Variation independent parameterizations of multivariate categorical distributions

Wicher P. Bergsma (w.p.bergsma@kub.nl)
Tamás Rudas (rudas@tarki.hu) *

Tilburg University, the Netherlands, and Eötvös Loránd University, Hungary

September 27, 2000

Abstract. A class of marginal log-linear parameterizations of distributions on contingency tables is introduced and necessary and sufficient conditions for variation independence are derived. Connections with the well-known marginal problem are discussed.

Keywords: decomposability, log-linear parameters, variation independence

AMS subject classification: 62H17, 62E99.

1. Introduction

In the past thirty years, log-linear methods have gained wide acceptance in categorical data modeling. More recently, the methods are being extended in order to allow the analysis of marginal distributions of contingency tables (see, for example, McCullagh and Nelder, 1989; Liang et al., 1992; Lang and Agresti, 1994; Glonek and McCullagh, 1995; and Bergsma, 1997). However, little attention has been paid to the feasibility of restrictions on marginals. In the present article, a class of parameterizations is defined which can be useful in the log-linear modeling of marginal distributions. Necessary and sufficient conditions for variation independence of these parameterizations are derived. Importantly, if the parameterization is variation independent, it can be arbitrarily restricted.

To see the importance of the above, consider a $2 \times 2 \times 2$ contingency table ABC. Assume that in the AB and BC marginals the cells $(1,1)$ and $(2,2)$ and in the AC table the cells $(1,2)$ and $(2,1)$ have probabilities equal to $1/2$. Although these marginals are (weakly) compatible, because they imply uniform one-way marginal distributions,

* Research supported by The Netherlands Organization for Scientific Research (NWO), Project Number 400-20-001 and the Hungarian National Science Foundation, Grants T-032213 and T-016032

there exists no three-way distribution with these two-way marginals. A (well-known) reason that this incompatibility can occur is that the set of marginals $\{AB, BC, AC\}$ is not decomposable (Kellerer, 1964). If the marginals are decomposable then weak compatibility (the given distributions coincide on the intersections of the marginals) implies strong compatibility (the existence of a joint distribution with the given marginal distributions). The apparent problem in the previous example is that parameterizations of the marginals which are prescribed are not variation independent. It will be shown in this paper that variation independence of a set of parameters pertaining to different marginals of the table depend on a generalization of the concept of decomposability.

Decomposability is only defined for incomparable (with respect to inclusion) marginals (Haberman, 1974). However, for many practical problems it is necessary to consider restrictions on comparable marginals also. Rather than fully prescribing marginals as in the example above, these restrictions usually pertain to the (marginal) dependence structure in the table.

In Section 2, set-theoretical concepts which are needed in the later sections, are defined. These include decomposability and ordered decomposability, where the latter is a generalization of the former. In Section 3, (marginal) log-linear parameters are introduced, and a class of useful parameterizations of distributions over a contingency table is defined. In Section 4, necessary and sufficient conditions for the variation independence of the parameterizations are given.

Extensions to continuous distributions, in particular the multivariate normal, are possible but are not considered in the present paper.

2. Decomposable and ordered decomposable hypergraphs

Let V be a finite set, called the *base set*. A *hypergraph* is a collection of subsets of V. An ordering $(h_1, \ldots, h_s)$ of the elements of a hypergraph is called *hierarchical* if $h_i \not\subseteq h_j$ if $i > j$. It satisfies the *running intersection* property if $s \leq 2$ or, for $k = 3, \ldots, s$, there exists a $j_k < k$ such that

$$\left(\cup_{i=1}^{k-1} h_i\right) \cap h_k = h_{j_k} \cap h_k$$

A hypergraph is called *reduced* if its subsets are pairwise incomparable in the sense that none is a subset of the other. A reduced hypergraph is called *decomposable* if there is an ordering of its elements satisfying the running intersection property. An arbitrary hypergraph is *ordered decomposable* if there is a hierarchical ordering of its elements, say $(h_1, \ldots, h_s)$, such that, for $k = 1, \ldots, s$, the maximal elements of $\{h_1, \ldots, h_k\}$ form a decomposable set. This ordering is also called ordered decomposable. Note that decomposable hypergraphs are also ordered decomposable.

The above definitions can be illustrated by some examples. Suppose $V = \{A, B, C, D\}$. Omitting braces and commas for the subsets of V, the reduced hypergraphs

$$\{AB, BC, CD\}, \{ABC, BCD\}$$

are decomposable (and therefore also ordered decomposable). On the other hand,

$$\{AB, BC, AC\}, \{AB, BC, CD, AD\}, \{ABC, ACD, BCD\}$$

are all non-decomposable. If a hypergraph is not reduced, i.e., it contains comparable subsets, the decomposability concept does not apply, but the ordered decomposability concept does. The hypergraphs

$$\{AB, BC, ABC\}, \{ABC, BCD, ABCD\}$$

are ordered decomposable, while

$$\{AB, BC, AC, ABC\}$$
$$\{AB, BC, CD, AD, ABCD\}$$
$$\{ABC, ACD, BCD, ABCD\}$$

are not.

A sufficient condition for ordered decomposability is that all subsets are decomposable. An example illustrating that there is no necessity is the hypergraph $\{AB, BC, ACD, ABC\}$, which has the non-decomposable subset $\{AB, BC, ACD\}$. However, the ordering (AB, BC, ABC, ACD) is an ordered decomposable one, so the hypergraph is ordered decomposable.

3. Marginal log-linear parameterizations

3.1. Log-linear parameters

Let V be a finite set of categorical variables, and for $v \in V$ let I_v be a finite index set. A contingency table T_V is defined as the Cartesian product $\times_{v \in V} I_v$. An element of T_V is called a *cell*. A probability distribution over T_V is defined by positive numbers π_i^V $(i \in T)$ for which $\sum_{i \in T_V} \pi_i^V = 1$. The number π_i is called a cell probability.

Log-linear parameters are defined as certain sums and differences of logarithms of cell probabilities. Marginal log-linear parameters are log-linear parameters calculated from marginal probabilities. A general (standard) definition can be found in Bergsma and Rudas (2001). Here, we suffice with a description of log-linear parameters when $V = \{A, B\}$ has only two elements. The log-linear decomposition for the cell probabilities in the usual notation is

$$\log \pi_{i\,j}^{AB} \;=\; \lambda + \lambda_i^A + \lambda_j^B + \lambda_{i\,j}^{AB}$$

For the present paper, log-linear parameters are taken from different marginal tables, and therefore the following notation is more convenient:

$$\log \pi_{i\,j}^{AB} \;=\; \lambda_{*\,*}^{AB} + \lambda_{i\,*}^{AB} + \lambda_{*\,j}^{AB} + \lambda_{i\,j}^{AB} \tag{1}$$

Here, the superscript indicates to which marginal table the parameters belong. An asterisk ($*$) in the subscript indicates that the parameter does not depend on the value of the corresponding variable in the superscript.

With the identifying restrictions

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

(where a '+' in the subscript denotes summation over the index) the λ parameters can be uniquely determined.

The A and B marginal probabilities are defined as

$$\begin{aligned} \pi_i^A &= \pi_{i\,+}^{AB} = \textstyle\sum_{j \in I_B} \pi_{i\,j}^{AB} \\ \pi_j^B &= \pi_{+j}^{AB} = \textstyle\sum_{i \in I_A} \pi_{i\,j}^{AB} \end{aligned}$$

respectively. The log-linear decompositions for the marginal probabilities are

$$\begin{aligned} \log \pi_i^A &= \lambda_*^A + \lambda_i^A \\ \log \pi_j^B &= \lambda_*^B + \lambda_j^B \end{aligned}$$

respectively, with identifying restrictions

$$\lambda_+^A = \lambda_+^B = 0$$

We now illustrate the calculation of the above log-linear parameters for the case $I_A = I_B = \{1, 2\}$. The one-variable marginal parameters are marginal logits:

$$\lambda_1^A = \tfrac{1}{2} \log \frac{\pi_1^A}{\pi_2^A} \quad \lambda_1^B = \tfrac{1}{2} \log \frac{\pi_1^B}{\pi_2^B}$$

(Note that the redundant parameters with index 2 are omitted.) The one variable parameters in T_{AB} are average conditional logits:

$$
\begin{aligned}
\lambda^{AB}_{1\,*} &= \frac{1}{2}\left(\tfrac{1}{2}\log\frac{\pi^{AB}_{1\,1}}{\pi^{AB}_{2\,1}} + \tfrac{1}{2}\log\frac{\pi^{AB}_{1\,2}}{\pi^{AB}_{2\,2}}\right)\\
\lambda^{AB}_{*\,1} &= \frac{1}{2}\left(\tfrac{1}{2}\log\frac{\pi^{AB}_{1\,1}}{\pi^{AB}_{1\,2}} + \tfrac{1}{2}\log\frac{\pi^{AB}_{2\,1}}{\pi^{AB}_{2\,2}}\right)
\end{aligned}
$$

Finally, the two variable parameter is the log odds-ratio:

$$
\lambda^{AB}_{1\,1} = \tfrac{1}{4}\log\frac{\pi^{AB}_{1\,1}\pi^{AB}_{2\,2}}{\pi^{AB}_{1\,2}\pi^{AB}_{2\,1}}
$$

3.2. Construction of parameterization

The (marginal) log-linear parameters defined above can be used to construct parameterizations of distributions over the contingency table T_V. The first step is to choose a set of marginals of interest (i.e., a hypergraph with V as the base set), and to order them hierarchically.

We illustrate the construction of a parameterization by an example. The general case presents no special additional difficulties, and is described formally in Bergsma and Rudas (2001). Suppose the marginals of interest are $\{AB, BC, ABC\}$. There are two hierarchical orderings, namely (AB, BC, ABC) and (BC, AB, ABC). For the ordering (AB, BC, ABC), the construction of a marginal log-linear parameterization is as follows:

$$
\begin{array}{llll}
(i) & \{\pi^{AB}_{i\,j}\}\cup & \{\pi^{BC}_{j\,k}\}\cup & \{\pi^{ABC}_{i\,j\,k}\}\\
(ii) & \{\pi^{AB}_{i\,j}\}\cup & \{\pi^{BC}_{j\,k}\}\cup & \{\lambda^{ABC}_{i\,*\,k}, \lambda^{ABC}_{i\,j\,k}\}\\
(iii) & \{\pi^{AB}_{i\,j}\}\cup & \{\lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k}\}\cup & \{\lambda^{ABC}_{i\,*\,k}, \lambda^{ABC}_{i\,j\,k}\}\\
(iv) & \{\lambda^{AB}_{*\,*}, \lambda^{AB}_{i\,*}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j}\}\cup & \{\lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k}\}\cup & \{\lambda^{ABC}_{i\,*\,k}, \lambda^{ABC}_{i\,j\,k}\}
\end{array}
$$

In (i), the marginal probabilities belonging to the tables T_{AB}, T_{BC}, and T_{ABC} are given. In the next steps, the (marginal) probabilities are replaced by (marginal) log-linear parameters, in order going from right to left. In (ii), $\{\pi^{ABC}_{i\,j\,k}\}$ has been substituted by the set of those log-linear parameters belonging to T_{ABC} for which the non-asterisked variables are not contained in AB or BC (i.e., the marginals appearing before ABC in the sequence (AB, BC, ABC)). For example, the non-asterisked set of variables of $\{\lambda^{ABC}_{i\,*\,k}\}$ is AC (omitting braces), which is not contained in either AB or BC. In (iii), be the same logic,

$\{\pi^{BC}_{j\,k}\}$ has been replaced by $\{\lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k}\}$. Note that the set of parameters $\{\lambda^{BC}_{*\,*}, \lambda^{BC}_{j\,*}\}$, which also belong to table BC, are omitted because $\emptyset \subseteq AB$ and $B \subseteq AB$. Finally, in (iv), $\{\pi^{AB}_{i\,j}\}$ is replaced by the set of all log-linear parameters belonging to T_{AB}. Since, omitting redundant parameters, every transformation is a homeomorphism, the final product is a proper parameterization of the probability distribution over table T.

The above procedure directly generalizes, in that arbitrary hierarchical orderings of marginals generate a marginal log-linear parameterization of the distribution.

4. Variation independence of parameterizations

A multidimensional parameter is called *variation independent* if its range is the Cartesian product of the separate ranges of its coordinates. In the previous section a class of parameterizations of probability distributions over a contingency table was given. The question now arises when such parameterizations are variation independent. Below, the problem is illustrated by two examples.

Consider the parameterization generated by the sequence of marginals (AB, BC, ABC), as discussed in Section 3.2. If the parameters have been assigned given values, the probability distribution over ABC can be reconstructed by following the steps (i) to (iv) in reverse order. The parameters in (iii) can be calculated from those in (iv) directly by applying formula (1). From $\{\pi^{AB}_{i\,j}\}$ we can immediately calculate $\{\pi^{B}_{j}\}$ by appropriate summation. Now $\{\pi^{B}_{j}, \lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k}\}$ forms a so-called mixed parameterization of $\{\pi^{BC}_{j\,k}\}$. A mixed parameterization has variation independent components (Barndorff-Nielsen, 1978), so $\{\pi^{BC}_{j\,k}\}$ in (ii) can be calculated from the parameters in (iii). The calculation can be carried out using the so-called iterative proportional fitting procedure (see, e.g., Agresti, 1990). Similarly, the parameters in (ii) form a mixed parameterization of $\{\pi^{ABC}_{i\,j\,k}\}$, and $\{\pi^{ABC}_{i\,j\,k}\}$ can be calculated using iterative proportional fitting.

The reason that the above reconstruction process can always be carried out, whatever the initial assignment of values to the λ parameters, is that the marginals which are calculated at the intermediate stages form a decomposable set, and hence have an extension. This is not always the case for parameterizations generated by arbitrary hierarchical orderings of marginals. For example, consider the parameterization based on the sequence of marginals (AB, BC, AC, ABC), which is

$$\{\lambda^{AB}_{i\,*}, \lambda^{AB}_{*\,j}, \lambda^{AB}_{i\,j}\} \cup \{\lambda^{BC}_{*\,k}, \lambda^{BC}_{j\,k}\} \cup \{\lambda^{AC}_{i\,k}\} \cup \{\lambda^{ABC}_{i\,j\,k}\} \tag{2}$$

If the λ parameters have been assigned given values, then an intermediate stage in the reconstruction of $\pi^{ABC}_{i\,j\,k}$ yields the marginals

$$\{\pi^{AB}_{i\,j}\} \cup \{\pi^{BC}_{j\,k}\} \cup \{\pi^{AC}_{i\,k}\} \tag{3}$$

Since the hypergraph $\{AB, BC, AC\}$ is non-decomposable, (3) may not have an extension. That is, it is possible to assign values to the parameters in (2) for which there does not exist a joint distribution. It follows that the parameters in (2) are not variation independent.

In general, we have the following theorem.

THEOREM 1. *A marginal log-linear parameterization generated by a hierarchical ordering of marginals is variation independent if and only if the ordering is ordered decomposable.*

A formal proof is given in Bergsma and Rudas (2001).

Note that in the context of log-linear modeling, decomposability is studied as a possible property of the log-linear effects in the model (Haberman, 1974), while in the above theorem ordered decomposability is a property of the marginals within which the effects are defined. In a log-linear parameterization there is only one marginal involved and ordered decomposability holds true.

References

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

Barndorff-Nielsen, O.: 1978, *Information and exponential families*. New York: Wiley.

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

Bergsma, W. P. and T. Rudas: 2001, 'Marginal models for categorical data'. *Tentatively accepted, Annals of Statistics* pp. 1–18.

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*. Chicago: University of Chicago Press.

Kellerer, H. G.: 1964, 'Verteilungsfunktionen Mit Gegebenen Marginalverteilungen'. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **3**, 247–270.

Lang, J. B. and A. Agresti: 1994, 'Simultaneously modelling the joint and marginal distributions of multivariate categorical responses'. *J. Am. Stat. Ass.* **89**, 625–632.

Liang, K. Y., S. L. Zeger, and B. Qaqish: 1992, 'Multivariate regression analyses for categorical data (with discussion)'. *J. Roy. Statist. Soc. Ser. B* **54**, 3–40.

McCullagh, P. and J. A. Nelder: 1989, *Generalized Linear Models*. London: Chapman and Hall.