Ann. Inst. Statist. Math.
Vol. 49, No. 2, 371–393 (1997)

CONVEX MODELS OF HIGH DIMENSIONAL DISCRETE DATA

MAX A. WOODBURY[1], KENNETH G. MANTON[1] AND H. DENNIS TOLLEY[2]

[1] *Duke University, Center for Demographic Studies, 2117 Campus Drive, Box 90408, Durham, NC 27708, U.S.A.*
[2] *Brigham Young University, Department of Statistics, Room 226, TMCB, Provo, UT 84602, U.S.A.*

(Received September 8, 1994; revised May 20, 1996)

Abstract. Categorical data of high (but finite) dimensionality generate sparsely populated J-way contingency tables because of finite sample sizes. A model representing such data by a "smooth" low dimensional parametric structure using a "natural" metric would be useful. We discuss a model using a metric determined by convex sets to represent moments of a discrete distribution to order J. The model is shown, from theorems on convex polytopes, to depend only on the linear space spanned by the convex set—it is otherwise measure invariant. We provide an empirical example to illustrate the maximum likelihood estimation of parameters of a particular statistical application (Grade of Membership analysis) of such a model.

Key words and phrases: Probability mixtures, convex sets, polytopes, convex duality, Grades of Membership.

1. Introduction

Often data on individual sampling units is gathered through closed ended questions. Each of the answers to these questions provides a response in one of a finite number of categories. When many questions are asked about each individual the data represent a sample from a high dimensional space. In this case though each individual can be uniquely classified into response categories, tabulations of these data often yield a sparsely filled table. In applications, the dimensionality is often reduced by aggregating data across questions and assuming that response variation within retained cells is due to heterogeneity.

Within cell heterogeneity can be modeled by specifying a distribution for individual response probabilities (e.g., Kleinman (1973)) using, for example, parametric empirical Bayes strategies (e.g., Laird (1978); Tsutakawa *et al.* (1985); Manton *et al.* (1989)) where the hyperparameters of the mixing distribution for individual parameters *implicitly* describe individual heterogeneity. We present here a model which explicitly represents this heterogeneity using convex subsets of a space of probability density functions to describe high, but finite, dimensional categorical data without such restrictions. The particular statistical application of this type of model illustrated here will be referred to as the Grade of Membership (GoM) model. In "GoM" convex scores are estimated for each sampled person indicating the degree, or "grade" of membership the individual has in each state defined by the analysis. Thus, the individual can be a partial (or fuzzy) member of several states. Grade of membership parameters implicitly exist in other models but are usually assumed a.) known, and b.) that each person is a member of only one discrete state, so they provide little additional information. When the membership in discrete states is not known, but a probabilistic assignment is made of each case into one of a set of discrete states the resulting model is often referred to as a latent class model (LCM; e.g., Lazarsfeld and Henry (1968)). In a model based on convex sets, such as GoM, parameters are *uniquely* identified with the extreme points of a convex polytope generated by restricting the space of probability distributions. This parameterization is not dependent on other measure theoretic assumptions and can be used to reduce the dimensionality of the parameter space needed to describe a given categorical data set.

In this paper, we illustrate the GoM model with data gathered on 4,525 nursing home residents. Each resident was assessed on 111 health and functional status measurements. If all questions are used to classify individuals, and each question had only two responses, the implied contingency table would have 2^{111} cells. Although a sample of 4,525 is practically large, the table for this example would have many empty cells. Reducing the number of cells requires assumptions about the classification variables. For example, one could reduce the number of classification variables by assuming that the cross classifications for the deleted variables were uninformative. Alternatively, one might assume that differences in responses across deleted classifications can be modeled by a less complex distribution with a relatively small number of parameters. Here we use a third method. We assumed that individuals have different response probabilities and show that the distribution of these probabilities can be naturally parameterized using convex sets. This represents within category heterogeneity by assuming an individual's probability of response is randomly sampled from a linear subspace of a discrete measurement space. The specification of that linear space uniquely determines the structural parameters because they are defined by the intersection of the linear measurement and convex space of probabilities.

Since the likelihood contains both global and latent random variables (local parameters) for each individual, increasing the number of individuals in the sample increases the number of parameters. This raises two issues; are global parameter estimates approximately normally distributed and are they consistently estimated. To demonstrate consistency, the conditions of Kiefer and Wolfowitz (1956) are generalized for the convex set conditions (Tolley and Manton (1992)). We show that constraints on the distribution of individual parameter estimates imposed in the likelihood by the structural parameters assures consistency of estimates of the moments of the distribution of the local parameters for individuals under suitable regularity conditions. Below, we define the general structure of a convex set model and show the identifiability of its parameters using theorems on convex polytopes. Next, we show sufficient conditions for identifying and estimating canonical model parameters. Third, we discuss estimator consistency. Finally, we provide an example using the GoM implementation of a convex set model.

2. A convex set model for categorical data: definitions

Define $\boldsymbol{Y}^i$ as a vector of binary random variables representing outcomes of J multinomial variables observed for person $i = 1, 2, \ldots, I$. $\boldsymbol{Y}^i$ has realizations, $\boldsymbol{y}^i = ((y^i_{jl}, l = 1, 2, \ldots, L_j), j = 1, 2, \ldots, J)$, where y^i_{jl} has the value of 0 or 1, with only one non-zero entry over l for each value of j. Assume $L_j > 1$ and finite for all j. The data can be viewed as an $N \times M$ table described by,

$$\boldsymbol{Y} = (((Y^i_{jl}, l = 1, 2, \ldots, L_j), j = 1, 2, \ldots, J), i = 1, 2, \ldots, N),$$

where $M = L_1 + L_2 + \cdots + L_J$. $E(Y^i_{jl})$ is the expected probability, p^i_{jl} of the i-th individual giving response l to question j. The row vector of probabilities (the probability "profile") for individual i is $\boldsymbol{p}^i = E(\boldsymbol{Y}^i) = ((p^i_{jl}, l = 1, \ldots, L_j), j = 1, \ldots, J)$. The distribution of $\boldsymbol{Y}^i$ given $\boldsymbol{p}^i$, is a joint multinomial distribution denoted $h(\boldsymbol{y}^i \mid \boldsymbol{p}^i)$. Standard multinomial models specify $\boldsymbol{p}^i$ as a parametric function of exogenous variables. Models representing individual heterogeneity allow $\boldsymbol{p}^i$ to vary over individuals, within classes defined by exogenous variables. Models with heterogeneity allow richer representations of $\boldsymbol{p}^i$. Models representing too little heterogeneity may be substantively invalid. We examine a model representing the heterogeneity consistent with a convex set parameterization.

Specifically, for each j ($j = 1, 2, \ldots, J$) a simplex, $\mathbb{M}_j$, can be defined with L_j extreme points and facets which contains the $\boldsymbol{p}^i_j = (p^i_{jl}, l = 1, 2, \ldots, L_j)$ for all i, where $p^i_{jl} \geq 0$ and $\sum_l p^i_{jl} = 1.0$. The set of all possible multinomial probabilities for variable j defines $\mathbb{M}_j$. Defining the independent vertices for $\mathbb{M}_j$ is the set of L_j unit vectors $\boldsymbol{u}_{j1} = (1, 0, 0, \ldots, 0), \boldsymbol{u}_{j2} = (0, 1, 0, \ldots, 0), \ldots, \boldsymbol{u}_{jL_j} = (0, 0, 0, \ldots, 1)$. Thus, the unique, barycentric coordinates of $\boldsymbol{p}_j \in \mathbb{M}_j$ with respect to the defining vertices are the values of $\boldsymbol{p}_j$ itself, i.e.,

$$\boldsymbol{p}_j = p_{j1}\boldsymbol{u}_{j1} + p_{j2}\boldsymbol{u}_{j2} + \cdots + p_{jL_j}\boldsymbol{u}_{jL_j}. \tag{2.1}$$

The space of all possible multinomial probabilitie for J variables is the Cartesian product of the $\mathbb{M}_j$. This is denoted by a direct sum, i.e.,

$$\mathbb{M} = \mathbb{M}_1 \oplus \mathbb{M}_2 \oplus \cdots \oplus \mathbb{M}_J, \tag{2.2}$$

with,

$$M = L_1 + L_2 + \cdots + L_J$$

coordinates.

Define a probability space $(\mathbb{M}, F, Pr)$, where F is a σ-field of measurable subsets of $\mathbb{M}$ and Pr is a σ-complete probability measure defined for each set in F. Let S be the set of measurable functions mapping $\mathbb{M}$ to $\boldsymbol{R}^1$ such as $w(\boldsymbol{p})$ below. Define $L_{\mathbb{M}}$ as the linear space spanned by $\boldsymbol{p} \in \mathbb{M}$. Specifically, one can choose elements (profiles) from $\mathbb{M}$, say, $\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^R$ and weights, w, to generate an element of $L_{\mathbb{M}}$.

$$\nu = w_1 \boldsymbol{p}^1 + w_2 \boldsymbol{p}^2 + \cdots + w_R \boldsymbol{p}^R$$

which can be extended to

$$\nu_w = \int_m \boldsymbol{p} w(\boldsymbol{p}) \operatorname{vol}(d\mathbb{M})$$

where $d\mathbb{M}$ is the differential volume element. To get $L_{\mathbb{M}}$ we need a basis for $\mathbb{M}$, $\{\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^R\}$ and the w. Using the expression for ν_w we can substitute the probability (Pr) for the volume in ($\mathbb{M}$) to get,

$$L_{\mathbb{B}} = \left\{ \nu \mid \nu = \int_{\mathbb{M}} \boldsymbol{p} w(\boldsymbol{p}) Pr[d\mathbb{M}] \right\}. \tag{2.3}$$

$L_{\mathbb{B}}$ is the reduction of $L_{\mathbb{M}}$ by F and Pr to the minimum dimension necessary to contain probability mass. Thus, $L_{\mathbb{B}}$ is the intersection of all (closed) linear subspaces containing all probability mass. Define the convex set,

$$\mathbb{B} = L_{\mathbb{B}} \cap \mathbb{M}. \tag{2.4}$$

$\mathbb{B}$ is of most interest since it is a subset of $\mathbb{M}$ containing all probability mass and often with reduced dimension. The objective is to analyze data subject to the constraint that the multinomial probabilities are realizations from $\mathbb{B}$. An acceptable model of heterogeneity will have a parametric structure within $\mathbb{B}$. In Section 3, we examine $\mathbb{B}$.

3. Results on convex sets and polytopes necessary to construct a model

LEMMA 3.1. *$\mathbb{M}$ is a polytope containing $\boldsymbol{p}$.*

PROOF. Since each $\mathbb{M}_j$ is a polytope with L_j extreme points, $\mathbb{M}$ has, at most, $L = \prod_j L_j$ extreme points. Let P be the polytope defined by the intersection of supporting half spaces of the subsets of extreme points of $\mathbb{M}$ which define its facets. Then, $\mathbb{M} \subseteq P$ (see Karlin and Shapley (1953)). For any $\boldsymbol{p} \in P$, define $\boldsymbol{p}_j$ as the vector made by taking elements numbered $\sum_{n=1}^{j-1} L_n + 1$ through $\sum_{n=1}^{j} L_n$ of $\boldsymbol{p}$ in order. Since P is a polytope, $\boldsymbol{p} \in P$ may be represented,

$$\boldsymbol{p} = \sum_{k=1}^{L} \alpha_k \lambda^k,$$

where λ^k are $1 \times M$ vectors of extreme points of $\mathbb{M}$ and $\{\alpha_k\}$ is a set of barycentric coordinates. The vector $\boldsymbol{p}_j$ may be represented as $\boldsymbol{p}_j = \sum_{k=1}^{L_j} \alpha_k^{(j)} \mu_k^{(j)}$ where $\alpha_k^{(j)} = p_{jk}$ and $\mu_k^{(j)}$, $k = 1, \ldots, L_j$ are basis vectors (as in (1)) $k = (1, \ldots, K)$ where the k-th element of μ_k is unity. Thus $\sum_{k=1}^{L_j} \alpha_k^j = 1$ and $\alpha_k^{(j)} \geq 0$ for all k and j implying $\boldsymbol{p}_j \in \mathbb{M}_j$, $\boldsymbol{p} \epsilon \mathbb{M}$ and consequently, $P \subseteq \mathbb{M}$. This proves Lemma 3.1.

COROLLARY 3.1. *$\mathbb{B}$ is a polytope.*

PROOF. Since $\mathbb{M}$ is a polytope, and $L_{\mathbb{B}}$ a linear space, the result follows because the intersection of a linear space with a polytope is a polytope (Weyl (1949)).

COROLLARY 3.2. *$\mathbb{B}$ is the convex hull of a finite number of unique extreme vectors $\lambda^1, \ldots, \lambda^K$, where ($\lambda^k = ((\lambda^k_{jl}; l = 1, \ldots, L_j), j = 1, \ldots, J)$, for $k = 1, 2, \ldots, K$).*

PROOF. The proof follows from Lemma 3.1 (see Weyl (1949)).

Corollary 3.2 suggests any element of $\mathbb{B}$ can be represented as a convex combination of λ^k vectors. Explicitly, for any $\boldsymbol{p}$ in $\mathbb{B}$ there exists at least one vector $g = (g_k, k = 1, 2, \ldots, K)$ such that $g_k \geq 0$ for all k and $\sum_k g_k = 1$. The j, l entry of $\boldsymbol{p}$ is,

$$p_{jl} = \sum_{k=1}^{K} g_k \lambda^k_{jl}. \tag{3.1}$$

The λ^k_{jl} is the j, l entry of the vector λ^k given in Corollary 3.2 and are determined by $L_{\mathbb{B}}$—*not* by the distribution of cases in the polytope. Specifically, the distributions of outcomes with the same support have the same $\mathbb{B}$.

The set of all λ^k_{jl} define $\mathbb{M}$ with $L = L_1 \times L_2 \times \cdots \times L_J$ extreme points and M facets. By construction, if (3.1) holds, all outcomes are in $\mathbb{B}$ with probability one. Heterogeneity in the data is determined by the distribution of multinomial probabilities in $\mathbb{B}$—in this case by Pr and F. Different measures of Pr on F define different models.

Discrete measures with all probability mass at *known* points λ^k_{jl}, described by cross-classification variables, are contingency tables that may be analyzed using log linear regression (e.g., Bishop *et al.* (1975)). When *all* mass is *at* λ^k_{jl} (i.e., g_k for all persons are forced to be 0 or 1 for one class), but the points are not described by cross-classifying or exogenous regression variables, the unknown λ^k_{jl} may be estimated using LCM as mentioned in the introduction (e.g., Lazarsfeld and Henry (1968)). A more general model also has latent λ^k_{jl} but, instead of requiring all probability mass to be concentrated at K points, allows individual heterogeneity by allowing the individual multinomial probabilities to be distributed *within* the convex set spanned by λ^k_{jl}. This model is described below and underlies the GoM statistical model.

The random sample of N probability vectors, $\boldsymbol{P}_N$, asymptotically defines $\mathbb{B}$ as N increases. Let each realization $\boldsymbol{p}_i$, $i = 1, \ldots, N$ define rows of the vector $\boldsymbol{P}_N$. For fixed N, $\boldsymbol{P}_N$ is a random set. However, the sample convex hull of columns in $\boldsymbol{P}_N$, because it is random, and highly multifaceted, is not useful. Whether the probability in $\mathbb{M}$ is contained in a few points, as in LCM, or is distributed over a continuum (as in GoM), $\boldsymbol{P}_N$ is rank $\boldsymbol{K}_N$, where $\boldsymbol{K}_N$ is asymptotically the dimension of $L_{\mathbb{B}}$.

Each point in $\mathbb{B}$ (or the asymptotic convex hull of $\boldsymbol{P}_N$) is a convex combination of the unique extreme points λ_{jl}^k. The matrix $\boldsymbol{P}_N$ of vectors $\boldsymbol{p}$ in $\mathbb{B}$ may be represented,

$$\boldsymbol{P}_N = G_N \Lambda, \tag{3.2}$$

where rows of Λ are vectors of extreme points of $\mathbb{B}$. G_N represent coefficients, g_k^i, guaranteed by Corollary 3.2 to represent each sample point as a convex combination of extreme points in Λ.

Although the λ_{jl}^k in (3.2) are unique, extreme points of the finite polytope, $\mathbb{B}$, the same may not be true for g_k^i. Under certain conditions, (e.g., $\mathbb{B}$ is a simplex) g_k^i are unique barycentric coordinates with respect to extreme points λ_{jl}^k. If $\mathbb{B}$ is not a simplex, constraints may be required to identify the g_k^i. One solution is presented in Theorem 3.1.

THEOREM 3.1. *There exists a triangulation of polytope $\mathbb{B}$ into simplicies $B_1, B_2, \ldots, B_R$ each of which is the convex closure of a subset of extreme points of $\mathbb{B}$ such that there exists a unique set of barycentric coordinates with respect to the triangulation.*

PROOF. From the definition of convex sets there exists a set of simple cuts of the polytope dividing the polytope into simplicies such that each $\boldsymbol{p}^* \in \mathbb{B}$ either lies in the interior of exactly one simplex or lies on the shared face of the two or more simplices of the triangulation. Barycentric coordinates for points interior to a simplex defined using only extreme points of the simplex are unique. A point not in the interior is in the convex closure of two or more simplicies. For a fixed triangulation the intersection of all simplicies containing $\boldsymbol{p}^*$ (assuming two or more such simplicies) is a face, $F\boldsymbol{p}^*$. $F\boldsymbol{p}^*$, is a simplex defined by the convex hull of extreme points common to all simplices in the triangulation which contain $\boldsymbol{p}^*$. Thus, $\boldsymbol{p}^*$ is a convex combination of extreme points of $F\boldsymbol{p}^*$ and there is a unique vector of barycentric coordinates $\boldsymbol{g}(\boldsymbol{p})$ such that $\boldsymbol{p}^* = \sum_k g_k(\boldsymbol{p}^*)\lambda^k$. This proves the theorem.

Table 1 contains the λ_{jl}^k estimates for several models applied to the example in Section 6. In this example there were $K = 11$ corners to the polytope. The different λ_{jl}^k estimates reflect different sets of model constraints. In one (i.e., LCM), all g_{ik} are constrained to be either zero or unity. In the second (basic GoM) the g_k^i vary between zero and unity and are constrained to sum to unity for each person. In the third, a generalization of GoM, individual heterogeneity may differ in responses to the j-th question. This heterogeneity determines how informative a question was in defining a dimension of the solution.

4. Construction of a likelihood for a specific convex set (GoM) model

Let $\eta = (\eta_{11}, \ldots, \eta_{JL_J})$ be the random vector consisting of $M = \sum L_j$ nonnegative random variables. Assume $\sum_{l=1}^{L_j} \eta_{jl} = 1$ for all j. Let $f(\boldsymbol{p})$ be the density of η, such that the domain of $f(\boldsymbol{p})$ is a subset of $\mathbb{B}$. If we assume $\boldsymbol{p}^i$ are independent realizations of η then the likelihood for $\underset{\sim}{y}^i$, $i = 1, 2, \ldots$, has a set of parameters for each individual i. This is the basic GoM parameterization. There is no constraint on the terms of $f(\underset{\sim}{p})$ since its domain is necessarily a subset of B. The realization $\underset{\sim}{p}^i$ of $\underset{\sim}{\eta}$ are represented as $\underset{\sim}{p}^i = \sum g_k^i \lambda_{jl}^k$, where the λ_{jl}^k parameters are fixed over i. In essence $f(\underset{\sim}{p})$ induces a distribution on g_k^i, $k = 1, \ldots, K$, say $H(\underset{\sim}{g})$. Following Little and Rubin (1986) or Orchard and Woodbury (1971), the likelihood in this case is formed by evaluating the expectation by integrating out $dH(\underset{\sim}{g})$, of $\prod_i \prod_l (\sum_k g_k^i \lambda_{jl}^k)^{y_{jl}^i}$. The likelihood may not be identifiable if the random variables do not uniquely characterize a distribution. The issue of identifiability is addressed in Theorem 4.1 (Woodbury *et al.* (1994)).

THEOREM 4.1. *The likelihood for the* GoM *form of a convex set model is,*

$$L = \prod_i \int \prod_j \prod_l \left(\sum_k g_k^j \lambda_{jl}^k \right)^{y_{ijl}} \cdot dH(\underset{\sim}{g}) \tag{4.1}$$

which is identifiable in that the λ are unique and moments of f are unique to order J.

The proof of Theorem 4.1 is based on results for finite convex polytopes (Brondsted (1983)). Before applying his results, we first need to demonstrate Lemma 4.1 and Corollary 4.1. As discussed in Theorem 3.1, any finite convex polytope can be divided into simplicies such that each point in a simplex can be represented as a convex combination of the extreme points of the simplex. The coefficients of this convex combination are unique barycentric coordinates with respect to extreme points of the simplex. There may be multiple such triangulations which can be indexed as $t = 1, \ldots, T$. For any point $\boldsymbol{p}_i$ in $\mathbb{B}$ define β_{it} as the vector of barycentric coordinates with respect to triangulation t.

LEMMA 4.1. *Let $\sum g_k^i \lambda_k$ be a convex representation of point $\boldsymbol{p}_i$ where g_k^i are scalars and λ_k are row vectors. For triangulation t with barycentric coordinates given by the row vector β_{it}, β_{it} can be uniquely represented as,*

$$\beta_{it} = G\Lambda, \tag{4.2}$$

where the i-th row of G, G_i, is the row vector of g_k^i.

PROOF. Let $\boldsymbol{p}_i = \sum g_k^i \lambda_k$ be represented by $\boldsymbol{p}_i = \sum \beta_{itk} \lambda_k$ with respect to t. Assume $\boldsymbol{p}_i$ is interior to one simplex of the triangulation. β_{itk} will be zero for k corresponding to extreme points not in the closure of the simplex. Let β_{it}^* denote the row vector of nonzero β_{itk}. Let G_i be a row vector of g_k^i arranged in the same order. Define the matrix Λ with vectors λ_k^i as rows with the first rows of Λ corresponding to entries in β^*_{it}. Assume the width of Λ (the length of the λ_k vectors) is s. Let the length of β^*_{it} be $u \leq s$. Partition Λ as

$$\Lambda = \begin{pmatrix} A \\ B \end{pmatrix}$$

where A is $u \times s$ and is made up of u extreme points (λ) of the simplex containing $\boldsymbol{p}_i$ for t. From results on polytopes if there is a simplex with barycentric coordinates β^*_{it} for interior point $\boldsymbol{p}_i$ then there is a permutation of columns of Λ such that,

$$\Lambda = \begin{pmatrix} A_1 & A_2 \\ B_1 & B_2 \end{pmatrix},$$

where A_1 is an $u \times u - 1$ matrix such that the $u - 1 \times u - 1$ matrix A_1^* formed by subtracting the last row of A_1 from every row of A_1, and then deleting the last row, is nonsingular (see Brondsted (1983)).

For any i, define β_{it} as the concatenation of β^*_{it} and a $s - u$ row vector of zeroes. Then,

$$\beta_{it} = [\beta^*_{it}, 0]\Lambda^* = G_i \Lambda^*. \tag{4.3}$$

Also,

$$\beta^*_{it} A_1 = G_i \begin{pmatrix} A_1 \\ B_1 \end{pmatrix}.$$

Define b_{it} as β^*_{it} with the last column deleted. Define G_i^* as G_i with the last column deleted. The matrix B_i^* is formed by subtracting the last row of A_1 from each row of B_1. From the constraint,

$$\sum_{k=1}^{u} \beta_{itk} = 1 = \sum_{k=1}^{s} g_{ik},$$

we have,

$$b_{it} A_1^* = G_i^* \begin{pmatrix} A_1^* \\ B_1^* \end{pmatrix}.$$

Since A_1^* is nonsingular,

$$b_{it} = G_i^* \begin{pmatrix} A_1^* \\ B_1^* \end{pmatrix} A_1^{*-1}.$$

For points $\boldsymbol{p}_i$ on a face of a simplex, the proof is altered by considering each face to be in exactly one simplex. The final β_{ijt} in β^*_{it} is determined by the convex constraint. The row vector β_{it} is as defined above. This completes the proof of Lemma 4.1.

COROLLARY 4.1. *Given raw moments of any convex representation of points in* $\mathbb{B}$, *of order* R *or less, the moments of order* R *or less of* $F(\cdot)$ *are uniquely determined with respect to a fixed* t.

PROOF. This follows from Lemma 4.1 by noting that each vector G_i maps uniquely onto the barycentric coordinates of the relevant simplex of the point.

Now we can prove Theorem 4.1.

PROOF OF THEOREM 4.1. Since $\mathbb{B}$ is a finite polytope, there exists a triangulation of $\mathbb{B}$ into simplicies which have no interior points in common whose extreme points are a subset of the λ. Define moments of $f(\cdot)$ with respect to a fixed coordinate system, e.g., a specific triangulation of the polytope (see Theorem 3.1). From Corollary 4.1, moments of $f(\cdot)$ are uniquely specified from the moments of any convex representation of points in the polytope. By corollaries to Lemma 3.1, extreme points of the polytope are also unique. Theorem 4.1 is shown by taking expectations noting that the y_{ijl} are indicator variables (see Woodbury *et al.* (1994)).

The importance of Theorem 4.1 is that although the representation of the $\boldsymbol{p}_i$ using a polytope is not necessarily unique, the polytope, as constructed above, is unique. Additionally, moments of points $\boldsymbol{p}_i$ relative to extreme points can be uniquely defined relative to a fixed triangulation of the polytope.

Tolley and Manton (1992) show that if (4.1) is identifiable in λ and moments to order J, then maximum likelihood estimates of λ_{jl}^k, and moments to order J of g_k^i, are consistent. Hence, under Theorem 4.1, maximization of (4.1) gives consistent estimates of λ_{jl}^k and moments of f up to order J.

5. A low dimensional geometric illustration

We have shown that if outcome probabilities were constrained to the measurable subspace resulting from the intersection of the polytope $\mathbb{M}$ with linear space $L_{\mathbb{B}}$, then all outcome probabilities are naturally represented as a convex sum of the unique extreme points of the resulting polytope $\mathbb{B}$. Coefficients of the convex sum may not be unique (depending on the relation of K to J) although moments up to order J of $f(\boldsymbol{g})$ are unique. The maximization of this likelihood produces consistent estimates of both the λ_{jl}^k and moments of $f(\boldsymbol{p})$. Consequently the model can be used to fit $\boldsymbol{p}$ by say, $\hat{\boldsymbol{p}}$. We now examine the geometry of the constraints on $\boldsymbol{p}$ in a low dimensional example. Suppose there are $J = 3$ multinomials, each with two outcomes, e.g., success and failure. Possible probabilities for outcomes of the three multinomials are represented by an unit cube (see Fig. 1).

The axes are labeled $M1$, $M2$, and $M3$ for the three multinomials. An individual with probabilities (p_1, p_2, p_3) for outcomes on $M1$, $M2$, and $M3$, respectively, represents one realization of ξ and can be plotted as a point in the cube in Fig. 1.

To construct $\mathbb{B}$ we need $L_{\mathbb{B}}$. The dimension of $L_{\mathbb{B}}$, is implicitly dictated by the probability distribution of ξ. Hence, only dimensions where $f(\boldsymbol{g})$ is nonzero are included. If $L_{\mathbb{B}}$ is assumed to be a plane, Fig. 2 illustrates its intersection with the cube. By assumption all multinomial probabilities are within the plane segment in the cube. These are not outcomes of the three questions but are the probabilities of response. This means that $f(\boldsymbol{g})$ is such that all probabilities $\boldsymbol{p}_i$, $i = 1, 2, 3$, calculated with realizations of ξ using (4.1) lie in the plane. Although each individual may have different probabilities of response, those probabilities

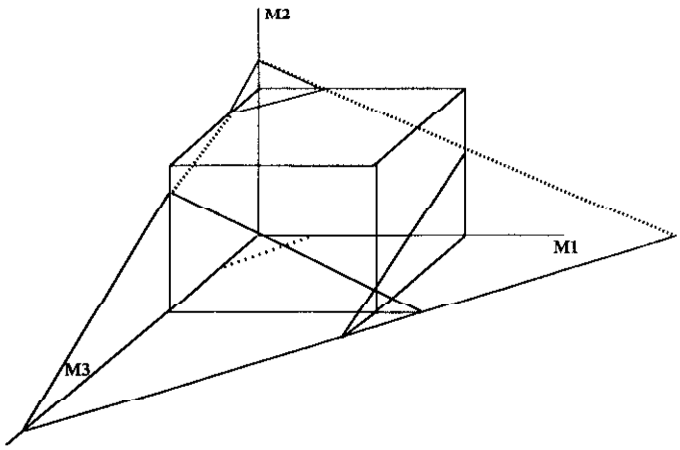

Fig. 1. Depiction of the range of probabilities for the three binomial outcomes.

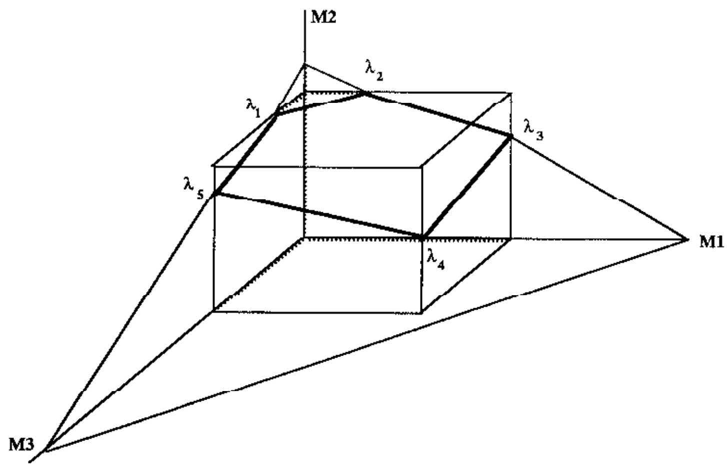

Fig. 2. Picture of plan intersecting the unit cube. Corners of the pentagon, labelled $\lambda_1, \ldots, \lambda_5$, are the extreme points of the five sided polytope.

must lie in the plane. In Fig. 2, $\mathbb{B}$ is a pentagon oriented with a negative slope toward $M1$ and $M3$. The λ are the five corners of the pentagon as $K = 5$.

A convex polytope can be represented as the intersection of its defining half spaces. For $\mathbb{M}$ in Fig. 2, (the cube) there are six half spaces ($M1 \geq 0$, $M1 \leq 1$, $M2 \geq 0$, $M2 \leq 1$, $M3 \geq 0$, and $M3 \leq 1$). The pentagon, however, requires only five constraints in the lower dimensional space $L_{\mathbb{B}}$. The constraining half spaces replace some of the defining half spaces. For example, the $M2$ dimension is bounded away from zero (so the $M_2 \geq 0$ constraint is not needed) whereas $M1$ and $M3$ are not. In general, the linear space $L_{\mathbb{B}}$, imposes constraints on the P_i. Experience with a wide range of data indicates that, unlike Fig. 2, $L_{\mathbb{B}}$ is usually of much loser dimension than $\mathbb{M}$.

6. Empirical application of the likelihood: estimation for a given observational plan

To estimate parameters, p^i_{jl} must be related to observations y^i_{jl} assumed to be generated by a specific observation plan. Assume responses are conditionally independent given g^i_k, since it is possible to construct a model for which this is true (see Suppes and Zanotti (1981)). To identify the space for p^i_{jl}, note $\mathbb{M}$ contains images of the Radon-Nicodym measures in $(\mathbb{M}, F, Pr)$ producing a map,

$$Pr : \omega \to ((p_{jl}(\omega), l = 1, 2, \ldots, L_j), j = 1, 2, \ldots, J).$$

For convex sets, the image on $\mathbb{M}$ can contain singular components. In a simplex, singularities are eliminated. The expectation is (see (2.4)),

$$E[\cdot] = \int_{\mathbb{M}} [\cdot] dPr_\theta(\boldsymbol{p}). \tag{6.1}$$

Using (6.1) the average profile $\boldsymbol{p}^o$ is,

$$\boldsymbol{p}^o = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{p}^i = E[\boldsymbol{p}]. \tag{6.2}$$

The expected cross products define Π, an $M \times M$ Gramian matrix, as

$$\Pi = \int_{\mathbb{M}} \boldsymbol{p}\boldsymbol{p}^T dPr_\theta[\boldsymbol{p}], \tag{6.3}$$

and, as can be shown (Woodbury *et al.* (1994)),

THEOREM 6.1. *The rank of* Π *is the dimension of* $\mathbb{B}$.

If $\mathbb{B}$ is dimension d, there are basis vectors $\boldsymbol{V}^1, \boldsymbol{V}^2, \ldots, \boldsymbol{V}^d$ where, for every $\boldsymbol{p} \in \mathbb{B}$, there exists an unique set of coefficients $\beta_m(\boldsymbol{p})$ such that $\sum_{m=1}^{d} \beta_m(\boldsymbol{p}) \boldsymbol{V}^m = \boldsymbol{p}$. For coordinates $\boldsymbol{V} = (\boldsymbol{V}^1, \boldsymbol{V}^2, \ldots, \boldsymbol{V}^d)$, $\boldsymbol{p} \leftrightarrow (\beta_k(\boldsymbol{p}), k = 1, 2, \ldots, d)$, with $\boldsymbol{V}_k$ an element in $\mathbb{B} \subseteq \mathbb{M}$. Thus,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} p^i_{jl} = p^o_{jl} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} y^i_{jl} \tag{6.4}$$

is consistently estimated, and defines a unique point in $\mathbb{B}$. The elements of Π,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} p^i_{jl} p^i_{rs} = \Pi^{ls}_{jr}, \tag{6.5}$$

exist and have rank d if $\boldsymbol{p}_i \epsilon \mathbb{B}$ for all i, $(i = 1, 2, \ldots, \ldots)$. Since p^i_{jl} are unobserved, for $j \neq r$, assuming conditional independence,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} y^i_{jl} y^i_{rs} = \Pi^{ls}_{jr}, \tag{6.6}$$

and for $j = r$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} y^i_{jl} y^i_{rs} = p^o_{jl}\delta_{ls} \neq \Pi^{ls}_{jj}. \tag{6.7}$$

Hence, the matrix of empirical second moments will *not* provide an estimate of Π. However, Π can be partitioned,

$$\Pi = \begin{pmatrix} \Pi^{11} & \Pi^{12} & \cdots & \Pi^{1J} \\ \Pi^{21} & \Pi^{22} & \cdots & \Pi^{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \Pi^{J1} & \Pi^{J2} & \cdots & \Pi^{JJ} \end{pmatrix}, \tag{6.8}$$

where $\Pi_{jr} = (\Pi^{ls}_{jr})(l = 1, 2, \ldots, L_j; s = 1, 2, \ldots, L_r)$. If Π_{jj} is replaced by a matrix with p^o_{jl} on the diagonal (i.e., $\Pi^{*ls}_{jj} = \delta_{ls} p^o_{jl}$) then,

$$\Pi^{*jr}_{ls} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} y^i_{jl} y^i_{rs}. \tag{6.9}$$

Given $\boldsymbol{V}$ and $\beta^i_m = \beta_m(\boldsymbol{p}^i)$ for all i, m, replace (6.4) by,

$$p^i_{jl} = \sum_{k=1}^{d} \beta^i_k V^k_{jl}, \tag{6.10}$$

and (6.5) can be written,

$$\frac{1}{N} \sum_{i=1}^{N} p^i_{jl} p^i_{rs} = \sum_{k=1}^{K} V^k_{jl} \sum_{m=1}^{d} V^m_{rs} \frac{1}{N} \sum_{i=1}^{d} \beta^i_k \beta^i_m. \tag{6.11}$$

Since the limit for $\boldsymbol{P}$ exists as $N \to \infty$ so does the limit on the right, i.e., $\Pi = \boldsymbol{VBV}^T$ where $\boldsymbol{B}$ is symmetric and non-negative definite. Consequently, Π is symmetric, non-negative definite and is of rank at most d since V is rank d. Π^* is not equal to Π in the diagonal blocks because $y^i_{jl} y^i_{jl'} = 0$ for $l \neq l'$. The data, however, can be used to estimate Π^*, i.e.:

THEOREM 6.2. *To estimate* Π *from* Π^* *it is sufficient that* $K < J/2$ *and a non-singular submatrix of* Π *of size* $K \times K$ *exists containing no elements in diagonal blocks.*

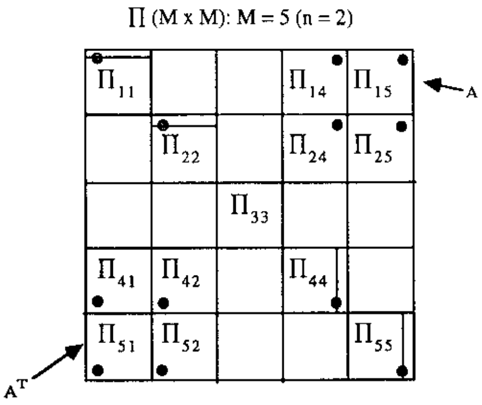

Fig. 3. Schematic of Π.

To prove Theorem 6.2, Woodbury *et al.* (1994) showed there exists a non-singular $K \times K$ submatrix of Π, and Π^*, $\boldsymbol{A}$ and mappings ψ_r and ψ_c such that ψ_r uniquely maps rows of $\boldsymbol{A}$ into rows of Π, and ψ_c maps columns of $\boldsymbol{A}$ into columns of Π. Rows of $\boldsymbol{A}$ have no elements in common with rows of $\boldsymbol{A}^T$ (otherwise an element of $\boldsymbol{A}$ is in a diagonal block where $\Pi^* \neq \Pi$). No two rows of $\boldsymbol{A}$ are in the same block of rows of Π and no two columns of $\boldsymbol{A}$ are in the same block of columns of Π. We can order variables so row blocks of Π containing rows of $\boldsymbol{A}$ come first and column blocks of Π containing columns of $\boldsymbol{A}$ come last if $K < J/2$.

In Fig. 3, $\boldsymbol{A}$ and $\boldsymbol{A}^T$, indicated by dots, are non-singular, rank d. We wish $\{\Pi_{ij}\} i \neq j$ to define Π_{jj} uniquely, with rank d. Select an element in Π_{jj} containing no rows of $\boldsymbol{A}$ and no columns of $\boldsymbol{A}^T$. Enlarge $\boldsymbol{A}$ to $\boldsymbol{A}^+$ where elements in Π_{jj} defines additional rows and columns of $\boldsymbol{A}^+$. $\boldsymbol{A}^+$ contains only one element in Π_{jj}, and is a rank d, singular submatrix of Π of size $(I+1) \times (I+1)$. All elements of $\boldsymbol{A}^+$ are known except, a_{rc}^{++}. Thus, $\boldsymbol{A}^+$ satisfies,

$$\det \begin{bmatrix} a_c^+ \boldsymbol{A} \\ a_{rc}^{++} a_r^+ \end{bmatrix} = 0 \quad \text{but} \quad \det \begin{bmatrix} a_c^+ \boldsymbol{A} \\ a_{rc}^{++} a_r^+ \end{bmatrix} = \det \boldsymbol{A} \det(a_{rc}^{++} - a_r^+ \boldsymbol{A}^{-1} a_c^+)$$

and $\det \boldsymbol{A} \neq 0$. Consequently,

$$a_{rc}^{++} = a_r^+ A^{-1} a_c^+ .$$

This defines Π_{jj} except for d rows, one in each Π_{jj} corresponding to rows of $\boldsymbol{A}$ and d columns in Π_{jj}, which are columns of $\boldsymbol{A}$ in Π. But $\Pi_{ij} = \Pi_{ij}^T$ so at most $2d$ elements in Π_{jj} are not evaluated. If other $d \times d$ matrices satisfy conditions for $\boldsymbol{A}$, the number of unidentified elements is reduced. These are in Π_{jj}. The principal minor of Π is $2d \times 2d$ with diagonal matrices C_r and C_c,

$$\boldsymbol{A}^* = \begin{bmatrix} C_r & \boldsymbol{A} \\ \boldsymbol{A}^T & C_c \end{bmatrix} .$$

$\boldsymbol{A}^*$ is rank K and size $2d \times 2d$ with C_r and C_c unknown. A non-negative, definite $M \times M$ matrix can be transformed to a matrix with rank $[\frac{M+1}{2}]$ or less with unique diagonal values (Albert (1944); Woodbury *et al.* (1994)). But $M = 2d$, so diagonal values can be found to make the rank of $\boldsymbol{A}^*$ at most $[\frac{2d+1}{2}]$. But $\boldsymbol{A}^*$ is of rank at least d by containing the non-singular submatrix $\boldsymbol{A}$ of rank d.

7. Properties of the likelihood function

Suppes and Zanotti (1981) show, for discrete variables, latent variables can be generated so that the J responses are conditionally independent. Let g_k^i denote such latent variables. Then,

$$(7.1) \qquad Pr[X_j^i = x_j^i, j = 1, 2, \dots, J \mid \boldsymbol{g}_k^i] = \prod_{j=1}^{J} Pr[X_j^i = x_j^i \mid \boldsymbol{g}_k^i],$$

where the data are parameterized by g_k^i and λ_{jl}^k, and K is the number of g_k^is needed for x_j^i to be independent. The conditional likelihood is,

$$(7.2) \qquad L_I^+ = \prod_{i=1}^{I} \prod_{j=1}^{J} \sum_{k_j=1}^{K} g_{k_j}^i \lambda_j^{k_j}(x_j^i).$$

The unconditional likelihood is derived from (7.2), using the law of total probability, and integrating $\{g_k^i\}$ with respect to H, their distribution function. Tolley and Manton (1992) replace g_k^i by variables of integration ξ_k^i (where $\xi_k^i \geq 0$, $\sum_k \xi_k^i = 1$) producing,

$$(7.3) \qquad L_I = \prod_{i=1}^{I} \int_{S_i} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \cdots \sum_{k_J=1}^{K} \prod_{j=1}^{j} (\xi_{k_j}^i) \prod_{j=1}^{J} \lambda_j^{k_j}(x_j^i) dH(\xi),$$

and $\xi^i = (\xi_1^i, \dots, \xi_K)^T$. From Theorem 4.1, the ξ^i are assumed to fall in $\mathbb{B}$. The g_k^i are determined from a map of realizations of ξ_k^i. For a fixed map $w \in W$ from $\mathbb{B}$ to G we determine a $\boldsymbol{g}$ for each realization of ξ. H denotes the induced distribution function of $\boldsymbol{g}$ from w and $F(\boldsymbol{p})$. If $\mathbb{S}_i$ is the regular simplex defined by constraints on ξ_k^i, the moments of order k, $\mu_{\boldsymbol{k}}$, of H, are,

$$(7.4) \qquad \mu_{\boldsymbol{k}} = \int_{\mathbb{S}_i} \prod_{j=1}^{J} g_{k_l}^i dH(g),$$

where $\boldsymbol{k} = (k_1, k_2, \dots, k_J)$. The $\mu = (\mu_{\boldsymbol{k}}, (k_1 = 1, 2, \dots, K), (k_2 = 1, 2, \dots, K), \dots, (k_J = 1, 2, \dots, K))$ are moments of $H(g)$ and invariant over permutations of $\boldsymbol{k}$. Hence, (7.4) may be written,

$$(7.5) \qquad L = \prod_{i=1}^{N} \sum_{\boldsymbol{k}} \mu_{\boldsymbol{k}} \Lambda^{\boldsymbol{k}}(\boldsymbol{x}^i)$$

where $\Lambda^k(\boldsymbol{x}^i) = \lambda_i^{k_1}(x_1^i) \otimes \lambda_2^{k_2}(x_2^i) \otimes \cdots \otimes \lambda^{k_J}(x_J^i)$ is the Kroneker product $\lambda_j^k(x_j^i)$, $(j = 1, 2, \ldots, J)$. Λ is dimension K^J by $L = L_1 \times L_2 \times \cdots \times L_J$. The rank of Λ is the product of the ranks of $\Lambda_j = (\lambda_{jl}^k(l = 1, 2, \ldots, L_j), (k = 1, 2, \ldots, K))$. Denoting the rank of Λ_j as R_j, where $R_j \leq K$ and $R_j \leq L_j$., the rank of Λ is the product $R_1 \times R_2 \times \cdots \times R_J$.

Because λ_{jl}^i are unique, moments of $f(\boldsymbol{p})$, μ_k, can be unbiasedly estimated by estimating sample moments of g_k^i and transforming (Woodbury *et al.* (1994)). Let μ_g be the vector of moments of g given a unique map of ξ_k^i to the g_k^i. Define a decomposition of Λ_j $(\lambda_{jl}^k(l = 1, 2, \ldots, L_j), (k = 1, 2, \ldots, K))$ to eliminate singularities, i.e.,

$$\Lambda = (U^{(1)} V_1^T) \otimes (U^{(2)} V_2^T) \otimes \cdots \otimes (U^{(J)} V_j^T),$$

which can be written,

$$\Lambda = (U^{(1)} \otimes U^{(2)} \otimes \cdots \otimes U^{(J)})(V_1^T \otimes V_2^T \otimes \cdots \otimes V_J^T) = UV^T,$$

where U and V^T have maximum rank $R_1 \times R_2 \times \cdots \times R_j = R$. The vector $\mu_{\boldsymbol{k}}$ is $1 \times K^J$ but has at most $R^* = (J + K - 1)!/J!(K - 1)!$ distinct values due to symmetries (Tolley and Manton (1992)). Δ is an $R^* \times K^J$ matrix with a one in each column, with rank R^* relating μ_g^* and μ_g. Δ is full rank and multiple row entries correspond to identical rows of U. Woodbury *et al.* (1994), show,

THEOREM 7.1. *The unconditional likelihood with $\Lambda^K = (UV^T)$, and Λ forming μ_g^* and μ_g, is;*

$$L_i = \mu_g^* \Delta(UV^T)(y_i). \tag{7.6}$$

One could evaluate the multinomial likelihood in a more general form, i.e.,

$$L = \left(\frac{\sum_k g_k^i \lambda_{jl}^k}{\sum_l \sum_k g_k^i \lambda_{jl}^k} \right)^{y_{jl}^i}. \tag{7.7}$$

Which differs by the term $\sum_k (g_k^i \sum_l \lambda_{jl}^k)$, i.e., we drop the constraint, $\sum_l \lambda_{jl}^k = 1$. In this model $\sum_l \lambda_{jl}^k$ is the weight of the variable on the K-th profile. As this approaches 0.0 the variable no longer contributes to p_{jl}^i through the K-th profile. Thus, the model can fit the data with fewer dimensions by allowing extreme points to have zero weight on irrelevant variables. This is one extension of basic GoM.

8. Example

To illustrate the GoM implementation of the convex set model we applied it to a data set on nursing home residents in six U.S. states. For 4,525 residents over age 65 in those nursing homes, 111 measurements were made to describe in detail their health and functional characteristics. The purpose of the data collection was to assess the levels of different services delivered to specific types of nursing home residents. One issue assessed was whether specific resident groups with very complex health problem received adequate levels and mixes of services. A continuation of the study will examine differences in outcomes for nursing home residents (Manton *et al.* (1995)).

To evaluate the 4,525 patients on the 111 measures we used three models which differ in how their membership functions represent heterogeneity. One model *not* investigated was the log linear model for categorical data (e.g., Bishop *et al.* (1975)). It was not possible to exploit such a large number of variables, or their interactions, with those procedures.

The three models applied were a.) latent class model (LCM) given by Lazarsfeld and Henry (1968), b.) simple GoM model given by (4.1), and c.) a GoM model with heterogeneity allowed between variables given by (7.7). In LCM the fundamental characteristic is that only one g_k^i for each person is allowed to be 1.0—all others are 0. Thus, in LCM the K groups are exclusively defined with no within group heterogeneity. In the simple GoM model the g_k^i is allowed to vary over individuals so that an individual's characteristics may be described as a convexly weighted function of the λs for more than one dimension. If, for a given value of K, there is no significant within group variation, then the g_k^is in simple GoM will be 0 or 1 within the level of statistical precision. In that case LCM and simple GoM describe the data equally well. The third model was an extension of GoM where additional heterogeneity was represented by the differential weighting of a variable by the term $(g_k^i \sum_l \lambda_{jl}^k)$ in (7.7).

In the analysis 111 variables were used. The difference in fit between LCM and simple GoM for $K = 11$ was highly significant (Wilson-Hilferty $t = 34.8$) so that the hypothesis that all $g_k^i = 0$ or 1 could be rejected, i.e., within class heterogeneity was significant. The difference between simple and extended GoM (Wilson-Hilferty $t = 135.7$) was also highly significant, i.e., the hypothesis that there were no differences in the weight of the contribution of individual variables to the K dimensions was rejected.

In Table 1 we present λ estimates for 22 of the 111 variables. These medical conditions reflect typical differences in λ_{jl}^k estimates for the three models. In Table 1 we also present a column containing the proportion (%) of the sample with an attribute. The solution for all three models used $K - 11$. The 11 dimensions or profiles in Table 1 are, for convenience, labelled by identifying the variables with the most salient λ_{jl}^ks for each. The λ_{jl}^ks estimated for each of the three models are referred to by LCM (latent class model), GoM I (simple GoM), and GoM II (extended GoM). The values in parentheses are the weight $(g_k^i \sum_l \lambda_{jl}^k = w_j)$ of each variable in defining a profile for extended GoM.

In Table 1, for LCM, most λ_{jl}^k for all 11 dimensions were nonzero, i.e., all

Table 1. The λ_{jl}^k estimates for three latent variable models estimated using a convex set parameterization applied to 4,525 elderly nursing home residents in six states.

Profile Number; K =			1	2	3	4	5	6	7	8	9	10	11
			Labels for Profiles of λ_{jl}^k										
Medical Condition	Type of Model	Proportion (%) of Sample	Major Stroke	Depressed	CVD Diabetes Dementia	Mid-Stage Dementia	Late Stage Dementias	Alzheimer's Disease	Comatose & Pulmonary	Multiple Disease, Osteoporosis & Fracture	Debilitated CVD	CVD Convalescence	Stable
1.) Arthritis	LCM	30.0	26.8	36.1	36.4	33.4	21.0	20.2	26.4	30.8	39.2	36.0	23.2
	GoM I		4.1	4.2	25.1	45.0	0.0	26.1	22.5	98.4	100.0	13.5	9.2
	GoM II		0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	$(w_1)^1$		(1.05)	(0.74)	(0.30)	(1.07)	(0.89)	(0.93)	(0.00)	(6.94)	(0.76)	(0.00)	(0.78)
2.) Anemia	LCM	11.1	10.6	12.8	9.7	13.7	7.0	10.6	8.7	14.2	14.9	11.9	8.9
	GoM I		0.0	47.4	7.6	3.4	0.9	19.6	5.5	12.9	38.5	7.9	6.0
	GoM II		0.6	0.0	0.0	0.0	9.0	0.0	0.0	100.0	10.1	0.0	11.0
	(w_2)		(0.95)	(0.95)	(1.02)	(0.95)	(1.01)	(0.97)	(1.02)	(1.27)	(0.97)	(0.99)	(1.01)
3.) Anxiety	LCM	4.6	1.8	5.8	5.3	15.7	3.1	1.2	4.2	4.5	3.9	6.9	1.1
	GoM I		0.0	0.0	0.0	66.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	GoM II		0.0	55.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	(w_3)		(1.01)	(1.07)	(0.86)	(0.95)	(1.00)	(1.00)	(1.07)	(1.04)	(1.02)	(0.96)	(0.98)
4.) Alzheimer's	LCM	13.1	5.4	0.9	3.5	2.4	24.7	19.2	41.2	12.5	1.3	7.8	30.9
	GoM I		0.0	0.0	0.0	0.0	34.0	0.0	62.8	0.0	0.0	0.0	34.8
	GoM II		0.0	0.0	0.0	14.4	0.0	100.0	0.0	0.0	0.0	0.0	23.8
	(w_4)		(1.20)	(1.04)	(1.30)	(0.76)	(0.78)	(1.03)	(1.89)	(1.10)	(0.82)	(1.48)	(0.90)
5.) Other Dementia	LCM	35.7	35.2	10.9	16.3	31.5	50.3	51.2	49.4	47.4	9.9	40.2	56.7
	GoM I		0.0	0.0	10.1	20.1	82.7	51.6	54.0	80.7	0.0	21.4	57.6
	GoM II		0.0	3.8	100.0	40.9	71.9	52.1	0.0	0.0	0.0	0.0	72.0
	(w_5)		(1.35)	(0.95)	(0.43)	(0.95)	(1.22)	(1.24)	(0.04)	(1.04)	(1.19)	(0.10)	(0.97)
6.) Aphasia	LCM	4.4	5.5	2.1	1.0	1.7	0.1	20.4	1.9	4.9	3.2	1.9	2.3
	GoM I		6.2	0.0	0.0	0.0	0.0	45.9	0.0	0.0	0.0	0.0	0.0
	GoM II		100.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	(w_6)		(0.72)	(1.02)	(1.03)	(0.77)	(1.00)	(0.96)	(1.39)	(1.03)	(1.00)	(1.22)	(0.98)

Table 1. (continued).

Profile Number; K =			1	2	3	4	5	6	7	8	9	10	11
			Label for Profiles of λ_{jl}^{k}										
Medical Condition	Type of Model	Proportion (%) of Sample	Major Stroke	Depressed	CVD Diabetes Dementia	Mid-Stage Dementia	Late Stage Dementias	Alzheimer's Disease	Comatose & Pulmonary	Multiple Disease. Osteoporosis & Fracture	Debilitated CVD	CVD Convalescence	Stable
7.) Cancer	LCM	6.4	6.2	9.6	6.7	5.6	5.0	4.7	4.5	10.0	8.3	6.7	2.7
	GoM I		0.0	100.0	5.7	0.0	0.0	3.7	2.9	0.0	0.0	0.0	0.0
	GoM II		0.2	0.0	0.0	14.3	0.3	9.5	0.0	67.7	11.3	0.0	0.0
	(w_7)		(0.91)	(0.91)	(1.38)	(0.93)	(0.88)	(0.93)	(1.43)	(0.91)	(0.91)	(1.56)	(1.00)
8.) Cataracts	LCM	6.3	1.8	9.5	7.0	7.1	5.0	2.6	5.4	9.3	8.5	7.4	6.4
	GoM I		0.0	0.0	2.1	5.0	0.0	0.0	0.0	45.8	28.5	0.0	0.0
	GoM II		0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	(w_8)		(1.39)	(1.02)	(0.76)	(0.97)	(1.03)	(1.01)	(0.71)	(1.45)	(1.04)	(0.74)	(0.99)
9.) Stroke	LCM	22.3	42.8	30.6	10.3	13.4	7.1	44.8	6.3	34.5	23.6	21.8	8.8
	GoM I		100.0	0.0	0.0	0.0	0.0	98.3	0.0	0.0	42.3	3.6	0.0
	GoM II		100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	(w_9)		(3.81)	(0.70)	(1.06)	(0.64)	(0.87)	(1.05)	(0.00)	(1.47)	(0.84)	(0.28)	(0.79)
10.) Congestive Heart Failure	LCM	24.0	24.0	26.8	26.3	30.0	13.7	21.8	14.7	25.3	30.3	30.6	18.8
	GoM I		0.0	0.0	12.1	41.1	0.0	35.1	0.0	100.0	100.0	6.2	3.5
	GoM II		0.0	4.2	82.3	2.8	0.0	0.0	100.0	100.0	0.0	100.0	16.9
	(w_{10})		(1.16)	(0.78)	(1.08)	(0.82)	(0.93)	(1.19)	(1.48)	(0.67)	(0.94)	(1.58)	(0.93)
11.) ASHD	LCM	21.1	18.6	21.5	21.4	26.3	16.6	19.9	21.1	23.0	21.4	22.9	20.9
	GoM I		0.0	0.0	13.5	35.2	2.5	27.7	11.2	88.7	61.1	9.5	6.7
	GoM II		0.0	6.0	98.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	20.2
	(w_{11})		(1.09)	(0.90)	(0.78)	(0.97)	(1.00)	(1.08)	(0.99)	(0.96)	(1.08)	(1.08)	(0.96)
12.) Peripheral Vascular Disease	LCM	8.0	7.5	13.8	7.3	8.1	4.6	6.9	3.4	12.1	9.9	10.7	4.2
	GoM I		20.6	34.1	3.0	14.3	0.0	14.1	0.0	0.0	16.9	11.5	0.0
	GoM II		23.4	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	(w_{12})		(0.88)	(0.93)	(0.96)	(0.92)	(0.97)	(1.03)	(0.91)	(1.58)	(0.98)	(0.99)	(1.02)

Table 1. (continued).

Profile Number; K =			1	2	3	4	5	6	7	8	9	10	11
			Labels for Profiles of λ_{jl}^{k}										
Medical Condition	Type of Model	Proportion (%) of Sample	Major Stroke	Depressed	CVD Diabetes Dementia	Mid-Stage Dementia	Late Stage Dementias	Alzheimer's Disease	Comatose & Pulmonary	Multiple Disease, Osteoporosis & Fracture	Debilitated CVD	CVD Convalescence	Stable
13.) Other Cardiovascular Disease	LCM	20.7	21.8	19.6	21.6	27.6	17.6	20.0	13.7	22.2	24.2	23.4	15.1
	GoM I		5.9	0.0	13.6	33.5	6.4	35.5	7.9	83.5	69.0	0.0	0.0
	GoM II		30.9	0.0	100.0	0.0	0.0	0.0	100.0	20.4	0.0	100.0	10.6
	(w_{13})		(0.87)	(0.88)	(0.92)	(0.99)	(0.99)	(1.02)	(1.34)	(0.64)	(1.06)	(1.32)	(0.99)
14.) Depression	LCM	9.1	7.2	18.2	6.4	23.0	4.4	4.2	4.7	15.9	10.0	9.7	1.8
	GoM I		20.2	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	GoM II		0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	(w_{14})		(0.74)	(1.17)	(1.26)	(1.00)	(0.97)	(0.93)	(1.21)	(0.86)	(0.90)	(1.45)	(1.00)
15.) Diabetes Mellitus	LCM	17.6	19.2	22.1	19.1	20.2	15.5	17.9	8.3	19.8	21.6	18.3	10.1
	GoM I		42.8	30.8	15.9	11.0	19.5	20.4	0.0	18.7	36.3	5.7	7.9
	GoM II		40.7	0.0	96.5	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
	(w_{15})		(0.75)	(0.91)	(0.83)	(0.95)	(1.01)	(1.11)	(1.54)	(0.98)	(1.04)	(1.00)	(0.95)
16.) Emphysema/Asthma/ COPD	LCM	10.7	8.5	15.7	11.4	20.4	8.4	4.4	6.1	11.8	15.0	11.9	6.3
	GoM I		0.0	0.0	0.0	77.3	0.0	17.8	0.0	0.0	47.2	0.0	0.8
	GoM II		0.0	0.0	0.0	90.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	(w_{16})		(1.12)	(0.59)	(0.99)	(4.29)	(0.91)	(0.99)	(1.15)	(0.67)	(0.87)	(0.88)	(0.95)
17.) Hypertension	LCM	26.9	27.8	34.6	33.1	30.6	24.6	22.6	16.3	27.0	31.6	30.1	14.6
	GoM I		60.6	15.5	30.9	25.9	20.7	33.6	14.6	31.2	53.5	16.0	4.0
	GoM II		86.8	0.0	100.0	0.0	0.0	0.0	100.0	41.4	0.0	100.0	0.0
	(w_{17})		(0.73)	(0.91)	(1.29)	(0.92)	(0.93)	(1.01)	(1.83)	(0.62)	(0.92)	(1.43)	(0.99)
18.) Hypothyroidism	LCM	7.1	5.2	8.4	6.7	9.7	8.2	6.3	9.9	5.4	7.9	7.2	4.4
	GoM I		0.0	76.1	5.5	0.0	20.9	13.1	0.0	0.0	0.0	0.0	0.0
	GoM II		0.0	0.0	57.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	4.1
	(w_{18})		(1.08)	(1.10)	(0.61)	(1.03)	(1.11)	(1.08)	(0.47)	(1.01)	(1.12)	(0.55)	(0.99)

Table 1. (continued).

Profile Number; K =			1	2	3	4	5	6	7	8	9	10	11
			Labels for Profiles of λ_{jl}^{k}										
Medical Condition	Type of Model	Proportion (%) of Sample	Major Stroke	Depressed	CVD Diabetes Dementia	Mid-Stage Dementia	Late Stage Dementias	Alzheimer's Disease	Comatose & Pulmonary	Multiple Disease, Osteoporosis & Fracture	Debilitated CVD	CVD Convalescence	Stable
19.) Osteoporosis	LCM	8.5	7.8	12.6	8.4	11.2	5.2	6.2	7.7	7.2	10.8	109	6.5
	GoM I		0.0	15.8	4.8	12.1	0.0	0.0	0.0	56.1	26.5	00	0.0
	GoM II		0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	00	0.0
	(w_{19})		(1.19)	(0.88)	(0.97)	(1.01)	(0.96)	(1.03)	(0.73)	(1.97)	(0.94)	(0.9)	(0.93)
20.) Parkinson's Disease	LCM	6.4	8.6	7.9	4.6	5.7	4.5	8.1	3.8	7.3	5.0	67	8.5
	GoM I		0.0	31.6	5.2	0.0	5.3	7.6	0.0	9.5	3.5	73	10.3
	GoM II		9.0	0.0	0.0	0.0	9.2	1.2	14.7	5.0	7.2	72	9.4
	(w_{20})		(0.98)	(0.96)	(1.10)	(0.99)	(0.97)	(0.98)	(1.13)	(0.95)	(0.98)	(1.55)	(1.00)
21.) Pneumonia	LCM	1.8	0.7	2.5	1.2	1.1	1.0	5.0	1.3	2.9	0.8	20	0.1
	GoM I		0.0	18.8	0.0	0.0	0.0	12.0	0.0	0.0	0.0	00	0.0
	GoM II		0.0	0.0	0.0	44.5	0.0	0.0	0.0	0.0	0.0	00	0.0
	(w_{21})		(0.89)	(0.98)	(0.95)	(1.45)	(1.01)	(1.00)	(0.93)	(0.99)	(1.02)	(0.91)	(0.99)
22.) Septicemia	LCM	0.3	0.5	0.6	0.0	0.0	0.0	1.0	0.0	1.1	0.0	01	0.0
	GoM I		0.4	4.4	0.0	0.0	0.0	1.5	0.0	0.0	0.0	00	0.0
	GoM II		2.1	0.0	0.0	0.0	0.0	0.0	4.6	0.0	0.0	00	0.0
	(w_{22})		(0.98)	(1.00)	(1.00)	(0.97)	(1.00)	(1.00)	(1.00)	(0.97)	(1.01)	(0.98)	(1.00)

Source: Nursing Home Case-Mix and Quality Demonstration (Manton *et al.* (1995)).
[1]Variable weights for extended GoM.

variables had effects on all 11 dimensions. Thus, the λ^k_{jl} estimated using LCM did not highly discriminate the profiles. For example, arthritis (variable 1) was reported by 30% of the sample. The λ^k_{jl} for LCM only varied between 20.2% (profile 6) and 39.2% (profile 9). For seven of 11 profiles the contribution of arthritis varied by only ±20% of the sample proportion (i.e., 30% ± 6%). The simple GoM, λ^k_{jl}s range from 0.0% (profile 5) to 100.0% (profile 9). Thus, λ^k_{jl} for simple GoM varied more across profiles—and better discriminated their substance. For extended GoM arthritis contributed 100% to profile 8. For eight other profiles $\lambda^k_{jl} = 0.0$—but with a $w_{j1} \neq 0$. The largest w_1 (6.94) was for profile 8 meaning arthritis contributed most heavily to the definition of that profile. For two profiles (7 and 10) $w_1 = 0.0$, i.e., arthritis did not provide any information to determine those two profiles.

The patterns of the λ^k_{jl}s for each model for the variables in Table 1 were similar. In LCM, because no within group heterogeneity is represented, the λ^k_{jl}s tend to be less discriminating (i.e., closer to the proportion of the sample reporting the condition). The λ^k_{jl} for GoM I are more discriminating, but there is still considerable diffusiveness—especially for conditions where the proportion reporting the condition in the sample is relatively large, i.e., the most frequent conditions. This is apparently due to the differential contribution of each variable to each profile. Hence, in GoM II, where variables are differentially weighted, the λ^k_{jl} profiles are more distinct and we see which variables contribute most to specific profiles. In some cases (e.g., anemia; variable 2) the w_2 are all near 1.0—though the λ^k_{jl} patterns are distinct. In other cases (e.g., stroke; variable 9) the variable dominated one profile (i.e., w_9 is 3.81 on profile 1).

The GoM II results were used in two ways (Manton *et al.* (1995)). First, the λ^k_{jl}s were examined by physician panels to identify categories of patients requiring specialized health services. Second, the 11 profiles were used to predict the actual consumption of services by nursing home residents. In the first application the analysis permitted more thorough examination of the inter-relation of the 111 variables than could be done by hand. In the second case there was considerable variation in the types and amounts of services predicted used by persons in different groups.

9. Summary

A convex set model representing high (but finite) dimensional categorical data is defined with an invariant metric. The intersection of the linear parameter space $L_{\mathbb{B}}$, and $\mathbb{M}$, is unique if $L_{\mathbb{B}}$ is unique (Weyl (1949)). $L_{\mathbb{B}}$ is uniquely defined by Π estimates (for $K < J/2$) derived from binary variables using a Nicodym-Radon integral for expected values. The intersection of $L_{\mathbb{B}}$ and $\mathbb{M}$ is a convex set, dually represented by extreme points and bounding half spaces (Weyl (1949)). Model parameters identify the extreme and bounding half spaces of the convex set $\mathbb{B} = L_{\mathbb{B}} \cap \mathbb{M}$ if values of λ^k_{jl} are nonnegative and bounded by 1.0 (i.e., they are in $\mathbb{M}$). This represents categorical data by unique individual scores and extreme points.

The model is useful when a large number of discrete measures on heterogeneous populations are made, e.g., describing the health and functioning of elderly populations where individuals may have multiple medical conditions and physical, mental and sensory impairments. This was illustrated in an analyses of 4,525 nursing home residents characterized by complex health and functional states. Extended GoM did better in describing the individual heterogeneity of health characteristics than did two alternate models.

In such cases, the identification of latent *homogeneous* classes, because of the high dimensionality of the data, suffers if large numbers of classes must be defined with memberships too small to be stable. In extended GoM, convex weights are used to represent individual parameters as combinations of a small number of extreme profiles. Thus, the smoothing conditions implied by the Neyman-Pearson lemma (see Keifer and Wolfowitz (1956) or discussion in Neyman and Scott (1948)) are derived, not by assuming homogeneous classes, but by estimating weights in a low dimensional sample space, i.e., by locating persons relative to extreme points in the convex sets.

REFERENCES

Albert, A. (1944). The matrices of factor analysis, *Proc. Nat. Acad. Sci.*, **30**, 90–95.

Bishop, Y., Fienberg, S. and Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Mass. and London, England.

Brondsted, A. (1983). *An Introduction to Convex Polytopes*, Springer-Verlag, New York.

Hahn, H. and Rosenthal, A. (1948). *Set Functions*, Albuquerque, New Mexico, University of New Mexico Press.

Karlin, S. and Shapley, L. S. (1953). Geometry of Moment Spaces, *Memoirs of the American Mathematical Society*, Providence, RI, American Mathematical Society.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many parameters, *Ann. Math. Statist.*, **27**, 887–890.

Kleinman, J. C. (1973). Proportions with extraneous variance: Single and Independent samples, *J. Amer. Statist. Assoc.*, **68**, 46–54.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution, *J. Amer. Statist. Assoc.*, **73**, 805–811.

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*, Houghton Mifflin, Boston.

Little, R. J. A. and Rubin, D. B. (1986). *Statistical Analysis with Missing Data*, Wiley, New York.

Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason, J. P. and Pellom, A. (1989). Empirical Bayes procedures for stabilizing maps of US cancer mortality rates, *J. Amer. Statist. Assoc.*, **84**, 637–650.

Manton, K. G., Cornelius, E. S. and Woodbury, M. A. (1995). Nursing home residents: A multivariate analysis of their medical, behavioral, psycho-social, and service use characteristics, *Journal of Gerontology: Biological Sciences & Medical Sciences*, **50**, M242–M251.

Neyman, J. and Scott, E. L. (1948). Consistent estimators based on partially consistent observations, *Econometrica*, **16**, 1–32.

Orchard, G. and Woodbury, M. A. (1971). A missing information principle: Theory and application, *Sixth Berkeley Symposium on Mathematical Statistics and Probability* (eds. L. M. LeCam, J. Neyman and E. L. Scott), pp. 697–715, Berkeley, University of California Press, California.

Suppes, P. and Zanotti, M. (1981). When are probabilistic explanations possible?, *Synthese*, **48**, 191–199.

Tolley, H. D. and Manton, K. G. (1992). Large sample properties of estimates of discrete grade of membership model, *Ann. Statist. Math.*, **44**, 85–95.

Tsutakawa, R. K., Shoop, G. L. and Marienfeld, C. J. (1985). Empirical bayes estimation of cancer mortality rates, *Statistics in Medicine*, **4**, 201–212.

Weyl, H. (1949). The elementary theory of convex polyhedra, *Ann. Math. Stud.*, **24**, 3–18.
Woodbury, M. A., Manton, K. G. and Tolley, H. D. (1994). A general model for statistical analysis using fuzzy sets: Sufficient conditions for identifiability and statistical properties, *Inform. Sci.*, **1**, 149–180.