TOWARDS BETTER UNDERSTANDING LATENT STRUCTURE ANALYSIS: A GEOMETRIC APPROACH (EXTENDED ABSTRACT)

MIKHAIL KOVTUN, IGOR AKUSHEVICH, AND ANATOLIY YASHIN

Abstract. Surveys and other similar data collection techniques are widely used in social sciences. In a survey, questions are not chosen arbitrarily, but to reflect the underlying *latent structure*, which cannot be observed directly. Latent structure analysis is a statistical technique for revealing such latent structures.

We develop a geometric view on latent structure analysis, which allows us to describe in simple terms relation between different branches of LSA (including latent class models, latent trait models, and linear latent structure models), clearly formulate conditions of identifiability of models, and provide guidelines for practical applications of LSA. A special attention is paid to the role and applicability of "local independence" assumption. An extensive example, based on National Long Term Care Survey data, is discussed from this point of view.

1. Overview

Surveys and other similar data collection techniques are widely used in social sciences. In a survey, questions are not chosen arbitrarily, but to reflect the underlying *latent structure*. Latent structure analysis is a statistical technique for revealing such latent structures. Although LSA is more than half a century old, it is still far from being clearly and unambiguously understood; many concepts were realized only behindhand (for example, it took several decades to recognize connection between LSA and a theory of mixed distributions; the same can be said about relation between latent class and latent trait models). We believe that our geometric approach would shed a new light on the methods of LSA and would help the applied researchers in selecting most appropriate models and interpreting their outcomes.

In this presentation we understand "latent structure analysis" in the sense of Lazarsfeld (Laz50b; Laz50a; LH68); that is, the observed values are represented by a number of categorical random variables, and latent structure under investigation governs distribution of these variables. Our mathematical treatment of latent structure analysis is based on the modern approach, which considers latent structure models as mixed distribution models (Bar02).

The summary of our approach is given by the following statements (required notions and notation are introduced below):

(1) Independent distributions (i.e., distributions being mixed to obtain the observed distribution) of observed random variables belong to $|L|$-dimensional linear space (for the case of 10 binary variables, this is 20-dimensional space).

2000 *Mathematics Subject Classification.* Primary 62G05; Secondary 62H25.

Key words and phrases. Mixed distributions, latent structure analysis.

The work of the first author was supported by the Office of Vice Provost for Research, Duke University.

The work of the second author was supported by NIH/NIA grant P01 AG17937-05.

The work of the third author was supported by NIH/NIA grants PO1 AG08761-01, U01 AG023712 and U01AG 07198.

Remark: In order to be a probabilistic distribution, a vector from $|L|$-dimensional space must satisfy J linear equations and a number of linear inequalities; together, these constraints define a $(|L| - J)$-dimensional convex polyhedron in $\mathbb{R}^{|L|}$ (for the case of 10 binary variables, the dimensionality of this polyhedron is 10).

Remark: We do not exclude redundant dimensions in order to preserve symmetry. The absence of symmetry significantly befogs the picture and complicates further considerations (and in our opinion, the previous researchers failed to paint a picture similar to ours just because the dimensionality was reduced as much as possible and as soon as possible).

(2) All probabilistic distributions of observed random variables belong to $|L^*|$-dimensional linear space (for the case of 10 binary variables, this is 1024-dimensional space).

Remark: In order to be a probabilistic distribution, a vector from $|L^*|$-dimensional space must satisfy 1 linear equation and a number of linear inequalities; together, these constraints define a unit simplex in $\mathbb{R}^{|L^*|}$.

(3) The natural imbedding of the space of independent distributions into the space of all distributions is not linear; its image, called *independence surface*, is an intersection of quadratic hypersurfaces.

(4) Latent structure model is a representation of a vector corresponding the observed distribution as a (generalized) convex linear combination (or *mixture*) of vectors belonging to the independence surface.

(5) *Identifiability* of a latent structure model means that such linear combination is unique.

(6) Latent structure model always exists but never is identifiable (except degenerated cases).

(7) To obtain identifiability, one needs to restrict the allowable family of linear combinations. Various branches of latent structure analysis differ by restrictions imposed on linear combinations.

(8) Latent class models (LCM) search for finite linear combinations. They are identifiable when number of independent distributions being mixed (number of *latent classes*) is sufficiently smaller than $|L^*|$.

(9) Latent trait models (LTM; also known as "item response theory," IRT) restrict distributions being mixed to a curve from a special class (one-dimensional LTM) or to a surface from a special class (multi-dimensional LTM). One-dimensional LTM are always identifiable, and multi-dimensional LTM are identifiable when dimensionality of the surface is sufficiently smaller than $|L^*|$.

(10) Linear latent structure (LLS) models restrict distributions being mixed to a linear subspace of the space of independent distributions. LLS models are identifiable when dimensionality of this subspace is sufficiently smaller than $|L|$.

(11) In the most cases, the existence of LCM or LTM implies existence of LLS model.

(12) If LLS model exists, the existence or nonexistence of LCM or LTM can be easily derived from the analysis of LLS model.

2. Linear spaces of distributions

The input of latent structure analysis consists of a number of categorical measurements made on individuals in a sample drawn from a population under investigation. The outcomes of these measurements are represented as realizations of random variables $X_1, \ldots, X_J$; random variable X_j takes values in a finite set $\{1, \ldots, L_j\}$.

The set of directly estimable from observations distribution characteristics consists of elementary probabilities

$$p_\ell = \mathsf{P}\left(X_1 = \ell_1 \text{ and } \ldots \text{ and } X_J = \ell_J\right) \tag{1}$$

Here $\ell = (\ell_1, \ldots, \ell_J)$ is a response pattern. The joint distribution of $X_1, \ldots, X_J$ is fully described by the set of elementary probabilities $\{p_\ell\}_\ell$, where ℓ ranges over the set of all possible response patterns, which we denote $\mathcal{L}$. As there exist $|L^*| = L_1 \times \cdots \times L_J$ different

response patterns, a joint distribution can be identified with a vector in linear space $\mathbb{R}^{|L^*|}$. In order to describe a probabilistic distribution, vector $(p_\ell)_\ell$ from $\mathbb{R}^{|L^*|}$ should satisfy

$$p_\ell \geq 0 \quad \text{for all } \ell; \qquad \sum_\ell p_\ell = 1 \tag{2}$$

Note that these conditions define a unit simplex in $\mathbb{R}^{|L^*|}$.

Among all joint distributions one can distinguish *independent* distributions, i.e. distributions, in which random variables $X_1, \ldots, X_J$ are mutually independent. This means that for every set of indices $j_1, \ldots, j_p$ and for every response pattern ℓ the relation

$$\mathsf{P}(X_{j_1} = \ell_{j_1} \text{ and } \ldots \text{ and } X_{j_p} = \ell_{j_p}) = \mathsf{P}(X_{j_1} = \ell_{j_1}) \cdot \ldots \cdot \mathsf{P}(X_{j_p} = \ell_{j_p}) \tag{3}$$

holds. Equation (3) allows us to describe an independent distribution using only $|L| = L_1 + \cdots + L_J$ parameters. Namely, let $\beta_{jl} = \mathsf{P}(X_j = l)$. Then for every response pattern ℓ,

$$p_\ell = \prod_{j=1}^{J} \beta_{j\ell_j} \tag{4}$$

Thus, every independent distribution can be identified with a point $\beta = (\beta_{jl})_{jl} \in \mathbb{R}^{|L|}$. Not every point $\beta \in \mathbb{R}^{|L|}$ corresponds to a probability distribution; to describe a distribution, β must satisfy the conditions:

$$\begin{cases} \sum_{l=1}^{L_j} \beta_{jl} = 1 & \text{for every } j \\ \beta_{jl} \geq 0 & \text{for every } j \text{ and } l \end{cases} \tag{5}$$

Conditions (5) define a convex $(|L| - J)$-dimensional polyhedron in $\mathbb{R}^{|L|}$, which we denote $\mathbb{S}^L$.

Equations (4) define a *natural embedding* of the space of independent distributions into the space of all distributions. Under this mapping, the image of $\mathbb{S}^L$, called *independence surface*, belongs to the unit simplex in $\mathbb{R}^{|L^*|}$. Note that the natural embedding *is not* a linear mapping.

To describe the independence surface, we need additional notation. Let $\mathcal{L}^0$ be a set of integer vectors of length J with j^{th} component from $\{0, 1, \ldots, L_j\}$, $j = 1, \ldots, J$ (in other words, $\mathcal{L}^0$ is obtained as extension of $\mathcal{L}$ by allowing some components of the vectors to be 0). For $\ell \in \mathcal{L}$ and $\ell' \in \mathcal{L}^0$ we say that ℓ *is a subpattern* of ℓ', or ℓ narrows ℓ', denoted $\ell \in \ell'$, if ℓ coincides with ℓ' in all components which are non-zero in ℓ' (for example, $(1,1) \in (1,0)$ and $(1,2) \in (1,0)$, but $(1,2) \notin (2,0)$). We say that vectors ℓ' and ℓ'' from $\mathcal{L}^0$ are disjoint, denoted $\ell' \perp \ell''$, if for every j either $\ell' = 0$ or $\ell'' = 0$ (for example, $(1,0,0) \perp (0,0,1)$ and $(2,0,2) \perp (0,1,0)$, but $(1,2,0) \not\perp (0,1,0)$).

The intended goal of introducing of $\mathcal{L}^0$ is to obtain notations for marginal probabilities; for example, with this notation $p_{(0,1,0)} = \mathsf{P}(X_2 = 1)$ and $p_{(2,0,1)} = \mathsf{P}(X_1 = 2 \text{ and } X_3 = 1)$. The usual summation conditions for marginals in a contingency table are written in our notation as $p_{\ell'} = \sum_{\ell \in \ell'} p_\ell$ (for example, in the case of 3 binary variables one obtains $p_{(2,0,1)} = p_{(2,1,1)} + p_{(2,2,1)}$).

Now the conditions for a distribution given by vector $(p_\ell)_\ell$ in $\mathbb{R}^{|L^*|}$ to be independent can be written as

$$p_{\ell'+\ell''} = p_{\ell'} \cdot p_{\ell''} \qquad \text{for every } \ell', \ell'' \in \mathcal{L}^0 \text{ such that } \ell' \perp \ell'' \tag{6}$$

Note that conditions (3) can be derived from conditions (6) and vice versa. Further, the conditions (6) can be rewritten in coordinates of $\mathbb{R}^{|L^*|}$ as

$$\sum_{\ell \in \ell'+\ell''} p_\ell = \left(\sum_{\ell \in \ell'} p_\ell\right) \cdot \left(\sum_{\ell \in \ell''} p_\ell\right) \tag{7}$$

The equation (7) is a quadratic equation; thus, the independence surface is an intersection of quadratic hypersurfaces. This is very attractive fact, as it suggests that the further investigation of properties of independence surface may be performed by means of analytic geometry.

3. Mixtures of independent distributions

The goal of latent structure analysis is to find a representation of the observed distribution as a mixture of independent distributions. In our approach, this means to represent the vector $p = (p_\ell)_\ell$, which describes the observed distribution, as a linear combination of vectors belonging to the independence surface:

$$p = \begin{pmatrix} p_{(1,\dots,1)} \\ \vdots \\ p_\ell \\ \vdots \\ p_{(L_1,\dots,L_J)} \end{pmatrix} = \sum_k \alpha_k \cdot \begin{pmatrix} q^{(k)}_{(1,\dots,1)} \\ \vdots \\ q^{(k)}_\ell \\ \vdots \\ q^{(k)}_{(L_1,\dots,L_J)} \end{pmatrix} = \sum_k \alpha_k \cdot \begin{pmatrix} \prod_j \beta^{(k)}_{j1} \\ \vdots \\ \prod_j \beta^{(k)}_{j\ell_j} \\ \vdots \\ \prod_j \beta^{(k)}_{jL_j} \end{pmatrix} \tag{8}$$

Here $q^{(k)}$, $k = 1, \dots, K$, are vectors belonging to the independence surface; consequently, there exist vectors $\beta^{(k)}$ in $\mathbb{R}^{|L|}$ such that $q^{(k)}_\ell = \prod_j \beta^{(k)}_{j\ell_j}$, which explains the last equality in (8).

The linear combination of vectors $q^{(k)}$ with coefficients α_k can be considered as a mixture with the mixing distribution concentrated in points $q^{(k)}$ (or in points $\beta^{(k)}$) with weights α_k. The natural way to generalize the representation (8) is to replace a mixing distribution concentrated in a finite number of points by an arbitrary mixing distribution. Such general form of mixing distribution may be given by its probability density function $f(\beta)$, cumulative distribution function $F(\beta)$, or measure μ_β defined on space $\mathbb{R}^{|L|}$. With this, the sought representation can be written as

$$p_\ell = \int \left(\prod_j \beta_{j\ell_j} \right) f(\beta)\, d\beta = \int \left(\prod_j \beta_{j\ell_j} \right) F(d\beta) = \int \left(\prod_j \beta_{j\ell_j} \right) \mu_\beta(d\beta) \tag{9}$$

One has to keep in mind, however, that p.d.f. $f(\beta)$ can be a generalized function.

The easiest way to explain usefulness of this generalization is an analogy with estimation of distribution law of real-valued random variable from finite number of observations: although the best that can be constructed based *only* on a finite number of observations is an empirical distribution, the nature of the applied domain often suggests that the true distribution is continuous, which the empirical distributions is an approximation to.

Now for every $\ell \in \mathcal{L}$ consider vector $e^{(\ell)}$ in $\mathbb{R}^{|L|}$, which has ℓ^{th} component equal to 1 and all other components equal to 0. On the one hand, all these vectors represent independent distributions with corresponding vectors $\beta^{(\ell)}$ having components $\beta^{(\ell)}_{jl} = 1$ if $l = \ell_j$ and $\beta^{(\ell)}_{jl} = 0$ otherwise. On the other hand, the vectors $e^{(\ell)}$ are vertices of the unit simplex in $\mathbb{R}^{|L^*|}$, and consequently every vector belonging to the unit simplex (i.e., every distribution) can be represented as a convex linear combination of vectors $e^{(\ell)}$

$$p = \sum_{\ell \in \mathcal{L}} p_\ell \cdot e^{(\ell)} \tag{10}$$

Thus, *every* distribution can be represented as a mixture of independent distributions. However, the representation (10) is totally useless, as it does not provide new knowledge. We precede the further discussion of the problem by a simple illustrative example.

Consider a case of two binary variables. Assume that the observed distribution is described by the vector of probabilities

$$p = \begin{pmatrix} p_{(1,1)} \\ p_{(1,2)} \\ p_{(2,1)} \\ p_{(2,2)} \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/6 \\ 1/6 \\ 1/3 \end{pmatrix} \tag{11}$$

It is easy to see that this distribution is not independent, as $p_{(1,1)} = 1/3 \neq 1/2 \cdot 1/2 = p_{(1,0)} \cdot p_{(0,1)}$. At the same time, this distribution can be represented as a linear combination of two independent distributions in multiple ways; two possible representations are:

$$p = 1/2 \begin{pmatrix} 0 \\ 1/3 \\ 0 \\ 2/3 \end{pmatrix} + 1/2 \begin{pmatrix} 2/3 \\ 0 \\ 1/3 \\ 0 \end{pmatrix} = 2/3 \begin{pmatrix} 1/12 \\ 1/4 \\ 1/6 \\ 1/2 \end{pmatrix} + 1/3 \begin{pmatrix} 5/6 \\ 0 \\ 1/6 \\ 0 \end{pmatrix} \tag{12}$$

Further, it can be represented in form (9). One possibility for this is to take the mixing distribution being uniformly distributed over segment connecting point $\beta^{(1)} = (1, 0, 1, 0)$ and $\beta^{(2)} = (0, 1, 0, 1)$ in the space of independent distributions $\mathbb{R}^{|L|}$. This means that the independent distributions being mixed can be parameterized by parameter t, uniformly distributed over interval $[0, 1]$, as

$$\beta(t) = \begin{pmatrix} \beta_{11}(t) \\ \beta_{12}(t) \\ \beta_{21}(t) \\ \beta_{22}(t) \end{pmatrix} = \begin{pmatrix} 1-t \\ t \\ 1-t \\ t \end{pmatrix} \tag{13}$$

Now probabilities of response patterns of the mixture can be calculated in accordance with (9) as

$$p_{(1,1)} = \int_0^1 \beta_{11}(t)\beta_{21}(t)\, dt = \int_0^1 (1-t)^2\, dt = 1/3 \tag{14}$$

and similarly for $p_{(1,2)}$, $p_{(2,1)}$, and $p_{(2,2)}$.

Similar examples can be readily constructed for higher dimensions.

The above considerations show that in such general settings (a) latent structure model for any observed distribution always exists; (b) latent structure model is never identifiable. As identifiability of the model is necessary for a practical usefulness of any statistical method, one need to employ additional ideas to make latent structure analysis a useful tool for the applied researcher.

The central idea is to search for the *simplest* latent structure model rather than for *any* model. The notion of simplicity may be introduced in many different ways, and here is the point where different branches of latent structure analysis emerge.

In all existing branches of latent structure analysis the "simplicity" is formulated as a set of restrictions imposed on the allowed linear combinations in (8) and (9)—or, in other words, by imposing restriction on the subset of the space of independent distributions that carries the mixing distribution. Below we describe the major branches of latent structure analysis from this point of view.

A drawback of imposing restrictions is that it is possible that the restricted model does not exist (or the restricted model does not fit data well).

3.1. **Latent Class Models (LCM).** In LCM, the class of allowed subsets is restricted to finite sets. Thus, the sought representation has the form (8). The vectors $q^{(k)}$ (or, equivalently, vectors $\beta^{(k)}$) are called *latent classes*.

To identify the model, one has to find K variables α_k plus $K \times |L|$ variables $\beta_{jl}^{(k)}$. For this, one has $|L^*|$ equations (8) plus $K \times J$ equations (5) (some of these equations are dependent, however). Note that equations $p_\ell = \sum_k \alpha_k \prod_j \beta_{j\ell_j}^{(k)}$ are nothing else than Lazarsfeld's "accounting equations."

This gives a rough way to formulate conditions for identifiability of LCM: the model is identifiable, if the number of variables does not exceed the number of independent equations. In practice, however, the situation is complicated by the fact that equations might be "almost dependent," which leads to ill-posed problem (one has to take into account that probabilities are *approximated* by frequencies, and thus one has to solve the system *approximately*).

In practice, the maximum likelihood methods are used to estimate LCM, as they have huge computational advantages (both in sense of performance and stability) in comparison with methods that solve the system directly.

3.2. **Latent trait models (LTM).** In fact, it is the whole spectrum of models. For the sake of simplicity, we restrict our attention here to the Rash models for binary variables.

LTM searches for representation (9) with mixing distribution carried by a one-dimensional curve. This curve is parameterized by parameter t (ranging over real line) as:

$$\beta_{j1}(t) = \frac{\exp(t - b_j)}{1 + \exp(t - b_j)}, \qquad j = 1, \dots, J \tag{15}$$

(As we consider here only binary variables, vector β for every j has only components β_{j1} and β_{j2}, and $\beta_{j2} = 1 - \beta_{j1}$.)

More detailed discussion of the properties of LTM requires detailed analysis of geometry of the curve (15), which is outside the scope of the present paper.

The estimation of model parameters b_j and distribution of t is performed by maximum likelihood methods (often employing some a priori information regarding distribution of t).

3.3. **Linear latent structure (LLS) analysis.** LLS analysis searches for representation (9) with the mixing distribution carried by a low-dimensional linear subspace of the space of independent distributions. Thus, one has to estimate K basis vectors of the supporting subspace, $\lambda^1, \dots, \lambda^K$, and the mixing distribution over this subspace.

It happens that basis vectors can be estimated independently from the mixing distributions—only based on the observed probabilities p_ℓ. The estimation of the basis is performed by methods similar to the methods of principal component analysis.

After the basis of the supporting subspace is estimated, the estimation of the mixing distribution can be done by solving a number of small systems of linear equations (one system per response pattern presented in the sample; each system is an overdetermined system with J equations).

The ability to estimate model using methods of linear algebra is a big advantage of LLS analysis: first, it allows to avoid the problem of multimodality (which maximum likelihood methods has to cope with), and second, it significantly increases dimensionality of the datasets that can be analyzed (for example, the authors successfully applied the prototype of LLS algorithm to the dataset involving 1,500 binary variables).

3.4. **Comparison of LCM and LLS models.** Our geometric approach allows us compare different latent structure models. We demonstrate it by the following properties, which can be easily derived from the above constructions:

- If LCM with K classes exists, then K-dimensional LLS model exists as well.
- The existence of LCM can be derived from the analysis of the mixing distribution in LLS model: if the mixed distribution has pronounced modality, then LCM exists.
- It is possible to construct a distribution, which has 2-dimensional LLS model and no LCM with the number of classes smaller than J.

3.5. **Parametric language.** A traditional exposition of latent structure analysis speaks about "latent variables" or "latent parameters." These notions are translated in our language in the following way.

The subset carrying the mixing distribution always can be parameterized—by a discrete parameter in the case of LCM, by one-dimensional real parameter in the case of LTM, or by K-dimensional vector parameter in the case of LLS analysis. This parameter can be considered as a random variable; the distribution of this random variable is the mixing distribution. This random variable is exactly what is traditionally called "latent variable" or "latent parameter."

4. Applications of latent structure models

One obvious application of a latent structure model is provided by interpretation of the discovered latent structure in terms of the applied domain. This topic, although very important, cannot be discussed in general framework.

Another important application is calculation of numerical values depending on higher-order moments of the observed distribution. We illustrate it by example. Suppose that in example of section 3 the outcome 1 of the first measurements denotes the event which causes an insurance company to pay a premium. Let Y be a random variable which equals 1 when $X_1 = 1$ and equals 0 otherwise. To design an optimal insurance policy, one need to know both expectation $\mathsf{E}\,Y$ (which is equal $p_{(1,0)} = 1/2$) and variance $\mathsf{D}\,Y$, which cannot be directly derived from the observed data. Without knowing latent structure, the only possible way to estimate $\mathsf{D}\,Y$ is to take it equal to $p_{(1,0)}(1 - p_{(1,0)}) = 1/4$. But if one knows that the observed distribution is governed by the latent structure (13), the variance can be calculated as

$$\mathsf{D}\,Y = \int_0^1 (1-t)t\,dt = 1/6 \tag{16}$$

5. Local independence assumption

The "local independence" assumption is always considered as essential part of the latent structure analysis. Different authors, however, express distinct opinions regarding its meaning. Her we present our point of view.

One has to distinguish *weak* and *strong* local independence assumptions.

The weak local independence assumption is "the observed distribution can be represented as a mixture of independent distributions." In fact, it is not assumption—rather, it is a property of a model being constructed. The question is not "whether the weak local independence assumption is valid," but "does LCM, LTM, or LLS model exists?" If one believes that a model correctly describes the distribution under consideration, he/she has to believe in all its corollaries (like described in section 4).

The strong local independence assumption is "the latent variable takes a single value on each individual." It implies that "conditional on individual" (which is stronger than "conditional on the value of latent parameter!") "the observed latent variables are independent." It is really an assumption, and it requires justification to validate any inference based on it. Fortunately, the strong local independence assumption is not required to derive properties of the *population*. It is crucial, however, for establishing *individual* properties (like probability to belong to a particular class in LCM, or individual scores in LLS analysis).

In applications, it is often possible to relax the strong local independence assumption to "conditional on individual, variance of latent variable is small." This would guarantee that individual properties, derived from the model, are held with "high" probability. Of course, the meaning of "small variance" and "high probability" should be quantified in each practical case.

6. Real-world example

We explain the above considerations by example of application of LLS analysis to National Long Term Care Survey data.

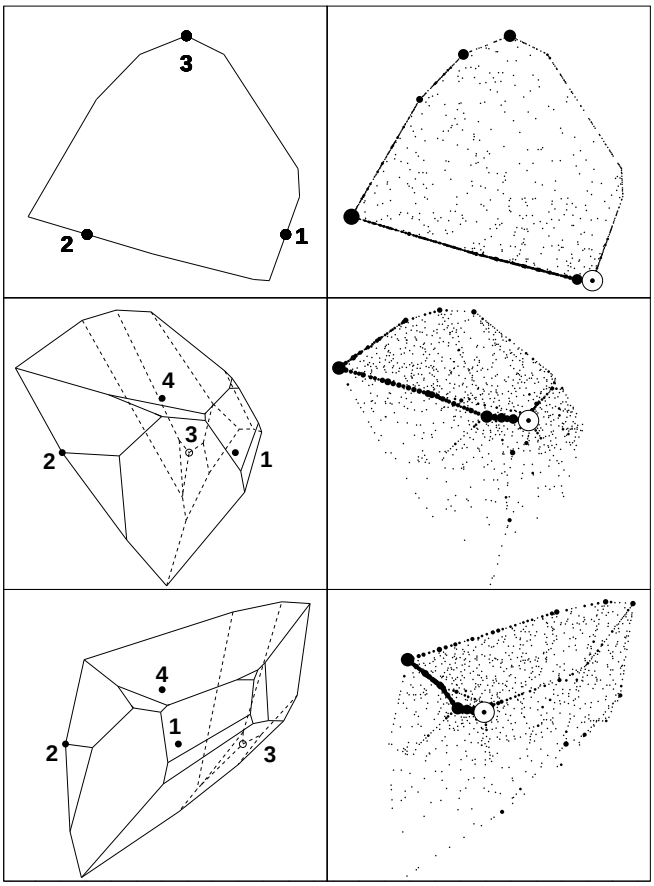

Figure 1. Mixing distribution of NLTCS data.

Figure 1 gives a pictorial representation of the mixing distribution. The mixing distribution is carried by a 3-dimensional polyhedron; to achieve a better impression of the volume, we provide three view from different angles. The left column contains pictures of polyhedron carrying the mixing distribution, the right column presents the distribution itself.

This picture clearly demonstrates the absence of LCM or LTM that reasonably fits the data (which coincides with findings of (Ero02)). At the same time, the LLS model fits data well.

7. Bibliographical notes

Latent structure analysis was introduced by Paul Lazarsfeld in (Laz50b; Laz50a); its development and current state is reflected in (LH68; Goo78; LR88; Hei96; BK99; MM02).

Mixed distribution models are described in (EH81; TSM85; Lin95).

Linear latent structure analysis was developed in (KAMT05a; KAMT05b). An algorithm for estimation LLS models is described in detail in (AKYM05). An important property of convergence of estimates of mixing distribution was proven in (KYA05).

References

[AKYM05] Igor Akushevich, Mikhail Kovtun, Anatoly Yashin, and Kenneth G. Manton, *Linear latent structure analysis: from foundations to algorithms and applications*, Submitted for publication to "Computational Statistics and Data Analysis." Available from e-Print archive arXiv.org at `http://www.arxiv.org`, arXiv code math.ST/0508299, 2005.

[Bar02] David J. Bartholomew, *Old and new approaches to latent variable modeling*, Latent Variable and Latent Structure Models (George A. Marcoulides and Irini Moustaki, eds.), Quantitative Methodology Series: Methodology for Buisness and Management, Lawrence Erlbaum Associates, Mahwah, NJ, 2002, pp. 1–14.

[BK99] D. J. Bartholomew and M. Knott, *Latent variable models and factor analysis*, 2 ed., Oxford University Press, New York, 1999.

[EH81] B. S. Everitt and D. J. Hand, *Finite mixture distributions*, Monographs on Applied Probability and Statistics, Chapman and Hall, London, 1981.

[Ero02] Elena A. Erosheva, *Grade of membership and latent structure models with application to disability survey data*, Ph.D. thesis, Carnegie Mellon University, Department of Statistics, 2002.

[Goo78] L. A. Goodman, *Analyzing qualitative/categorical data: Log-linear models and latent-structure analysis*, Abt Books, Cambridge, MA, 1978.

[Hei96] T. Heinen, *Latent class and discrete latent trait models: Similarities and differences*, SAGE Publications, Thousand Oaks, 1996.

[KAMT05a] Mikhail Kovtun, Igor Akushevich, Kenneth G. Manton, and H. Dennis Tolley, *Grade of membership analysis: One possible approach to foundations*, To appear in *Focus on Probability Theory*, Nova Science Publishers. Available from e-Print archive arXiv.org at `http://www.arxiv.org`, arXiv code math.PR/0403373, 2005.

[KAMT05b] ———, *Linear latent structure analysis: Mixture distribution models with linear constraints*, Submitted for publication to "Statistical Methodology." Available from e-Print archive arXiv.org at `http://www.arxiv.org`, arXiv code math.PR/0507025, 2005.

[KYA05] Mikhail Kovtun, Anatoliy Yashin, and Igor Akushevich, *Convergence of estimators in LLS analysis*, Submitted for publication to "Journal of Nonparametric Statistics." Available from e-Print archive arXiv.org at `http://www.arxiv.org`, arXiv code math.ST/0508297, 2005.

[Laz50a] P. F. Lazarsfeld, *The interpretation and computation of some latent structures*, Studies in Social Psychology in World War II, vol. IV, ch. 11, pp. 413–472, Princeton University Press, Princeton, NJ, 1950.

[Laz50b] ———, *The logical and mathematical foundations of latent structure analysis*, Studies in Social Psychology in World War II, vol. IV, ch. 10, pp. 362–412, Princeton University Press, Princeton, NJ, 1950.

[LH68] P. F. Lazarsfeld and N. W. Henry, *Latent structure analysis*, Houghton Mifflin Company, Boston, MA, 1968.

[Lin95] Bruce G. Lindsay, *Mixture models: Theory, geometry and applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 5, Institute of Mathematical Statistics, Hayward, CA, 1995.

[LR88] R. Langeheine and J. Rost (eds.), *Latent trait and latent class models*, Plenum Press, New York, 1988.

[MM02] George A. Marcoulides and Irini Moustaki (eds.), *Latent variable and latent structure models*, Methodology for Business and Management, Lawrence Erlbaum Associates, Mahwah, NJ, 2002.

[TSM85] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical analysis of finite mixture distributions*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Chichester, 1985.

Mikhail Kovtun: Duke University, CDS, 2117 Campus Drive, Durham, NC 27708
E-mail address: mkovtun@cds.duke.edu

Igor Akushevich: Duke University, CDS, 2117 Campus Drive, Durham, NC 27708
E-mail address: aku@cds.duke.edu

Anatoliy Yashin: Duke University, CDS, 2117 Campus Drive, Durham, NC 27708
E-mail address: yashin@cds.duke.edu