A Strategy for Designing Telescoping Models for Analyzing Multiway Contingency Tables Using Mixed Parameters

EDWARD H. IP
University of Southern California
YUCHUNG J. WANG
Rutgers University

In the analysis of cross-classified data, sociologists often focus on building flexible models for the marginal distributions of a selected set of variables. One strategy for achieving flexible modeling is to design models for telescoping marginal distributions. As an illustration of telescoping distributions, consider a joint distribution of four cross-classifying variables: occupational attainment, education level, race, and gender. A set of telescoping distributions would be the univariate occupation distribution, the bivariate occupation by education, the trivariate occupation by education by race, and the entire joint distribution. A methodology that enables the telescoping modeling strategy is mixed parameterization, which has its roots in the statistics and sociology literatures. In this article, the authors develop a scheme of multilevel mixed parameterization that can be applied to a hierarchy of marginal distributions of reducing dimension. An example from the General Social Survey illustrates mixed parameters for telescoping models.

Keywords: *multilevel partition; marginal models; association-marginal models; log-linear models.*

1. INTRODUCTION

While the log-linear model remains one of the most useful tools for analyzing cross-classified data in sociological research, the demand for alternative tools is growing. This need to expand and supplement conventional log-linear models is understandable. Sociological hypotheses often focus on the association among a smaller set of variables, but the parameters of a log-linear model are always related to conditional associations involving all of the classification variables. In other words, as pointed out by Becker, Minick, and Yang (1998),

SOCIOLOGICAL METHODS & RESEARCH, Vol. 31, No. 3, February 2003 291-324
DOI: 10.1177/0049124102239078
© 2003 Sage Publications

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

when specific marginalized data tables and corresponding models are points of interest, log-linear models are not ideally suited for addressing substantive questions. These authors also observed that "a better approach is the simultaneous modeling of associations and univariate distributions."

The purpose of this article is to describe a framework for extending the notion of modeling associations and univariate distributions simultaneously. The framework is based on the statistical theory of mixed parameterization (Barndorff-Nielsen and Cox 1994) and related but independent work in sociology (Rudas 1998). In its simplest form, mixed parameterization decomposes a two-way contingency table into an association component of local odds ratios and a mean component of univariate marginal distributions. For example, a probability distribution for a table of counts of occupation attainment by education level decomposes into the individual marginal distribution of occupation attainment and education level, plus the odds ratios between the two variables at various levels. This makes possible explicit modeling of associations and univariate marginal distributions.

To broaden simultaneous modeling beyond the scope of associations and univariate marginal distributions, we propose a framework of mixed parameterization to enable a strategy for building a collection of "telescoping" submodels and models. Under this strategy, a researcher can "zoom in" to a selected set of variables, fit a model; "zoom out" to a larger set of selected variables, fit a second model; and continue telescoping until a joint model is fitted to the entire cross-classified table. This strategy is particularly meaningful when the researcher wants to preserve (or closely approximate) structures of some tables of reduced dimensions while maintaining flexibility in fitting parsimonious yet adequate models to the entire joint distribution.

In the sociological literature, two related approaches have been proposed to augment the log-linear model: the association-marginal (AM) models and the marginal models. Lang and Agresti (1994) and Lang and Eliason (1997) proposed using the AM model to fit the entire joint table by a log-linear (conditional association) model while treating marginal tables as constraints on the joint model. On the other hand, Becker (1994) proposed the marginal model, which is a purely

Request Permissions / Order Reprints
powered by RIGHTSLINK

unconditional approach. The marginal model uses marginal associations and logits to specify the joint distribution. Unlike the log-linear model, in which marginal models are implied by constraints on the log-linear representation, both the AM and the marginal approaches directly model the marginal distributions.

In this article, we present an alternative approach: the mixed parameterization of joint distribution and selected marginal distributions to facilitate flexible modeling. Suppose a table of cross-classified cell counts is available. The parameterization (designing a system of equations to describe a space of probability distributions, as described in Becker et al. 1998:512) and the subsequent modeling activities (placing various constraints on the space of probability distributions) can be summarized into four procedures. First, a parameterization procedure embeds the multinomial distribution for cross-classified data into a regular exponential family and transforms cell probabilities to canonical parameters. Second, a partition procedure specifies the joint distribution in terms of a hybrid of canonical and mean parameters. Third, an iterative procedure repeats the first two steps to further decompose distributions of reduced dimensions into hybrid representations. This procedure exploits the fact that mean parameters typically correspond to low dimensional marginal distributions. Finally, a model-fitting procedure infers models at each level of partition, starting with the tables of lowest dimension. The outcomes of the four-step process are a set of telescoping marginal models of selected order and the fitted joint density.

The remainder of this article is organized as follows. In Section 2, we set up notations for the exponential family and mixed parameterization. The embedding of multinomials into the exponential family is discussed in Section 3, in which we also describe a difference operator and show how to use it to systematically represent the potentially large number of parameters. In Section 4, we highlight how multilevel mixed parameterization can be employed to create commonly used sociological models such as the uniform association model (Goodman 1979), marginal homogeneity models, and marginal shift models. We discuss one-, two-, three-, and four-dimensional marginal modeling in this section. To illustrate the multilevel approach, Section 5 provides an analysis of a sample of the General Social Survey data. Section 6 analyzes the connection between mixed parameterization,

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

the log-linear model, the AM model, and the marginal model. The computational algorithm is outlined in Section 7. Finally, we provide some concluding remarks in Section 8.

2. *EXPONENTIAL FAMILY AND MIXED PARAMETERIZATION*

Let $y = (y_1, \ldots, y_q)$ be a q-dimensional random variable. Suppose that there are q unknown parameters, denoted by $\theta = (\theta_1, \ldots, \theta_q)$, and q statistics, denoted by $S = (S_1, \ldots, S_q)$, such that the density of y is given by

$$g(y; \theta) = f(S, \theta) = h(S) \exp\{S\theta^T - K(\theta)\}. \tag{1}$$

The values of θs for which there is a density of form (1) are called the canonical parameters. Assume the ranges of variation for both S and θ are nondegenerate, that is, open subsets of R^q. The above family is then called the (q, q) exponential family. The term $K(\theta)$ is the cumulant-generating function that generates the mean and covariance of S with the following equations (Lehmann 1991:29):

$$\mu_i = E_\theta(S_i) = \frac{\partial K(\theta)}{\partial \theta_i},$$

and

$$\sigma_{ij} = \text{cov}\,(S_i, S_j) = \frac{\partial^2 K(\theta)}{\partial \theta_i \partial \theta_j},$$

where $\mu = (\mu_1, \ldots, \mu_q) = E_\theta S$ is called the mean parameter. It can be proved that the correspondence between θ and μ is one-to-one and invertible; that is, for any valid value μ_0 of μ, there exists a unique θ_0 consistent with $\mu_0 = E_{\theta_0}(S)$ (Barndorff-Nielsen and Cox 1994:7).

Definition: Let θ and μ denote, respectively, the canonical and mean parameters of the distribution defined in (1), and let $\theta^{(1)}$ and $\theta^{(2)}$ be subsets of θ. If $\theta^{(1)} \cup \theta^{(2)} = \theta$ and $\theta^{(1)} \cap \theta^{(2)} = \phi$, then $\theta^{(1)}$and $\theta^{(2)}$ are said to be complementary. Moreover, if $(\mu^{(1)}, \mu^{(2)})$ is the corresponding partition of μ, then both $(\theta^{(1)}, \mu^{(2)})$ and $(\mu^{(1)}, \theta^{(2)})$ are called complementary mixed parameterizations for the distribution defined

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

in (1). Complementary mixed parameters such as $(\theta^{(1)}, \mu^{(2)})$ uniquely specify a distribution.

Barndorff-Nielsen and Cox (1994:64) proved two useful results for complementary mixed parameterization of the exponential family:

a. The canonical parameter and the mean parameters are variation independent, that is, the range of variation for $(\theta^{(1)}, \mu^{(2)})$ is the Cartesian product of the ranges of $\theta^{(1)}$ and $\mu^{(2)}$: $\Re(\theta^{(1)}, \mu^{(2)}) = \Re(\theta^{(1)}) \times \Re(\mu^{(2)})$, where $\Re$ denotes the range space.
b. The canonical parameter and the mean parameter are orthogonal (in the likelihood sense).

Variation independence implies that models for the canonical parameter θ^i impose no restriction on the model for mean parameter $\mu^{(2)}$. Orthogonality implies that the asymptotic inference based on profile likelihood (say, $\theta^{(1)}$ conditioned on $\mu^{(2)}$) converges to the asymptotic inference based on the full likelihood parameterized by $(\theta^{(1)}, \mu^{(2)})$ (Cox and Reid 1987).

3. *REPARAMETERIZATION OF THE MULTINOMIAL DISTRIBUTION*

To take advantage of the above separation properties, it is necessary to embed the multinomial distribution for an n-way table into the exponential family. Let q denote the number of cells in the n-way table, and let $\pi = (\pi_1, \ldots, \pi_q)$ and $n = (n_1, \ldots, n_q)$ denote, respectively, the row vectors of strictly positive expected cell probabilities and the observed cell frequencies. By convention, the cell probabilities are arranged in lexicographical order. That is, the first index changes fastest. For example, in a $2 \times 2 \times 2$ table, $\pi = (\pi_{111}, \pi_{211}, \pi_{121}, \pi_{221}, \pi_{112}, \pi_{212}, \pi_{122}, \pi_{222})$. The logarithm of the multinomial probability function is

$$\ell = b(n) + n \log \pi^T, \tag{2}$$

where $\log \pi = (\log \pi_1, \ldots, \log \pi_q)$. The q components of $\log \pi$ are not linearly independent because they must satisfy the constraint $\sum \pi_i = 1$. Also, no cumulant-generating function is shown in (2).

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

One way to express (2) (together with the constraint) in the form of (1) is to transform π_i into $\exp\{\theta_i\}/\sum_{i=1}^{q}\exp\{\theta_i\}$ for $1 \le i \le q$. Now, the canonical parameter θ has full range, and (2) is embedded into the exponential family as

$$l = b(n) + n\theta^T - \left(\sum_{l}^{q} n_i\right) \log\left(\sum_{i=1}^{q} \exp\{\theta_i\}\right). \tag{3}$$

Here, $K(\theta) = n_+ \log(\exp\{\theta_1\} + \cdots + \exp\{\theta_q\})$ is the cumulant-generating function, where$_+$ indicates summation over the subscript it replaces. It is easy to see that the mean parameter corresponding to θ_i in (3) is $\mu_i = E_\theta(n_i) = n_+\pi_i$.

Wang (1986) introduced a parameterization in which mean parameters correspond to a hierarchy of marginal totals. His idea was to select an incidence square matrix A in such a way that nA is equivalent to cumulative margins and $A^{-1}\log\pi^T$ becomes local odds ratios. For $2^2 = 2 \times 2$ and $2^3 = 2 \times 2 \times 2$ cross-classifications, the respective incidence matrices and inverses are as follows:

$$A_{2^2} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad A_{2^2}{}^{-1} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A_{2^3} = \begin{bmatrix} A_{2^2} & A_{2^2} \\ 0 & A_{2^2} \end{bmatrix} \quad \text{and} \quad A_{2^2}{}^{-1} = \begin{bmatrix} A_{2^2}{}^{-1} & -A_{2^2}{}^{-1} \\ 0 & A_{2^2}{}^{-1} \end{bmatrix}.$$

Rewrite $n\log\pi^T = n(AA^{-1})\log\pi^T$. The exponential family parameterization of (2) is

$$\begin{aligned} l &= b(n) + (nA)(A^{-1}\log\pi^T) \\ &= b(n) + S\theta^T - K(\theta), \end{aligned} \tag{4}$$

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

where $S = nA$, $\theta^T = A^{-1} \log \pi^T$, and

$$K(\theta) = n_+ \log\left(\sum_{i=1}^{q} \exp\left\{\sum_{j=1}^{q} a_{ij}\theta_j\right\}\right)$$

$$= n_+ \log\left(\sum_{i=1}^{q} \exp\{A_{(i)}\theta^T\}\right),$$

with $A_{(i)}$ being the i^{th} row of A. For the detailed derivation of (4) and $K(\theta)$, see Wang (1986).

3.1. ONE- AND TWO-WAY TABLES

The following two examples show how A matrices are created to generate desirable mixed parameters.

Example 3.1. When the q categories represent a one-dimensional ordered classification with cells arranged in ascending order, the incidence matrix A is an upper triangular matrix. That is,

$$A = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ & & \dots & \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

Its inverse is

$$A^{-1} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0\dots & 0 \\ & & \ddots & & \\ & 0's & & 1 & -1 \\ & & & & 1 \end{bmatrix}, \tag{6}$$

with 1 in the diagonal, -1 one step above diagonal, and zeros filling the rest of the matrix. The resulting canonical parameters are the consecutive odds

$$\theta_i = \log \frac{\pi_i}{\pi_{i+1}} \quad \text{for} \quad 1 \le i \le (q-1), \quad \theta_q = \log \pi_q, \tag{7}$$

Request Permissions / Order Reprints
powered by RIGHTSLINK

and

$$K(\theta) = n_+ \log \left(\sum_{i=1}^{q} \exp \left\{ \sum_{j=i}^{q} \theta_j \right\} \right).$$

The mean parameter corresponding to θ_i, $1 \leq i \leq q-1$ is the cumulative count:

$$\mu_i = E_\theta S_i = n_+(\pi_1 + \cdots + \pi_i) \quad \text{for} \quad 1 \leq i \leq (q-1).$$

The mean parameter corresponding to the last cell (the highest or last category) $\theta_q = \log \pi_q$ is $\mu_q = E(S_q) = E(n_+)$, which is the multinomial sampling total.

Example 3.2. This example illustrates both the construction of the A matrix for a two-way table and a simple case of mixed parameterization. For a two-dimensional $I \times J$ cross-classification, with $q = I \times J$, the incidence matrix A is a tensor product $A_1 \otimes A_2$, where A_1 and A_2 are, respectively, $I \times I$ and $J \times J$ upper triangular matrices that have the same form as (5), and $C \otimes B$ is defined as

$$\begin{pmatrix} b_{11}C & . & . & b_{1J}C \\ b_{21}C & . & . & b_{2J}C \\ & & & \\ b_{J1}C & . & . & b_{JJ}C \end{pmatrix},$$

$C = (c_{rs})$, $B = (b_{pq})$, $1 \leq r, s \leq I$; $1 \leq p, q \leq J$. For example,

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \; A_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \; A_1 \otimes A_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Wang (1986) showed that the set of canonical parameters $\theta^T = A^{-1} \log \pi^T$ is given by

$$\theta_{ij} = \log(\pi_{ij}\pi_{i+1j+1}/\pi_{i+1j}\pi_{ij+1}) \quad \text{for} \quad 1 \leq i \leq (I-1), 1 \leq j \leq (J-1), \tag{8}$$

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

$$\theta_{iJ} = \log(\pi_{iJ}/\pi_{i+1J}), \quad \text{for} \quad 1 \le i \le (I-1), \tag{9}$$

$$\theta_{Ij} = \log(\pi_{Ij}/\pi_{Ij+1}), \quad \text{for} \quad 1 \le j \le (J-1), \tag{10}$$

$$\theta_{IJ} = \log \pi_{IJ}, \quad \text{and} \tag{11}$$

$$K(\theta) = n_{++} \log \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \exp \left\{ \sum_{k=1}^{i} \sum_{l=1}^{j} \theta_{kl} \right\} \right). \tag{12}$$

The formulation of the joint density of cross-classifying variables as an exponential family allows the theory of mixed parameterization to be applied. Let m_{ij} denote the mean cell count $n_{++}\pi_{ij}$. Suppose that $\theta^{(1)}$ is the $(I-1)(J-1)$ vector of the logarithms of local odds ratios of (8) and that $\theta^{(2)}$ is the complementary parameter $(\theta_{Ij}, \theta_{iJ}, \theta_{IJ})$ for $l \le j \le (J-1)$ and $1 \le i \le (I-1)$. The mean parameter corresponding to $\theta^{(2)}$ then comprises univariate marginal cumulative cell counts and the sampling total:

$$\mu^{(2)} = \left(\sum_{k=1}^{j} m_{+k}, \sum_{k=1}^{i} m_{k+}, m_{++} \right). \tag{13}$$

The components $\theta^{(1)}$ of (8) and $\mu^{(2)}$ of (13) form the mixed parameters of the joint distribution of the $I \times J$ table in such a way that $(\theta^{(1)}, \mu^{(2)})$ uniquely specifies the two-way table. The mathematical proof for this specific result was first obtained by Sinkhorn (1967).

In most cases of data analysis, the expected total count m_{++} is constrained to be identical to the observed total count: $m_{++} = n_{++}$. The sampling constraint was called *samp(m)* by Lang and Agresti (1994). In general, the expected total count is the mean parameter corresponding to the last cell. To observe the sampling constraint, by convention we include the mean parameter of expected total count in $\mu^{(2)}$. Thus, $\mu^{(2)}$ is always nonempty.

3.2. n-*WAY TABLE AND DIFFERENCE OPERATOR*

We proceed to extend the result from Wang (1986). For an n-dimensional table, A is the tensor product of square matrices. Each of them is an upper-triangular matrix in the same form as (5). In general, it can be shown that $A = A_1 \otimes A_2 \otimes \cdots \otimes A_n$ implies $A^{-1} = A_1^{-1} \otimes A_2^{-1} \otimes \cdots \otimes A_n^{-1}$ (Zwillinger 1996:136). Each A_i^{-1} takes the form of (6). If the size of each A_i is q_i, then both A and A^{-1}

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

are square matrices of the size $q_1 \ldots . q_n$, which can be quite large. To systematically provide an interpretation of each parameter in the potentially long vector $A^{-1} \log \pi^T$, we introduce the following partial difference operator (Whittaker 1990:35):

$$\nabla_i g(x_1, \ldots x_i \ldots, x_n) = g(\ldots, x_i, \ldots) - g(\ldots, x_i + 1, \ldots). \tag{14}$$

In our case, the function of interest takes the form of $\log \pi_{ij \ldots l}$, and the difference operator applies to the indexing variables. As an example, $\nabla_1 \log \pi_{11}$ is the adjacent category logit $\log \pi_{11} - \log \pi_{21}$, and $\nabla_2 \log \pi_{11} = \log \pi_{11} - \log \pi_{12}$. Higher order mixed difference operators are recursively defined. For example,

$$\begin{aligned} \nabla_1 \nabla_2 \log \pi_{ij} &= \nabla_1 (\log \pi_{ij} - \log \pi_{ij+1}) \\ &= \log(\pi_{i+1j+1} \pi_{ij} / \pi_{i+1j} \pi_{ij+1}), 1 \leq i \leq I - 1, \\ &\quad 1 \leq j \leq J - 1. \end{aligned}$$

The representations using the matrix A^{-1} and the difference operators are equivalent. From a one-dimensional distribution (Example 3.1), we have

$$\begin{aligned} A^{-1} \log \pi^T &= (\log\{\pi_1/\pi_2\}, \ldots, \log\{\pi_{q-1}/\pi_q\}, \log \pi_q) T \\ &= (\nabla_1 \log \pi_1, \ldots, \nabla_1 \log \pi_{q-1}, \log \pi_q)^T. \end{aligned}$$

We can represent matrix A^{-1} in condensed form by a vector of the partial difference operators: $(\nabla, \ldots, \nabla, I)^T$, where I is the identity operator, that is, $Ig(x_1, \ldots, x_n) = g(x_1, \ldots, x_n)$. For $2^2 = 2 \times 2$ and $2^3 = 2 \times 2 \times 2$ cross-classifications, matrices $A^{-1}_{2^2}$ and $A^{-1}_{2^3}$ are equivalent to $(\nabla_1 \nabla_2, \nabla_2, \nabla_1, I)^T$ and $(\nabla_3 \nabla_2 \nabla_1, \nabla_3 \nabla_2, \nabla_3 \nabla_1, \nabla_2 \nabla_1, \nabla_3, \nabla_2, \nabla_1, I)^T$, respectively. These two vectors of operators can be expressed as $(\nabla_1, I)^T \otimes (\nabla_2, I)^T$ and $(\nabla_1, I)^T \otimes (\nabla_2, I)^T \otimes (\nabla_3, I)^T$, respectively. In general, for n-dimensional tables, the canonical parameter can be written in either the matrix or the difference operator representation:

$$\theta = A^{-1} \log \pi^T,$$

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

or

$$\theta = [(\nabla_1, \ldots, \nabla_1, I)^T \otimes (\nabla_2, \ldots, \nabla_2, I)^T \otimes \cdots\cdots \otimes (\nabla_n, \ldots, \nabla_n, I)^T] \log \pi^T.$$

The right-hand side of the last equation consists of all the logarithms of consecutive odds according to $\nabla_i I^{(n-1)}$, all of the logarithms of the local odds ratios of order 2 according to $\nabla_i \nabla_j I^{(n-2)}$, logarithms of ratios of odds ratios of order 3, and so on up to order n. All of the odds and odds ratios are measures of conditional dependence of various orders.

When it is necessary to generate *complete* marginal distributions from $\theta^{(2)}$, the partition of θ into $(\theta^{(1)}, \theta)^{(2)}$ must follow a hierarchy. Rudas (1998) described the hierarchy principle, which states that if a marginal distribution is known, then all of its derived lower dimensional marginals must also be known. Formally, let $N = \{1, \ldots, n\}$ denote the set of n categorical variables, a be a subset of N, and ∇_a be the mixed difference operator with regard to the indexing variables in a: $(\Pi_{i \in a} \nabla_i)$. For example, $N = \{1, 2, 3\}$, $a = \{1, 3\}$, $\nabla_a = \nabla_1 \nabla_3 I$. If $(\theta^{(1)}, \mu^{(2)})$ is a complementary mixed parameterization, then the set $\Delta = \{a, \nabla_a \in \theta^{(2)}\}$ must be closed with respect to inclusion that is, if $b \in \Delta$, then any subset of b must also be included in Δ. Rudas called the set Δ a descending system of subsets (see also Glonek 1996).

3.3 TELESCOPING MODELS

Mixed parameterization enables a strategy for designing telescoping models by the multilevel partitioning of the parameter space. As a generic example of multilevel partitioning, consider an n-way table in which $n \geq 2$. At the first level, mixed parameterization produces $(\theta^{(1)}, \mu^{(2)})$. Suppose $\mu^{(2)}$ corresponds to a complete set of mean parameters that specifies a marginal table of lower dimension. For the specific table of reduced dimension, we can then apply a second level of mixed parameterization to decompose it into yet another partition, say $(\xi^{(1)}, \nu^{(2)})$, of canonical component $\xi^{(1)}$ and mean component $\nu^{(1)}$, where $\nu^{(2)}$ corresponds to some still lower dimensional margins. Clearly, this process of partitioning and mixed

Request Permissions / Order Reprints
powered by RIGHTSLINK

parameterization can be generalized to more than two levels, resulting in a sequence of "telescoping" models and submodels.

The practical implication of the telescoping model for sociological research is its ability to broaden the scope of conventional methods for modeling associations and univariate marginal distribtutions (e.g., Plewis 1985) to include refined models at selected levels of data aggregation. For example, a researcher may have accumulated sufficient knowledge about some variables from past social surveys but may have only limited information on newly collected variables from a more recent survey. The telescoping scheme allows him or her to "zoom in" to the marginal distribution formed by a few existing variables. An accurate model (possibly a saturated one) can be fitted to the observed table of lower dimension. The joint model can then be fitted by parsimonious and adequate models that do not alter the previously fitted table of reduced dimension. Note that while multilevel partitioning starts with the entire joint distribution, the fitting process starts "inside-out," that is, from fitting tables of the lowest dimension all the way up to the entire joint distribution.

4. *MIXED PARAMETERIZATION FOR* n-*WAY TABLES*

In this section, we demonstrate multilevel partitions for two-way, three-way, and four-way tables with reference to some commonly used sociological models. To facilitate subsequent discussions, we differentiate distribution parameters from model parameters. For example, in the log-linear model $\log \pi^T = D\lambda^T$, where the model function D consists of ANOVA-type indicators, $\log \pi$ is the distribution parameter and λ is the model parameter. This distinction is important because interpretation of model parameters often depends on the definition of and the relationship between the two. For example, marginal models and AM models use different sets of distribution parameters. This implies that their model parameters have different interpretations, even if their model functions look similar.

4.1. TWO-WAY TABLE

Goodman's (1979) uniform association model (UAM), having one degree of freedom more than independence, has sometimes replaced

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

independence as the benchmark of association models in sociological studies. We shall use UAM to illustrate mixed parameterizations and to show how model complexity evolves. UAM is characterized by the following restriction imposed on the mixed parameters of

(1) the canonical component $\theta^{(1)}$:

$$\theta^{(1)} = (1, \ldots, 1)\theta^{RC}, \tag{15}$$

where $\theta^{(1)}$ is the row vector of local odds ratios (8), which corresponds to the $\nabla_1\nabla_2$ group of parameters under the difference operator representation, and

(2) the mean component $\mu^{(2)}$ of (13):

$$m_{i+} = n_{i+}, \tag{16}$$

$$m_{+j} = n_{+j}, \tag{17}$$

and $samp(m)$: $m_{++} = n_{++}$. Note that $\mu^{(2)}$ is equivalent to two univariate marginal distributions. One can further proceed to mixed parameterize each univariate distribution to obtain the following canonical components: $\xi^{(1)} = (\xi_R^{(1)}, \xi_C^{(1)})$; $\nu^{(2)} = (\nu_R^2, \nu_C^{(2)})$, where $\xi_R^{(1)} = \theta_i^R, \xi_C^{(1)} = \theta_j^C$ and mean component $\nu_R^{(2)} = \nu_C^{(2)} = m_{++}$. Here, $\theta_i^R = \nabla_1 \log \pi_{i+}$ and $\theta_j^C = \nabla_2 \log \pi_{+j}$. Equations (16) and (17) can now be expressed in terms of the canonical component:

$$\theta_i^R = \alpha_i \quad \text{and} \quad \theta_j^C = \beta_j. \tag{18}$$

This example illustrates a simple case of two-level partitioning of mixed parameters. When no confusion arises, we shall use $(\xi^{(1)}, \nu^{(2)})$ to represent the mixed components at the second level of mixed parameterization.

To compare various parameterizations, we list in Table 1 five possible ways to formulate UAM: Model 4 uses exclusively canonical parameters, Model 3 uses exclusively mean parameters, and Model 2 uses mixed parameters. Constant odds ratios can be read directly from Model 2 and Model 4 but are implicit in other models. For UAM, the M part for Model 5 is redundant because the log-linear model in the A part satisfies the constraints in M. Of course, this may not be true for other models. We also note that Model 4 requires the sampling constraint $samp(m)$: $n_{++} = m_{++}$, although this condition was not explicitly stated in the original article of Becker (1994:242).

An appealing feature of mixed parameterization for specifying models such as UAM is that unlike the λ parameter in log-linear

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

TABLE 1: Five Equivalent Modeling Strategies for Uniform Association

Model	*Reference*	*Equations*	*Remark*
1. Conventional log-linear	Clogg and Shihadeh (1994:21)	$log\ m_{ij} = \lambda + \lambda_i^R + \lambda_j^c + u_i v_i \lambda^{RC}$	Marginal model is implicit
2. Mixed parameters		$\theta^{(1)}:\ \theta_{ij} = \theta^{RC}$ $\mu^{(2)}$: $m_{i+} = \alpha_i, 1 \le i \le (I-1)$, $m_{+j} = \beta_j, 1 \le j \le (J-1)$, and $m_{++} = n_{++}$;	
3. All mean parameters	Clogg and Shihadeh (1994:31)	$\Sigma\Sigma u_i v_j m_{ij} = \Sigma\Sigma u_i v_j n_{ij}$, $m_{i+} = n_{i+}$ and $m_{+j} = n_{+j}$;	
4. All canonical parameters	Becker (1994:242)	$\theta_{ij} = \theta^{RC}$ and no restriction on θ_i^R and θ_j^C, i.e., $\theta_i^R = \alpha_i, 1 \le i \le (I-1)$ and $\theta_j^C = \beta_j, 1 \le j \le (J-1)$;	
5. Association-marginal (AM)	Lang and Agresti (1994)	A: $log\ m_{ij} = \lambda + \lambda_i^R + \lambda_j^C + u_i v_j \lambda^{RC}$ M: $log\ m_{i+} = a_i, 1 \le i \le (I-1)$, $log\ m_{+j} = b_j, 1 \le j \le (J-1)$ $samp(m)$: $m_{++} = n_{++}$, where $\{u_i\}$ and $\{v_j\}$ are equally space known scores.	M part redundant in this special case

models, each parameter has an interpretation that is not changed by how the values of other parameters are set (Becker et al. 1998). For example, in equation (15), the estimated value of $\exp\{\theta^{RC}\}$ is the uniform cross-product ratio of the fitted table, and this interpretation does not change, despite the fact that other parameters may be set to zero.

Using UAM as a starting point, one can proceed in two directions to search for better models: increase the number of model parameters for association and reduce the number of model parameters for the univariate marginal distributions. For example, Goodman's R, C, $R+C$, and RC models increase the number of parameters, and hence model complexity, for the canonical component. On the other hand, marginal homogeneity and marginal shift models reduce the number of parameters for the mean (marginal) component.

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

In a study of square mobility tables, in which the comparison of two marginal distributions (structural mobility) is of substantive interest, Lang and Eliason (1997) considered UAM for association, plus marginal homogeneity. That is, equations (16) and (17) were replaced by the following constraint for $1 \leq i \leq I$:

$$m_{i+} = m_{+i} = \frac{(n_{i+} + n_{+i})}{2}. \tag{19}$$

In terms of mixed parameters, model (19) is equivalent to setting $\theta_i^R = \theta_i^C$ in (18). Alternatively, the AM model specifies (18) with two separate equations (Lang and Eliason 1997:195):

$$\log m_{i+} = \beta + \beta_i + \eta_1, \tag{20}$$

$$\log m_{+j} = \beta + \beta_j + \eta_2. \tag{21}$$

An additional constraint, $\eta_1 = \eta_2$, is required to ensure that $\Sigma m_{i+} = \Sigma m_{+j}$.

Another important model that reduces complexity in the marginal distributions is the marginal shift model (Becker et al. 1998:520), which is specified by imposing the following shift constraint on (18):

$$\theta_i^R - \theta_i^C = c. \tag{22}$$

When the constant c is set to zero, (22) reduces to marginal homogeneity.

Implicit in the above discussion is the fact that searching for an optimal joint model can be accomplished in two separate steps: search for an optimal marginal model (e.g., marginal homogeneity, marginal shift) and search for an optimal association model (e.g., $R, C, R + C$). Indeed, the assumption is valid. Because of the orthogonality between canonical and mean parameters, a misspecified model of association has little impact on the search for a correct model for the marginal distribution, and vice versa (Liang and Zeger 1986; Fitzmaurice and Laird 1993). This is a potentially useful property of mixed parameterization.

4.2. THREE-WAY TABLES

Consider a $3 \times 3 \times 3$ table for the three classification variables X, Y, and Z. The incidence matrix is given by $A = A_1 \otimes A_2 \otimes A_3$, and A_i^{-1} is $(\nabla_i, \nabla_i, I)^T$, $i = 1, 2, 3$. The canonical parameters $\theta = A^{-1} \log \pi^T$ consist of

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

(a) the $\nabla_1\nabla_2\nabla_3$ group: $\{\nabla_1\nabla_2\nabla_3 \log \pi_{ijk}$ for $1 \le i, j, k \le 2\}$;
(b) the $\nabla_i\nabla_j I$ group: $\{\nabla_1\nabla_2 \log \pi_{ij3}, 1 \le i, j \le 2\}$, $\{\nabla_1\nabla_3 \log \pi_{i3k}, 1 \le i, k \le 2\}$, and $\{\nabla_2\nabla_3 \log \pi_{3jk}, 1 \le j, k \le 2\}$;
(c) the $\nabla_i I^2$ group: $\{\nabla_1 \log \pi_{i33}, 1 \le i \le 2\}$, $\{\nabla_2 \log \pi_{3j3}, 1 \le j \le 2\}$, and $\{\nabla_3 \log \pi_{33k}, 1 \le k \le 2\}$; and
(d) the reference cell I: $\log \pi_{333}$.

Except for $\log \pi_{333}$ in (d), the canonical parameters measure various orders of conditional associations. Group (a) measures three-way associations. For example, $\nabla_3\nabla_2\nabla_1 \log \pi_{111} = \log(\pi_{111}\pi_{221}/\pi_{211}\pi_{121}) - \log(\pi_{112}\pi_{222}/\pi_{212}\pi_{122})$ measures the three-way association among the eight cells surrounding the $(1, 1, 1)$ cell. It indicates the gap between the conditional log odds ratios $(X, Y|Z)$ at two specific values $(= 1, 2)$ of the layering variable Z. For example, in Section 5, we describe categorical data cross-classified by occupational attainment, educational attainment, and race (gender is a fourth cross-classifying variable, but is ignored here for brevity). The three-way interaction indicates the difference between the associations between occupation and education for two races, say, Black and White. Group (b) measures conditional associations. For example, $\nabla_3\nabla_2 \log \pi_{311}$ measures the Y-Z association among the four cells surrounding $(3, 1, 1)$, conditional on $X = 3$. In the occupation example, this group contains association terms such as log odds ratios between occupation and education conditional upon a specific race. On the other hand, the mean parameters map to unconditional probabilities. For example, the mean parameters that correspond to the three canonical parameters $\nabla_3\nabla_2\nabla_1 \log \pi_{111}$, $\nabla_3\nabla_2 \log \pi_{311}$ and $\nabla_3 \log \pi_{331}$ map to the probabilities $\pi_{111} = \Pr(X = 1, Y = 1, Z = 1)$, $\pi_{+11} = \Pr(Y = 1, Z = 1)$ and $\pi_{++1} = P(Z = 1)$, respectively.

4.2.1. One-Dimensional Marginal Modeling

We apply a multilevel partition to a three-way table and show how the approach leads to several commonly used sociological models. At the first level, we partition θ into

$$\theta^{(1)} = (\nabla_1\nabla_2\nabla_3, \nabla_1\nabla_2, \nabla_1\nabla_3, \nabla_2\nabla_3)$$
$$\theta^{(2)} = (\nabla_1, \nabla_2, \nabla_3, I).$$

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

The corresponding mean parameter of $\theta^{(2)}$ comprises necessary elements to generate the individual univariate distributions:

$$\mu^{(2)} = (m_{1++}, m_{1++} + m_{2++}, m_{+1+}, m_{+1+} + m_{+2+}, m_{++1}, m_{++1} + m_{++2}, m_{+++}).$$

A full-rank transformation on $\mu^{(2)}$ yields the following three one-way marginal distributions: $\{m_{i++}, 1 \leq i \leq 3\}$, $\{m_{+j+}, 1 \leq j \leq 3\}$, and $\{m_{++k}, 1 \leq k \leq 3\}$. In the occupation example, this set consists of the marginal distributions of occupation, education, and race.

A possible second-level mixed parameterization for $\mu^{(2)}$ is $\xi^{(1)} = (\nabla_1 \log m_{i++}, \nabla_2 \log m_{+i+}, \nabla_3 \log m_{++i}, 1 \leq i \leq 2)$ and $\nu^{(2)} = m_{+++}$, where $\nabla_1 \log m_{i++}$ are the adjacent-category conditional logits for the first variable X, and so on. This configuration can accommodate marginal models such as marginal shift, which is defined by $\nabla_1 \log m_{i++} - \nabla_2 \log m_{+i+} = \beta_1$ and $\nabla_2 \log m_{+i+} - \nabla_3 \log m_{++i} = \beta_2$ for $1 \leq i \leq 2$. On the other hand, models for associations in the joint distribution can be defined using the canonical component $\theta^{(1)}$. For example, the following equations characterize a completely homogeneous UAM:

$$\nabla_1\nabla_2 \log m_{ij3} = \alpha_1, \nabla_1\nabla_3 \log m_{i3k} = \alpha_2, \nabla_2\nabla_3 \log m_{3jk} = \alpha_3,$$

with $\alpha_1 = \alpha_2 = \alpha_3$, and $\nabla_1\nabla_2\nabla_3 \log m_{ijk} = 0, 1 \leq i, j, k \leq 2$. In the occupation example, this would imply that conditional associations are uniform across different levels of occupation and education, and so on. If conditional associations are heterogeneous across the conditioning variable, then we can set $\nabla_1\nabla_2\nabla_3 \log m_{ijk} = \gamma_k$, $1 \leq i, j, k \leq 2$. The assumption $\alpha_1 = \alpha_2 = \alpha_3$ can also be relaxed. For example, if the α s are not all necessarily equal, we call the model "dimensional UAM."

4.2.2. Two-Dimensional Marginal Modeling

By selecting a sufficiently large set of parameters for the mean component, multilevel partitioning can lead to more complex and interesting telescoping models. Suppose Z is a moderating variable and the interest is in studying the marginal association structure of

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

response variables X and Y. One possible partition of θ at the first level would be

$$\theta^{(1)} = (\nabla_1\nabla_2\nabla_3, \nabla_1\nabla_3, \nabla_2\nabla_3), \text{ and}$$
$$\theta^{(2)} = (\nabla_1\nabla_2, \nabla_1, \nabla_2, \nabla_3, I).$$

The mean parameters $\mu^{(2)}$ are equivalent to the two-way marginal distribution $\{m_{ij+}\}$ plus the univariate marginal $\{m_{++k}\}$. In the occupation example, this set could be the marginal distributions of occupation and education and of race.

A second-level mixed parameterization for the X-Y margin produces unconditional associations for the canonical component: $\xi^{(1)} = \nabla_1\nabla_2 \log m_{ij+}, i, j = 1, 2$, and univariate marginal distributions for the mean component: $\nu^{(2)} = \{m_{i++}, m_{+j+}, m_{++k}\}$. In the occupation example, the second-level mixed parameterization would produce the unconditional log odds between occupation and education.

Inference of parameters in the two-level partition can be used to address substantive questions of unconditional nature. For example, Goodman's R association model for the X-Y margin can be specified by $\nabla_1\nabla_2 \log \pi_{ij+} = \alpha_i, 1 \leq i, j \leq 2$. The null hypothesis of no moderating effect is specified by $\nabla_1\nabla_2\nabla_3 \log \pi_{ijk} = 0$. Finally, a third-level mixed parameterization on the univariate marginal distribution in $\nu^{(2)}$ (Section 4.2.1) produces logit models such as (18). As a result, the overall joint model consists of telescoping submodels of both one and two dimensions.

4.2.3. Marginal Independence and Conditional Independence

It is interesting that the two-level partitions for one- and two-dimensional marginal distributions can be used to illuminate the distinction between marginal independence and conditional independence. Conditional independence between X and Y given Z, or $X \perp Y | Z$, is specified by restrictions on $\theta^{(1)}$(Section 4.2.1):

$$\nabla_1\nabla_2\nabla_3 \log \pi_{ijk} = 0 \quad \text{and} \tag{23}$$
$$\nabla_1\nabla_2 \log \pi_{ij3} = 0, \tag{24}$$

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

whereas the marginal independence model for $X \perp Y$ is specified by restrictions on $\xi^{(1)}$ (Section 4.2.2):

$$\nabla_1 \nabla_2 \log m_{ij+} = 0.$$

See Whittaker (1990:36). Simultaneously specifying $X \perp Y | Z$ and $X \perp Y$ requires that the $\nabla_1 \nabla_2$ group of parameters be shared by both $\theta^{(1)}$ and $\theta^{(2)}$, which results in noncomplementary partitions. Thus, unless a table is collapsible along variable Z, simultaneously requiring $X \perp Y | Z$ and $X \perp Y$ may produce model conflict.

Similar situations occur for the multivariate normal distribution. Cox and Wermuth (1993) showed that conditional independence and marginal independence are defined using zero (x, y) entry in the concentration (canonical parameters) matrix and the covariance (mean parameters) matrix, respectively. For an arbitrarily given mean and variance structure, a normal distribution that simultaneously satisfies both types of independence may not exist.

Using multilevel mixed parameterization as a tool, we list in Table 2 several possible complementary independence models for three-way tables. Table 2 can be used to provide researchers in sociology with some guidelines for building parsimonious independence models.

4.3. FOUR-WAY TABLE

For n-way tables, where $n > 3$, multilevel partitions become more interesting, and more complex as well. For example, pairwise marginal association and pairwise association conditioned on one, two, and up to $(n-2)$ variables need to be distinguished. In a four-way cross-classification of (X_1, X_2, X_3, X_4), three possible associations can be defined between X_1 and X_2: (X_1, X_2) marginally, (X_1, X_2) conditioned on X_3 or X_4, and (X_1, X_2) conditioned on both X_3 and X_4.

To fix our notation, we let X_1, X_2, X_3 and X_4 have, respectively, I, J, K, and L categories, with $\{\pi_{ijkl}\}$ representing the joint distribution. The canonical parameter $\theta = A^{-1} \log \pi^T$ consists of the following five groups of parameters:

1. the $\nabla_1 \nabla_2 \nabla_3 \nabla_4$ group;
2. the $\nabla_i \nabla_j \nabla_k$ group, $1 \leq i \neq j \neq k \leq 4$;
3. the $\nabla_i \nabla_j$ group, $1 \leq i \neq j \leq 4$;

Request Permissions / Order Reprints
powered by RIGHTSLINK

TABLE 2: Five Complementary Partitions for Conditional Independence in $I \times J \times K$ Table

$\theta^{(1)}$	$\mu^{(2)}$	*Possible Conditionally Independent Models*	*Possible Marginal Models*
1) $\nabla_3\nabla_2\nabla_1$, $\nabla_1\nabla_2$, $\nabla_1\nabla_3$, $\nabla_2\nabla_3$	m_{i++}, m_{+j+}, m_{++k}	$X \perp Y \mid Z$, $X \perp Z \mid Y$, and $Y \perp Z \mid X$ ($\nabla_1\nabla_2\nabla_3 = 0$, $\nabla_1\nabla_2 = \nabla_1\nabla_3 = \nabla_2\nabla_3 = 0$)	M: logistic regression model, for univariate distribution
2) $\nabla_3\nabla_2\nabla_1$, $\nabla_2\nabla_3$, $\nabla_1\nabla_3$, ∇_3	m_{ij+}	$X \perp Z \mid Y$, ($\nabla_1\nabla_2\nabla_3 = \nabla_1\nabla_3 = 0$) $Y \perp Z \mid X$, ($\nabla_1\nabla_2\nabla_3 = \nabla_2\nabla_3 = 0$) and/or constant odds for Z given $X = I$, $Y = J$ ($\nabla_3 = const.$)	$X \perp Y$, M, association model for (X, Y)
3) $\nabla_3\nabla_2\nabla_1$, $\nabla_2\nabla_3$	m_{ij+}, m_{i+k}	$Y \perp Z \mid X$ $\nabla_3\nabla_2\nabla_1 = \nabla_2\nabla_3 = 0$)	$X \perp Y$, $X \perp Z$ plus M, association models for (X, Y) and (X, Z)
4) $\nabla_3\nabla_2\nabla_1$	m_{ij+}, m_{+jk}, m_{j+k}	Second-order Markov property ($\nabla_3\nabla_2\nabla_1 = 0$)	$X \perp Y$, $X \perp Z$, $Y \perp Z$ plus M, association models for (X, Y), (X, Z), (Y, Z)

4. the ∇_i group, $1 \leq i \leq 4$; and
5. the reference cell $\log \pi_{IJKL}$.

In the following subsection, we illustrate multilevel mixed parameterizations that focus on the association (X_1, X_2) conditioned on X_3.

4.3.1. *Modeling Association Between* X_1 *and* X_2 *Given* X_3

Consider the following hypothetical example of voter preference. Let X_1 and X_2 denote the polytomous responses (such as strongly agree, agree, disagree, and strongly disagree) on two issues on an election ballot, let X_3 denote different counties, and let X_4 denote party affiliation. Two questions of interest are the following: "Does county of residence affect the association pattern among responses, regardless of party affiliation?" and "What is the effect of a voter's county on each individual response?" These questions involve only three of the four variables, and so a log-linear model is not suitable.

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

One possible mixed parameterization designed to answer the above questions would be to use, at the first level, the partition that takes $\theta^{(1)}$ as $\{\nabla_1\nabla_2\nabla_3\nabla_4, \nabla_1\nabla_2\nabla_4, \nabla_1\nabla_3\nabla_4, \nabla_2\nabla_3\nabla_4, \nabla_1\nabla_4, \nabla_2\nabla_4, \nabla_3\nabla_4, \nabla_4\}$, and the complementary $\theta^{(2)}$ as $\{\nabla_1\nabla_2\nabla_3, \nabla_1\nabla_2, \nabla_1\nabla_3, \nabla_2\nabla_3, \nabla_1, \nabla_2, \nabla_3, \log \pi_{IJKL}\}$.

The corresponding $\mu^{(2)}$ is equivalent to

$$M = \{m_{ijk+} \quad \text{for} \quad 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K\}.$$

The (X_1, X_2, X_3) marginal distribution M needs to be further mixed parameterized into the following partition of the canonical parameter:

$$\xi^{(1)} = \{\nabla_1\nabla_2\nabla_3 \log m_{ijk+}, \nabla_1\nabla_2 \log m_{ijK+}\},$$

and the complementary mean parameter

$$\nu^{(2)} = \left\{ \sum_{a=1}^{i}\sum_{c=1}^{k} m_{a+c+}, \ \sum_{b=1}^{j}\sum_{c=1}^{k} m_{+bc+} \right\}.$$

From $\xi^{(1)}$, we can derive all $K(I-1)(J-1)$ pairwise associations between X_1 and X_2, given $X_3 = 1, 2, \ldots, K$. Because $\nabla_1\nabla_2\nabla_3 \log m_{ijk+}$ indicates the step from the conditional association at $X_3 = k$ to $X_3 = k-1$ (e.g., $\nabla_1\nabla_2 \log m_{ij(K-1)+} = \nabla_1\nabla_2\nabla_3 m_{ij(K-1)+} + \nabla_1\nabla_2 \log m_{ijK+}$), the association among responses is homogeneous across counties if and only if $\nabla_1\nabla_2\nabla_3 \log m_{ijk+} = 0$. Furthermore, one could consider a simpler model, such as uniform association among responses for individual counties, which would be given by $\nabla_1\nabla_2 \log m_{ijk+} = \alpha_k^{RC}$.

Further telescoping partitioning of the parameters beyond two levels is possible for four-way cross-classified data. For example, a possible third-level mixed parameterization is to decompose distributions in $\nu^{(2)}$ to answer the question, "What is the effect of X_3 on $X_1(X_2)$?" The solution can be based on inference of the first (second) components of $\nu^{(2)}$, using the models described in Sections 4.1 and 4.2.

5. EXAMPLE

Here is an example to illustrate the telescoping partitioning scheme. The data were originally collected from the 1972-1990 General Social

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

TABLE 3: Observed and Fitted Values of the Four-Way Table

	White		*Black*	
	Nonmanual	*Manual*	*Nonmanual*	*Manual*
Male				
College or above	590 (563.5)	66 (49.2)	46 (24.6)	10 (12.3)
High school or above	638 (700.6)	776 (710.5)	64 (96.0)	173 (177.0)
Less than high school	107 (70.2)	379 (417.7)	14 (3.2)	122 (160.2)
Female				
College or above	550 (577.0)	36 (44.6)	63 (84.2)	5 (11.1)
High school or above	1,372 (1,308.5)	568 (642.9)	632 (598.5)	172 (160.1)
Less than high school	203 (239.2)	416 (378.0)	25 (35.9)	183 (145.0)

NOTE: Numbers in parentheses indicate fitted values of the final model.

Survey (Davis and Smith 1990). We adopt a data set that was extracted by Wong (1995), and we aggregate several tables across time. Some categories are also collapsed. As a result, the data used in the analysis consist of a $2 \times 3 \times 2 \times 2$ table of counts of respondents cross-classified by occupational attainment (X_1, nonmanual and manual), educational attainment (X_2, college or greater, high school or greater, and less than high school), race (X_3, White and Black), and gender (X_4, male and female). One primary interest is in studying the relationship between occupational and educational attainments and how this relationship varies with race, gender, or both.

An initial glance at the data (see Table 3) and various odds ratios reveals that the associations between variables differ quite widely. For example, the odds ratio between occupational and educational attainment (at less than high school level) within Black females is 26.9, while within White females it is 4.9. Clearly, fitting a simple model such as independence to the entire distribution does not work well. More refined modeling efforts are required.

To preserve certain association structures that appear to be inherent within the data set, we design telescoping marginal models to provide increasingly accurate fits with reducing dimension. Suppose a researcher would like to fit a joint model in such a way that (1) the two-way table for occupational and educational attainments fits exactly or almost exactly with observed data, (2) the three-way table for occupation by education by race fits accurately with observed data, and (3) the joint distribution allows some degrees of freedom to

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

TABLE 4: Partitioning of Canonical Parameters at First Level of Mixed Parameterization

$\theta^{(1)}$ *for Canonical Component*			$\theta^{(2)}$ *for Mean Component*		
Difference Operator	*Log (cell)*	*Label for Parameter*	*Difference Operator*	*Log (cell)*	*Corresponding Margin*
$\nabla_1\nabla_2\nabla_3\nabla_4$	$\log\pi_{1111}$	β_{1234}	$\nabla_1\nabla_2\nabla_3$	$\log\pi_{1112}$	m_{ijk+}
$\nabla_2\nabla_3\nabla_4$	$\log\pi_{2111}$	β_{234}	$\nabla_2\nabla_3$	$\log\pi_{2112}$	
$\nabla_1\nabla_2\nabla_3\nabla_4$	$\log\pi_{1211}$	α_{1234}	$\nabla_1\nabla_2\nabla_3$	$\log\pi_{1212}$	
$\nabla_2\nabla_3\nabla_4$	$\log\pi_{2211}$	α_{234}	$\nabla_2\nabla_3$	$\log\pi_{2212}$	
$\nabla_1\nabla_3\nabla_4$	$\log\pi_{1311}$	β_{134}	$\nabla_1\nabla_3$	$\log\pi_{1312}$	
$\nabla_1\nabla_2\nabla_4$	$\log\pi_{1121}$	β_{124}	∇_3	$\log\pi_{2312}$	
$\nabla_3\nabla_4$	$\log\pi_{2311}$	β_{34}	$\nabla_1\nabla_2$	$\log\pi_{1122}$	
$\nabla_1\nabla_2\nabla_4$	$\log\pi_{1221}$	α_{124}	∇_2	$\log\pi_{2122}$	
$\nabla_1\nabla_4$	$\log\pi_{1321}$	β_{14}	$\nabla_1\nabla_2$	$\log\pi_{1222}$	
$\nabla_2\nabla_4$	$\log\pi_{2121}$	β_{24}	∇_2	$\log\pi_{2222}$	
$\nabla_2\nabla_4$	$\log\pi_{2221}$	α_{24}	∇_1	$\log\pi_{1322}$	
			∇_4	$\log\pi_{2321}$	m_{+++l}
			I	$\log\pi_{2322}$	m_{++++}

accommodate certain outstanding features in the entire joint distribution, given (1) and (2).

A three-level partition is designed to accomplish the above modeling objective. At the first level, we select the mean component $\mu^{(2)}$ so that its parameters completely specify the three-way distribution of occupation by education by race. Table 4 shows the partition $(\theta^{(1)}, \theta^{(2)})$. For the ease of description in later analysis, the parameters of the canonical component are labeled as β. Because educational attainment X_2 has three levels, we use α for associations of various orders at the less than high school level and β for associations at the high school or above level. A second-level partition further transforms the three-way distribution of occupation by education by race into a canonical component $\xi^{(1)}$ and a mean component $\nu^{(2)}$, the latter containing parameters necessary to specify the two-way marginal distribution of education by occupation, plus the univariate marginal distribution of race. Finally, a third-level mixed parameterization decomposes the education by occupation table within $\nu^{(2)}$ into log odds ratio and univariate distributions (Section 3.1).

The fitting procedure begins with the table of the lowest dimension, the two-way table of occupation by education. The independence model fits poorly ($G^2 = 1,468, 1df$) and is clearly not a viable

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

alternative. Because of the rather limited degrees of freedom for a 2×3 table, we proceed to consider three-way models for occupation by education by race that fit exactly with the two-way margin of occupation by education. For the three-way table, associations are indicated by (α^*, β^*), the definitions of which are analogous to their counterparts of the four-way table in Table 4. For example, α^*_{12} indicates the conditional association between occupation and education at the less-than-high-school level given that race is Black. The canonical component $\xi^{(1)}$ contains the elements ($\alpha^*_{123}, \beta^*_{123}, \alpha^*_{23}, \beta^*_{23}, \beta^*_{13}$), and the mean component $\nu^{(2)}$ corresponds to the two-way margin of occupation by education.

Table 5 shows the results of the analysis. Models 2 through 5 set a selected subset of $\xi^{(1)}$ to zero. Only competitive models with relatively good fits are reported. It can be seen that almost all models show rather large values of G^2, which could be a result of the fact that the three-way table contains relatively large counts. Furthermore, most models are not nested, making the use of G^2 for comparison inappropriate. Therefore, the Bayesian Information Criterion (BIC) statistic is also used to support the decision in model selection.

We also consider models that do not fit exactly with the two-way table of occupation by education. Specifically, we fit UAM to the three-way table (Models 6 through 8, Table 5). UAM models do not fit well and do especially poorly in fitting to the occupation by education 2×3 marginal table, suggesting that marginal association patterns do differ across various levels of educational attainment.

To assess whether the fit could be improved by using gender instead of race as the moderating variable, we repeat the above analysis for the three-way table based on occupation by education by gender. The several models that provide the best fits are shown in Table 5 (Models 9 through 12). The fits show a surprising similarity to the three-way analysis of occupation by education by race, suggesting that gender has a similar moderating effect on the association pattern between occupation and education. We also fit UAM models to the three-way table. The fits are poor and are not reported here.

Among the two classes of models, we select the conditional association Model IVA (Model 5, $G^2 = 4.4$, $1df$, BIC $= -8$), although a researcher may want to use other models for substantive reasons. This model fits exactly with the two-way margin of occupation by

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

TABLE 5: Telescoping Model Applied to the Three-Way Cross-Classification of Occupation by Education by Race (Models 1-8) and Occupation by Education by Gender (Models 9-12)

Model	*Equations*	df	G^2	*BIC*
1. Complete independence	$\alpha^*_{12} = \beta^*_{12} = \beta^*_{13} = \alpha^*_{23} = \beta^*_{23} = 0$ $\alpha^*_{123} = \beta^*_{123} = 0$	7	1,698	787
2. Conditional association IA: homogeneity across race	$\alpha^*_{123} = \beta^*_{123} = 0$	2	63	13
3. Conditional association IIA	$\alpha^*_{23} = \beta^*_{23} = 0$	2	28	–4
4. Conditional association IIIA	$\alpha^*_{23} = 0$	1	20	–8
5. Conditional association IVA	$\beta^*_{23} = 0$	1	4.4	–8
6. Completely homogeneous UAM	$\alpha^*_{12} = \beta^*_{12} = \beta^*_{13} = \alpha^*_{23} = \beta^*_{23}$ $\alpha^*_{123} = \beta^*_{123} = 0$	6	767	330
7. Heterogeneous UAM	$\alpha^*_{12} = \beta^*_{12} = \beta^*_{13} = \alpha^*_{23} = \beta^*_{23}$ $\alpha^*_{123} = \beta^*_{123}$	5	750	329
8. Dimensional UAM	$\alpha^*_{12} = \beta^*_{12}, \alpha^*_{23} = \beta^*_{23}$ $\alpha^*_{123} = \beta^*_{123} = 0$	4	144	37
9. Conditional association IB: homogeneity across gender	$\alpha^{**}_{24} = \beta^{**}_{24} = 0$	2	40	2
10. Conditional association IIB	$\alpha^{**}_{24} = \beta^{**}_{24}$	2	38	6
11. Conditional association IIIB	$\alpha^{**}_{24} = 0$	1	30	6
12. Conditional association IVB	$\beta^{**}_{24} = 0$	1	3.5	–7

NOTE: UAM = uniform association model; BIC = Bayesian Information Criterion.

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

TABLE 6: Table of Observed and Expected Counts for Occupation by Education by Race From Model IVA (Table 5)

	White		*Black*	
	Nonmanual	*Manual*	*Nonmanual*	*Manual*
College or above	1,140 (1,139.9)	102 (93.7)	109 (109.0)	15 (23.3)
High school or above	2,010 (2,010.0)	1,344 (1,352.4)	696 (696.0)	345 (336.9)
Less than high school	310 (310.0)	795 (795.2)	39 (39.1)	305 (305.0)

NOTE: Numbers in parentheses indicate fitted values.

TABLE 7: Parameter Estimates in Telescoping Models

Description of Parameter	*Parameter*	*Estimate*	*ASE*
Three-way table			
Interaction between occupation, education (less than high school), and race	α^*_{123}	–1.43	0.20
Occupation and education (less than high school) given race = Black	α^*_{12}	2.78	0.18
Occupation and race given education = less than high school	α^*_{13}	1.11	0.18
Education (less than high school) and race given occupation = manual	α^*_{23}	0.43	0.09
Occupation and education (high school or above) given race = Black	β^*_{12}	0.82	0.28
Interaction between occupation, education (high school and above), and race	β^*_{123}	1.29	0.30
Four-way table			
Interaction between occupation, education, and gender given race = Black	β_{124}	1.82	0.18
Interaction between occupation, race, and gender given education = less than high school	β_{134}	1.20	0.30
Race and gender given occupation (manual) and education = less than high school	β_{14}	–2.53	0.35

NOTE: ASE is asymptotic standard error of maximum likelihood estimate based on large-sample theory. See Agresti (1990:54). All estimates are on log scale (e.g., α^*_{12} is conditional log odds ratio between occupation and education).

education and rather accurately to the expanded three-way table (Table 6). Maximum likelihood estimates of this model are presented in Table 7.

Next, we proceed to fit to the entire four-way table a joint distribution under the constraint that it matches the previously fitted three-way table of occupation by education by race (see Table 6).

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

Our aim is to find a parsimonious and adequate representation for the joint distribution through inference of the canonical component $\theta^{(1)}$. Because the canonical parameters all have conditional interpretations, it might not be straightforward to design meaningful models for $\theta^{(1)}$ based on substantive grounds. To provide some structure to $\theta^{(1)}$, we impose the restriction that the two conditional associations across the various levels of education are uniform, that is, $\alpha_{1234} = \beta_{1234}, \alpha_{234} = \beta_{234}, \alpha_{124} = \beta_{124}, \alpha_{24} = \beta_{24}$. The results of the analysis are summarized in Table 8. We observe that three empirical associations, β_{1234}, β_{234}, and β_{24}, are close to zero. Accordingly, in the subsequent analysis, their values are set to zero to reduce the total number of models to be fitted. We report the results from only a sample of competitive models that fit relatively well in terms of BIC statistics. The best model is one that uses three additional degrees of freedom (Model 4, $G^2 = 146$, $9df$, BIC $= -7$). The model suggests that occupation and gender are conditionally dependent, given race and educational attainment levels. In other words, gender provides additional information about occupational attainment beyond that provided by race and education. Table 7 provides the parameter estimates, and Table 3 presents the fitted values of the joint table.

6. RELATIONSHIP WITH OTHER PARAMETERIZATIONS

6.1. RELATIONSHIP WITH LOG-LINEAR MODEL

Let $\log \pi = D\lambda$ be a saturated log-linear model, where D is a non-full-rank design matrix and λ is the interaction parameter. The log-linear model is usually overparameterized and thus additional identifiability constraints are required. One such constraint is to set the parameter that corresponds to the first category to zero. When all classifications are binary and a dummy coding scheme (set-first-to-zero) is applied to D, the matrix D reduces to A in (4). In other words, the log-linear model parameter λ is identical to its canonical counterpart θ. However, when any classification has more than two categories, θ in (4) is different from λ. For example, consider a polytomous two-dimensional log-linear model:

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}.$$

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

TABLE 8: Telescoping Model Applied to the Four-Way Cross-Classification of Occupation by Education by Race by Gender

Model	*Equations*	df	G^2	*BIC*
1. Conditional independence	All available parameters set to zero	12	696	241
2. Saturated canonical	All available parameters free	5	90	1
3. Conditional association I	$\beta_{1234} = \beta_{234} = \beta_{24} = 0$	8	131	–6
4. Conditional association II	$\beta_{1234} = \beta_{234} = \beta_{24} = 0$; $\beta_{34} = 0$	9	146	–7
5. Conditional association III	$\beta_{1234} = \beta_{234} = \beta_{34} = \beta_{124} = 0$	9	240	35
6. Conditional association IV	$\beta_{1234} = \beta_{234} = \beta_{24} = 0$; $\beta_{34} = \beta_{124} = 0$	10	255	39
7. Conditional association V	$\beta_{1234} = \beta_{234} = \beta_{24} = 0$; $\beta_{14} = \beta_{124} = 0$	10	420	121
8. Conditional association VI	$\beta_{1234} = \beta_{234} = \beta_{34} = \beta_{124} = 0$; $\beta_{134} = 0$	10	307	65
9. Conditional association VII	$\beta_{1234} = \beta_{234} = \beta_{34} = \beta_{124} = 0$; $\beta_{134} = \beta_{24} = 0$	11	449	126
10. Conditional association VIII	$\beta_{1234} = \beta_{234} = \beta_{34} = \beta_{124} = 0$; $\beta_{14} = \beta_{24} = 0$	11	674	240

NOTE: BIC = Bayesian Information Criterion. The constraints $\alpha_{1234} = \beta_{1234}, \alpha_{234} = \beta_{234}, \alpha_{124} = \beta_{124}, \alpha_{24} = \beta_{24}$ are present in each fitted model.

The relationships between λ and θ of (8), (9), and (10) are as follows:

$$\begin{aligned} \nabla_1(\lambda_i^A + \lambda_{iJ}^{AB}) &= \theta_{iJ}, \\ \nabla_2(\lambda_j^B + \lambda_{Ij}^{AB}) &= \theta_{Ij}, \quad \text{and} \\ \nabla_1\nabla_2\lambda_{ij}^{AB} &= \theta_{ij}. \end{aligned}$$

6.2. RELATIONSHIP WITH AM MODEL

The objective of mixed parameterization is similar to that of AM models. Both approaches aim to simultaneously design models for some of the selected marginal and the entire joint distribution. Formally, the AM model (Lang and Eliason 1997) is formed by

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

simultaneously specifying a log-linear (A) model for the joint distribution, $\log m = W\alpha$, and a regression type (M) model for the margin, $\log Mm = X\beta$. The distinction between an AM model and mixed parameterization is that in the latter, the A (canonical) and M (mean) parts are mutually exclusive components of the joint distribution. While the AM model is convenient to use because its A model adopts an established tool—the log-linear model—the A and M models are required to satisfy certain compatibility conditions (Lang and Eliason 1997:208). One of them is that "the range space of the A model matrix W contains the range space of the M model, in symbol $\Re(W) \supseteq \Re(M^T)$." In other words, the A model should be broad enough to avoid inducing any undue constraints on the margin Mm. When the condition $\Re(W) \supseteq \Re(^T)$ does not hold, there may not exist a solution for m that satisfies both the A and M models.

Compared with the log-linear model, both mixed parameterization and the AM approaches share several common advantages. For illustration, we use the analysis of mobility table as an example. First, modeling and comparison of marginal distributions (e.g., origin and destination marginal distributions) are made explicit and direct. Standard log-linear models, on the other hand, model the marginal distribution of structural mobility implicitly through induced constraints on the log-linear model parameters λ. Second, even when the A component in both AM and mixed parameterization (e.g., modeled using quasi-symmetry) does not fit the data well, structural mobility (marginal) change can still be summarized without bias. Third, both of the simultaneous approaches often result in a model that combines the best of both the A and the M models.

6.3. RELATIONSHIP WITH MARGINAL MODEL

The marginal model approach can be viewed as a special case of multilevel mixed parameterization. For any n-way table, it uses all of the marginal associations ranging from $(n - 1)$ to 2 dimensions as distribution parameters, with the partitions being strictly hierarchical. For the multilevel partition, one proceeds as follows: recursively apply mixed parameterization $(n - 1)$ times, each time retaining the highest order canonical association and transforming the remaining parameters into mean parameters. In other words, the marginal model

Request Permissions / Order Reprints
powered by RIGHTSLINK

can be viewed as an extreme case of applying mixed parameterization to every combination of variables: zooming in to the lowest order marginal distributions, then to the second-order margins, and so forth. Although each application of mixed parameterization to a marginal distribution is complementary, in general, the variation independence between parameters in a marginal model does not hold (the same can be said for parameters within the mean component of mixed parameterization). The likelihood for a marginal model is constructed in reverse order, moving from n one-dimensional marginal distributions, to $n(n-1)/2$ two-dimensional marginal distributions, and so on. In the process, $2^n - 1$ marginal distributions are constructed. The process is computationally intensive if covariates are involved.

7. MODELS AND COMPUTATIONAL ALGORITHMS

We outline an algorithm for obtaining maximum likelihood estimates for models formulated in terms of mixed parameters. Consider the following models:

$$\text{Canonical model:} \quad \theta^{(1)} = g(B_1\alpha) \quad \text{and}$$
$$\text{mean model:} \quad \mu^{(2)} = h(B_2\beta),$$

where B_1 and B_2 are either known design matrices or matrices of observed covariates and both g and h are smooth functions. Given some starting values, $\alpha^{(0)}, \beta^{(0)}$, the estimation proceeds with the following two steps:

1. Based on the current value $(\alpha^{(t)}, \beta^{(t)})$ search for the next β that gives the largest increase in likelihood. Denote the updated value by $\beta^{(t+1)}$.
2. Based on the value of $(\alpha^{(t)}, \beta^{(t+1)})$ search for the next α that gives the largest increase in likelihood. Denote the updated value by $\alpha^{(t+1)}$.

The procedure is iterated for $t > 0$ until a measure of the change in $(\alpha^{(t)}, \beta^{(t+1)})$ does not exceed a specified threshold. At each iteration of the algorithm, evaluation of the vector of cell probabilities $\pi(\theta^{(1)}(\alpha), \mu^{(2)}(\beta))$ is necessary (Fitzmaurice and Laird 1993) because the likelihood is the object to be maximized and its value depends on π and the observed data n (see equation (2)). Cell probabilities are computed by the iterative proportional fitting (IPF) algorithm. IPF starts with a q-dimensional probability table $\pi(\theta^{(1)}(\alpha^{(t)}),$

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

$\theta^{(2)} = (0, \ldots, 0)^T$) (not necessarily a probability distribution) and iteratively scales each of the marginal distributions identified in the set $\mu^{(2)}$ until its mean matches $\mu^{(2)}(\beta^{(t)})$. The value of $\theta^{(1)}$ is not altered after the IPF step (Darroch and Ratcliff 1972). In other words, the "starting values" define the association structure. The algorithm for fitting margins of order higher than 1 is also outlined in Gange (1995).

Incidentally, the partition of a q-dimensional θ into p-dimensional $\theta^{(1)}$ and $(q-p)$-dimensional $\mu^{(2)}$ offers some computational advantages in the estimation procedure. For example, if Newton-Raphson is used to update the parameters in Steps (a) and (b), then we invert two smaller covariance matrices instead of a single large one.

Two commonly used methods can be employed to obtain maximum likelihood estimates in Steps (a) and (b). The first method is to apply Newton-Raphson–type algorithms to search for the next point of iteration. An alternative way is to use a one-dimensional search method or Newton approximation for updating the parameters. The latter method is consistent with the spirit of Goodman (1979) and Becker (1990), who suggested using the unidimensional Newton algorithm to circumvent the problem of inverting large Hessian matrices.

8. DISCUSSION

While the theory and methodology of mixed parameterization used in this article are built on established work in statistics and sociology, we believe the article offers several important contributions to the sociology literature. The following summarizes these contributions and discusses possible future research.

1. Mixed parameterization generalizes the conventional method of modeling associations and univariate marginal distributions. For example, the mean component in mixed parameterization can include bivariate marginal distributions. Indeed, one can even fit a model to a selected set of aggregated counts that corresponds to a set of transformed mean parameters. The theory of mixed parameterization of Barndorff-Nielsen and Cox (1994) is rather general and allows such models. The fitting of models (including saturated ones) to selected sets of counts is not uncommon in sociology. For example, in some

Request Permissions / Order Reprints
powered by RIGHTSLINK COPYRIGHT CLEARANCE CENTER, INC.

mobility models, the diagonal cells are required to be fitted exactly (e.g., see discussion in Becker et al. 1998:518). Possible extensions of mixed parameterization to include such application would provide researchers in sociology with an even more flexible tool.

2. Multilevel partitioning of the parameter space makes possible the design of a sequence of compatible and meaningful telescoping models that describes fine marginal structures for cross-classified data. The multilevel partitioning scheme may be particularly useful when the joint table is sparse (i.e., contains less information on average per variable) but the marginal tables of interest are relatively dense (i.e., contain more information on average per variable). To take advantage of the differential information present in the data, one modeling strategy would be to fit increasingly accurate models to a telescoping set of judiciously chosen marginal tables of decreasing dimension. The same principle might be applied to nontelescoping models. While nontelescoping models may be meaningful in sociological research, certain issues arise, such as the fact that one marginal model may induce constraints on another. Further investigations are necessary.
3. A system of matrix and difference operators makes possible parameterization of the joint density into the exponential family, which is a powerful vehicle for further exploring the statistical properties of mixed parameters (Barndorff-Nielsen 1978). Such a system also defines local conditional associations of various orders. When other types of parameters are required, the matrix representation needs to be modified. For example, instead of local odds ratio, which is used in this article, Glonek (1996) used multivariate logistic contrasts or global odds ratios as the distribution parameters in characterizing the mean component $\mu^{(2)}$. Therefore, one possible direction of future research is to extend the system of matrix and difference operators to incorporate different types of parameters (e.g., global odds ratios) and cross-classifying variables (e.g., non-equally-spaced categories and partially ordered categories).
4. The comparison of three prominent methodologies in the literature—the log-linear model, the AM model, and the marginal model—to mixed parameterization provides insights into the relationship among these important modeling approaches. Furthermore, this article demonstrates how multilevel mixed parameterization accommodates commonly used models. Both could be meaningful for practitioners who seek to extend or improve existing methods to broaden the scope of their applications in sociology.

Request Permissions / Order Reprints
powered by RIGHTSLINK

REFERENCES

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

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

Barndorff-Nielsen, O. E. and David R. Cox. 1994. *Inference and Asymptotics*. London: Chapman and Hall.

Becker, Mark P. 1990. "Quasisymmetric Models for the Analysis of Square Contingency Tables." *Journal of the Royal Statistical Society* 52(Series B):369-78.

———. 1994. "Analysis of Cross-Classification of Counts Using Models for Marginal Distributions—An Application to Trends in Attitudes on Legalized Abortion." *Sociological Methodology* 24:229–65.

Becker, Mark P., Susan Minick, and Ilsoon Yang. 1998. "Specifications of Models for Cross-Classified Counts," *Sociological Methods & Research* 26:511-29.

Clogg, Clifford C. and E. S. Shihadeh. 1994. *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage.

Cox, David R. and Nancy Reid. 1987. "Parameter Orthogonality and Approximate Conditional Inference (With Discussion)." *Journal of the Royal Statistical Society* 49(Series B):1-39.

Cox, David R. and Nanny Wermuth. 1993. "Linear Independence Represented by Chain Graph." *Statistical Science* 8:204-18.

Darroch, J. N. and Ratcliff, D. 1972. "Generalized Iterative Scaling for Log-Linear Models." *The Annals of Mathematical Statistics* 43:1470-80.

Davis, J. A. and T. W. Smith. 1990. *General Social Surveys: 1972-1990* (Machine-readable data file). Chicago: National Opinion Research Center.

Fitzmaurice, Garrett M. and Nan Laird. 1993. "A Likelihood-Based Method for Analyzing Longitudinal Binary Responses." *Biometrika* 80:141-51.

Gange, Stephen J. 1995. "Generating Multivariate Categorical Variates Using the Iterative Proportional Fitting Algorithm." *The American Statistician* 49:134-38.

Glonek, G. F. V. 1996. "A Class of Regression Models for Multivariate Categorical Response." *Biometrika* 83:645-55.

Goodman, Leo A. 1979. "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 74:537-52.

Lang, Joseph B. and Alan Agresti. 1994. "Simultaneously Modeling Joint and Marginal Distributions of Multivariate Categorical Responses." *Journal of the American Statistical Association* 89:625-32.

Lang, Joseph B. and Scott R. Eliason. 1997. "Application of Association-Marginal Models to the Study of Social Mobility." *Sociological Methods & Research* 26:183-212.

Lehmann, Eric L. 1991. *Theory of Point Estimation*. Belmont, CA: Wadsworth.

Liang, Kung-Yee and Scott Zeger (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73:13-22.

Plewis, Ian. 1985. *Analyzing Change: Measurement and Explanation Using Longitudinal Data*. New York: Wiley.

Rudas, Tamas. 1998. *Odds Ratios in the Analysis of Contingency Tables*. (Quantitative Applications in the Social Sciences Series, 07-119). Thousand Oaks, CA: Sage.

Sinkhorn, Richard. 1967. "Diagonal Equivalence to Matrices With Prescribed Row and Column Sums." *American Mathematical Monthly* 74:402-405.

Wang, Yuchung J. 1986. "Ordered-Dependent Parameterization of Multinomial Distributions." *Scandinavian Journal of Statistics* 13:199-205.

Request Permissions / Order Reprints
powered by RIGHTSLINK
COPYRIGHT CLEARANCE CENTER, INC.

Whittaker, Joe. 1990. *Graphical Models in Applied Mathematical Multivariate Statistics*. West Sussex, UK: Wiley.

Wong, Raymond S. 1995. "Extensions in the Use of Log-Multiplicative Scaled Association Models in Multiway Contingency Tables." *Sociological Methods & Research* 23:507-38.

Zwillinger, Daniel, ed. 1996. *Standard Mathematical Tables and Formulae*. New York: CRC Press.

Edward H. Ip is an assistant professor in the Information and Operations Management Department, Marshall School of Business, University of Southern California. His research interest is in psychometrics and the analysis of large-scale databases. Relevant recent publications include "Empirical Bayes and Item Clustering Effects in a Latent Variable Hierarchical Model: A Case Study From the National Assessment of Educational Progress" (with Steven Scott, Journal of the American Statistical Association, *2002), and "Locally Dependent Latent Trait Model and the Dutch Identity Revisited"* (Psychometrika, *2002).*

Yuchung J. Wang is an associate professor in the Department of Mathematical Sciences, Rutgers University, Camden, New Jersey. He is interested in the development of methods for the analysis of ordered categorical data, social network, and quality engineering. Specifically, he is working on incorporating theory and the method developed for exponential family into discrete data analysis. One of his recent publications is "Multivariate Normal Integral and Contingency Tables With Ordered Categories" (Psychometrika, *1997*).

Request Permission or Order Reprints Instantly

Interested in copying, sharing, or the repurposing of this article? U.S. copyright law, in most cases, directs you to first get permission from the article's rightsholder before using their content.

To lawfully obtain permission to reuse, or to order reprints of this article quickly and efficiently, click on the "Request Permission/ Order Reprints" link below and follow the instructions. For information on Fair Use limitations of U.S. copyright law, please visit Stamford University Libraries, or for guidelines on Fair Use in the Classroom, please refer to The Association of American Publishers' (AAP).

All information and materials related to SAGE Publications are protected by the copyright laws of the United States and other countries. SAGE Publications and the SAGE logo are registered trademarks of SAGE Publications. Copyright © 2003, Sage Publications, all rights reserved. Mention of other publishers, titles or services may be registered trademarks of their respective companies. Please refer to our user help pages for more details: http://www.sagepub.com/cc/faq/SageFAQ.htm

Request Permissions / Order Reprints