

A New Efficient Algorithm for Construction of GoM Models*

(Extended Abstract)

Mikhail Kovtun Igor Akushevich Kenneth G. Manton

Center for Demographic Studies, Duke University, Durham, NC

H. Dennis Tolley

Department of Statistics, Brigham Young University, Provo, UT

September 23, 2004

Abstract

We present a new efficient algorithm for construction of grade of membership (GoM) models. This algorithm reduces a problem of estimation of model parameters to a sequence of problems of linear algebra, which assures a low computational complexity and ability to handle on desktop computers data that involve up to thousands of variables.

The class of grade of membership (GoM) models belongs to a family of latent structure models, which, in turn, is a subfamily of a family of mixed distribution models. Such models naturally occur when a population of interest is supposed to be heterogeneous.

The most widely used methods for estimation latent structure models are based on maximization of the likelihood function. These are well established methods possessing many good properties. Nevertheless, they have limitations, which may restrict or even prevent their usage. First, the number of parameters to be optimized is proportional to the number of variables (measurements), which in practice limits the number of variables used in the analysis to several dozens. Second, the likelihood function in the case of latent structure analysis is often multimodal, which requires usage of additional techniques to ensure that the absolute maximum is found.

Our algorithm is based on methods of linear algebra, which eliminates the problem of multimodality and allows us to analyze up to thousands of variables. The time spent by the algorithm is proportional to the cube of the number of variables.

Basic notions. GoM analysis considers J discrete measurements, represented by a vector of random variables $X = (X_1, \dots, X_J)$, with the set of outcomes of j^{th} measurement (i.e. the set of possible values of random variable X_j) being $\{1, \dots, L_j\}$. We consider a distribution law of this random vector as a mixture of *independent* distribution laws, i.e. distribution laws satisfying

$$\Pr(X_1 = l_1 \wedge \dots \wedge X_J = l_J) = \prod_j \Pr(X_j = l_j) \quad (1)$$

Representation of the observed distribution law as a mixture of independent distribution laws is standard for latent structure analysis (and it is its defining characteristic).

Due to (1), description of independent distribution law requires only knowing $\Pr(X_j = l_j)$. Thus, an independent distribution law may be described by $|L| = L_1 + \dots + L_J$ -dimensional vector $\beta = (\beta_{jl})_{jl}$, where j ranges from 1 to J , and for every j , l ranges from 1 to L_j ; $\beta_{jl} = \Pr(X_j = l)$. Let μ_β be a mixing measure producing the observed distribution.

*This research was supported by grants from National Institute of Aging.

The main GoM assumption is that for some K , μ_β is supported by a K -dimensional linear subspace Q of \mathbb{R}^{L_l} . Later, we refer to this K as to the *dimensionality of GoM problem*.

This assumption is essentially equivalent to the assumption that there exists a K -dimensional random vector G such that for every j a regression of Y_j on G is linear. Here Y_j is an L_j -dimensional random vector, $Y_j = \mathbf{1}_l$ if $X_j = l$ (where $\mathbf{1}_l$ denotes a vector which has l^{th} component equal to 1, and all other components equal to 0.) Namely, let $\Lambda = \{\lambda^1, \dots, \lambda^K\}$, $\lambda^k = (\lambda_{jl}^k)_{jl}$, be any basis of Q , and for $\beta \in Q$, let $g = (g_k)_{k=1, \dots, K}$ be its coordinates in basis Λ . Then the random vector G is the random vector β (distributed according μ_β) written in coordinates g , and matrices $\Lambda_j = (\lambda_{jl}^k)_{kl}$ are linear regression matrices.

The linear regression assumption is crucial for understanding the meaning of the GoM model and gives guidelines for its applicability. It essentially means that the measurements are not chosen arbitrarily but rather to reflect in some degree a hidden property, or a hidden state, represented by the random vector G . GoM analysis is about how to discover this hidden state and describe it as precisely as possible.

Let μ_g be a measure μ_β written in coordinates g .

Let $\ell = (\ell_1, \dots, \ell_J)$ be an integer vector with $0 \leq \ell_j \leq L_j$. Such a vector represents the outcome of J measurements, and $\ell_j = 0$ means that we do not take into account the outcome of the j^{th} measurement. Thus, a value of $\ell_j = 0$ in a vector ℓ means that the vector is a marginal vector across all values of the j^{th} measurement. Let \mathcal{L}^0 be a set of all such vectors, and for every $\mathcal{J} \subseteq \{1, \dots, J\}$ let $\mathcal{L}^{[\mathcal{J}]}$ be a set of vectors having 0's exactly on places from \mathcal{J} . Let $v = (v_1, \dots, v_K)$ be an integer vector with $v_k \geq 0$, and for every integer $J' \geq 0$ let $\mathcal{V}[J']$ be a set of such vectors satisfying the additional condition $\sum_k v_k = J'$.

In this language, the values of interest are unconditional moments of the distribution μ_β

$$M_\ell(\mu_\beta) = \int \prod_{j: \ell_j \neq 0} \beta_{j\ell_j} \mu_\beta(d\beta) \quad (2)$$

and conditional moments of distribution μ_g ,

$$\mathcal{E}(G^v | X = \ell) = \int \prod_k g_k^{v_k} \frac{\prod_{j: \ell_j \neq 0} \sum_k g_k \lambda_{jl}^k}{M_\ell(\mu_\beta)} \mu_g(dg) \quad (3)$$

The unconditional moments $M_\ell(\mu_\beta)$ are the probabilities of obtaining the response pattern ℓ (under assumptions of the model.) Thus, frequencies of response patterns ℓ in a sample, denoted f_ℓ , are consistent and efficient estimators for unconditional moments $M_\ell(\mu_\beta)$.

The conditional moments $\mathcal{E}(G^v | X = \ell)$ express our knowledge of the state of the individual (represented by random vector G) based on the outcomes of the measurements. These values are not directly estimable from the observations. The goal of GoM analysis is to obtain estimates for these conditional moments.

The most important relation connecting unconditional moments, conditional moments and the basis Λ (in which conditional moments are calculated) is:

$$\sum_k \lambda_{jl}^k \cdot (M_\ell(\mu_\beta) \cdot \mathcal{E}(G^{v+\mathbf{1}_k} | X = \ell)) = M_{\ell+\mathbf{l}_j}(\mu_\beta) \cdot \mathcal{E}(G^v | X = \ell + \mathbf{l}_j) \quad (4)$$

The main system of equations. We have shown in [Kovtun et al., 2004b] that the GoM model defined above is fully described by a system of equations (with respect to variables α_{jl}^k and h_ℓ^v)

$$\left\{ \begin{array}{ll} \sum_k \alpha_{jl}^k h_\ell^{v+\mathbf{1}_k} = h_{\ell+\mathbf{l}_j}^v, & J' \in [0..J-1], \quad v \in \mathcal{V}[J'], \\ & \mathcal{J} \subseteq [1..J] : |\mathcal{J}| > J', \quad \ell \in \mathcal{L}^{[\mathcal{J}]}, \\ & j \in \mathcal{J}, \quad l \in [1..L_j] \\ h_\ell^{(0, \dots, 0)} = M_\ell, & \ell \in \mathcal{L}^0 \\ \sum_{v \in \mathcal{V}[J']} \frac{(\sum_k v_k)!}{\prod_k v_k!} h_{(0, \dots, 0)}^v = 1, & J' \in [0..J] \end{array} \right. \quad (5)$$

In this system, the first group of equations corresponds to the main relation between moments (4), and the last two equations are normalization conditions.

We have proven the following properties of the main system:

1. Any basis Λ of Q together with conditional moments $\mathcal{E}(G^v \mid X = \ell)$ calculated in this basis give a solution of (5) ($\lambda_{j\ell}^k$ should be substituted for $\alpha_{j\ell}^k$, and $M_\ell(\mu_\beta) \cdot \mathcal{E}(G^v \mid X = \ell)$ should be substituted for h_ℓ^v .)
2. Under mild conditions, *every* solution of (5) gives a basis of Q and conditional moments calculated in this basis.

As the main system of equations fully describes the model, the important property of the GoM analysis follows: the mixing distribution is not fully identifiable. Only a finite number of moments may be found by solving the system, and any mixing distribution that have these moments would satisfy the system. The fact of nonidentifiability also follows from the general theorem about identifiability of mixtures, because the family of distributions contained in Q is not linearly independent.

The attractive feature of the GoM analysis is that it can discover a number of useful invariants of the mixture. The supporting plane of the mixing distribution is defined uniquely, and low-order moments are identifiable as well. This information is sufficient to make practically substantial conclusions about the population under consideration.

The main system of equations provides a means for consistent estimation of model parameters. The solution of this system continuously depends on unconditional moments M_ℓ ; thus, substitution of frequencies f_ℓ for moments M_ℓ gives a system, which solutions converge to the true values of parameters when frequencies converge to the true moments.

One good property of the main system of equations is that it is linear with respect to variable h_ℓ^v . Thus, if the supporting plane of distribution is known, the conditional moments (3) may be obtained by solving a linear system of equations. It happens that the supporting plane may be found independently by analysis of the moment matrix, which we describe in the next subsection.

The moment matrix. Let us write a vector of moments $(M_{l_j})_{j\ell}$ together with incomplete vectors of moments $(M_{l'_j+l_j})_{j\ell: j \neq j'}$, etc., as columns of a matrix, with places for which we do not have moments filled by question marks. We refer to this incomplete matrix as the *moment matrix*. The moment matrix contains a column for every $\ell \in \mathcal{L}^0$. Figure 1 gives an example of a portion of a moment matrix for the case $J = 3$, $L_1 = L_2 = L_3 = 2$. Columns in this matrix correspond to $\ell = (000), (100), (200), (010), (020), (001), (002), (110)$; other columns are not shown.

$$\begin{pmatrix} M_{(100)} & ? & ? & M_{(110)} & M_{(120)} & M_{(101)} & M_{(102)} & ? & \cdots \\ M_{(200)} & ? & ? & M_{(210)} & M_{(220)} & M_{(201)} & M_{(202)} & ? & \cdots \\ M_{(010)} & M_{(110)} & M_{(210)} & ? & ? & M_{(011)} & M_{(012)} & ? & \cdots \\ M_{(020)} & M_{(120)} & M_{(220)} & ? & ? & M_{(021)} & M_{(022)} & ? & \cdots \\ M_{(001)} & M_{(101)} & M_{(201)} & M_{(011)} & M_{(021)} & ? & ? & M_{(111)} & \cdots \\ M_{(002)} & M_{(102)} & M_{(202)} & M_{(012)} & M_{(022)} & ? & ? & M_{(112)} & \cdots \end{pmatrix}$$

Figure 1: Example of moment matrix

Note that certain moments (which are replaced by question marks in the moment matrix) are not observable. The reason for this is that we are not able to perform a measurement on an individual multiple times independently, and since individuals are heterogeneous (have different probabilities of outcomes of measurements), we do not have multiple realizations of independent identically distributed random variables.

For a moment matrix M let its completion \bar{M} be a matrix obtained from M by replacing question marks by arbitrary numbers. We have shown that the moment matrix always has a completion in which all columns belong to the supporting plane Q . Thus, if the moment matrix has sufficient rank (which is the case in non-degenerate situations,) a basis of Q may be obtained from this matrix. As we have a consistent estimator of the moment matrix in form of a frequency matrix, the supporting plane may be consistently estimated.

This property of the moment matrix suggests an efficient algorithm to obtain GoM estimates. First, a basis of the supporting plane can be obtained from the moment matrix (a way to do this is described in the next section), and second, conditional moments can be found by solving a linear system of equations.

Algorithm. As it is suggested by a structure of the main system of equations (5) and by properties of the moment matrix, the algorithm is naturally decomposed into two parts. On the first step, a basis of the supporting plane should be constructed; the input for this step is the frequency matrix. On the second step, a system of linear equations should be solved to obtain estimates for conditional expectations.

Step 1: Finding the supporting plane. As for the model distribution all columns of the moment matrix belong to the supporting plane, and as the frequency matrix is an approximation of the moment matrix, the natural way to search for the supporting plane is to search for a plane that minimizes the sum of distances from it to the columns of the frequency matrix. In our case, however, this way is complicated by at least three obstacles: (a) a sought basis Λ should *exactly* satisfy conditions $\sum_l \lambda_{jl}^k = 1$ for every k and j ; (b) the statistical inaccuracy of approximation of moments M_ℓ by frequencies f_ℓ varies considerably over elements of frequency matrix; (c) the moment matrix (and, correspondingly, the frequency matrix) is incomplete.

The suggested algorithm for estimating the supporting plain consists of the following steps.

- (i) The computational rank of the frequency matrix is estimated. For this, we take the biggest minor of the frequency matrix that does not contain question marks. (For the example given in Figure 1, it is the left bottom minor of size 3×3 .) Then we calculate the singular value decomposition (SVD) and take K_0 (the first approximation of dimensionality of the GoM problem) equal to the number of singular values that are greater than standard deviation of the norm of columns involved in the minor. (The final value for dimensionality of GoM problem will be chosen on the step (v).) As one of requirements for applicability of GoM model is $K \ll |L|$, nothing can be done further if all (or too many) singular values are greater than the standard deviations.
- (ii) We construct a completion of the frequency matrix by means of the following procedure. For every column c of the frequency matrix and row jl of a question mark in c , we select K_0 columns $c^{(1)}, \dots, c^{(K_0)}$ satisfying: (a) all columns $c^{(i)}$ contain a value (not a question mark) in row jl ; (b) there exist $p \geq K_0$ rows such that all columns $c, c^{(1)}, \dots, c^{(K_0)}$ contain values in these rows. Let $c[p]$ be a subcolumn containing only selected rows. Then we solve a linear system $\alpha_1 c^{(1)}[p] + \dots + \alpha_{K_0} c^{(K_0)}[p] = c[p]$ and replace a question mark at the position c_{jl} by $\alpha_1 c_{jl}^{(1)} + \dots + \alpha_{K_0} c_{jl}^{(K_0)}$. The system to be solved is overdetermined; we solve it by minimization of residuals using SVD. When K_0 is sufficiently smaller than $|L|$, the required selection of the columns is possible for every column c which contains at least K_0 values; columns containing less than K_0 values are discarded from further consideration. According to [Kovtun et al., 2004b], the moment matrix always has a completion of rank equal to the dimensionality of GoM problem; thus, this method should give good results.
- (iii) Columns are normalized, so that the condition $\sum_l c'_{jl} = 1$ holds for every j . This is always possible, as for every column c we have $\sum_l c_{jl} = s$, where s does not depend on j . Thus, we take $c' = \frac{1}{s}c$.
- (iv) Next, we remove the restriction $\sum_l c_{jl} = 1$ by reducing number of rows by J (1 for every group of indices $j1, \dots, jL_j$). For this, we use a linear map from $\mathbb{R}^{|L|}$ to $\mathbb{R}^{|L|-J}$ given by a block-diagonal matrix A with J blocks

$$A_j = \begin{pmatrix} -\frac{\sqrt{L_j-1}}{L_j-1} & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -\frac{\sqrt{L_j-1}}{L_j-1} & 0 & 0 & \dots & 1 \end{pmatrix} \quad (6)$$

of size $L_j \times (L_j - 1)$. This map is an isometry of the subspace of $\mathbb{R}^{|L|}$ defined by equations $\sum_l c_{jl} = 1$ to $\mathbb{R}^{|L|-J}$ (every block A_j defines a rotation of a unit simplex in L_j -dimensional space around hypersurface opposite to the first vertex; the angle of this rotation is such that the first vertex moves to the point with the first coordinate equals 0).

- (v) Now we have n points y^1, \dots, y^n (images of columns of frequency matrix) in $m = |L| - J$ -dimensional space. The problem is to find an affine plane that minimally deviates from these points. First, we find the center of gravity of this system

$$y^0 = \frac{1}{n} \sum_i y^i \quad (7)$$

and then consider a new set of points $x^i = y^i - y^0$. We need to find a linear subspace in \mathbb{R}^n that minimally deviates from this set of points. The solution of this problem is well-known (see, for example, chapter 43 of [Kendall and Stuart, 1977]): one has to consider an $m \times m$ matrix $X = (X_{rs})_{rs}$ with components $X_{rs} = \sum_i x_r^i \cdot x_s^i$; this matrix is symmetrical and positively defined, and thus it possess an orthonormal basis of eigenvectors. Let $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_m \geq 0$ be eigenvalues of matrix X , and let z^1, \dots, z^m be corresponding them eigenvectors. The plane of dimensionality p that minimizes the sum of squared distances from point x^1, \dots, x^n is spanned by z^1, \dots, z^p , and the sum of squared distances is $\text{tr}X - \sum_{k=1}^p \gamma_k$. This gives us a criterion for the selection of the dimensionality K of the GoM problem: one has to take K to be the smallest integer such that eigenvalues $\gamma_K, \dots, \gamma_m$ are smaller than inaccuracy in input data. Vectors $y^0, y^0 + z^1, \dots, y^0 + z^{K-1}$ give us an affine basis of the sought affine plane.

- (vi) Lastly, we apply inverses of transformation (6) to $y^0, y^0 + z^1, \dots, y^0 + z^{K-1}$ to obtain the sought basis $\lambda^1, \dots, \lambda^K$ of the subspace Q .

The above algorithm solves the problems (a)–(c) listed in the beginning of the subsection, and it possesses two important properties that are crucial to its usefulness: (a) if the input of the algorithm are true moments of a distribution generated by K -dimensional GoM model, the output of the algorithm is the true supporting plane; (b) there exists an open neighbourhood of the true moment matrix in which the output of the algorithm continuously depend on its input.

The preliminary experiments with the prototype of the algorithm performed by the applicants demonstrated that it restores the supporting plane with a good degree of precision.

Step 2: Calculation of conditional expectations. When a basis of the supporting plane is found, the conditional expectations can be found from the main system of equations (5), which is a linear system after substituting the basis. This is a sparse overdetermined system; methods for solving such systems are well-elaborated—see, for example, [Forsythe et al., 1977, Kahaner et al., 1988].

More detailed discussion of the algorithms and its properties (including results of simulation and application to real-world data) are given in the full version of the paper.

References

- [Bartholomew and Knott, 1999] Bartholomew, D. J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis*. Oxford University Press, New York, 2 edition.
- [Bishop et al., 1975] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis*. MIT Press, Cambridge, MA.
- [Everitt and Hand, 1981] Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions*. Monographs on Applied Probability and Statistics. Chapman and Hall, London.
- [Forsythe et al., 1977] Forsythe, G. E., Malcolm, M. A., and Moler, C. B. (1977). *Computer Methods for Mathematical Computations*. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- [Goodman, 1978] Goodman, L. A. (1978). *Analyzing Qualitative/Categorical Data: Log-linear Models and Latent-Structure Analysis*. Abt Books, Cambridge, MA.
- [Heinen, 1996] Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. SAGE Publications, Thousand Oaks.

- [Kahaner et al., 1988] Kahaner, D., Moler, C., and Nash, S. (1988). *Numerical Methods and Software*. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- [Kendall and Stuart, 1977] Kendall, M. G. and Stuart, A. (1977). *Design and Analysis, and Time-Series*, volume 3 of *The Advanced Theory of Statistics*. C. Griffin, London, 4 edition.
- [Kovtun et al., 2004a] Kovtun, M., Akushevich, I., Manton, K. G., and Tolley, H. D. (2004a). Grade of membership analysis: Newest development with application to national long-term care survey data. In *2004 PAA Annual Meeting*, Boston, MA.
- [Kovtun et al., 2004b] Kovtun, M., Akushevich, I., Manton, K. G., and Tolley, H. D. (2004b). *Grade of Membership Analysis: One Possible Approach to Foundations*. To appear in “Focus on Probability Theory”. Nova Science Publishers, New York.
- [Langeheine and Rost, 1988] Langeheine, R. and Rost, J., editors (1988). *Latent Trait and Latent Class Models*. Plenum Press, New York.
- [Lazarsfeld and Henry, 1968] Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*. Houghton Mifflin Company, Boston, MA.
- [Lindsay, 1995] Lindsay, B. G. (1995). *Mixture models: Theory, Geometry and Applications*, volume 5 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA.
- [Manton et al., 1994] Manton, K. G., Woodbury, M. A., and Tolley, H. D. (1994). *Statistical applications using fuzzy sets*. John Wiley and Sons, New York.
- [Titterton et al., 1985] Titterton, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Chichester.
- [Woodbury and Clive, 1974] Woodbury, M. A. and Clive, J. (1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics*, 4:111–121.
- [Woodbury et al., 1978] Woodbury, M. A., Clive, J., and Garson, A. (1978). Mathematical typology: a grade of membership technique for obtaining disease definition. *Computers and Biomedical Research*, 11:277–298.