the Conditions (C_1) and (C_2) for each $W \subseteq \{1, \ldots, n-1\}$. Indeed, the following procedure simplifies the task.

Table 3 Number of accidents corresponding to the categories A, B, C and D

B	1				2			
C	1		2		1		2	
	D		D		D		D	
A	1	2	1	2	1	2	1	2
1	8314	263	4188	50	42476	3440	19495	552
2	313	37	291	6	1841	383	1678	98
3	189	30	155	8	1214	290	1058	80
4	24	30	14	10	146	51	91	15

(i) Check first if the Condition (C_2) holds for sets W_1 of cardinality unity. If (C_2) holds for some W_1, then it holds for all sets in $\mathcal{D}_1 = \{W' | W' \supset W\}$.
(ii) Repeat Step (i) for sets $W_2 \notin \mathcal{D}_1$, with $|W_2| = 2$, and so on.
(iii) Finally, for those W's for which (C_2) is not satisfied, check for Condition (C_1). If all those W's satisfy (C_1), then the full model is hierarchical. If (C_1) is not satisfied for some W, then stop the procedure and the model is non-hierarchical.
(iv) If, during the Steps (i)–(ii), the Condition (C_2) is not satisfied for some W^*, then immediately check if W^* satisfies (C_1). If W^* does not satisfy (C_1), then stop the procedure and the model is non-hierarchical.

We remark that, if the conditional Models (3) are hierarchical for each $i_n \in \bar{I}_n$, then the Condition $C_2(b)$ is automatically satisfied. Therefore, there is no need to check Condition $C_2(b)$.

We now apply Theorems 1–3, and the above procedure to analyze a real-life example.

Example 3 The 4-dimensional contingency table (Table 3) is from Jobson (1992). The accidents were classified according to the following characteristics:

A: Injury Level (None(1), Minimal(2), Minor(3), Major/Fatal(4)); B: Seatbelt Usage (Yes (1), No(2)); C: Sex of the Driver (Male(1), Female(2)); and D: Drivers Condition (Normal(1), Drinking(2)).

Table 4 gives the estimated values of u-terms for each level of A. These values are calculated using layer tables. Also, the associated values of test statistic (χ^2-statistic) and the corresponding p-values are also calculated using SAS.

Note that a small value of χ^2 (or large p-value) indicates the corresponding interaction parameter is zero. It is clear from the tables that u_{23} and u_{234} can be taken to be zero and hence are equal for each level of A. Therefore, from Corollary 1, $\tau_{123} = \tau_{1234} = 0$. Also, using (7), we get $\tau_{23} = \tau_{234} = 0$. Thus, the highest non-zero interaction factors for the full model are τ_{124} and τ_{134}. It can be seen that $W = \{2, 3\}$ satisfies Condition (C_2) of Theorem 3. Also, the remaining sets $\{2\}, \{3\}, \{4\}, \{2, 4\}, \{3, 4\}$ satisfy Condition (C_1) of Theorem 3. Therefore, the full model is hierarchical with the generating class $\mathcal{G} = \{\{A, B, D\}, \{A, C, D\}\}$.

Table 4 Estimated values of the parameters using SAS

Parameter(u)	Levels	Estimate	Chi-square	Pr > ChiSq
A=1				
B	1	−1.0177	2517.82	< 0.0001
C	1	0.6193	932.35	< 0.0001
B×C	1, 1	−0.0328	2.62	0.1055
D	1	1.7449	7401.87	< 0.0001
B×D	1, 1	0.2255	123.57	< 0.0001
C×D	1, 1	−0.2532	155.81	< 0.0001
B×C×D	1, 1, 1	0.00956	0.22	0.6374
A=2				
B	1	−1.0818	349.74	< 0.0001
C	1	0.4185	52.34	< 0.0001
B×C	1, 1	0.0545	0.89	0.3458
D	1	1.3034	507.73	< 0.0001
B×D	1, 1	0.2008	12.05	0.0005
C×D	1, 1	−0.3771	42.49	< 0.0001
B×C×D	1, 1, 1	−0.0595	1.06	0.3037
A=3				
B	1	−1.0440	372.10	< 0.0001
C	1	0.3682	46.28	< 0.0001
B×C	1, 1	0.0118	0.05	0.8269
D	1	1.1023	414.83	< 0.0001
B×D	1, 1	0.0988	3.33	0.0678
C×D	1, 1	−0.2842	27.58	< 0.0001
B×C×D	1, 1, 1	0.00336	0.00	0.9505
A=4				
B	1	−0.5767	60.72	< 0.0001
C	1	0.4168	31.71	< 0.0001
B×C	1, 1	−0.00736	0.01	0.9207
D	1	0.3710	25.13	< 0.0001
B×D	1, 1	−0.3427	21.44	< 0.0001
C×D	1, 1	−0.1638	4.90	0.0269
B×C×D	1, 1, 1	0.0239	0.10	0.7465

Observe that Theorem 3 is of rather practical importance. Based on the given conditional data sets, we can first estimate the conditional interaction parameters, and then using Theorem 3, we can check if the full log-linear model is hierarchical or not. Thus, Theorem 3 is useful from the point of view of model selection for the full model.

2.3 Relationships between the conditional and the full model

In this subsection, we obtain the explicit relationships between the full log-linear models and log-linear structures for the conditional/layer tables.

Define, for any $Z \subseteq V = \{1, \ldots, n\}$, $V_Z = V \setminus Z$ and

$$\tilde{l}_Z^{(n)}(i_Z) = \frac{1}{\prod_{j \in V_Z} I_j} \sum_{i_j : j \in V_Z} l(i_1, \ldots, i_n), \tag{15}$$

where the superscript n denotes the dimension of the table.

Similarly, for any $W \subseteq D = \{1, \ldots, n-1\}$, let $D_W = D \backslash W$, and

$$\tilde{l}_W^{(n-1)}(i_W | i_n) = \frac{1}{\prod_{j \in D_W} I_j} \sum_{i_j : j \in D_W} l(i_1, \ldots, i_{n-1} | i_n). \tag{16}$$

Obviously, the log-linear model for the full table, from (1), is

$$\tilde{l}_V^{(n)}(i_V) = l(i_1, \ldots, i_n) = \sum_{Z \subseteq V} \tau_Z(i_Z) \tag{17}$$

and log-linear model for the conditional table, from (3), is

$$\tilde{l}_D^{(n-1)}(i_D | i_n) = l(i_1, \ldots, i_{n-1} | i_n) = \sum_{W \subseteq D} u_W(i_W | i_n). \tag{18}$$

Observe that, if $Y \subseteq V$, it follows from (17),

$$\tilde{l}_Y^{(n)}(i_Y) = \sum_{Z \subseteq Y} \tau_Z(i_Z). \tag{19}$$

The following result gives the general relationships between the log-linear models for the full table and the conditional tables. In particular, when $A = D$, the result (iii) of Theorem 4 gives the explicit representation for the full log-linear model.

Theorem 4 *Consider the log-linear models defined in* (17) *and* (18). *Let* $D = \{1, \ldots, n-1\}$, *and* $A \subseteq D$. *Then*

(i) $\tilde{l}_\phi^{(n)} = \frac{1}{I_n} \sum_{i_n \in \bar{I}_n} \{\ln p(i_n) + \tilde{l}_\phi^{(n-1)}(i_\phi | i_n)\}$,

(ii) $\tilde{l}_n^{(n)}(i_n) = \ln p(i_n) + \tilde{l}_\phi^{(n-1)}(i_\phi | i_n)$,

(iii) $\tilde{l}_{A \cup n}^{(n)}(i_A, i_n) - \tilde{l}_n^{(n)}(i_n) = \tilde{l}_A^{(n-1)}(i_A | i_n) - \tilde{l}_\phi^{(n-1)}(i_\phi | i_n)$, *and*

(iv) $\tilde{l}_A^{(n)}(i_A) - \tilde{l}_\phi^{(n)} = \frac{1}{I_n} \sum_{i_n \in \bar{I}_n} (\tilde{l}_A^{(n-1)}(i_A | i_n) - \tilde{l}_\phi^{(n-1)}(i_\phi | i_n))$,

where $\tilde{l}_\phi^{(n-1)}(i_\phi | i_n) = u_\phi(i_\phi | i_n)$.

The proof of the above theorem is omitted in view of Theorem 8.

Remark 3 The n-dimensional full model can be obtained from the conditional models, if we also know the marginal distribution ($i.e.$, $\ln p(i_n)$) of the conditioned variable. Note also that the results in (iii) and (iv) hold for the layer tables also, when the $\tilde{l}_A(i_A|i_n)$ is replaced by the corresponding quantity $\tilde{l}_A(i_A; i_n)$. Thus, the above results are also useful for analyzing full model in terms of the layer models. For example, these results can be used to interpret various hierarchical log-linear models (see, Santner and Duffy 1989, p. 154).

3 Log-linear modeling based on conditional (set of variables) structures

In the earlier sections, we analyzed log-linear models for conditional tables where only one variable (X_n) is conditioned. However, there are many practical situations where the data is obtained, when a set of variables is conditioned. For this case of multiple conditional variables, Kim (2006) explored the relationship between conditional graphical models and the full graphical model. In the same spirit, we analyze, in this section, the full (general) log-linear model based on conditional structures (models), when a set of variables is conditioned.

Consider an n-dimensional contingency table corresponding to the variables $X_1, \ldots, X_n$, and $V = \{1, 2, \ldots, n\}$. Also, let C be the set of indices of the conditional variables. Let $\tau_Z(i_Z)$ and $u_W(i_W|i_C)$ henceforth respectively denote the $|Z|$-factor and $|W|$-factor interactions for the full and the conditional models. As seen earlier, the log-linear model for full (n-dimensional) table can be represented as

$$l(i_1, \ldots, i_n) = \sum_{Z \subseteq V} \tau_Z(i_Z). \tag{20}$$

Assume hereafter, for simplicity, $C = \{d+1, \ldots, n\}$ so that the support of $(X_{d+1}, \ldots, X_n)$ is $\bar{I}_C = \bar{I}_{d+1} \times \cdots \times \bar{I}_n$, where $\bar{I}_k = \{1, \ldots, I_k\}$. Then, the log-linear model for the conditional table, for fixed level of $X_C = i_C \in \bar{I}_C$, is given by

$$l(i_1, \ldots, i_d | i_{d+1}, \ldots, i_n) = \sum_{W \subseteq D} u_W(i_W|i_C), \tag{21}$$

where, here and henceforth, W is any subset of $D = \{1, \ldots, d\}$. Let the log-linear model for the marginal $|C|$-dimensional table of the conditional variables be

$$\ln p(i_C) = \sum_{Z \subseteq C} \delta_Z(i_Z). \tag{22}$$

We now introduce the following notation. Let $C_A = C \setminus A$, for $A \subseteq C$, and define $\tilde{u}_{W|A}$ as

$$\tilde{u}_{W|A}(i_W|i_A) = \frac{1}{\prod_{j \in C_A} I_j} \sum_{i_j : j \in C_A} u_W(i_W|i_C). \tag{23}$$

When $A = C$, we get

$$\tilde{u}_{W|C}(i_W|i_C) = u_W(i_W|i_C). \tag{24}$$

Also, define $\tilde{T}_{W|A}$ as

$$\tilde{T}_{W|A}(i_W|i_A) = \sum_{Y\subseteq A}(-1)^{|A-Y|}\tilde{u}_{W|Y}(i_W|i_Y), \tag{25}$$

where $|B|$ denotes the cardinality of the set B. Note that $\tilde{u}_{W|A}$ and $\tilde{T}_{W|A}$ are defined for any subset W (including the null set) of D.

We now need the following result, known as Möbius inversion formula. For more details, see, (Lauritzen, 1996, p. 239).

Lemma 1 *Let f and g be functions defined on the set of all subsets of a finite set V. Then*

$$f(A) = \sum_{Z\subseteq A} g(Z), \quad \textit{for all } A \subseteq V \iff g(A) = \sum_{Z\subseteq A}(-1)^{|A-Z|} f(Z), \quad \textit{for all } A \subseteq V. \tag{26}$$

The next result is a generalization of Theorem 1, when a set of variables is conditioned.

Theorem 5 *Consider the log-linear Models (20), (21) and (22), and let $\tilde{T}_{W|A}$ be defined as in (25). Then,*

$$\text{(i) } \tau_A(i_A) = \delta_A(i_A) + \tilde{T}_{\phi|A}(i_\phi|i_A), \tag{27}$$

for any subset $A \subseteq C$, and

$$\text{(ii) } \tau_{W\cup A}(i_W, i_A) = \tilde{T}_{W|A}(i_W|i_A), \tag{28}$$

for $W \subseteq \{1, \ldots, d\}$, $W \neq \phi$, and $A \subseteq C$.

When $A = \phi$, Eqs. (27) and (28) reduce to

$$\begin{aligned}\tau_\phi(i_\phi) &= \delta_\phi(i_\phi) + \tilde{T}_{\phi|\phi}(i_\phi|i_\phi)\\ &= \frac{1}{\prod_{j\in C} I_j}\sum_{i_j:j\in C}\{\ln p(i_C) + u_\phi(i_\phi|i_C)\},\end{aligned} \tag{29}$$

and

$$\tau_W(i_W) = \tilde{T}_{W|\phi}(i_W|i_\phi) = \frac{1}{\prod_{j\in C} I_j}\sum_{i_j:j\in C} u_W(i_W|i_C). \tag{30}$$

It is clear now that the τ_A-terms, $A \subseteq C$, can not be obtained based on conditional u-terms only. Indeed, we need the knowledge of the marginal distribution of X_C (or the marginal interaction parameters) also. This is precisely reflected in (27). Observe also that Eqs. (27) and (28) specify all the τ-terms of the full model.

The following result, which is a direct consequence of Theorem 5, gives the conditions under which a τ-term is zero in the full model.

Corollary 2 *Let $A \subseteq C$. Then, $\tau_{W\cup A}(i_W, i_A) = 0$ if and only if $\tilde{T}_{W|A}(i_W|i_A) = 0$, where $W \subseteq \{1, \ldots, d\}$ and $W \neq \phi$.*

The following result, which is an analogue of Theorem 2 and easy to prove, gives connections between the interaction parameters of the layer model and that of the conditional model.

Theorem 6 *Let $\ln p(i_1, \ldots, i_d; i_C) = \sum_{W\subseteq\{1,\ldots,d\}} \eta_W(i_W; i_C)$ denote log-linear model for the d-dimensional layer table corresponding to level i_C of X_C. Also, let the conditional log-linear model for $X_1, \ldots, X_d$ be given by (21). Then,*

$$\eta_W(i_W; i_C) = u_W(i_W|i_C) \tag{31}$$

for each nonempty subset W of $\{1, \ldots, d\}$, and

$$\eta_\phi(i_C) = u_\phi(i_\phi|i_C) + \ln p(i_C). \tag{32}$$

Next, we present a necessary and sufficient condition for the full Model (20) to be hierarchical.

Theorem 7 *Consider the full and the conditional models specified in (20) and (21). Assume $\tau_A \neq 0$, for all $A \subseteq C = \{d+1, \ldots, n\}$. Then the full log-linear model is hierarchical if and only if for each nonempty $W \subseteq \{1, \ldots, d\}$ and $\phi \neq A \subseteq C$, the u-terms satisfy one of the following two conditions:*

(C_1): $\tilde{T}_{W|A}(i_W|i_A) \neq 0 \implies \sum_{i_C\in\bar{I}_C} u_W(i_W|i_C) \neq 0.$

(C_2): (a) $\tilde{T}_{W|A}(i_W|i_A) = 0 \implies \tilde{T}_{W'|A'}(i_{W'}|i_{A'}) = 0$ *for all* $W' \supseteq W$ *and* $A \subseteq A' \subseteq C$; *and*

(b) $\tilde{T}_{W|A}(i_W|i_A) = 0$ *and* $\sum_{i_C\in\bar{I}_C} u_W(i_W|i_C) = 0 \implies \tilde{T}_{W'|A'}(i_{W'}|i_{A'}) = 0$ *and* $\sum_{i_C\in\bar{I}_C} u_{W'}(i_{W'}|i_C) = 0$ *for all* $W' \supseteq W$ *and* $A \subseteq A' \subseteq C$.

Note that, if Condition (C_2) of Theorem 7 holds for some $W \subseteq \{1, \ldots, d\}$ and for some $A \subseteq C$, then it is also true for all $W' \supset W$ and for all $A' \supset A$.

Finally, we present the relationships between the log-linear model of n-dimensional table and the conditional d-dimensional table, where $d = n - |C|$. Let, for $A \subseteq C$,

$$\tilde{l}_A^{(n-d)}(i_A) = \frac{1}{\prod_{j\in C_A} I_j} \sum_{i_j:j\in C_A} l(i_{d+1}, \ldots, i_n) \tag{33}$$

and, for $W \subseteq D = \{1, \ldots, d\}$,

$$l^{*(d)}_{W|A}(i_W|i_A) = \frac{1}{\prod\limits_{j \in C_A} I_j} \sum_{i_j : j \in C_A} \tilde{l}^{(d)}_{W|C}(i_W|i_C), \tag{34}$$

where,

$$\tilde{l}^{(d)}_{W|C}(i_W|i_C) = \frac{1}{\prod\limits_{j \in D_W} I_j} \sum_{i_j : j \in D_W} \ln p(i_1 \ldots, i_d|i_{d+1}, \ldots, i_n), \tag{35}$$

and $D_W = D \backslash W$, as used earlier.

The next result is an extension of Theorem 4, for the case when a set of variables is conditioned. As mentioned earlier, this result is useful to interpret various hierarchical log-linear models also.

Theorem 8 *Consider the log-linear Models* (20) *and* (21)*. Then, for each* $A \subseteq C = \{d+1, \ldots, n\}$

(i) $\tilde{l}^{(n)}_A(i_A) = \tilde{l}^{(n-d)}_A(i_A) + l^{*(d)}_{\phi|A}(i_\phi|i_A)$*; and*
(ii) $\tilde{l}^{(n)}_{W \cup A}(i_W, i_A) - \tilde{l}^{(n)}_A(i_A) = l^{*(d)}_{W|A}(i_W|i_A) - l^{*(d)}_{\phi|A}(i_\phi|i_A),$

where $l^{*(d)}_{\phi|A}(i_\phi|i_A) = \tilde{u}_{\phi|A}(i_\phi|i_A)$ *and* W *is any nonempty subset of* $D = \{1, \ldots, d\}$.

Observe that when $W = D$ and $A = C$, we obtain the log-linear model for the full table from Part (ii) of Theorem 8.

Appendix

Proof of Theorem 1. Note that, for each level of $X_n = i_n$,

$$l(i_1, \ldots, i_n) = l(i_1, \ldots, i_{n-1}|i_n) + \ln p(i_n).$$

Using (1) and (3), we get

$$\sum_{Z \subseteq \{1,2,\ldots,n\}} \tau_Z(i_Z) = \sum_{W \subseteq \{1,2,\ldots,n-1\}} u_W(i_W|i_n) + \ln p(i_n). \tag{36}$$

Summing over i_j, for every $j \in \{1, \ldots, n-1\}$ and using (4), we obtain

$$\tau_\phi + \tau_n(i_n) = u_\phi(i_\phi|i_n) + \ln p(i_n), \tag{37}$$

which proves Part (ii).

Summing now over $i_n \in \bar{I}_n$, in (37) and using (4) again, Part (i) follows. Substituting (37) in (36), we get

$$\sum_{\substack{Z\subseteq\{1,\cdots,n\}\\ Z\neq\phi, Z\neq n}} \tau_Z(i_Z) = \sum_{\substack{W\subseteq\{1,\cdots,n-1\}\\ W\neq\phi}} u_W(i_W|i_n)$$

which is equivalent to

$$\sum_{\substack{W\subseteq\{1,\cdots,n-1\}\\ W\neq\phi}} \{\tau_W(i_W) + \tau_{W\cup n}(i_W, i_n)\} = \sum_{\substack{W\subseteq\{1,\cdots,n-1\}\\ W\neq\phi}} u_W(i_W|i_n). \tag{38}$$

Summing now over i_j in (38), for every $j \in \{1, \ldots, n\}$, except $j = r \neq n$, and using (4), we get

$$\tau_r(i_r) = \frac{1}{I_n} \sum_{i_n\in\bar{I}_n} u_r(i_r|i_n). \tag{39}$$

Similarly, summing over i_j in (38), for every $j \in \{1, \ldots, n-1\}$, except $j = r$, and using (4) gives,

$$\tau_{r\cup n}(i_r, i_n) = u_r(i_r|i_n) - \tau_r(i_r). \tag{40}$$

Since r is arbitrary, Eqs. (39) and (40) prove Parts (iii) and (iv) respectively for the subset W with $|W| = 1$.

Assume now

$$\tau_W(i_W) = \frac{1}{I_n} \sum_{i_n\in\bar{I}_n} u_W(i_W|i_n); \quad \text{and} \tag{41}$$

$$\tau_{W\cup n}(i_W, i_n) = u_W(i_W|i_n) - \tau_W(i_W) \tag{42}$$

are true for any subset W with $|W| \leq t$, where $1 < t \leq n-2$.

Let $B \subseteq \{1, \ldots, n-1\}$ be such that $|B| = t+1$. Summing over i_j in (38), for every $j \in \{1, \ldots, n\} \setminus B$, and using (4), we get

$$\sum_{\substack{W\subseteq B\\ W\neq\phi}} \tau_W(i_W) = \frac{1}{I_n} \sum_{i_n\in\bar{I}_n} \sum_{\substack{W\subseteq B\\ W\neq\phi}} u_W(i_W|i_n),$$

which, using the assumption (41), gives

$$\tau_B(i_B) = \frac{1}{I_n} \sum_{i_n\in\bar{I}_n} u_B(i_B|i_n).$$

Also, summing now over i_j in (38), for every $j \in \{1, \ldots, n-1\} \setminus B$, and using (4), we get

$$\sum_{\substack{W \subseteq B \\ W \neq \phi}} \{\tau_W(i_W) + \tau_{W \cup n}(i_W, i_n)\} = \sum_{\substack{W \subseteq B \\ W \neq \phi}} u_W(i_W | i_n)$$

which, using (42), leads to

$$\tau_B(i_B) + \tau_{B \cup n}(i_B, i_n) = u_B(i_B | i_n).$$

Thus, (41) and (42), are true for $W = B$ also. By induction, Parts (iii) and (iv) follow. □

Proof of Theorem 3. First suppose that the Model (1) is hierarchical. We need to show that either (C_1) or (C_2) holds. Consider now $u_W(i_W | i_n)$, where W is any subset of $\{1, \ldots, n-1\}$. Then either

$$u_W(i_W | i_{n_1}) \neq u_W(i_W | i_{n_2}) \quad \text{for at least one pair} \;\; (i_{n_1}, i_{n_2}) \tag{43}$$

or

$$u_W(i_W | i_{n_1}) = u_W(i_W | i_{n_2}) \quad \text{for all} \;\; (i_{n_1}, i_{n_2}). \tag{44}$$

Suppose now (43) holds. In this case, we show (C_1) holds. Suppose now

$$\sum_{i_n \in \bar{I}_n} u_W(i_W | i_n) = 0. \tag{45}$$

Then, using Corollary 1, the Eq. (43) implies $\tau_{W \cup n} \neq 0$. Also, (45) and (7) imply that $\tau_W = 0$, which contradicts the assumption that the full model is hierarchical. Thus, (C_1) holds.

Suppose, instead of (43), (44) holds. Suppose there exists a set $W' \supset W$ for which

$$u_{W'}(i_{W'} | i_{n_1}) \neq u_{W'}(i_{W'} | i_{n_2}) \quad \text{for some } i_{n_1}, i_{n_2} \in \bar{I}_n.$$

Using Corollary 1, the above equations imply that $\tau_{W \cup n} = 0$ and $\tau_{W' \cup n} \neq 0$, contradicts the assumption. Therefore, $(C_2(a))$ holds.

Finally, let $u_W(i_W) = 0$. Suppose there exists a set $W' \supset W$ for which $u_{W'}(i_{W'}) \neq 0$. Then, from (7), $\tau_W = 0$ and $\tau_{W'} \neq 0$. Again a contradiction arises. Hence, $(C_2(b))$ holds.

Conversely, let Condition (C_1) or (C_2) holds for each $W \subseteq \{1, \ldots, n-1\}$. Suppose the model is non hierarchical. Then, for $W \subseteq W' \subseteq \{1, \ldots, n-1\}$, one of the following three conditions must be true.

$$\begin{aligned} &\text{(a)} \;\; \tau_W = 0, \;\; \text{but } \tau_{W'} \neq 0. \\ &\text{(b)} \;\; \tau_W = 0, \;\; \text{but } \tau_{W \cup n} \neq 0. \\ &\text{(c)} \;\; \tau_{W \cup n} = 0, \;\; \text{but } \tau_{W' \cup n} \neq 0. \end{aligned} \tag{46}$$

Suppose (a) holds. Then, from (7), we have

$$\sum_{i_n \in \bar{I}_n} u_W(i_W|i_n) = 0, \quad \text{and} \tag{47}$$

$$\sum_{i_n \in \bar{I}_n} u_{W'}(i_{W'}|i_n) \neq 0. \tag{48}$$

Note that (47) is true under one of the following two cases.

(i) $u_W(i_W|i_{n_1}) \neq u_W(i_W|i_{n_2})$, for some $i_{n_1}, i_{n_2} \in \bar{I}_n$; or (ii) $u_W(i_W|i_n) = 0$ for each $i_n \in \bar{I}_n$.

If (i) holds then, using Condition (C_1), $\sum_{i_n \in \bar{I}_n} u_W(i_W|i_n) \neq 0$, which contradicts (47). Note also that $u_W(i_W|i_{n_1}) \neq u_W(i_W|i_{n_2})$, for some $i_{n_1}, i_{n_2} \in \bar{I}_n$, shows that Condition (C_2) can't be satisfied.

If (ii) holds, then using Condition $(C_2(b))$, $u_{W'}(i_{W'}|i_n) = 0$, for each $i_n \in \bar{I}_n$ and $W' \supset W$. This contradicts (48). Since $u_W(i_W|i_{n_1}) = u_W(i_W|i_{n_2})$, for all $i_{n_1}, i_{n_2} \in \bar{I}_n$, Condition (C_1) can not be satisfied. Hence, (a) does not hold.

Similarly, if (b) is true, then, from (7), $\tau_W = 0$ implies $\sum_{i_n \in \bar{I}_n} u_W(i_W|i_n) = 0$. Also, using Corollary 1, $\tau_{W\cup n} \neq 0$ implies that there exist at least one pair $i_{n_1}, i_{n_2} \in \bar{I}_n$ for which $u_W(i_W|i_{n_1}) \neq u_W(i_W|i_{n_2})$ which contradicts Condition (C_1). Therefore, (b) can not hold. Also, since $u_W(i_W|i_{n_1}) \neq u_W(i_W|i_{n_2})$, Condition (C_2) can not hold.

Assume now Condition (c) of (46) is true. Then, from Corollary 1, we have $u_W(i_W|i_{n_1}) = u_W(i_W|i_{n_2})$, for all $i_{n_1}, i_{n_2} \in \bar{I}_n$ but $u_{W'}(i_{W'}|i_{n_1}) \neq u_{W'}(i_{W'}|i_{n_2})$ for at least one pair $i_{n_1}, i_{n_2} \in \bar{I}_n$. This contradicts Condition (C_2). Since $u_W(i_W|i_n) = u_W(i_W|i_n)$, Condition (C_1) can not be satisfied. Therefore, (c) does not hold. Thus, we have shown that none of the Conditions (a)–(c) holds. This is equivalent to saying that the full model is hierarchical. This completes the proof. □

Proof of Theorem 5. First note that,

$$l(i_1, \ldots, i_n) = l(i_1, \ldots, i_d|i_C) + \ln p(i_C). \tag{49}$$

Substituting (20) and (21) in (49), we get

$$\sum_{Z \subseteq \{1,2,\ldots,n\}} \tau_Z(i_Z) = \sum_{W \subseteq \{1,2,\ldots,d\}} u_W(i_W|i_C) + \ln p(i_C). \tag{50}$$

Summing over i_j, for every $j \in \{1, \ldots, d\}$ and using (4), we obtain

$$\sum_{Z \subseteq C} \tau_Z(i_Z) = u_\phi(i_\phi|i_C) + \ln p(i_C). \tag{51}$$

(i): Now from (22) and (51), we get

$$\sum_{Z \subseteq C} \{\tau_Z(i_Z) - \delta_Z(i_Z)\} = u_\phi(i_\phi|i_C).$$

Summing over $i_j : j \in C_A = C \setminus A$ and using (4), we get

$$\sum_{Z \subseteq A} \{\tau_Z(i_Z) - \delta_Z(i_Z)\} = \frac{1}{\prod_{j \in C_A} I_j} \sum_{i_j : j \in C_A} u_\phi(i_\phi | i_C) = \tilde{u}_{\phi|A}(i_\phi | i_A), \quad \text{(using (23))}$$

for every $A \subseteq C$. Using now Möbius inversion given in (26), we get

$$\begin{aligned}\tau_A(i_A) &= \delta_A(i_A) + \sum_{Z \subseteq A} (-1)^{|A-Z|} \tilde{u}_{\phi|Z}(i_\phi | i_Z) \\ &= \delta_A(i_A) + \tilde{T}_{\phi|A}(i_\phi | i_A).\end{aligned}$$

This proves Part (i).
(ii): Substituting (51) in (50), we get

$$\sum_{\substack{Z \subseteq \{1,\ldots,n\} \\ Z \not\subseteq C}} \tau_Z(i_Z) = \sum_{\substack{W \subseteq \{1,\ldots,d\} \\ W \neq \phi}} u_W(i_W | i_C)$$

which can be written as

$$\sum_{\substack{W \subseteq \{1,\ldots,d\} \\ W \neq \phi}} \sum_{Y \subseteq C} \tau_{Y \cup W}(i_Y, i_W) = \sum_{\substack{W \subseteq \{1,\ldots,d\} \\ W \neq \phi}} u_W(i_W | i_C). \tag{52}$$

Let A be an arbitrary subset of C. Summing over i_j, for every $j \in C \setminus A$ in (52), and using (23), we get

$$\sum_{\substack{W \subseteq \{1,\ldots,d\} \\ W \neq \phi}} \sum_{Y \subseteq A} \tau_{Y \cup W}(i_Y, i_W) = \sum_{\substack{Z \subseteq \{1,\ldots,d\} \\ Z \neq \phi}} \tilde{u}_{W|A}(i_W | i_A). \tag{53}$$

Using the arguments similar to the proof of Theorem 1, we get

$$\sum_{Y \subseteq A} \tau_{Y \cup W}(i_Y, i_W) = \tilde{u}_{W|A}(i_W | i_A), \quad \forall\, W \subseteq \{1, \ldots, d\}, \;\; W \neq \phi. \tag{54}$$

Since (54) is true for all $A \subseteq C$, we get by Möbius inversion formula (26),

$$\tau_{A \cup W}(i_A, i_W) = \sum_{Y \subseteq A} (-1)^{|A-Y|} \tilde{u}_{W|Y}(i_W | i_Y).$$

The proof of Part (ii) now follows from (25). □

Proof of Theorem 7. First suppose the Model (20) is hierarchical. We need to show that either (C_1) or (C_2) holds. Consider now $\tilde{u}_{W|A}(i_W | i_A)$, where $W \subseteq D$ and $A \subseteq$

$C, A \neq \phi$. Then either

$$\text{(i)}\quad \tilde{T}_{W|A}(i_W|i_A) \neq 0 \quad \text{or} \quad \text{(ii)}\quad \tilde{T}_{W|A}(i_W|i_A) = 0.$$

If (i) holds, we need to show that $\sum_{i_C \in \bar{I}_C} u_W(i_W|i_C) \neq 0$. Suppose now

$$\sum_{i_C \in \bar{I}_C} u_W(i_W|i_C) = 0. \tag{55}$$

Using Corollary 2, we get, from (i), $\tau_{W \cup A} \neq 0$. Also, (30) and (55) imply that $\tau_W = 0$. This contradicts that the full model is hierarchical. Hence, Condition (C_1) holds.

Assume next (ii) holds, which implies, using Corollary 2, $\tau_{W \cup A} = 0$. Suppose now $(C_2(a))$ does not hold. Then $\tilde{T}_{W'|A'} \neq 0$, for some $W' \supseteq W$ and $A' \supseteq A$, which, using again Corollary 2, implies that $\tau_{W' \cup A'} \neq 0$, which shows that the full model is non-hierarchical. Hence, $(C_2(a))$ holds.

The proof of $(C_2(b))$ follows similarly.

Conversely, let for each $W \subseteq \{1, \ldots, d\}$ and for each nonempty subset $A \subseteq C$, the Conditions (C_1) or (C_2) hold. Suppose that the full model is not hierarchical. Then there must exist at least one W for which one of the following conditions must be satisfied.

(a_1) $\tau_W = 0$, and $\tau_{W'} \neq 0$,
(a_2) $\tau_W = 0$, and $\tau_{W \cup A} \neq 0$,
(a_3) $\tau_{W \cup A} = 0$, and $\tau_{W' \cup A'} \neq 0$,

for $A \subseteq B \subseteq C$ and $W \subset W' \subseteq \{1, \ldots, d\}$.

Suppose (a_1) is true, which, using (30), is equivalent to

$$\sum_{i_C \in \bar{I}_C} u_W(i_W|i_C) = 0, \tag{56}$$

$$\sum_{i_C \in \bar{I}_C} u_{W'}(i_{W'}|i_C) \neq 0. \tag{57}$$

Suppose now W satisfies (C_1). Then $\sum_{i_C \in \bar{I}_C} u_W(i_W|i_C) \neq 0$, which contradicts (56). Therefore, Condition (C_1) can not hold. Similarly, if W satisfies (C_2), then $(C_2(b))$ and (56) imply that $\sum_{i_C \in \bar{I}_C} u_{W'}(i_{W'}|i_C) = 0$, which contradicts (57). Therefore, W does not satisfy (C_2) also. Hence, Condition (a_1) cannot hold.

Next suppose (a_2) is true. Then, from (30), we have $\sum_{i_C \in \bar{I}_C} u_W(i_W|i_C) = 0$. Also, $\tau_{W \cup A} \neq 0$ and Corollary 2 imply $\tilde{T}_{W|A}(i_W|i_A) \neq 0$. This shows Condition (C_1) cannot hold. It is clear that Condition (C_2) is not applicable for this case. Hence, (a_2) does not hold.

Finally, suppose (a_3) holds. Then, from Corollary 2, $\tilde{T}_{W|A}(i_W|i_A) = 0$. Also, $\tau_{W' \cup A'} \neq 0$ and Corollary 2 imply that $\tilde{T}_{W'|A'}(i_{W'}|i_{A'}) \neq 0$. Hence, (C_2) cannot hold. Obviously (C_1) also does not hold if (a_3) holds.

Thus, none of the Conditions (a_1)–(a_3) holds. Therefore, the full model is hierarchical. This completes the proof. □

Proof of Theorem 8. Recall from (51),

$$\sum_{Z\subseteq\{d+1,\ldots,n\}} \tau_Z(i_Z) = u_{\phi|C}(i_\phi|i_C) + \tilde{l}_C^{(n-d)}(i_C).$$

Summing over $i_j : j \in C_A$, we get

$$\sum_{Z\subseteq A} \tau_Z(i_Z) = \frac{1}{\prod_{j\in C_A} I_j} \sum_{i_j:j\in C_A} [u_{\phi|C}(i_\phi|i_C) + \tilde{l}_C^{(n-d)}(i_C)]$$
$$= \tilde{u}_{\phi|A}(i_\phi|i_A) + \tilde{l}_A^{(n-d)}(i_A) = l_{\phi|A}^{*(d)}(i_\phi|i_A) + \tilde{l}_A^{(n-d)}(i_A),$$

using (34). This proves Part (i).

Next, it follows from (19),

$$\begin{aligned}\tilde{l}_{W\cup A}^{(n)}(i_W, i_A) &= \sum_{Z\subseteq W\cup A} \tau_Z(i_Z)\\ &= \sum_{Y\subseteq W}\sum_{Z\subseteq A} \tau_{Y\cup Z}(i_Y, i_Z)\\ &= \sum_{\substack{Y\subseteq W\\ Y\neq\phi}}\sum_{Z\subseteq A} \tau_{Y\cup Z}(i_Y, i_Z) + \sum_{Z\subseteq A}\tau_Z(i_Z).\end{aligned}$$

Using (54) and (19), the above equation becomes

$$\begin{aligned}\tilde{l}_{W\cup A}^{(n)}(i_W, i_A) &= \sum_{\substack{Y\subseteq W\\ Y\neq\phi}} \tilde{u}_{Y|A}(i_Y|i_A) + \tilde{l}_A^{(n)}(i_A)\\ &= l_{W|A}^{*(d)}(i_W|i_A) - l_{\phi|A}^{*(d)}(i_\phi|i_A) + \tilde{l}_A^{(n)}(i_A)\end{aligned}$$

which proves Part (ii). □

Acknowledgments Authors thank the Associate Editor and the Referees for some helpful comments. The second author wishes to thank the Council of Scientific and Industrial Research (CSIR), India, for the award of a research fellowship.

References

Agresti, A. (2002). *Categorical data analysis* (2nd Ed.). New York: Wiley.

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

Christensen, R. (1997). *Log-linear models and logistic regressions* (2nd Ed.). New York: Springer.

Christensen, R. (2000). Linear and log-linear models. *Journal of the American Statistical Association*, *95*, 1290–1292.

Darroch, J. N., Lauritzen, S. L., Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Annals of Statistics*, *8*, 522–539.

Fienberg, S. E., Kim, S. H. (1999). Combining conditional log-linear structures. *Journal of the American Statistical Association*, *94*, 229–239.

Gilula, Z., Haberman, S. J. (1994). Conditional log-linear models for analyzing penal categorical data. *Journal of the American Statistical Association*, *89*, 645–656.

Jobson, J. D. (1992). *Applied multivariate data analysis*, Vol. II. New York: Springer.

Kim, S. H. (2006). Conditional log-linear structures for log-linear modeling. *Computational Statistics and Data Analysis*, *50*, 2044–2064.

Lauritzen, S. L. (1996). *Graphical models*. Oxford: Oxford University Press.

Santner, T. J., Duffy, D. E. (1989). *The statistical analysis of discrete data*. New York: Springer.

Vellaisamy, P., Vijay, V. (2007). Some collapsibility results for n-dimensional contingency tables. *Annals of the Institute of Statistical Mathematics* (to appear).

Whittemore, A. S. (1978). Collapsibility of multidimensional contingency tables. *Journal of the Royal Statistical Society*, *Ser. B, 40*, 328–340.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. New York: Wiley.