Developments in Statistics
Andrej Mrvar and Anuška Ferligoj (Editors)
Metodološki zvezki, 17, Ljubljana: FDV, 2002

Restricting the Marginals of Contingency Tables: Potential Problems and How Can They be Avoided

Wicher P. Bergsma[1] and Tamás Rudas[2]

Abstract

Statistical models defined by imposing restrictions on marginal distributions of contingency tables have received considerable attention recently. While these models are flexible and useful, certain theoretical questions have remained open in the literature. These include, firstly, the existence of a joint distribution with certain restrictions on some of its marginals, or general conditions under which the existence of such distributions is guaranteed; secondly, the determination of the dimension of a model; thirdly, a formal argument for the applicability of large sample results for maximum likelihood estimates. A general theory for answering these questions is discussed in this article.

1 Introduction

Several recent papers discuss the theory and application of models for contingency tables which impose restrictions on marginal distributions of the contingency table, see, for example, McCullagh and Nelder (1989); Liang, Zeger, and Qaqish (1992); Becker (1994); Lang and Agresti (1994); Glonek and McCullagh (1995); and Bergsma (1997). However, certain theoretical questions have remained open in the literature. These include, firstly, what are the conditions for restrictions on marginals to be feasible. That is, it is possible to specify contradictory constraints, but no simple method for determining whether or not constraints are contradictory is known. A second question concerns the determination of the dimension of a model. In many cases, some of the constraints imposed on certain marginals are redundant in a not so obvious way. This makes it difficult to determine the dimension of the model. A third question is whether standard large sample theory is applicable. Most

[1] Department of Methodology and Statistics, Tilburg University, P. O. Box 90153, 5000 LE Tilburg, The Netherlands.

[2] Eötvös Loránd University, Budapest, Hungary.

authors assume it is (e.g., Lang and Agresti (1994); Glonek and McCullagh (1995)), while this has not been proven, nor is this always so.

A theory for answering the above questions has been developed by the authors of the present article in a theoretical paper Bergsma and Rudas (2002). Since these results may be of interest to many applied researchers, we summarize them here in an application oriented manner, omitting all the proofs and some of the more technical details.

Two concepts play an important role in answering the questions above: smoothness and variation independence of parameters. A parameter is smooth if it satisfies certain differentiability conditions. This is important for the interpretation of the parameter. Furthermore, if the parameters belong to an exponential family of distributions, standard maximum likelihood theory can be applied if the model of interest is defined by linear restrictions imposed on a smooth parameter. Two parameters are variation independent if the range of possible values of one of them does not depend on the other's value. If two parameters are not variation independent, this causes problems in their interpretation, it leads to the possibility of the definition of non-existing models, and frequently also causes problems in various computations. Note that the multivariate logistic parameters of Glonek and McCullagh (1995) are not variation independent if there are more than two variables.

Marginal models are studied in this paper by studying marginal log-linear parameters. Both are defined in Section 2. In Section 3, some examples are given where problems occur when imposing restrictions on marginal distributions. In Section 4, smoothness and variation independence are explained, and are shown to be important concepts in the understanding of marginal log-linear parameters. In Section 5, the properties of marginal log-linear parameters are used to provide a useful methodology for checking whether models exist and whether standard asymptotic theory can be applied to maximum likelihood estimates.

2 Marginal models

In Section 2.1, marginal log-linear parameters are introduced. In Section 2.2, log-affine marginal models are defined.

2.1 Marginal log-linear parameters

For a given set of categorical variables, a marginal frequency $\mu_{\mathcal{M}}(\mathbf{i})$ is defined by a subset $\mathcal{M}$ of the variables and an index $\mathbf{i}$ for the cell to which the marginal frequency belongs. For example, for a given $I \times J$ contingency table AB, the frequency belonging to cell (i, j) of the table is denoted as $\mu_{AB}(i, j)$.

A marginal log-linear parameter $\lambda_{\mathcal{L}}^{\mathcal{M}}(\mathbf{i})$ is defined by two subsets of the variables: $\mathcal{M}$ denotes the marginal variables to which the parameter belongs, $\mathcal{L}$ denotes a subset of these marginal variables. For example, $\lambda^{ABC}_{AB}(i,j)$ belongs to the marginal ABC, and relates to the interaction effect between A and B in that table for categories i and j, respectively.

For a marginal table AB, the marginal log-linear decomposition of expected cell frequencies $\mu_{AB}(i,j)$ is

$$\log \mu_{AB}(i,j) = \lambda^{AB} + \lambda^{AB}_{A}(i) + \lambda^{AB}_{B}(j) + \lambda^{AB}_{AB}(i,j)$$

Note the difference with the standard notation of log-linear parameters: the superscript AB refers to the *marginal* distribution from which the parameters are calculated. The subscript refers to the log-linear effect, which in the standard notation is in the superscript.

The marginal log-linear parameters have not been identified. As identifying restrictions, we may use

$$\lambda^{AB}_{A}(+) = \lambda^{AB}_{B}(+) = \lambda^{AB}_{AB}(+,j) = \lambda^{AB}_{AB}(i,+) = 0$$

where a "+" in place of an index denotes summation over that index. Analogously, for the marginal tables A and B, the marginal log-linear decompositions are

$$\begin{aligned} \log \mu_{A}(i) &= \lambda^{A} + \lambda^{A}_{A}(i) \\ \log \mu_{B}(j) &= \lambda^{B} + \lambda^{B}_{B}(j) \end{aligned}$$

with possible identifying restrictions $\lambda^{A}_{A}(+) = \lambda^{B}_{B}(+) = 0$.

Since summing a marginal log-linear parameter over an index yields zero, some are redundant. We define $\lambda^{\mathcal{M}}_{\mathcal{L}}$ as the set of parameters $\lambda^{\mathcal{M}}_{\mathcal{L}}(\mathbf{i})$, omitting redundant elements by not including the last value of each index. For example,

$$\lambda^{AB}_{AB} = \{\lambda^{AB}_{AB}(i,j) \mid 1 \leq i \leq I-1, 1 \leq j \leq J-1\}$$

There are two well-known sets of parameters for a contingency table: the ordinary log-linear parameters and the multivariate logistic parameters (Glonek and McCullagh, 1995). The multivariate logistic parameters are marginal parameters: for every subset of k variables they can be viewed as measuring k-dimensional marginal "association." The ordinary log-linear parameters are conditional parameters: for every subset of k variables they can be viewed as measuring k-dimensional "association" *conditionally on the remaining variables.* The ordinary log-linear parameters have all the variables in the superscript (i.e. the superscript is maximal), for the multivariate logistic parameters the superscript equals the subscript (i.e. the superscript is minimal). For 3-way tables ABC they are:

$$\begin{aligned} &(\lambda^{ABC}_{\emptyset}, \lambda^{ABC}_{A}, \lambda^{ABC}_{B}, \lambda^{ABC}_{C}, \lambda^{ABC}_{AB}, \lambda^{ABC}_{BC}, \lambda^{ABC}_{AC}, \lambda^{ABC}_{ABC}) \\ &(\lambda^{\emptyset}_{\emptyset}, \lambda^{A}_{A}, \lambda^{B}_{B}, \lambda^{C}_{C}, \lambda^{AB}_{AB}, \lambda^{BC}_{BC}, \lambda^{AC}_{AC}, \lambda^{ABC}_{ABC}) \end{aligned}$$

respectively. These two types of parameters form the endpoints of a broad range of different types of parameters. One can replace the superscript of any of the parameters by something in between the subscript and the complete set of variables. For example, replacing λ_{AC}^{AC} in the multivariate logistic transform by λ_{AC}^{ABC} yields the parameter

$$(\lambda_{\emptyset}^{\emptyset}, \lambda_{A}^{A}, \lambda_{B}^{B}, \lambda_{C}^{C}, \lambda_{AB}^{AB}, \lambda_{BC}^{BC}, \lambda_{AC}^{ABC}, \lambda_{ABC}^{ABC}) \tag{2.1}$$

2.2 Log-affine marginal models

A log-linear marginal model is obtained by imposing linear restrictions on marginal log-linear parameters. A log-affine marginal model is obtained by imposing affine restrictions on marginal log-linear parameters. Affine restrictions are linear restrictions with a constant added. Such restrictions occur, for example, when a marginal log-linear parameter is given a specific non-zero value.

To illustrate the utility of marginal log-linear models, consider an experiment designed to measure a treatment effect. The subjects are divided into two groups: an experimental and a control group. Let $X = 1$ stand for subjects in the former, and $X = 2$ for subjects in the latter group. A categorical variable A is measured at time point one, then the experimental group receives a treatment, after which the variable B is measured on both groups. This design is used in many fields of study, such as medicine or psychology. A first hypothesis which may be tested is whether the experimental group has changed over time. With A_e and B_e denoting variables A and B for the experimental group, this hypothesis has the form

$$\lambda_{A_e}^{A_e}(i) = \lambda_{B_e}^{B_e}(i)$$

for all i. This hypothesis cannot be tested using standard log-linear techniques, because the outcomes at the two points in time are correlated: they involve the same subjects.

A more sophisticated hypothesis is whether the difference between the experimental and control groups, which is the inhomogeneity of their distributions in the present setup, is the same before and after the treatment. This hypothesis is the equality of marginal odds ratios:

$$\lambda_{XA}^{XA}(t, i) = \lambda_{XB}^{XB}(t, i)$$

for all i and t.

In the next section, it is shown that certain marginal hypotheses may lead to unexpected problems.

3 Potential problems with marginal models

In Section 3.1, seemingly harmless constraints imposed on marginal tables are shown to be contradictory, i.e. there is no joint distribution satisfying the marginal constraints. In Section 3.2, it is shown that certain constraints on marginals are redundant, i.e. implied by others. In Section 3.3, it is shown that certain marginal constraints lead to models for which standard asymptotic theory does not necessarily apply. A general method for determining when these problems can occur is lacking.

3.1 Infeasible marginal constraints

Consider a $2 \times 2 \times 2$ contingency table ABC. We impose the restrictions on the one-dimensional marginals A, B, and C that they are symmetric, i.e. $P(A=1) = P(A=2)$, $P(B=1) = P(B=2)$, and $P(C=1) = P(C=2)$. In terms of restrictions on marginal log-linear parameters, this is

$$\lambda_A^A(1) = \lambda_B^B(1) = \lambda_C^C(1) = 0 \tag{3.1}$$

We may further hypothesize that the odds ratios in tables AB, BC, and AC are 5, 5, and 1/5, respectively. In terms of restrictions on marginal log-linear parameters, this is

$$\lambda_{AB}^{AB}(1,1) = \lambda_{BC}^{BC}(1,1) = -\lambda_{AC}^{AC}(1,1) = \tfrac{1}{4}\log 5 \tag{3.2}$$

It is easy to verify that the approximate probability distributions of the two-dimensional marginals are

	B	
A	0.40	0.10
	0.10	0.40

	C	
B	0.40	0.10
	0.10	0.40

	C	
A	0.10	0.40
	0.40	0.10

It can be verified that there exists no joint distribution for table ABC with these two-dimensional marginals; for example, the product moment correlation matrix is not positive definite. In general, however, this positive definiteness is not a sufficient condition, and generally a linear programming problem has to be solved to verify existence of a joint distribution.

The example can have practical use when different censuses provide the bivariate marginals. If these are incompatible, the censuses must come from different populations.

3.2 Redundancy of constraints

The *marginal homogeneity* (MH) model for a table AB is used to test for equality of the marginal distributions A and B. In terms of marginal log-linear parameters, it is specified as

$$\lambda_A^A(i) = \lambda_B^B(i) \tag{3.3}$$

The *symmetry* (S) model is specified as

$$\lambda_{AB}^{AB}(i,j) = \lambda_{AB}^{AB}(j,i) \tag{3.4}$$

$$\lambda_{A}^{AB}(i) = \lambda_{B}^{AB}(i) \tag{3.5}$$

It is easy to verify that S implies MH. However, equations (3.3) and (3.5) are only equivalent if (3.4) holds, and this is not easy to see[2]. For example, consider the table (with marginals included)

B

	1	2	3
A	3	4	7
	4	6	10

For this asymmetric table, the one-variable effect parameters are different when the effect pertains to the one-dimensional marginals or to the whole table. In particular, the marginal effects are

$$\lambda_A^A(1) = \tfrac{1}{2}\log(3/7) \qquad \lambda_B^B(1) = \tfrac{1}{2}\log(2/3)$$

while the effects pertaining to the complete table are

$$\lambda_A^{AB}(1) = \log(3/8) \qquad \lambda_B^{AB}(1) = \log(1/6)$$

A general method for checking whether or not there are redundant constraints, and if so which ones, is lacking.

3.3 Failure of standard asymptotic theory

Consider the $2 \times 2 \times 2$ table ABC. Marginal independence of A and B is specified as

$$\lambda_{AB}^{AB}(1,1) = 0 \tag{3.6}$$

Conditional independence of A and B given C is specified as

$$\lambda_{ABC}^{ABC}(1,1,1) = \lambda_{AB}^{ABC}(1,1) = 0 \tag{3.7}$$

Dawid (1980) showed that the simultaneous model (3.6) and (3.7) is equivalent to A being independent of both B and C, or B being independent of both A and C (or both). In terms of prescriptions for log-linear parameters, this is

$$\left(\lambda_{ABC}^{ABC}(1,1,1) = \lambda_{AB}^{ABC}(1,1) = \lambda_{AC}^{ABC}(1,1) = 0\right) \tag{3.8}$$

or

$$\left(\lambda_{ABC}^{ABC}(1,1,1) = \lambda_{AB}^{ABC}(1,1) = \lambda_{BC}^{ABC}(1,1) = 0\right) \tag{3.9}$$

[2]Note that this corrects a condition given by Bishop, Fienberg, and Holland (1975) - end of Section 8.2.1. The verification is by straightforward algebra.

In fact, if either (3.8) or (3.9) holds, but not both, standard asymptotic theory can be applied. However, if both models hold simultaneously, i.e. if A, B, and C are mutually independent, it cannot. This happens because in this case for any sample size there is positive probability of two local maxima, namely on both (3.8) and (3.9), and the likelihood of each maximum goes to a fixed value greater than zero as the sample size goes to infinity. In contrast for large classes of models commonly used in practice, if the model holds the likelihood of all but one of the local maxima goes to zero as the sample size goes to infinity. It follows that in the present case the likelihood ratio statistic does not have a large sample chi-square distribution, but rather is distributed as the minimum of two chi-square distributions.

Of particular importance is that the conditional likelihood ratio statistic for testing mutual independence of A, B, and C against the alternative that A and B are both marginally independent and conditionally independent given C does not have a large sample chi-square distribution if the mutual independence model is true. Rather, it is distributed as the maximum of two chi-square distributions.

4 Basic properties of marginal log-linear parameters

In Sections 4.1 and 4.2, the two key concepts for understanding marginal log-linear parameters are discussed: smoothness and variation independence, respectively. It is shown how these properties facilitate the interpretation of parameters.

4.1 Smoothness of marginal log-linear parameters

A parameter is smooth if it is twice continuously differentiable with full rank Jacobian. Knowledge that a parameter is smooth facilitates its interpretation. For example, for a dichotomous variable A for which $P(A = 1) = \pi$ and $P(A = B) = 1-\pi$ (for some $0 < \pi < 1$), consider the parameters

$$\alpha = \pi(1-\pi) \qquad \beta = \frac{\pi}{1-\pi} \tag{4.1}$$

Then α is the variance of A and β is the odds of being in category 1 of A rather than in category 2 of A. Differentiation with respect to π yields the Jacobians

$$\frac{\partial\alpha}{\partial\pi} = 1 - 2\pi \qquad \frac{\partial\beta}{\partial\pi} = \frac{1}{(1-\pi)^2} \tag{4.2}$$

Note that the Jacobian of α is singular (i.e. zero in the present case) when $\pi = \frac{1}{2}$, and therefore α is not smooth. On the other hand, the Jacobian of β is nonsingular (strictly positive for all $0 < \pi < 1$ in the present case) and continuously differentiable, and is therefore smooth. For the interpretation of a given value of the variance of A, the fact that it is not smooth has to be taken into account. For example, suppose $\alpha = 0.21$. Then we find that either $\pi = 0.3$ or $\pi = 0.7$. On the other hand, prescribing a value to β always yields one value for π. Hence, the interpretation of β is more straightforward than the interpretation of α.

Generally, smoothness is guaranteed for a broad class of marginal log-linear parameters:

Theorem 1 *A marginal log-linear parameter is smooth if there is an ordering $\lambda_{\mathcal{L}_1}^{\mathcal{M}_1}, \ldots, \lambda_{\mathcal{L}_s}^{\mathcal{M}_s}$ of its components such that*

$$\mathcal{L}_i \subseteq \mathcal{M}_j \quad \Leftrightarrow \quad \mathcal{M}_i \subseteq \mathcal{M}_j \tag{4.3}$$

$$\mathcal{L}_i = \mathcal{L}_j \quad \Leftrightarrow \quad i = j \tag{4.4}$$

As examples, ordinary log-linear parameters and multivariate logistic transform parameters are smooth. Mixtures of these parameters are smooth if the log-linear parameters are taken from higher marginals, and the multivariate logistic parameters are taken from lower marginals. However, if there are components with the same subscript but different superscripts, a marginal log-linear parameter is not smooth:

Theorem 2 *A marginal log-linear parameter is not smooth if it has components $\lambda_{\mathcal{L}}^{\mathcal{M}_1}$ and $\lambda_{\mathcal{L}}^{\mathcal{M}_2}$ with $\mathcal{M}_1 \neq \mathcal{M}_2$.*

An example of a parameter which is not smooth is $(\lambda_{AB}^{ABC}, \lambda_{AB}^{AB})$, because the subscript AB appears twice. In most practical cases, smoothness holds if no subscript appears twice with different superscripts.

4.2 Variation independence of marginal log-linear parameters

Besides smoothness, a second useful property of parameters is variation independence. A (multidimensional) parameter is variation independent if its range is the Cartesian product of the ranges of its coordinates.

Variation independence is of major importance in the interpretation of parameters. This can be illustrated as follows. Consider again the example of Subsection 2.2 and suppose the marginal tables AX and BX are (in percentages)

	X		
	10	5	15
A	40	45	85
	50	50	100

	X		
	30	20	50
B	20	30	50
	50	50	100

Let ε_A be the proportion of subjects for which $A = 1$ who were going to receive treatment minus the proportion of subjects for which $A = 1$ who were not going to receive treatment. Define ε_B for B as a comparison of those who have and those who have not received treatment. Then

$$\varepsilon_A = \tfrac{1}{10} \qquad \varepsilon_B = \tfrac{1}{5}$$

The fact that ε_B is twice as large as ε_A suggests that the treatment had a large effect on subjects. However, given the one-dimensional marginals, the maximum value of ε_A is 0.3 and the maximum value of ε_B is 1, i.e. the actual values are $\frac{1}{3}$ and $\frac{1}{5}$ the maximum values, respectively. In fact, the difference in the values of ε_A and ε_B can be attributed to the differences in the marginal distributions of A and B. This can be seen from the value of the odds ratio, which equals 9/4 in both tables. The odds ratio, or a function of it, is the only measure which is variation independent of the marginals. Since ε_A and ε_B are not variation independent of the A and B marginals, respectivelyx, one should take care in comparing their values in tables AX and BX if A and B have different marginal distributions. As far as can be seen from the above tables, and disregarding the difference in the A and B marginals, the treatment had no effect.

As is clear from the above, it is important to establish variation independence of parameters of interest. The ordinary log-linear parameters are well-known to be variation independent, but when log-linear parameters are taken from various marginal distributions, the situation has not been clarified in the literature. For this reason, a solution to the problem is given below.

In the study of variation independence of marginal log-linear parameters, the concept of *decomposability* plays a central role. A class of incomparable finite sets $\{\mathcal{M}_1, \ldots, \mathcal{M}_s\}$ is called decomposable if it has at most two elements or if there is an ordering $\mathcal{M}_1, \ldots, \mathcal{M}_s$ of its elements such that, for $k = 3, \ldots, s$, there exists a $j_k < k$ such that

$$\left(\cup_{i=1}^{k-1} \mathcal{M}_i\right) \cap \mathcal{M}_k = \mathcal{M}_{j_k} \cap \mathcal{M}_k$$

(Haberman, 1974). For example, the set $\{AB, BC\}$ (where, for example, AB is short for the set containing A and B) is decomposable, but the set $\{AB, BC, AC\}$ is not.

A class of arbitrary finite sets is *ordered decomposable* if it has at most two elements or if there is an ordering $\mathcal{M}_1, \ldots, \mathcal{M}_s$ of its elements such that $\mathcal{M}_i \not\subseteq \mathcal{M}_j$ if $i > j$, and, for $k = 3, \ldots, s$, the maximal elements of $\{\mathcal{M}_1, \ldots, \mathcal{M}_k\}$ form a decomposable set. The ordering $\mathcal{M}_1, \ldots, \mathcal{M}_s$ is then also called ordered decomposable. Note that ordered decomposability is a generalization of decomposability to apply to incomparable subsets as well. For comparable subsets, the two concepts are identical. A sufficient (but not necessary) condition for a set of subsets to be ordered decomposable is that all subsets containing only comparable sets are decomposable. For example, the set $\{AB, BC, ABC\}$ is ordered decomposable, but the set $\{AB, BC, AC, ABC\}$ is not.

Ordered decomposability is useful for establishing variation independence of marginal log-linear parameters:

Theorem 3 *A marginal log-linear parameter is variation independent if there is an ordering* $\lambda_{\mathcal{L}_1}^{\mathcal{M}_1}, \ldots, \lambda_{\mathcal{L}_s}^{\mathcal{M}_s}$ *of its components such that, in addition to (4.3) and (4.4),* $\mathcal{M}_1, \ldots, \mathcal{M}_s$ *is an ordered decomposable ordering.*

As mentioned above, the ordinary log-linear parameters are variation independent and the conditions of the theorem hold. However, the multivariate logistic transform parameters are not variation independent if the number of variables exceeds two. The parameter (2.1) is variation independent. In fact, for three variables, it can be shown that a parameter satisfying the conditions of Theorem 1 is variation independent if and only if it does not contain the components

$$(\lambda_{AB}^{AB}, \lambda_{BC}^{BC}, \lambda_{AC}^{AC})$$

In summary, if a parameter is smooth and, additionally, the marginals to which the parameter pertains form an ordered decomposable set, the parameter is variation independent.

5 Feasibility of constraints on marginal frequencies and applicability of standard asymptotic theory

In the previous section, the basic properties of marginal log-linear parameters were discussed. The properties of marginal models can to a large extent be understood by understanding marginal log-linear parameters. Below, a useful methodology for checking whether models exist and whether standard asymptotic theory can be applied to maximum likelihood estimates is presented.

Linear restrictions on marginal log-linear parameters are always feasible as it can be shown that the uniform distribution satisfies them. However, the determination of the dimension of such a model may not be simple; for example, the uniform distribution may be the only one satisfying the model, yielding a dimension of zero. The question of the feasibility of affine restrictions is a very difficult one. However, a sufficient condition can be given: affine restrictions are feasible if the parameter to be restricted is variation independent. This yields the following theorem.

Theorem 4 *If a marginal log-linear parameter* λ *is variation independent, then any affine restrictions are feasible.*

In Section 3.1 an example was given where parameters which are not variation independent are restricted, yielding an infeasible model. A different problem is to determine whether restrictions on parameters which are not variation independent are feasible.

If the feasibility question has been resolved, the model may be tested by drawing a sample from the population. It is convenient if standard asymptotic theory can be applied:

Theorem 5 *Suppose a marginal log-linear parameter λ is smooth, and a Poisson or multinomial sampling scheme is used. Then under feasible affine restrictions,*

1. *The probability that $\hat{\mu}$ exists uniquely tends to one as the sample size goes to infinity.*
2. *If $\hat{\mu}$ exists, it is a stationary point of the likelihood in the model*
3. *With sample size N, $N^{-\frac{1}{2}}(\hat{\mu} - \mu)$ has an asymptotic multivariate normal distribution with mean zero.*
4. *The likelihood ratio statistic has an asymptotic chi square distribution with* $\dim(\mathcal{H}) - \dim(\lambda)$ *degrees of freedom.*

In Section 3.2, the parameters which are restricted are not smooth. Even though in this case standard asymptotic theory can be applied, the number of degrees of freedom of the likelihood ratio statistic is not equal to the number of constraints, since the marginal homogeneity constraints are redundant. In Section 3.3, an example was given of restrictions on a parameter which is not smooth, and leading to a model for which standard asymptotic theory does not apply for all parameter values.

Finally, the parameters restricted in Section 2.2 are both smooth and variation independent, and hence the model is guaranteed to be feasible, while standard asymptotic theory can be applied to maximum likelihood estimates.

References

[1] Becker, M. P. (1994): Analysis of repeated categorical measurements using models for marginal distributions: an application to trends in attitudes on legalized abortion. In P.V. Marsden (Ed.): *Sociological Methodology*, **24**, 229-265. Oxford, UK: Blackwell.

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

[3] Bergsma, W.P. and Rudas, T. (2002): Marginal models for categorical data. *Annals of Statistics*, **30**, 140-159.

[4] Bishop, Y. V.V., Fienberg, S.E., and Holland, P.W. (1975): *Discrete Multivariate Analysis.* Cambridge, MA: MIT Press.

[5] Dawid, A.P. (1980): Conditional independence for statistical operations. *Ann. Stat.*, **8**, 598-617.

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

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

[8] Lang, J.B. and Agresti, A. (1994) Simultaneously modelling the joint and marginal distributions of multivariate categorical responses. *J. Am. Stat. Ass.*, **89**, 625-632.

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

[10] McCullagh, P. and Nelder, J.A. (1989): *Generalized Linear Models.* London: Chapman and Hall.