Conditional and marginal association in contingency tables

Wicher Bergsma and Tamás Rudas
Tilburg University and Eötvös Loránd University

October 18, 2002

Abstract

Standard tools for the analysis of the (average) conditional association structure of the distribution on a multiway contingency table are log-linear models. A different association concept is that of marginal association and this paper describes how marginal log-linear parameters can be used to measure this aspect of association. The paper gives a non-technical discussion of these two aspects of association by discussing their complementary nature and also describes how conditional association is naturally incorporated in the framework provided by marginal log-linear parameters. The properties and interpretation of these parameters are discussed, including the variation independence of hierarchically related marginal log-linear parameters, and the modeling implications of these results are indicated.

1 Introduction

The first aim of this paper is to give an intuitive description of various aspects of association in contingency tables and of some consequences for parameter and model definition. Our approach will be introduced by sharply distinguishing between conditional and marginal aspects of association. However, it will be shown that what is traditionally considered as conditional association, can be incorporated in a generalized framework of marginal association by the introduction of marginal log-linear parameters. It is hoped that by this understanding of the relationship between conditional and marginal association, new insight can be gained into the association structure of multidimensional contingency tables.

The second aim of this paper, which in part overlaps with the first, is to provide an accessible exposition of some of the results of Bergsma and

Rudas (2002). This includes variation independence and smoothness conditions for marginal log-linear parameters. Furthermore, a simple notation for marginal log-linear parameters is introduced which should facilitate their use, especially in applied papers.

Section 2 of the paper briefly reviews the basic facts and ideas regarding the association structure of a multidimensional contingency table. The conditional and marginal aspects of association are discussed with reference to log-linear analysis. Section 3 considers log-linear parameters that are standard tools to measure conditional association and introduces marginal log-linear parameters that are able to represent a more general aspect of association. Marginal log-linear parameters are simply log-linear parameters computed from marginals of the table.

Section 4 discusses parameterizations of the joint distribution using marginal log-linear parameters. This is a large class of flexible parameterizations, with the log-linear parameterization being a simple special case. Conditions for such desirable properties as the parameterization being smooth and its components being variationally independent are also given. These conditions are formulated in terms of simple combinatorial properties of the subsets of variables involved. Section 5 considers statistical models obtained by imposing affine restrictions on marginal log-linear parameters. The conditions that assure existence and standard asymptotic behavior of such models are the same combinatorial properties.

The paper contains no proofs and most of the detailed arguments that support the claims made here can be found in Bergsma and Rudas (2002).

2 Conditional and marginal association

In real-life applications of statistics, the relevant problems are almost all multivariate. In such situations, it is not so much the separate behavior of the variables observed but rather the association among them which is of primary interest. Association, of course, can have many different forms and subject matter knowledge can often be used to postulate a particular association structure.

A definition of association among a group of variables, which is not related to any specific type of association, is obtained by considering the difference of information contained in the joint distribution and that of the lower order marginal distributions. Here, again, no particular technical meaning of information is used, rather it is said that if all lower order marginal distributions are known, the *additional* information needed to reconstruct the joint distribution is the association among the variables. The *additional* information should be sufficient to reconstruct the joint distribution, but it also should be necessary (non-redundant) in the sense that association is only information not contained in the lower order marginal distributions. This latter requirement is best characterized by the concept of variation independence, that is, the joint range of the marginals and of the measure of association should be the Cartesian product of the separate ranges.

As an example, consider a 2×2 contingency table. Here, the lower order marginal distributions are represented by the marginal probabilities π_{1+} and π_{+1}. There are several expressions of the cell probabilities that carry enough information to reconstruct the joint distribution. For example $\pi_{11}/(\pi_{1+}\pi_{+1})$ is intuitively appealing and is sometimes used as a measure of the strength of association. This quantity, however, is not variationally independent from the one way marginals, that is, its range is effected by the actual marginals. Therefore, it lacks calibration and its values, other than 1, may be difficult to compare across different tables. It is only the odds ratio

$$\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}},$$

and its one-to-one functions that are both sufficient and necessary in the above sense to reconstruct the joint distribution. That is, every measure of association (if defined as information not contained in the marginals) is a one-to-one function of the odds ratio. For a detailed exposition of this argument see Edwards (1963) or Rudas (1998).

The log-linear association term is a one-to-one function of the odds ratio and is therefore an appropriate measure of association. In the multivariate case, the argument above generalizes, in parallel to the theory of log-linear representation (Bishop, Fienberg and Holland, 1975), in a hierarchical way (see Rudas, 1998).

A crucial aspect in understanding and modeling the association structure of a multiway table, is the way of defining subsets of variables of which the strength of association is measured. From a technical point of view, there are two ways of deriving lower dimensional subsets from a set of variables: conditioning and marginalization. In conditioning, some of the variables are fixed at certain categories, and the strength of association is measured for the remaining variables. The parameter values obtained will depend on the actual categories of the fixed variables and refer to association in a subset of the population. Log-linear parameters can be used to measure the "average" conditional association over the categories of the fixed variable (more precisely, the log-linear parameters refer to a geometric mean over the categories). Analysis of this conditional association structure can therefore be done by means of log-linear analysis (Bishop, Fienberg and Holland, 1975, pages 33-34). For example, the log-linear model of no-three-variable interaction is the model of constant conditional association between any two variables given the third.

Marginalization, on the other hand, considers subsets of variables without paying attention to the remaining variables, no selection is involved and the association for a group of variables refers to the entire population. This marginal approach to measuring and modeling association cannot be implemented in standard log-linear analysis and it is the aim of the present paper to illustrate how the theory developed in Bergsma and Rudas (2002) can be used to analyze the association structure of a table, including marginal associations, but also conditional associations and certain mixtures of these. The approach presented here also contains the log-linear approach, as a simple special case.

This general methodology is based on the introduction of a very flexible class of parameters of association that will be discussed in the next section.

3 Parameters of association

For simplicity, log-linear and marginal log-linear parameters will be introduced here for an $I \times J \times K$ contingency table ABC, but the definitions extend in a natural way to higher dimensional tables.

The decomposition of log cell probabilities $\log \pi_{i\,j\,k}^{ABC}$ as a sum of log-linear parameters is as follows (see Bishop, Fienberg and Holland, 1975 or Agresti, 1990):

$$\log \pi_{i\,j\,k}^{ABC} = \lambda_{*\,*\,*}^{ABC} + \lambda_{i\,*\,*}^{ABC} + \lambda_{*\,j\,*}^{ABC} + \lambda_{*\,*\,k}^{ABC} + \lambda_{i\,j\,*}^{ABC} + \lambda_{i\,*\,k}^{ABC} + \lambda_{*\,j\,k}^{ABC} + \lambda_{i\,j\,k}^{ABC} \quad (1)$$

In our notation, the superscript of a log-linear parameter identifies the variables and the subscript shows to which variables the parameter refers to (the ones not replaced by an asterisk) and these variables are represented by their relevant indices. For example, $\lambda_{*\,j\,k}^{ABC}$ is usually denoted as $\lambda_{j\,k}^{BC}$. The log-linear parameters are not yet identified and cannot therefore be interpreted. Many identification methods exist, but a common one is the so-called effect coding, obtained by setting the sum over any subscript to zero that is,

$$\lambda_{+*\,*}^{ABC} = \lambda_{*\,+*}^{ABC} = \lambda_{*\,*\,+}^{ABC} = 0$$
$$\lambda_{i\,+*}^{ABC} = \lambda_{+j\,*}^{ABC} = \lambda_{i\,*\,+}^{ABC} = \lambda_{+*\,k}^{ABC} = \lambda_{*\,j\,+}^{ABC} = \lambda_{*\,+k}^{ABC} = 0$$

$$\lambda^{ABC}_{i\,j\,+} = \lambda^{ABC}_{i\,+k} = \lambda^{ABC}_{+j\,k} = 0,$$

where a "+" in the subscript means summation over the corresponding index. Another popular method is the so-called dummy coding, where identification is obtained by setting certain log-linear parameters to zero.

It is well known (Bishop, Fienberg and Holland, 1975, pages 33-34) that the log-linear parameters measure the average strength of conditional association. For example, in the effect coding scheme, if B and C are binary, $\lambda^{ABC}_{*\,j\,k}$ is equal to constant times the average of the values of the odds ratios of B and C, conditioned on the different values of A.

More generally (in the above notation), a log-linear parameter represents the strength of association between the non-asterisked variables, conditioned on and then averaged over the categories of the asterisked variables. If the higher order parameters are (close to) zero then the values averaged do not differ (too much) and the log-linear parameters can be interpreted as partial associations (Hagenaars, 1990). Notice however, that the assumption that for a multiway contingency table the higher order interaction parameters are zero is a very strong one that essentially removes the most important differences between the otherwise unrestricted distribution on the contingency table and a multivariate normal distribution, in the sense that a basic property of the latter is the lack of higher than first order interactions. Such an assumption is usually called a log-linear model (Haberman, 1974; Rudas, 1998).

The two-dimensional marginal cell probabilities are defined as

$$\pi^{AB}_{i\,j} = \sum_k \pi^{ABC}_{i\,j\,k} \quad \pi^{BC}_{i\,j} = \sum_i \pi^{ABC}_{i\,j\,k} \quad \pi^{AC}_{i\,k} = \sum_j \pi^{ABC}_{i\,j\,k}$$

and the one-dimensional marginal cell probabilities as

$$\pi^{A}_{i} = \sum_{j,k} \pi^{ABC}_{i\,j\,k} \quad \pi^{B}_{j} = \sum_{i,k} \pi^{ABC}_{i\,j\,k} \quad \pi^{C}_{k} = \sum_{i,j} \pi^{ABC}_{i\,j\,k}\,.$$

Analogously to the joint probabilities, the two-dimensional marginal ones can be decomposed as

$$\begin{aligned}
\log \pi^{AB}_{i\,j} &= \lambda^{AB}_{*\,*} + \lambda^{AB}_{i\,*} + \lambda^{AB}_{*\,j} + \lambda^{AB}_{i\,j} \\
\log \pi^{BC}_{j\,k} &= \lambda^{BC}_{*\,*} + \lambda^{BC}_{j\,*} + \lambda^{BC}_{*\,k} + \lambda^{BC}_{j\,k} \\
\log \pi^{AC}_{i\,k} &= \lambda^{AC}_{i\,k} + \lambda^{AC}_{*\,*} + \lambda^{AC}_{i\,*} + \lambda^{AC}_{*\,k}
\end{aligned}$$

and the one-dimensional ones as

$$\log \pi^{A}_{i} = \lambda^{A}_{*} + \lambda^{A}_{i}$$

$$\begin{aligned}\log \pi_j^B &= \lambda_*^B + \lambda_j^B \\ \log \pi_k^C &= \lambda_*^C + \lambda_k^C \,.\end{aligned}$$

The log-linear parameters used in the above representations are computed from a marginal of the original table and are therefore, called marginal log-linear parameters. The marginal to which a parameter pertains is indicated in the superscript. The parameters have an interpretation similar to the classical log-linear parameters discussed above. They measure the strength of the average conditional association among a certain group of variables (the effect) , with some of the other variables omitted and some others fixed (i.e. conditioned on). To define a marginal log-linear parameter, one has to choose a subset of the variables (the marginal – and the variables not selected are omitted) and within this marginal another subset (the effect) and the variables in the marginal but not in the effect are conditioned upon. Such a marginal log-linear parameter takes on different values depending on the actual categories of the variables in the effects and the term parameter refers to all these values (see Bergsma and Rudas, 2002, for details).

The definition of marginal log-linear parameters opens up the possibility of defining a large number of parameters, depending on the choice of the marginal and of the effect and the rest of this section in concerned with discussing a strategy of defining marginal log-linear parameterizations. Certain properties of the parameterizations and of the statistical models obtained by restricting the ranges of these parameters will be investigated later on.

It is always the substantial problem at hand that determines which groups of variables (marginals) of the table are of interest. These must be ordered *hierarchically*, that is, in such a way that a later marginal is not contained in an earlier one. For example, for a table ABC one may be interested in the three marginals AB, BC, and ABC. There are two possible hierarchical orderings:

$$(AB, BC, ABC) \text{ and } (BC, AB, ABC)$$

If the ordering of the marginals has been established, the following inductive scheme can be used to construct a set of parameters:

1. Calculate the log-linear parameters all the effects in the first marginal

2. For $k = 2, \ldots, n$, calculate the log-linear parameters for those effects of the kth marginal that have not been used before,

where n is the number of marginals involved.

To illustrate for the ordering (AB, BC, ABC), the first step is to calculate the log-linear parameters in table AB from the probabilities $\pi_{i\,j}^{AB}$:

$$\{\pi_{i\,j}^{AB}\} \rightarrow \{\lambda_{*\,*}^{AB}, \lambda_{i\,*}^{AB}, \lambda_{*\,j}^{AB}, \lambda_{i\,j}^{AB}\} \tag{2}$$

Next, for the BC marginal the effects that have not been used before are included, i.e., the C and BC effects:

$$\{\pi_{j\,k}^{BC}\} \rightarrow \{\lambda_{*\,k}^{BC}, \lambda_{j\,k}^{BC}\} \tag{3}$$

Finally, for the ABC marginal, the only effects that have not been included yet are the AC and ABC effects. Hence

$$\{\pi_{i\,j\,k}^{ABC}\} \rightarrow \{\lambda_{i\,*\,k}^{ABC}, \lambda_{i\,j\,k}^{ABC}\} \tag{4}$$

Thus, combining the sets obtained in (2), (3), and (4), the parameters generated by the sequence (AB, BC, ABC) are

$$\{\lambda_{*\,*}^{AB}, \lambda_{i\,*}^{AB}, \lambda_{*\,j}^{AB}, \lambda_{i\,j}^{AB}, \lambda_{i\,*\,k}^{ABC}, \lambda_{i\,j\,k}^{ABC}\} \tag{5}$$

Note that in (5) all subsets are included, as effects, in the set of generated parameters. It is easy to verify that this only happens if the whole table is included in the sequence of marginals. Such a sequence is called *complete.* Because of hierarchy, the whole table appears at the end of the sequence. Notice that if there are several hierarchical orderings of the marginals possible, the one selected will determine which subsets appear as effects within the marginals. Marginal log-linear parameters generated by a complete sequence form a parameterization of the distribution on the contingency table.

Two specific sets of parameters generated by complete hierarchical sequences of marginals have received special attention in the literature. The first is the set of ordinary log-linear parameters that is generated by the whole table as the only marginal involved: all log-linear effects pertain to the full table. The second is what is called by Glonek and McCullagh (1995) the multivariate logistic transform, and is generated by a hierarchical sequence of all the subsets of the variables as marginals.

For three-way tables ABC, the ordinary log-linear and multivariate logistic parameterization are generated by

$$(ABC),\ (\emptyset, A, B, C, AB, BC, AC, ABC)$$

respectively, yielding the complete hierarchical sets of parameters

$$\{\lambda_{*\,*\,*}^{ABC}, \lambda_{i\,*\,*}^{ABC}, \lambda_{*\,j\,*}^{ABC}, \lambda_{*\,*\,k}^{ABC}, \lambda_{i\,j\,*}^{ABC}, \lambda_{*\,j\,k}^{ABC}, \lambda_{i\,*\,k}^{ABC}, \lambda_{i\,j\,k}^{ABC}\} \tag{6}$$

$$\{\lambda^{\emptyset}, \lambda_i^{A}, \lambda_j^{B}, \lambda_k^{C}, \lambda_{i\,j}^{AB}, \lambda_{j\,k}^{BC}, \lambda_{i\,k}^{AC}, \lambda_{i\,j\,k}^{ABC}\} \tag{7}$$

respectively. The ordinary log-linear parameters have all the variables in the superscript (i.e., the superscript is maximal), and the multivariate logistic parameters have a minimal superscript. The latter contain no asterisks in the subscript. In this sense, the ordinary log-linear and multivariate logistic parameterizations form the end-points of all hierarchical marginal log-linear parameterizations.

A huge number of sequences of marginals is possible even for a moderate number of variables. For one variable, say A, there are two possible complete sequences. The sequences and the parameters they generate are:

$$\begin{array}{lcl} (A) & \rightarrow & \{\lambda^{A}_{*}, \lambda^{A}_{i}\} \\ (\emptyset, A) & \rightarrow & \{\lambda^{\emptyset}, \lambda^{A}_{i}\}\,. \end{array}$$

Note that these are the log-linear and univariate logistic parameters, respectively. For two variables, say A and B, the nine possible complete sequences and the parameters they generate are:

$$\begin{array}{lcl} (AB) & \rightarrow & \{\lambda^{AB}_{*\,*}, \lambda^{AB}_{i\,*}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j}\} \\ (\emptyset, AB) & \rightarrow & \{\lambda^{\emptyset}, \lambda^{AB}_{i\,*}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j}\} \\ (A, AB) & \rightarrow & \{\lambda^{A}_{*}, \lambda^{A}_{i}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j}\} \\ (B, AB) & \rightarrow & \{\lambda^{B}_{*}, \lambda^{B}_{j}, \lambda^{AB}_{i\,*}, \lambda^{AB}_{i\,j}\} \\ (\emptyset, A, AB) & \rightarrow & \{\lambda^{\emptyset}, \lambda^{A}_{i}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j}\} \\ (\emptyset, B, AB) & \rightarrow & \{\lambda^{\emptyset}, \lambda^{B}_{j}, \lambda^{AB}_{i\,*}, \lambda^{AB}_{i\,j}\} \\ (A, B, AB) & \rightarrow & \{\lambda^{A}_{*}, \lambda^{A}_{i}, \lambda^{B}_{j}, \lambda^{AB}_{i\,j}\} \\ (B, A, AB) & \rightarrow & \{\lambda^{B}_{*}, \lambda^{B}_{j}, \lambda^{A}_{i}, \lambda^{AB}_{i\,j}\} \\ (\emptyset, A, B, AB) & \rightarrow & \{\lambda^{\emptyset}, \lambda^{A}_{i}, \lambda^{B}_{j}, \lambda^{AB}_{i\,j}\}\,. \end{array}$$

The first and last sets form the log-linear and bivariate logistic parameters, respectively. Note that the seventh and eighth sequences are the same except for the A and B marginals that are interchanged, yielding two different sets of parameters. In the last sequence A and B can be interchanged but this would yield the same set of parameters.

4 Marginal log-linear parameterizations

A mixed parameterization of a distribution on a contingency table consists of certain marginal probabilities and all the higher order log-linear effects within the table (see Rudas, 1998). For example, for the ABC table, a mixed parameterization may consist of the marginal probabilities $\{\pi^{AB}_{i\,j}\}$ and $\{\pi^{BC}_{j\,k}\}$, and the higher order log-linear effects $\{\lambda^{ABC}_{i\,*\,k}\}$ and $\{\lambda^{ABC}_{i\,j\,k}\}$.

This is a large and flexible class, with the log-linear parameterization being a simple special case. As is well-known, if the marginal probabilities and log-linear parameters in a mixed parameterization have prescribed values, the iterative proportional fitting (IPF) algorithm can be used to reconstruct the joint probability distribution.

In exponential family terminology, the marginal probabilities are called mean value parameters, and the log-linear parameters are called canonical parameters. Barndorff-Nielsen (1978) proved several important properties of mixed parameterizations in terms of mean value and canonical parameters. Firstly, they are obtained from the distributions via a one-to-one transformation that satisfies certain differentiability conditions. Such a parameterization is called *smooth*. Secondly, the mean value and canonical parameters are variation independent. This means that, provided both the mean value and the canonical parameters are compatible within themselves, then they can always be combined to form a joint distribution. More formally, two (possibly vector valued) components of a parameterization are variation independent, if their joint range is the Cartesian product of their separate ranges.

The absence of variation independence can lead to problems in estimation, and may lead to difficulties in the interpretation of parameters (see the example in section 2).

Appropriately selected marginal log-linear parameters constitute a parameterization of the joint distribution. This is a generalization of the result concerning mixed parameterizations.

Consider the marginal log-linear parameters generated by a complete hierarchical sequence $(M_1, \ldots, M_k)$, where M_k contains all the variables. For $1 \leq j < i \leq k$, let $M_j^{(i)}$ consist of those variables, if any, in M_j which also belong to M_i, and let $\lambda(i)$ be the set of marginal log-linear parameters belonging to M_i. Then, $\lambda(i)$ and the marginal probabilities over $M_1^{(i)}, \ldots, M_{i-1}^{(i)}$ form a mixed parameterization of the distribution over M_i, because $\lambda(i)$ contains classical log-linear parameters in the marginal M_i (Bergsma and Rudas, 2002). This fact allows the following recursive scheme to reconstruct the distribution over M_k to be established:

1. Calculate the probability distribution over M_1 directly from the (marginal) log-linear parameters.

2. For $i = 2, \ldots, k$, calculate, using IPF, the probability distribution over M_i from $\lambda(i)$ and the distributions over $M_1^{(i)}, \ldots, M_{i-1}^{(i)}$.

Note that it is assumed that, for each i, the distributions over $M_1^{(i)}, \ldots, M_{i-1}^{(i)}$

are compatible. Conditions for this to be the case will be discussed later on. To illustrate the reconstruction procedure, consider the parameters generated by the marginals (AB, BC, ABC):

$$\{\lambda^{AB}_{*\,*}, \lambda^{AB}_{i\,*}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j}, \lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k}, \lambda^{ABC}_{i\,*\,k}, \lambda^{ABC}_{i\,j\,k}\}$$

To reconstruct the distribution over ABC, we first obtain the AB distribution by direct calculation:

$$\pi^{AB}_{i\,j} = \exp(\lambda^{AB}_{*\,*} + \lambda^{AB}_{i\,*} + \lambda^{AB}_{*\,j} + \lambda^{AB}_{i\,j})$$

Hence, we are left with the reduced set

$$\{\pi^{AB}_{i\,j}, \lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k}, \lambda^{ABC}_{i\,*\,k}, \lambda^{ABC}_{i\,j\,k}\} \tag{8}$$

Now $\pi^B_j = \pi^{AB}_{+j}$ so the mixed parameterization

$$\{\pi^B_j, \lambda^{BC}_{*\,k}\} \tag{9}$$

of $\{\pi^{BC}_{j\,k}\}$ is included in (8). From (9), $\{\pi^{BC}_{j\,k}\}$ can be reconstructed using IPF. Replacement in (8) yields:

$$\{\pi^{AB}_{i\,j}, \pi^{BC}_{j\,k}, \lambda^{ABC}_{i\,*\,k}, \lambda^{ABC}_{i\,j\,k}\}$$

Now this is a mixed parameterization of $\{\pi^{ABC}_{i\,j\,k}\}$, that can again be found using IPF. Hence, the complete distribution over ABC can be reconstructed by applying IPF to a sequence of mixed parameterizations.

As long as the original set of parameters is compatible, the construction can always be performed. As shown by Barndorff-Nielsen (1978), each step is a one-to-one and differentiable transformation. The following theorem follows directly:

Theorem 1 *The set of parameters generated by a hierarchical complete sequence of marginals of a contingency table $\mathcal{T}$, excluding the null effect, forms a smooth parameterization of the distributions over $\mathcal{T}$.*

Note that the null effect is redundant because the probabilities must sum to one.

It is important, that the previous theorem starts with parameters derived from an existing distribution. If one starts the reconstruction process with arbitrarily selected parameter values, the reconstruction process can sometimes fail. Consider the parameterization generated by the marginals (AB, BC, AC, ABC):

$$\{\lambda^{AB}_{*\,*}, \lambda^{AB}_{i\,*}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j},\ \ \lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k},\ \ \lambda^{AC}_{i\,k},\ \ \lambda^{ABC}_{i\,j\,k}\}$$

The first two steps are the same as above and yield

$$\{\pi^{AB}_{i\,j}, \pi^{BC}_{j\,k}, \lambda^{AC}_{i\,k}, \lambda^{ABC}_{i\,j\,k}\}$$

Now $\{\pi^{A}_{i}, \pi^{C}_{k}, \lambda^{AC}_{i\,k}\}$ forms a mixed parameterization of $\{\pi^{AC}_{i\,k}\}$, that can be reconstructed using IPF. So, one obtains the following:

$$\{\pi^{AB}_{i\,j}, \pi^{BC}_{j\,k}, \pi^{AC}_{i\,k}, \lambda^{ABC}_{i\,j\,k}\} \tag{10}$$

However, now there may be a problem. The AB, BC, and AC marginals have been constructed, but they may not be compatible. For example, there is no three-way distribution, that would have the following two-way marginals.

	B	
A	0.35	0.15
	0.15	0.35

	C	
B	0.35	0.15
	0.15	0.35

	C	
A	0.15	0.35
	0.35	0.15

Notice that this is an example of lack of variation independence. The values given belong to the respective ranges of the parameters but they do not belong to the joint range. The well-known reason that a set of marginals may have prescribed values that are incompatible, is that they do not form a so called decomposable set. A set of marginals is decomposable if there is an ordering that satisfies the so-called running intersection property (see Haberman, 1974). This means that for any marginal in the ordering, all those variables which have appeared in any of the marginals before it, also appear in a *single* marginal before it. For example, for the set of marginals $\{AB, CD, AC\}$ the ordering (AB, CD, AC) does not satisfy the running intersection property, since the variables from AC which have appeared before are A and C, but they do not both appear in either AB or CD. However, the ordering (AB, AC, CD) does satisfy the running intersection property, hence the set $\{AB, AC, CD\}$ is decomposable. On the other hand, the elements of the set $\{AB, BC, AC\}$ do not have an ordering satisfying the running intersection property, so $\{AB, BC, AC\}$ is non-decomposable.

The decomposability concept can be used to give explicit conditions for marginal log-linear parameters generated by a sequence of marginals to be variation independent. In particular, all the marginals that are constructed at any given step in the reconstruction process have to form a decomposable set (Kellerer, 1964). That is, for all $k \leq n$, the maximal elements of the first k marginals in the sequence must be decomposable. Such sequences are called *ordered decomposable*. For example, the ordering $(AB, BC, CD, ABCD)$ is ordered decomposable, since $\{AB\}$, $\{AB, BC\}$, $\{AB, BC, CD\}$, and $\{ABCD\}$ are all decomposable, respectively. The ordering (AB, BC, AC, ABC), however, is not ordered decomposable, since the maximal elements of the first three marginals are $\{AB, BC, AC\}$, that is not a decomposable set. The following theorem follows immediately from the construction process.

Theorem 2 *The marginal log-linear parameters generated by a hierarchical sequence of marginals are variation independent if and only if the sequence is ordered decomposable.*

It follows that the ordinary log-linear parameters (excluding the redundant ones) are variation independent. This is easy to see, since given any prespecified values, a distribution can immediately be found by appropriate additions such as in (1). On the other hand, the multivariate logistic transform is not variation independent if there are more than two variables. However, for the three variable case, replacing $\lambda_{i\,k}^{AC}$ in (7) by $\lambda_{i\,*\,k}^{ABC}$ yields the set

$$\{\lambda^{\emptyset}, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{i\,j}^{AB}, \lambda_{j\,k}^{BC}, \lambda_{i\,*\,k}^{ABC}, \lambda_{i\,j\,k}^{ABC}\} \tag{11}$$

This set is generated by the sequence $(\emptyset, A, B, C, AB, BC, ABC)$, that is ordered decomposable. Hence, the parameterization is variation independent.

5 Restricting marginal log-linear parameters

A wide range of interesting statistical models are obtained by imposing affine restrictions on marginal log-linear parameters. Two fundamental questions that will be dealt with in this section are, firstly, when those restrictions are feasible, and, secondly, how to test, using a randomly drawn sample, the hypothesis that the restrictions hold true for a population.

An example of infeasible restrictions was given above, where prescribed AB, BC, and AC marginals of a table ABC turned out to be incompatible. The restrictions can be obtained by prescribing, for example, the marginal log-linear parameters generated by the sequence (AB, BC, AC). Since the three marginals are not ordered decomposable, the generated parameters are not variation independent by Theorem 2. Hence, it is possible that restrictions on them are infeasible. Note that no ordering of these marginals makes them ordered decomposable. In general, if restrictions are placed on parameters generated by an ordered decomposable sequence of marginals, those restrictions will be feasible. Furthermore, as shown by Bergsma and Rudas (2002), linear restrictions on parameters generated by any sequence of marginals are always feasible because the uniform distribution satisfies them.

It may be noted that the question of feasibility is also important in a related field, namely that of conditionally specified distributions (Arnold, B. C., E. Castillo and J-M. Sarabia, 1999). However, the marginal log-linear parameters discussed here are most suitable for specifying average conditional parameters or for specifying constancy of conditional distributions, rather than for the complete specification of the conditional distributions. Therefore the feasibility problems arising in the two fields are of a different nature, and some of the complex issues arising in general conditionally specified distributions do not occur here.

If it has been determined that a particular model is feasible, it may be tested by drawing a sample from the population and assessing its goodness-of-fit to the model. The most widely used sampling schemes are multinomial and Poisson. From Theorem 1, it follows that such models form a so-called curved exponential family, to which standard asymptotic theory is applicable (Lauritzen, 1996). A detailed discussion of this topic is given by Bergsma and Rudas (2002).

6 Acknowledgments

Bergsma's research was supported by The Netherlands Organization for Scientific Research (NWO), Project Number 400-20-001. Rudas's research was supported in part by Grant T-016032 from the Hungarian National Science Foundation.

7 References

Agresti, A. (1990) *Categorical Data Analysis.* Wiley: New York.

Arnold, B. C., E. Castillo and J-M. Sarabia (1999) *Conditional Specification of Statistical Models.* New-York: Springer-Verlag

Barndorff-Nielsen, O. E. (1978) *Information and Exponential Families in Statistical Theory.* Wiley: New York.

Bergsma, W. P. (1997) *Marginal Models for Categorical Data.* Tilburg University Press: Tilburg.

Bergsma, W. P., Rudas, T. (2002) Marginal models for categorical data. *Annals of Statistics*, to appear.

Bishop, Y. M. M., Fienberg, S. E., Holland, P. W. (1975) *Discrete Multivariate Analysis: Theory and Practice.* M. I. T. Press: Boston.

Colombi, R. (1998) A multivariate logit model with marginal canonical association. *Comm. Statist., Theory & Methods*, 27: 2953-2972.

Colombi, R and A. Forcina, (2001). Marginal regression models for the analysis of positive association of ordinal response variables. *Biometrika,* 88: 1007-1019.

Edwards, A. W. F., (1963). The measure of association in a 2×2 table. *J. Roy. Statist. Soc., Ser. A,* 126: 109-114.

Glonek, G. J. N. and P. McCullagh, (1995). Multivariate logistic models. *J. Roy. Statist. Soc., Ser. B,* 57: 533-546.

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

Haberman, S. J., (1981). Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction. *Ann. Statist.* 9: 1178-1186.

Hagenaars, J. A., (1990). *Categorical Longitudinal Data.* Sage: Newbury Park.

Kellerer, H. G., (1964). Verteilungsfunktionen mit gegebenen Marginalverteilungen. *Z. Wahrscheinlichkeitstheorie u. verw. Gebiete*, 3: 247-70.

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

Rudas, T., (1998). *Odds Ratios in the Analysis of Contingency Tables.* Sage: Thousand Oaks.

Rudas, T. and W. P. Bergsma, (2001). On generalized symmetry. *The Quasi Symmetry Project.* http://www.ups-tlse.fr/PROJET_QS/PAPERS/Rudas2/Rudas2.html