Published in final edited form as:
ESAIM Probab Stat. 2014 January ; 18: 207–232. doi:10.1051/ps/2011166.

ON IDENTIFIABILITY OF MIXTURES OF INDEPENDENT DISTRIBUTION LAWS*, **, ***

Mikhail Kovtun[1], **Igor Akushevich**[2], and **Anatoliy Yashin**[2]

Mikhail Kovtun: mikhail.kovtun@duke.edu; Igor Akushevich: igor.akushevich@duke.edu; Anatoliy Yashin: anatoly.yashin@duke.edu

[1]Duke University, Dept. of Biology, Benfey Lab Durham, 27708 NC, 27708, USA

[2]Center for Population Health and Aging Durham, 27708 NC, USA

Abstract

We consider representations of a joint distribution law of a family of categorical random variables (*i.e.*, a multivariate categorical variable) as a mixture of independent distribution laws (*i.e.* distribution laws according to which random variables are mutually independent). For infinite families of random variables, we describe a class of mixtures with identifiable mixing measure. This class is interesting from a practical point of view as well, as its structure clarifies principles of selecting a "good" finite family of random variables to be used in applied research. For finite families of random variables, the mixing measure is never identifiable; however, it always possesses a number of identifiable invariants, which provide substantial information regarding the distribution under consideration.

Keywords and phrases

Latent structure analysis; mixed distributions; identifiability; moment problem

1. Introduction

Linear latent structure (LLS) analysis is aimed at deriving properties of a population as a whole and properties of individuals from a large number of categorical measurements made on each individual in a sample. An exposition of LLS analysis is given in [8].

LLS analysis searches for a representation of the observed joint distribution of random variables (representing measurements) as a mixture of *independent distributions*, *i.e.* distributions, in which random variables are mutually independent. Such an approach is common for all branches of latent structure analysis. The specific LLS assumption is that the mixing measure is supported by a low-dimensional linear subspace of the space of independent distributions.

*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 NIA grants R01AG030198, R01AG028259, and R01AG032319.
***The work of the third author was supported by NIA grants R01AG030198, R01AG028259, and R01AG032319.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

In this article we investigate in detail the question of identifiability of LLS models, *i.e.*, what are conditions under which the observed distribution uniquely defines the mixing measure.

It is known that mixing distribution is in general not identifiable. More or less substantial results exist for finite mixtures and for a few special classes (see, for example [2, 5, 11–14]). From this point of view, the present article has interest as it discovers a new family of identifiable mixture invariants.

The results obtained in this article are of great importance for practial applications of LLS analysis. These results show that given the observations of a multivariate categorical variable (*e.g.*, outcome of a survey) one can estimate:

- dimensionality of the supporting subspace of mixing measure;
- a basis of the supporting subspace;
- stability level of the observed distribution;
- higher-order (up to stablity level) moments of mixing measure;
- characteristics of individuals in terms of expectation of position in β-space conditional on answers given by an individual (this topic is not covered in the present article).

The results we will show suggest that estimation of various parameters of an LLS model can be derived from the moment matrix, which, in turn, can be consistently estimated by frequences. There exists a very efficient estimation algorithm, which can be used to analyze thousands of categorical variables observed on millions of individuals.

The detailed exposition of statistical estimators and estimation algorithms is given in [1, 8]. These articles also discuss other statistical properties of LLS estimators and provide many examples of application of LLS analysis, both artificial and practical.

The question of identifiability of a mixing measure that produces the observed joint distribution can easily be formulated without explicitly referencing the LLS analysis. This allows us to make the article self-contained. We provide such formulation in Section 2.

In Section 3 we develop the notion of the moment matrix of a mixing measure, which we will use as a primary tool in our subsequent investigation.

Using this tool, we prove in Section 4 that the supporting subspace of a mixing measure is a mixture invariant in the class of stable essential distributions. We also prove an important Theorem 4.18, which shows that there are no stable low-dimensional mixing measures except essential ones if the observed distribution has a sufficiently high level of stability. We discuss the importance of this theorem for practical applications in greater detail in Section 7.

In Section 5 we describe another set of mixture invariants – low-order moments of mixing measure. The existence of these invariants allows us to prove that in the infinite-dimensional case the essential mixing measure is identifiable.

In Section 7 we discuss implications of the results we have obtained for practical applications.

Most of the proofs in the present article are exercises in linear algebra. Although in most cases they are straightforward, the necessity to consider high-dimensional cases (β-space should have dimensionality at least 6 to observe non-trivial behavior) and complexity of notations may make understanding the proofs difficult. To help the reader, we have created a number of examples. It might seem natural to have examples spread out over the text. However, in most cases the examples depend on definitions and theorems that appear in the article after the first reference to an example; all examples share a common notation; often subsequent examples use constructions defined in the previous ones; some examples are referenced in the article multiple times. Taking all of this into account, we have decided to put all of the examples in a separate section (Sect. 6).

2. The problem

We consider an infinite family of random variables $\{X_j\}_j$[3]; variable X_j takes values in a finite set $\{1, \ldots, L_j\}$. Without loss of generality, these variables can be considered as being defined on a probability space $\mathscr{A}_\infty = \prod_{j=1}^{\infty} \{1, \ldots, L_j\}$ with $X_j(a) = X_j(a_1, \ldots) = a_j$. The space $\mathcal{A}_\infty$, endowed with Tikhonov topology, is compact and metrizable by metric $\rho(a, a') = 1/\inf\{j \mid a_j \neq a'_j\}$; it is complete with respect to this metric[4]. The Borel σ-algebra $\mathcal{B}(\mathcal{A}_\infty)$ on $\mathcal{A}_\infty$ coincides with the σ-algebra, generated by random variables $\{X_j\}_j$; thus, joint distributions of $\{X_j\}_j$ are probability Borel measures on $\mathcal{A}_\infty$. We use $\mathcal{P}(\mathcal{A}_\infty)$ to denote the space of all probability measures on $\mathcal{A}_\infty$.

Remark 2.1

Here, we use the term "random variable" as a synonym for "measurable function"; no distribution law is assumed implicitly. Thus, the specification of family $\{X_j\}_j$ is merely the specification of space $\mathcal{A}_\infty$. In fact, the whole exposition can be conducted just as a discussion of some properties of $\mathcal{P}(\mathcal{A}_\infty)$, without explicit introduction of random variables $\{X_j\}_j$. The sole purpose of introducing random variables is to provide better insight for the subsequent constructions.

We always consider $\mathcal{P}(\mathcal{A}_\infty)$ with topology of weak convergence. Topology of weak convergence on $\mathcal{P}(\mathcal{A}_\infty)$ is metrizable, $\mathcal{P}(\mathcal{A}_\infty)$ is compact, and thus separable ([4], IV.6.3; [7], IV.3.4.4,5). $\mathcal{P}(\mathcal{A}_\infty)$ is a subset of $\mathfrak{M}(\mathcal{A}_\infty)$, the linear space of all signed Radon measures on $\mathcal{A}_\infty$. When we speak about a linear structure of $\mathcal{P}(\mathcal{A}_\infty)$, we always assume the linear structure inherited from $\mathfrak{M}(\mathcal{A}_\infty)$.

[3]As we have to use complex (and, probably, cumbersome) notations, we are trying to be as unambiguous as possible. In this particular case, we have to distinguish between a set consisting of a *single* random variable X_j (which we denote as $\{X_j\}$) and a set consisting of *multiple* random variables X_j where j ranges over set $\mathcal{J}$. The full notation in the latest case is $\{X_j \mid j \in \mathcal{J}\}$, or $\{X_j\}_{j\in\mathcal{J}}$. We abbreviate the last one to $\{X_j\}_j$ if $\mathcal{J}$ is obvious from the context or is insignificant.

[4]Finite spaces $\{1, \ldots, L_j\}$ are endowed with discrete topology.

We also consider finite subfamilies $\{X_j\}_{j=1,\dots,n}$. Random variables in such subfamilies are considered as defined on a space $\mathscr{A}_n = \prod_{j=1}^{n}\{1,\dots,L_j\}$. The space $\mathcal{A}_n$ is finite, and its Tikhonov topology is the discrete one. The set of probability measures on $\mathcal{A}_n$, $\mathcal{P}(\mathcal{A}_n)$, is a simplex in a finite-dimensional Euclidean space (with dimensionality $|\mathcal{A}_n| = L_1 \cdot \ldots \cdot L_n$), and the topology of weak convergence on $\mathcal{P}(\mathcal{A}_n)$ coincides with the Euclidean topology.

There are natural projections $\pi_n : \mathcal{A}_\infty \to \mathcal{A}_n$ and $\pi_{m,n} : \mathcal{A}_m \to \mathcal{A}_n$ (for $m \geq n$), defined in an obvious manner: $\pi_n(a_1, \dots, a_n, a_{n+1}, \dots) = (a_1, \dots, a_n)$, and $\pi_{m,n}(a_1, \dots, a_n, a_{n+1}, \dots, a_m) = (a_1, \dots, a_n)$. These projections induce mappings of corresponding spaces of measures $\pi_n : \mathcal{P}(\mathcal{A}_\infty) \to \mathcal{P}(\mathcal{A}_n)$ and $\pi_{m,n} : \mathcal{P}(\mathcal{A}_m) \to \mathcal{P}(\mathcal{A}_n)$, defined as $(\pi(\mathrm{P}))(A) = \mathrm{P}(\pi^{-1}(A))$. It is important for our purposes that these mappings of measure spaces are linear.

Most of the considerations in this section apply to both $\mathcal{A}_\infty$ and $\mathcal{A}_n$. In such contexts we use $\mathcal{A}$ to denote either $\mathcal{A}_\infty$ or $\mathcal{A}_n$; the index is included only when a statement reads differently for finite and infinite cases.

We shall often use *elementary cylinders* in $\mathcal{A}$, *i.e.* subsets of $\mathcal{A}$, whose projections on the factors of $\mathcal{A}$ are either the whole factor or a single point (the latter is possible only for finitely many factors). For such cylinders, we will use notation

$$\mathrm{Cyl}[j_1{:}l_1, \dots, j_p{:}l_p] \overset{\mathrm{def}}{=} \{a \in \mathscr{A} \,|\, a_{j_1} = l_1 \wedge \cdots \wedge a_{j_p} = l_p\}$$

(this notation also implicitly assumes that $j_1, \dots, j_p$ are distinct integers).

Among all joint distributions of $\{X_j\}_j$, we distinguish *independent* ones, *i.e.* those distributions, in which random variables $\{X_j\}_j$ are mutually independent. To specify an independent distribution, one needs to specify only probabilities $\beta_{jl} = \mathrm{P}(X_j = l)$. Then, due to independence, one has for elementary cylinders $\mathrm{P}(\mathrm{Cyl}[j_1 : l_1, \dots, j_p : l_p]) = \mathrm{P}(X_{j_1} = l_1 \wedge \cdots \wedge X_{j_p} = l_p) = \beta_{j_1 l_1} \dots \beta_{j_p l_p}$. Since elementary cylinders compose a topology base, this uniquely extends to $\mathcal{B}(\mathcal{A})$. Thus, any independent distribution is uniquely described by an infinite-dimensional vector $\beta = (\beta_{11}, \dots, \beta_{1L_1}, \dots, \beta_{j1}, \dots, \beta_{jL_j}, \dots)$. To specify a probability distribution, such a vector must satisfy conditions:

$$\begin{cases} 0 \leq \beta_{jl} \leq 1 & \text{for all } j \text{ and } l \\ \sum_{l=1}^{L_j} \beta_{jl} = 1 & \text{for all } j. \end{cases} \quad (2.1)$$

The set P of vectors satisfying (2.1) is convex and bounded; it is closed and compact in Tikhonov topology. We use P_β to denote the independent measure on $\mathcal{A}$ corresponding to a vector $\beta \in P$. We refer to the linear space containing P as *β-space*. The β-space is either an infinite-dimensional space (when an infinite family of random variables is considered) or a finite-dimensional space of dimensionality $L_1 + \cdots + L_n$ (if a finite family $X_1, \dots, X_n$ is considered).

The mapping $\beta \mapsto P_\beta$ is one among many possible parametrizations of the family of independent distributions. This particular parametrization, however, possesses a number of good properties, one of which is the convexity of P.

Remark 2.2

Without introduction of random variables X_j, one can define independent measures on $\mathcal{A}$ as those which are products of their marginals, $P = \Pi_j P_j$, where P_j is a measure on a finite set $\{1, \ldots, L_j\}$. The vector $(\beta_{jl})_{jl}$ that corresponds to measure $\Pi_j P_j$ is defined by letting $\beta_{jl} = P_j(\{l\})$.

We shall investigate a class of mixtures of independent distributions, *i.e.* those distributions P on $\mathcal{A}$ which can be represented in form:

$$\mathrm{P}(A)=\int_P \mathrm{P}_\beta(A)\mu(\mathrm{d}\beta) \quad \text{for all } A \in \mathscr{B}(\mathscr{A}) \tag{2.2}$$

where μ is a probability measure on P. For this definition of mixture to be correct, one needs to show that the mapping $\beta \mapsto P_\beta(A)$ is measurable for every Borel $A \subseteq \mathcal{A}$. We show a stronger fact (which is also important by itself, as it clarifies a relation between β-space and the space of probability measures $\mathcal{P}(\mathcal{A})$):

Proposition 2.3

The mapping $\beta \mapsto P_\beta$ is a homeomorphism of P onto a set of independent distributions in $\mathcal{P}(\mathcal{A})$, with respect to Tikhonov topology on $\mathbb{R}^\infty$ and topology of weak convergence on $\mathcal{P}(\mathcal{A})$.

Proof—The fact that the mapping $\beta \mapsto P_\beta$ is one-to-one follows from Remark 2.2.

To prove the continuity in both directions, it is sufficient to show that a sequence $\{\beta^n\}_n$ in P converges to $\beta \in P$ if and only if the sequence P_{β^n} weakly converges to P_β. The convergence $P_{\beta^n} \to P_\beta$ is equivalent to the convergence $P_{\beta^n}(C) \to P_\beta(C)$ for every elementary cylinder C ([10], III.1.5). (Note that in $\mathcal{A}$ every cylinder is an open-closed set; thus, its boundary is empty.) But for a cylinder $C = \mathrm{Cyl}[j_1 : l_1, \ldots, j_p : l_p]$ one has $\mathrm{P}_{\beta^n}(C)=\prod_q \beta^n_{j_q l_q}$, and required convergence is obviously equivalent to the convergence $\beta^n \to \beta$.

Corollary 2.4

For every $A \in \mathcal{B}(\mathcal{A})$, the mapping $\beta \mapsto P_\beta(A)$ is measurable (w.r.t. corresponding Borel σ-algebras).

Proof—By ([3], III.55, 60), the mapping $P \mapsto P(A)$ is measurable. Thus, our mapping is measurable as a composition of two measurable mappings.

Remark 2.5

The Tikhonov topology on $\mathbb{R}^\infty$ is the topology of pointwise convergence. Proposition 2.3 may fail to be true if P is considered with other topologies.

Remark 2.6

Note that the mapping $\beta \mapsto \mathrm{P}_\beta$ is *not* a linear mapping. In a finite-dimensional case, the image of *P* under this mapping is a (part of a) polynomial surface (more precisely, it is an intersection of quadratic hypersurfaces). The family of mixtures of independent distributions is the convex hull of the image of *P*.

We abbreviate equation (2.2) to P = Mix(μ). We also use P_μ to denote Mix(μ).

The question that we are interested in is what are the conditions for identifiability of mixtures of independent distributions (in sense of [12]), *i.e.* under what conditions does the mixture Mix(μ) uniquely define the mixing measure μ or some of its invariants. It is easy to show, however, that without additional restrictions *any* distribution on $\mathcal{A}$ can be represented as a mixture of independent ones, and (except degenerate cases) every distribution has infinitely many such representations.

The additional restrictions that we discuss below are restrictions on the dimensionality of the support of a mixing measure in β-space.

Remark 2.7

Note that these restrictions are expressed in terms of the chosen parametrization of the family of independent distributions. Example 6.1 demonstrates that the image of one-dimensional subset of *P* under the mapping $\beta \mapsto \mathrm{P}_\beta$ can be infinite-dimensional.

For a measure μ on *P*, let $\mathcal{L}(\mu)$ denote the smallest linear subspace of β-space supporting μ (this subspace is a linear span of supp(μ)). We refer to this subspace as *the supporting subspace* of μ.

Remark 2.8

Due to restrictions (2.1), the mixing measure is carried by a (bounded convex subset of a) linear manifold that never passes through the origin. The supporting subspace is the subspace spanned by this manifold (and thus has 1 more dimension than the manifold has). This additional “unnecessary” dimension serves a good purpose as it allows us to avoid separate consideration of special cases. The motivation to use supporting subspace instead of supporting manifold is very similar to the motivation for “homogeneous coordinates” in projective geometry or the motivation for the use of “barycentric coordinates” to specify points within a simplex.

Definition 2.9

A rank of mixing measure μ is the dimensionality of its supporting subspace,

$$\mathrm{rank}(\mu) \stackrel{\mathrm{def}}{=} \dim(\mathscr{L}(\mu)).$$

Definition 2.10

A rank of a distribution $\mathrm{P} \in \mathcal{P}(\mathcal{A})$ is the smallest rank of a mixing measures generating P,

$$\mathrm{rank}(\mathrm{P}) \stackrel{\mathrm{def}}{=} \min\{\mathrm{rank}(\mu) | \mathrm{P} = \mathrm{Mix}(\mu)\}.$$

Note that according to our definitions of rank, one has rank(P_μ) ≤ rank(μ), but it is possible to have a strict inequality (see Example 6.6). This suggests the following:

Definition 2.11

A mixing measure μ is *essential*, if rank(μ) = rank(P_μ).

Obviously, there is no identifiability if nonessential mixing measures are allowed. In subsequent considerations, we shall establish identifiability conditions within the class of essential mixing measures. For the applied researcher, one motivation for such a restriction is Occam's razor. We shall also show that a natural class of "stable mixing measure with ∞-stable support" does not contain nonessential mixing measures (see Thms. 4.18 and 4.20).

The question of identifiability now can be split into two subquestions:

- For a distribution P of rank K, what are conditions for identifiability of the supporting subspace of an essential mixing measure producing P? More precisely: what are conditions for the following implication

  $$\begin{array}{l}\mathrm{rank}(\mathrm{P})=K\\ \mathrm{P}=\mathrm{Mix}(\mu)=\mathrm{Mix}(\mu^{'})\\ \mathrm{rank}(\mu)=\mathrm{rank}(\mu^{'})=K\end{array}\Rightarrow \mathscr{L}(\mu)=\mathscr{L}(\mu^{'})$$

  to be true?

- For a fixed K-dimensional subspace $\mathscr{L}$ of $\mathscr{B}$, what are conditions for identifiability of mixing measure carried by this subspace? More precisely: what are conditions for the following implication

  $$\begin{array}{l}\mathrm{Mix}(\mu)=\mathrm{Mix}(\mu^{'})\\ \mathscr{L}(\mu)=\mathscr{L}(\mu^{'})=\mathscr{L}\end{array}\Rightarrow \mu=\mu^{'}$$

  to be true?

We shall show that the answer to the first question (identifiability of the supporting subspace) is positive even in a finite-dimensional case, provided that the dimensionality of the supporting subspace is sufficiently smaller than the dimensionality of β-space (alongside with some non-degenerality conditions).

In contrast, the second question (identifiability of the mixing measure) has a positive answer only in the infinite-dimensional case. However, in the finite-dimensional case, there exists a number of identifiable invariants (*i.e.* properties which are the same for all mixing measures that produce the observed mixture) – namely, the low-order moments of the mixing measure.

Here, we give a semi-formal formulation of the main results of this paper (the strict formulation will be deferred until we introduce more notions).

The *visible rank* of distribution P (see Def. 4.9) is what can be concluded from direct observation of its *moment matrix*. We say that a distribution of rank K is k-stable if, after removing arbitrary k random variables from consideration, its visible rank is still K. Similarly, a supporting subspace is k-stable if its dimensionality does not decrease after removing arbitrary k random variables, and a mixing measure μ is k-stable if $\mathcal{L}(\mu)$ is k-stable. We say "stable" instead of "1-stable".

k-stability essentially means that any dimension of a supporting subspace is confirmed by most random variables – not just one variable or a few of them.

The main results of the present article are:

Theorem [semi-formal version of Thm. 4.12]

Supporting subspace of a stable distribution is identifiable.

The above theorem speaks for itself. The next theorem justifies restricting our attention to essential mixtures. The theorem states that any possible non-essential mixture is "very far" from the essential one, and thus can hardly be considered as a "good alternative" to an essential mixture (as it is described in detail in Sect. 7, in practice one would expect $k \gg K$).

Theorem [semi-formal version of Thm. 4.18]

Let P be a distribution of rank K with a k-stable supporting subspace. Then there is no stable mixing measure μ producing P with $K < \text{rank}(\mu) \le k$.

The infinite-dimensional version of this theorem has an even more elegant form:

Theorem [semi-formal version of Thm. 4.20]

Let P be a distribution of rank K with ∞-stable supporting subspace. Then there is no stable mixing measure μ producing P with $\text{rank}(\mu) > K$.

The following theorems clarify how much can be learned about a mixing measure.

Theorem [semi-formal version of Thm. 5.3]

For a stable distribution with k-stable supporting subspace, all moments of a mixing measure up to order $k + 1$ are identifiable.

Theorem [semi-formal version of Thm. 5.4]

For a stable distribution with an ∞-stable supporting subspace, all moments of a mixing measure are identifiable; consequently, the mixing measure itself is identifiable.

General remark on dimensionality

All definitions and theorems below are applicable both to the case of finite and infinite numbers of random variables. This follows from the fact that all definitions and proofs are made in terms of *finite* submatrices (which may be selected from either finite or infinite matrices).

However, we consider distributions and measures only of *finite* rank. Most of our results are not applicable to distributions and measure of infinite rank (and it is not the purpose of this article to investigate the case of infinite rank).

3. Moment matrix

The main tool for the subsequent reasoning is the *moment matrix*, which is introduced in this section. For a cylinder C = Cyl[$j_1 : l_1, \ldots, j_p : l_p$] equation (2.2) gives

$$\mathrm{P}_{\mu}(C)=\int_{P}\mathrm{P}_{\beta}(C)\mu(\mathrm{d}\beta)=\int_{P}\prod_{q=1}^{p}\beta_{j_q l_q}\mu(\mathrm{d}\beta). \quad (3.1)$$

Equation (3.1), although obvious, is of great importance. The left-hand side of this equation is the measure of a cylinder. For the identifiability problem, these measures are known quantities, and the measures of cylinders uniquely define the measure P_μ on the whole σ-algebra $\mathcal{B}(\mathcal{A})$.

The right-hand side of equation (3.1) is a moment of mixing measure. The support of the mixing measure μ is always compact (as it is contained in a compact set P). If in addition it is finite-dimensional, the set of all moments uniquely determines μ ([6], 5.5.2).

Unfortunately, equation (3.1) does not allow us to find *all* moments of the mixing measure. For example, for moments $\int\beta_{11}^2\mu(\mathrm{d}\beta)$ and $\int\beta_{11}\beta_{12}\mu(\mathrm{d}\beta)$ there are no corresponding cylinders in $\mathcal{A}$.

We shall find conditions under which the known moments of a mixing measure uniquely determine the unknown ones, which gives us the conditions of identifiability of the mixing measure with a finite-dimensional support.

The first step is to develop a way of indexing moments of μ. A *moment index* is an integer vector v with components indexed by pairs j, l–exactly in the same way as components of vectors in β-space. All components of a moment index v are nonnegative integers, and only finitely many components are distinct from 0. The *order* of moment index v is $|v| = \Sigma_{j,l} v_{jl}$.

Let $\mathcal{V}$ be a set of all moment indices. For $v \in \mathcal{V}$, the moment of μ with index v, denoted $m_v(\mu)$ (or simply m_v when μ is obvious from the context), is:

$$m_v=m_v(\mu) \overset{\mathrm{def}}{=} \int_{P}\prod_{j,l}\beta_{jl}^{v_{jl}}\ \mu(\mathrm{d}\beta). \quad (3.2)$$

Note that the product under the integral in the right-hand side of the equation contains $|v|$ factors, and thus is finite.

Moment indices of order 1 are used often in the rest of the paper, so we introduce a special notation for them. Namely, we use $v[j, l]$ to denote a moment index that has 1 in position j, l and 0s in all other positions.

The *height* of moment index v is $h(v) = \Sigma_j (\Sigma_l v_{jl} - 1)^+$ (where $(x - y)^+$ denotes "positive part", *i.e.* $\max(x - y, 0)$). Informally, the height of a moment index v describes "how far" v is from moments (3.1), all of which have height 0.

By slightly abusing the language, we will say "height of moment m_v " etc. instead of "height of moment index v of moment m_v " *etc*.

All moments in (3.1) (and only such moments) have height 0; thus, we can define a cylinder that corresponds to a moment index v of height 0. We denote this cylinder by Cyl[v]; if v has 1 at positions $j_1l_1,\ldots, j_pl_p$ (and zeroes at all other positions), then Cyl[v] = Cyl[j_1: l_1, …, j_p: l_p].

We arrange all moments of μ in the *moment matrix*. Rows of the moment matrix are indexed by moment indices of order 1; columns of the moment matrix are indexed by all moment indices. The element of the moment matrix in row v' and column v'' is the moment of μ with index $v' + v''$ (addition of moment indices is the usual addition of vectors).

To unambiguously write a moment matrix, we have to impose some ordering on $\mathcal{V}$.

First, we order moment indices of order 1. We say that $v' = v[j', l']$ is smaller than $v'' = v[j'', l'']$, denoted $v' \prec v''$, if either $j' < j''$ or $j' = j''$ and $l' < l''$. Informally, $v' \prec v''$, if 1 in v' is on the left from 1 in v''.

Second, we order moments of order k for $k = 2, \ldots$. For this, note that if $|v| = k$, there exists a unique representation $v = v^{(1)} + \cdots + v^{(k)}$ such that $|v^{(1)}| = \cdots = |v^{(k)}| = 1$ and $v^{(1)} \preceq \cdots \preceq v^{(k)}$. Now for arbitrary v' and v'' of order k with representations $v' = v'^{(1)} + \cdots + v'^{(k)}$ and $v'' = v''^{(1)} + '' + v''^{(k)}$ we say that $v' \prec v''$ if representation of v' is lexicographically smaller than representation of v'', *i.e.* if for some $p < k$ one has $v'^{(1)} = v''^{(1)}, \ldots, v'^{(p)} = v''^{(p)}$ and $v'^{(p+1)} \prec v''^{(p+1)}$.

Finally, for arbitrary $v', v'' \in \mathcal{V}$ we say that $v' \prec v''$, if either $|v'| < |v''|$ or $|v'| = |v''|$ and $v' \prec v''$ in the above sense.

The moment matrix always has infinitely many columns. It has finitely or infinitely many rows depending on whether we consider a finite or infinite family of random variables.

The following is an example of the moment matrix for the case of two binary variables:

$$\begin{array}{c} \\ (1,0;0,0) \\ (0,1;0,0) \\ (0,0;1,0) \\ (0,0;0,1) \end{array} \left(\begin{array}{cccccccc} (0,0;0,0) & (1,0;0,0) & (0,1;0,0) & (0,0;1,0) & (0,0;0,1) & (2,0;0,0) & (1,1;0,0) & \cdots \\ m_{(1000)} & m_{(2000)} & m_{(1100)} & m_{(1010)} & m_{(1001)} & m_{(3000)} & m_{(2100)} & \cdots \\ m_{(0100)} & m_{(1100)} & m_{(0200)} & m_{(0110)} & m_{(0101)} & m_{(2100)} & m_{(1200)} & \cdots \\ m_{(0010)} & m_{(1010)} & m_{(0110)} & m_{(0020)} & m_{(0011)} & m_{(2010)} & m_{(1110)} & \cdots \\ m_{(0001)} & m_{(1001)} & m_{(0101)} & m_{(0011)} & m_{(0002)} & m_{(2001)} & m_{(1101)} & \cdots \end{array} \right).$$

Note that the indices of the rows of the moment matrix correspond to indices of coordinates in β-space. This allows us to consider columns of the moment matrix as vectors in β-space. We use M_v to denote the column of the moment matrix indexed by moment index v. We also use M_0 to denote the first column of the moment matrix (instead of lengthy $M_{(0,\ldots,0)}$).

The importance of the above structure of the moment matrix is explained by the following:

Theorem 3.1

If measure μ is supported by a finite-dimensional subspace $\mathcal{L}$ of β-space, then each column of the moment matrix belongs to $\mathcal{L}$.

Proof—Let $\lambda_1, \ldots, \lambda^K$ be a basis of $\mathcal{L}$. Then the mapping $(g_1, \ldots, g_K) \mapsto \Sigma_k g_k \lambda^k$ is a linear isomorphism of arithmetic space $\mathbb{R}^K$ and $\mathcal{L}$. This isomorphism allows us to consider an equivalent measure μ_g on $\mathbb{R}^K$ (which is an image of measure μ under this isomorphism).

For arbitrary $v \in \mathcal{V}$ one obtains:

$$\begin{aligned} m_{v+v[j_0,l_0]} &= \int \left(\prod_{j,l} \beta_{jl}^{v_{jl}}\right) \beta_{j_0 l_0}\, \mu(\mathrm{d}\beta) \\ &= \int \left(\prod_{j,l} \left(\sum_k g_k \lambda_{jl}^k\right)^{v_{jl}}\right) \left(\sum_k g_k \lambda_{j_0 l_0}^k\right) \mu_g(\mathrm{d}g) \\ &= \sum_k \lambda_{j_0 l_0}^k \int \left(\prod_{j,l} \left(\sum_k g_k \lambda_{jl}^k\right)^{v_{jl}}\right) g_k \mu_g(\mathrm{d}g). \end{aligned} \quad (3.3)$$

Let $b_k = \int \left(\prod_{j,l} \left(\sum_k g_k \lambda_{jl}^k\right)^{v_{jl}}\right) g_k \mu_g(\mathrm{d}g)$. Note that b_k does not depend on j_0, l_0. Combining equations (3.3) for all possible j_0, l_0, one obtains a vector equation:

$$M_v = \sum_k b_k \lambda^k$$

which means exactly that the column v of the moment matrix belongs to $\mathcal{L}$.

On the other hand, the submatrix of the moment matrix consisting of columns indexed by moment indices of order 1 is a shifted covariance matrix of μ. Namely, let C_{jl} denote a column of the covariance matrix, $C_{jl} = (\mathrm{Cov}(\beta_{jl}, \beta_{j'l'}))_{j',l'}$. Then one has:

$$C_{jl} = M_{v[j,l]} - m_{v[j,l]} \cdot M_0. \quad (3.4)$$

It is a well-known fact that a measure ν in Euclidean space is carried by a linear manifold m + Lin(Cov(ν)), where m is a vector of means and Lin(Cov(ν)) is a linear subspace spanned by columns of the covariance matrix of ν. Furthermore, Lin(m, Cov(ν)) is the smallest linear subspace carrying ν. In addition, the linear manifold carrying a mixing measure never passes through the origin (due to the property $\Sigma_l \beta_{jl} = 1$). Thus, taking into account equation (3.4), we obtain:

Theorem 3.2

The supporting subspace of μ coincides with the linear span of all columns of the moment matrix and coincides with the linear span of the columns indexed by moment indices of order 1,

$$\mathscr{L}(\mu)=\mathrm{Lin}(\{M_v\}_{v\in\mathscr{V}})=\mathrm{Lin}(\{M_v\}_{|v|=1}).$$

4. Identifiability of the supporting subspace

Equation (3.1) allows us to conclude that equality $P_\mu = P_{\mu'}$ implies

$$m_v(\mu)=m_v(\mu') \quad \text{for all } v \text{ of height } 0. \quad (4.1)$$

It follows from Theorem 3.2 that equality

$$m_v(\mu)=m_v(\mu') \quad \text{for all } v \text{ of order} \leq 2 \quad (4.2)$$

is sufficient for $\mathscr{L}(\mu) = \mathscr{L}(\mu')$.

We shall show that (4.1) implies (4.2) provided that μ and μ' are essential and the mixture $P = P_\mu = P_{\mu'}$ is stable in the sense defined later on. To proceed, we need tools for referring various minors of moment matrices. The following definitions serve this purpose.

Definition 4.1

A *row selector* of size k is a sequence of k moment indices of order 1, $r = (v^1, \ldots, v^k)$, satisfying $v^1 \prec \cdots \prec v^k$.

Definition 4.2

A *column selector* of size k is a sequence of k moment indices (of arbitrary order), $c = (v^1, \ldots, v^k)$, satisfying $v^1 \prec \cdots \prec v^k$.

Definition 4.3

A *minor selector* is a pair of a row selector and a column selector of equal size, $s = (r, c)$.

Definition 4.4

The *height* of a minor selector $s = ((v^1, \ldots, v^k), (w^1, \ldots, w^k))$ is $h(s) = \max_{i,j} h(v^i + w^j)$.

Notation 4.5

For a mixing measure μ and a minor selector $s = ((v^1, \ldots, v^k), (w^1, \ldots, w^k))$, $s(\mu)$ denotes a minor of the moment matrix of μ:

$$s(\mu)=\begin{pmatrix} m_{v^1+w^1}(\mu) & \ldots & m_{v^1+w^k}(\mu) \\ m_{v^k+w^1}(\mu) & \ldots & m_{v^k+w^k}(\mu) \end{pmatrix}.$$

Notation 4.6

For a distribution P and a minor selector $s = ((v^1, \ldots, v^k), (w^1, \ldots, w^k))$ of height 0, s(P) denotes a matrix constructed from probabilities of cylinders:

$$s(\mathrm{P})=\begin{pmatrix} \mathrm{P}(\mathrm{Cyl}[v^1+w^1]) & \ldots & \mathrm{P}(\mathrm{Cyl}[v^1+w^k]) \\ \ldots & & \\ \mathrm{P}(\mathrm{Cyl}[v^k+w^1]) & \ldots & \mathrm{P}(\mathrm{Cyl}[v^k+w^k]) \end{pmatrix}.$$

Notation 4.7

For a moment index v, let $\mathcal{J}(v)$ denote a set of js such that v has a nonzero component at position j, l for some l, $\mathcal{J}(v) = \{j \mid \exists l : v_{jl} > 0\}$. We also extend this notation to sets of moment indices: $\mathcal{J}(v^1, \ldots, v^k) = \mathcal{J}(v^1) \cup \cdots \cup \mathcal{J}(v^k)$

Definition 4.8

A minor selector s (respectively, row selector r, column selector c) *touches* variable X_j (or, simply, touches j), if $j \in \mathcal{J}(s)$ (respectively, $\mathcal{J}(r)$, $\mathcal{J}(c)$).

Definition 4.9

A distribution P has *visible rank k* if there exist a minor selector s of size k and height 0 such that s(P) is nondegenerate.

The following definition is a key property of distribution P that is required to prove various identifiability results.

Definition 4.10

A distribution P of rank K is *k-stable* if it retains visible rank K after removing from consideration any arbitrary subfamily of k random variables.

More formally, this means that for every $j_1, \ldots, j_k$ there exists a minor selector s of size K and height 0 such that s does not touch $j_1, \ldots, j_k$ and s(P) is nondegenerate.

We say *stable* instead of 1-stable.

Remark 4.11

Note that according to our definition there might be distributions which are not even 0-stable –and Example 6.5 presents such a distribution.

With the above definitions at hand, we can formulate and prove the main theorem about identifiability of the supporting subspace of an essential mixing measure.

Theorem 4.12

Let $\mathrm{P} \in \mathcal{P}(\mathcal{A})$ be a stable distribution of rank K, and μ, μ' be essential mixing measures producing P (i.e. $\mathrm{P}_\mu = \mathrm{P}_{\mu'} = \mathrm{P}$).

Then $\mathcal{L}(\mu) = \mathcal{L}(\mu')$.

Proof—Due to Theorem 3.2, it is sufficient to show that under the assumptions of the theorem one has $m_v(\mu) = m_v(\mu')$ for all moment indices v of order 2. It follows from $P_\mu = P_{\mu'}$ that equality $m_v(\mu) = m_v(\mu')$ holds for moment indices v of height 0; thus, we need to prove this equality for moment indices of height 1. Figure 1 illustrates the proof.

Consider an arbitrary moment index v of height 1. It can be represented as $v = v^0 + w^0$ with $v^0 = v[j, l]$ and $w^0 = v[j, l']$. As P is stable, there exists a minor selector $s = ((v^1, \ldots, v^K), (w^1, \ldots, w^K))$ that does not touch j, consists of moments of height 0 and is nondegenerate.

Note that as s does not touch j, moment indices $v^1 + w^0, \ldots, v^K + w^0$ and $v^0 + w^1, \ldots, v^0 + w^K$ also have height 0. Moments with these indices are equal for mixing measures μ and μ'; therefore, we omit their mixing measure in Figure 1.

As the rank of the moment matrix is K for both μ and μ', there exists a unique set of coefficients $\alpha_1, \ldots, \alpha_K$ such that

$$\begin{pmatrix} m_{v^1+w^0} \\ \vdots \\ m_{v^K+w^0} \end{pmatrix} = \alpha_1 \begin{pmatrix} m_{v^1+w^1} \\ \vdots \\ m_{v^K+w^1} \end{pmatrix} + \cdots + \alpha_K \begin{pmatrix} m_{v^1+w^K} \\ \vdots \\ m_{v^K+w^K} \end{pmatrix}.$$

Consequently, we have for columns of moment matrices of μ and μ':

$$M_{w^0}(\mu) = \sum_k \alpha_k M_{w^k}(\mu) \quad \text{and} \quad M_{w^0}(\mu') = \sum_k \alpha_k M_{w^k}(\mu')$$

and in particular:

$$m_{v^0+w^0}(\mu) = \sum_k \alpha_k m_{v^0+w^k}(\mu) = \sum_k \alpha_k m_{v^0+w^k}(\mu') = m_{v^0+w^0}(\mu')$$

which proves the required equality.

According to this theorem, a supporting subspace of any essential mixing measure is uniquely determined if distribution P is stable. This allows us to speak about *the supporting subspace of a stable distribution* P. We use $\mathcal{L}(P)$ to denote this subspace.

Example 6.2 demonstrates applications of this theorem. Example 6.7 shows that the nonstability of distribution P may lead to nonidentifiability of the supporting subspace.

The important question arising from Theorem 4.12 is whether the case of stable distribution is *generic, i.e.* whether “almost all” (in some strict sense) distributions are stable? We shall give an affirmative answer to this question elsewhere.

Now we show that a stronger identifiability statement holds. To do that, we need a couple more notions.

Notation 4.13

Let $\Lambda = \{\lambda^1, \ldots, \lambda^k\}$ be a family of vectors in β-space, and let $r = (v^1, \ldots, v^k)$ be a row selector. Then $r(\Lambda)$ denotes a matrix:

$$r(\Lambda)=\begin{pmatrix} \lambda^1_{v^1} & \cdots & \lambda^k_{v^1} \\ \vdots & & \vdots \\ \lambda^1_{v^k} & \cdots & \lambda^k_{v^k} \end{pmatrix}.$$

(Recall that a row selector consists of moment indices of order 1, and every moment index of order 1 has form $v[j, l]$ for some j and l; further, $\beta_{v[j,l]}$ is the same as β_{jl}.)

Definition 4.14

A family of vectors $\Lambda = \{\lambda^1, \ldots, \lambda^K\}$ is *k-stable* if it retains full rank after removing components corresponding to any arbitrary subfamily of k random variables.

A subspace $\mathcal{L}$ of β-space is *k-stable*, if it has a k-stable basis.

A mixing measure μ is *k-stable*, if $\mathcal{L}(\mu)$ is k-stable.

More formally, family Λ is k-stable, if for arbitrary distinct $j_1, \ldots, j_k$ there exists a row selector r of size K such that r does not touch $j_1, \ldots, j_k$ and $r(\Lambda)$ is nondegenerate.

Note that if *some* basis of a subspace is k-stable, then *every one* of its bases is k-stable. Also, as k-stability is defined in terms of $\mathcal{L}(\mu)$, any two mixing measures with the same support have the same level of stability.

Note that the statements “P is k-stable” (Def. 4.10) and “ $\mathcal{L}$(P) is k-stable” (Def. 4.14) are not equivalent. The first statement obviously implies the second one, but the opposite implication is not true. Corollary 4.17 and the subsequent discussion shed some light on what one can say about the level of stability of P given the stability level of $\mathcal{L}$(P).

We need to clarify a possible structure of a row selector that selects a nondegenerate minor. A system of vectors $\Lambda = (\lambda^1, \ldots, \lambda^K)$ in β-space can be thought of as being constructed from blocks Λ_j, which correspond to individual random variables:

$$\Lambda=\begin{pmatrix} \Lambda_1 \\ \vdots \\ \Lambda_J \end{pmatrix}, \quad \Lambda_j=\begin{pmatrix} \lambda^1_{j1} & \cdots & \lambda^K_{j1} \\ \vdots & & \vdots \\ \lambda^1_{jL_j} & \cdots & \lambda^K_{jL_j} \end{pmatrix}. \quad (4.3)$$

Due to restrictions (2.1), each matrix Λ_j has rank at least 1, and any such matrix of rank 1 consists of equal columns. Consequently, any minor $r(\Lambda)$ that touches only submatrices Λ_j of rank 1, itself has rank 1.

Proposition 4.15

Let $\Lambda = (\lambda^1, \ldots, \lambda^K)$, $K > 1$, be a system of linearly independent vectors in β-space, and let r be a row selector of size K that selects a nondegenerate minor.

Then there exists a row selector $\tilde{r}$ such that $\tilde{r}(\Lambda)$ is nondegenerate, $\mathcal{J}(\tilde{r}) \subseteq \mathcal{J}(r)$, and $|\mathcal{J}(\tilde{r})| < K$.

Proof—If $|\mathcal{J}(r)| < K$, it is sufficient to take $\tilde{r} = r$. Assume that $|\mathcal{J}(r)| = K$. At least one of matrices Λ_j, $j \in \mathcal{J}(r)$, has rank 2 or greater (otherwise, $r(\Lambda)$ has rank 1; see paragraph before this proposition). Take 2 linearly independent rows from this matrix and add $K - 2$ rows from matrices Λ_j, $j \in \mathcal{J}(r)$, to obtain a linearly independent system of K rows. This gives the required row selector $\tilde{r}$.

The statement of the above proposition cannot be strengthened. To show this, consider the case when all random variables are binary (which means that all matrices Λ_j consist of two rows). If row selector $\tilde{r}$ touches $K - 2$ variables or less, it has to select both rows from at least two matrices, say Λ_j and $\Lambda_{j'}$. But in such a case one has $(\lambda^1_{j1}, \ldots, \lambda^K_{j1}) + (\lambda^1_{j2}, \ldots, \lambda^K_{j2}) = (1, \ldots, 1) = (\lambda^1_{j'1}, \ldots, \lambda^K_{j'1}) + (\lambda^1_{j'2}, \ldots, \lambda^K_{j'2})$ which implies that $\tilde{r}(\Lambda)$degenerate.

The following technical result is a key to subsequent theorems regarding identifiability.

Proposition 4.16

Let μ be a k-stable measure with rank$(\mu) = K \le k$ and let Λ be some basis of $\mathcal{L}(\mu)$. Then for each row selector r of size K such that $r(\Lambda)$ is nondegenerate and for every set $\mathcal{J}_0$ such that $\mathcal{J}_0 \cap \mathcal{J}(r) = \emptyset$ and $|\mathcal{J}_0| \le k - |\mathcal{J}(r)|$ there exists a column selector c such that the minor $(r, c)(\mu)$ is nondegenerate and c does not touch $\mathcal{J}_0 \cup \mathcal{J}(r)$ (which implies that the height of (r, c) is 0).

Proof—Figure 2 illustrates the proof. Without loss of generality, one can assume that $\mathcal{J}_0$ consists of the first indices, and $\mathcal{J}(r)$ touches the next indices (which justifies placement of elements in the figure). For our purposes it is sufficient to consider the part of the moment matrix consisting of moments of order 2, $M_2(\mu)$.

As μ is k-stable and $|\mathcal{J}_0| + |\mathcal{J}(r)| \le k$, there exists row selector $\tilde{r}$ such that $\tilde{r}(\Lambda)$ is nondegenerate and $\tilde{r}$ does not touch $\mathcal{J}_0 \cup \mathcal{J}(r)$. We claim that one can take $c = \tilde{r}$.

As $r(\Lambda)$ is nondegenerate, the rank of the horizontal dashed strip in $M_2(\mu)$ is K, and as $M_2(\mu)$ is symmetric, the same is true for the vertical dashed strip. Thus, the vertical dashed strip spans $\mathcal{L}(\mu)$. As $\tilde{r}(\Lambda)$ is nondegenerate, minor $A = (r, \tilde{r})(\mu)$ is also nondegenerate, and as $M_2(\mu)$ is symmetric, minor $A^T = (r, c)(\mu)$ is nondegenerate as well, q.e.d.

Corollary 4.17

Let μ be an essential mixing measure of rank K. Assume that μ is K-stable. Then P_μ is stable.

Proof—Let Λ be a some basis of the supporting subspace of μ.

Take an arbitrary j. As μ is stable, there exists a row selector r such that $r(\Lambda)$ is nondegenerate and r does not touch j. Due to Proposition 4.15, we cat take r so that it touches only $K - 1$ variables.

Now, by applying Proposition 4.16 to μ, $\mathcal{J} = \{j\}$, r, we obtain a nondegenerate minor of height 0 that does not touch j. Thus, P_μ is stable.

As Example 6.9 shows, the statement of the above proposition cannot be strengthened: if the level on stability of μ is smaller than its rank, some row selectors may have no required column selectors.

Example 6.9 also shows that for a stable distribution P of rank K its supporting subspace $\mathcal{L}(\mathrm{P})$ need not to be K-stable. This, in particular, means that the following theorem is not a generalization of Theorem 4.12.

Theorem 4.18

Let P be a stable distribution with $\mathcal{L}(\mathrm{P})$ being k-stable. Then there is no stable mixing measure $\bar\mu$ such that:

- $P_{\bar\mu} = \mathrm{P}$;
- $\mathrm{rank}(\bar\mu) \le k$;
- $\mathcal{L}(\bar\mu) \ne \mathcal{L}(\mathrm{P})$.

Proof—Note that if $k < \mathrm{rank}(\mathrm{P})$, there is no measure $\bar\mu$ such that $P_{\bar\mu} = \mathrm{P}$ and $\mathrm{rank}(\bar\mu) \le k$, and the statement of the theorem trivially holds. Thus we have to prove the theorem for the case $k \ge \mathrm{rank}(\mathrm{P})$.

Assume $P_{\bar\mu} = \mathrm{P}$ and $\mathrm{rank}(\bar\mu) \le k$. Let μ be any essential mixing measure generating P. We shall show that $\mathcal{L}(\bar\mu) = \mathcal{L}(\mu)$. As in Theorem 4.12, to prove this fact we have to show that $m_\nu(\bar\mu) = m_\nu(\mu)$ for all moment indices of order 2 and height 1.

Let $K = \mathrm{rank}(\mu) = \mathrm{rank}(\mathrm{P})$ and $\bar K = \mathrm{rank}(\bar\mu)$. Let $\Lambda = (\lambda^{(1)}, \dots, \lambda^{(K)})$ be a basis of $\mathcal{L}(\mu)$ and $\bar\Lambda = (\bar\lambda^{(1)}, \dots, \bar\lambda^{(\bar K)})$ be a basis of $\mathcal{L}(\bar\mu)$. Take an arbitrary moment index ν of order 2 and height 1. Without loss of generality, we may assume that $\nu = \nu[1, 1] + \nu[1, 1]$.

As $\bar\mu$ is stable, one can find a row selector $\bar r$ such that it does not touch the first random variable and $\bar r(\bar\Lambda)$ is nondegenerate. By Proposition 4.15, $\bar r$ can be selected in such a way that it touches only $\bar K - 1$ random variables.

Next, as Λ is k-stable, one can find a row selector r such that $r(\Lambda)$ is nondegenerate and r does not touch $\mathcal{J}(\bar r) \cup \{1\}$. Again, one can find r that touches only $K - 1$ random variables.

By Proposition 4.16, there exists a column selector c such that c does not touch X_1 and $A = (r, c)(\mathrm{P})$ is nondegenerate and contains only moments of height 0. As $M_2(\mathrm{P})$ is symmetric, $A^T = (c, r)(\mathrm{P})$ is also nondegenerate.

Let $r'' = \bar{r} \cup c$ and $r' = r'' \cup \{v[1, 1]\}$. From now on, we restrict our attention to rows in r' (the rows above dotted line in Fig. 3). For use in subsequent notations, we let $r'' = (v_1, \ldots, v_p)$. (In the figure, c is shown to lie inside $\bar{r}$. This is just for convenience of drawing; c and $\bar{r}$ may overlap arbitrarily, or may even be disjoint).

The submatrix $(r', r)(\mathrm{P})$ (in the figure, the block on the right surrounded by a dashed line) consists only from moments of height 0 and has rank K. Thus, the linear span of columns of this submatrix coincides with $\mathcal{L}(r'(\Lambda))$. On the other hand, it is a subspace of $\mathcal{L}(r'(\bar{\Lambda}))$. Thus, $\mathcal{L}(r'(\Lambda)) \subseteq \mathcal{L}(r'(\bar{\Lambda}))$, and without loss of generality we can assume that

$$\begin{cases} r'(\overline{\lambda}^1) = r'(\lambda^1) \\ \cdots \\ r'(\overline{\lambda}^K) = r'(\lambda^K). \end{cases} \quad (4.4)$$

Now look at the first column of $r'(M_2(\mathrm{P}))$. Except for the topmost element, m_v, it consists of moments of height 0. As both $r''(\Lambda)$ and $r''(\bar{\Lambda})$ are nondegenerate, there are two *unique* linear combinations

$$\begin{pmatrix} m_{v_1} \\ \vdots \\ m_{v_p} \end{pmatrix} = \alpha_1 \cdot \begin{pmatrix} \lambda^1_{v_1} \\ \vdots \\ \lambda^1_{v_p} \end{pmatrix} + \cdots + \alpha_K \cdot \begin{pmatrix} \lambda^K_{v_1} \\ \vdots \\ \lambda^K_{v_p} \end{pmatrix}$$

and

$$\begin{pmatrix} m_{v_1} \\ \vdots \\ m_{v_p} \end{pmatrix} = \overline{\alpha}_1 \cdot \begin{pmatrix} \overline{\lambda}^1_{v_1} \\ \vdots \\ \overline{\lambda}^1_{v_p} \end{pmatrix} + \cdots + \overline{\alpha}_K \cdot \begin{pmatrix} \overline{\lambda}^K_{v_1} \\ \vdots \\ \overline{\lambda}^K_{v_p} \end{pmatrix} + \overline{\alpha}_{K+1} \cdot \begin{pmatrix} \overline{\lambda}^{K+1}_{v_1} \\ \vdots \\ \overline{\lambda}^{K+1}_{v_p} \end{pmatrix} + \cdots + \overline{\alpha}_{\overline{K}} \cdot \begin{pmatrix} \overline{\lambda}^{\overline{K}}_{v_1} \\ \vdots \\ \overline{\lambda}^{\overline{K}}_{v_p} \end{pmatrix}.$$

But due to equation (4.4) we have $\bar{\alpha}_1 = \alpha_1, \ldots, \bar{\alpha}_K = \alpha_K, \bar{\alpha}_{K+1} = 0, \ldots, \bar{\alpha}_{\bar{K}} = 0$, and consequently

$$m_v(\mu) = \alpha_1 \lambda^1_{v[1,1]} + \cdots + \alpha_K \lambda^K_{v[1,1]} = \overline{\alpha}_1 \overline{\lambda}^1_{v[1,1]} + \cdots + \overline{\alpha}_{\overline{K}} \overline{\lambda}^{\overline{K}}_{v[1,1]} = m_v(\overline{\mu})$$

which completes the proof of the theorem.

As Example 6.8 shows, the requirement for $\bar{\mu}$ to be stable is important.

To clarify the meaning of the above theorem, consider the case of 100 random variables and stable distribution P of rank 4 with $\mathcal{L}(\mathrm{P})$ being 90-stable (see Sect. 7 for motivations of importance of this case). Then there is no stable mixing measure of rank 5, 6, …, 90 that produces P. One may hope only to find a stable mixing distribution producing P of rank 91.

Theorems 4.12 and 4.18 are true in both finite and infinite cases. It appears that in infinite case there is a simpler and more natural version of Theorem 4.18.

Definition 4.19

A finite family of vectors Λ in an infinite-dimensional β-space (finitely-dimensional subspace $\mathcal{L}$, mixing measure μ with finitely-dimensional support) is ∞-*stable*, if it is k-stable for every k.

The following theorem is a direct corollary of Theorem 4.18.

Theorem 4.20

Let P be a stable distribution of rank K with $\mathcal{L}$(P) being ∞-stable.

Then there is no stable mixing measure $\bar{\mu}$ of finite rank such that $P_{\bar{\mu}} = P$ and rank($\bar{\mu}$) > K.

5. Identifiability of the mixing measure

In the case of finite family of random variables, the mixing measure is not identifiable in general. To see this, note first that a distribution P_a which puts all the mass at point $a = (a_1, \ldots, a_n) \in \mathcal{A}_n$ is independent. The corresponding vector $\beta[a]$ has coordinates $\beta_{jl}[a] = 1$ if $l = a_j$ and $\beta_{jl}[a] = 0$ if $l \neq a_j$. Thus, any joint distribution P can be represented as a *trivial finite mixture*:

$$\mathrm{P} = \sum_{a \in \mathscr{A}} p_a \mathrm{P}_a = \sum_{a \in \mathscr{A}} p_a \mathrm{P}_{\beta[a]} \quad (5.1)$$

where $p_a = \mathrm{P}(\{a\})$. Now take arbitrary measure μ on P which does not have form 5.1. Then for the distribution $\mathrm{P} = \mathrm{P}_\mu$ there exist at least two mixing measures producing P: first, measure μ and, second, trivial finite mixing measure.

Even if we restrict the set of mixing measure to essential ones, there is no identifiability – see Example 6.4.

In rare cases, the mixing measure can be identifiable (see Example 6.10). But in the case of finite family of random variables this is rather an exclusion than a rule.

In the absence of identifiability, one would like to learn those properties of mixing measures which are common for all mixing measures producing a given joint distribution P. This suggests the following:

Definition 5.1[12]

A functional F defined on the family of mixing measures is a *mixture invariant* of distribution P, if for arbitrary mixing measures μ and μ' the equality $\mathrm{P}_\mu = \mathrm{P}_{\mu'} = \mathrm{P}$ implies $F(\mu) = F(\mu')$.

This definition, unfortunately, is not satisfactory: in general, a joint distribution P has no mixture invariants other than moments of height 0. One possibility to improve the situation is to restrict attention to essential mixing measures.

Definition 5.2

A functional F defined on the family of mixing measures is an *essential mixture invariant* of distribution P, if for arbitrary essential mixing measures μ and μ' the equality $P_\mu = P_{\mu'} = P$ implies $F(\mu) = F(\mu')$.

In this terminology, Theorem 4.12 states that for a stable distribution P the supporting subspace is an essential mixture invariant. Other important essential mixture invariants are given by the following:

Theorem 5.3 (finite-dimensional case)

Let P be a stable distribution with $\mathcal{L}(P)$ being k-stable. Then all moments of order $\leq k + 1$ are essential mixture invariants of P.

Proof—The proof uses the method similar to the method of Theorem 4.12. Example 6.3 illustrates the proof.

Take arbitrary $p \leq k + 1$. Moments of order p occupy columns of the moment matrix indexed by moment indices of order $p - 1$. Moment indices of order $p - 1$ can be split into p groups: indices of height 0, …, indices of height $p - 1$. We use induction over the height of a moment index to show that all moments of order p are identifiable.

Basis of induction: all moments of height 0 are identifiable due to equation (3.1).

Induction step. Assume that all moments of height q are known and let us show that we can identify all moments of height $q + 1$. Note that all moments of height $q + 1$ appear in columns indexed by moment indices of height q. Take an arbitrary column index w of height q. The moments in the corresponding column consist of moments of height q and moments of height $q + 1$. For a moment in this column to have height $q + 1$, it is required that its row index v touches one of the random variables that are touched by w. Thus, if we exclude random variables $\mathcal{I}(w)$ from consideration, the remaining moments have height q–*i.e.*, are already known. But $|\mathcal{I}(w)| \leq p - 1 \leq k$, and, due to k-stability of $\mathcal{L}(P)$, a basis Λ of $\mathcal{L}(P)$ keeps full rank after exclusion of the random variables $\mathcal{I}(w)$. Thus, we can determine the unique linear combination of Λ that produces column w – which gives us moments of height $q + 1$.

In the infinite-dimensional case, this theorem has a more elegant form:

Theorem 5.4 (infinite-dimensional case)

Let P be a joint distribution of an infinite family of random variables X_1, …. Assume that P has finite rank, is stable, and $\mathcal{L}(P)$ is ∞-stable. Then all moment of the mixing distribution are essential mixture invariants of P.

Corollary 5.5

Let P be a joint distribution of an infinite family of random variables X_1, …. Assume that P has finite rank, is stable, and $\mathcal{L}$(P) is ∞-stable. Then an essential mixing distribution μ producing P is identifiable.

Proof—Follows from the fact that measure supported by a compact subset of finite-dimensional Euclidean space is uniquely determined by its moments ([6], 5.5.2).

6. Examples

Most of the examples in this section deal with the case of 3 binary random variables and rely on the moment matrix. The (part of the) moment matrix shown in such examples usually contains only moments of order 1 and 2, and these moments always are listed in the following order:

$$\begin{array}{c} \\ (1,0;0,0;0,0) \\ (0,1;0,0;0,0) \\ (0,0;1,0;0,0) \\ (0,0;0,1;0,0) \\ (0,0;0,0;1,0) \\ (0,0;0,0;0,1) \end{array} \begin{pmatrix} (0,0;0,0;0,0) & (1,0;0,0;0,0) & (0,1;0,0;0,0) & (0,0;1,0;0,0) & (0,0;0,1;0,0) & (0,0;0,0;1,0) & (0,0;0,0;0,1) \\ m_{(100000)} & \mathbf{m}_{(\mathbf{200000})} & \mathbf{m}_{(\mathbf{110000})} & m_{(101000)} & m_{(100100)} & m_{(100010)} & m_{(100001)} \\ m_{(010000)} & \mathbf{m}_{(\mathbf{110000})} & \mathbf{m}_{(\mathbf{200000})} & m_{(011000)} & m_{(010100)} & m_{(010010)} & m_{(010001)} \\ m_{(001000)} & m_{(101000)} & m_{(011000)} & \mathbf{m}_{(\mathbf{002000})} & \mathbf{m}_{(\mathbf{001100})} & m_{(001010)} & m_{(001001)} \\ m_{(000100)} & m_{(100100)} & m_{(010100)} & \mathbf{m}_{(\mathbf{001100})} & \mathbf{m}_{(\mathbf{000200})} & m_{(000110)} & m_{(000101)} \\ m_{(000010)} & m_{(100010)} & m_{(010010)} & m_{(001010)} & m_{(000110)} & \mathbf{m}_{(\mathbf{000020})} & \mathbf{m}_{(\mathbf{000011})} \\ m_{(000001)} & m_{(100001)} & m_{(010001)} & m_{(001001)} & m_{(000101)} & \mathbf{m}_{(\mathbf{000011})} & \mathbf{m}_{(\mathbf{000002})} \end{pmatrix}$$

(commas and semicolons are not shown in the moment indices in order to save space).

Occasionally, the moment matrix in an example will contain, in addition, a column with moments of order 3. In such cases, the column index of the included column is given in the text.

The moments of height 0 (which are common for all mixing measures producing the observed distribution P) are typeset in plain font; moments of greater height are typeset in **bold**. Slightly abusing language, we speak sometime about *the moment matrix of distribution* P, which means that we include only moments of height 0 in the moment matrix and replace all other moments with question marks.

Example 6.1 (dimensionality of a set in β-space is not equal to dimensionality of its image in $\mathcal{P}(\mathcal{A})$)

This examples clarifies some properties of the parametrization $\beta \mapsto \mathrm{P}_\beta$. We show that the image of a low-dimensional subset of P may be infinite-dimensional (in $\mathcal{P}(\mathcal{A})$).

Consider an infinite family of binary random variables (which means that $\mathcal{A} = \{1,2\}^{\mathbb{N}}$). Let

$$\beta' = (1,0;1,0;\ldots), \quad \beta'' = (0,1;0,1;\ldots).$$

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Further, let P_0 be a closed interval with endpoints at β' and β'' and let $\mathscr{P}_0$ be its image under mapping $\beta \mapsto \mathrm{P}_\beta$:

$$P_0=\{t \cdot \beta' + (1-t) \cdot \beta'' | t \in [0,1] | \}, \quad \mathscr{P}_0=\{\mathrm{P}_\beta | \beta \in P_0\}$$

P_0 is a one-dimensional subset of P. We shall show that $\mathscr{P}_0$ is infinite-dimensional. For this, we show a stronger fact that for every n the natural projection of $\mathscr{P}_0$ on $\mathscr{P}(\mathcal{A}_n)$, denoted $\mathscr{P}_0^{(n)}$, has dimensionality at least n.

Recall that $\mathscr{P}(\mathcal{A}_n)$ is the unit simplex in 2^n-dimensional space; we use elements of $\mathcal{A}_n$ to index coordinates in this space. A distribution $\mathrm{P} \in \mathscr{P}(\mathcal{A}_n)$ is represented by the vector $p = (p_a)_{a \in \mathcal{A}_n} \in \mathbb{R}^{2^n}$ with coordinates $p_a = \mathrm{P}(a)$. To show that $\mathscr{P}_0^{(n)}$ is at least n-dimensional, it is sufficient to present $n + 1$ linearly independent vectors belonging to $\mathscr{P}_0^{(n)}=\pi_n(\mathscr{P}_0)$.

For this, take arbitrary $0 \le t_0 < t_1 < \cdots < t_n \le 1$, and let

$$\beta^{(i)}=t_i \cdot \beta' + (1-t_i) \cdot \beta'', \quad i=0,1,\ldots,n.$$

Further, let $\mathrm{P}^{(i)} = \mathrm{P}_{\beta^{(i)}}$, and let $p^{(i)}$ be a vector from $\mathbb{R}^{2^n}$ representing measure $\pi_n(\mathrm{P}^{(i)})$. We have to show that vectors $p^{(0)}, p^{(1)}, \ldots, p^{(n)}$ are linearly independent – or, equivalently, that the $2^n \times (n + 1)$ matrix, which has these vectors as columns, is nondegenerate. To show the latter, it is sufficient to present a nondegenerate $(n + 1) \times (n + 1)$ minor of this matrix. Let us choose rows with indices $a^{(0)} = (1, \ldots, 1)$, $a^{(1)} = (1, \ldots, 1, 2)$, …, $a^{(n-1)} = (1, 2, \ldots, 2)$, $a^{(n)} = (2, \ldots, 2)$. Then the corresponding minor is:

$$\begin{pmatrix} t_0^n & t_1^n & \ldots & t_n^n \\ t_0^{n-1}(1-t_0) & t_1^{n-1}(1-t_1) & \ldots & t_n^{n-1}(1-t_n) \\ \vdots & & & \\ t_0(1-t_0)^{n-1} & t_1(1-t_1)^{n-1} & \ldots & t_n(1-t_n)^{n-1} \\ (1-t_0)^n & (1-t_1)^n & \ldots & (1-t_n)^n \end{pmatrix}.$$

By applying elementary transformation “add row $(i-1)$ to row (i)” subsequently for $i = n, n-1, \ldots, 2, 1$; $n, n-1, \ldots, 2$; …; $n, n - 1$; n, one obtains:

$$\begin{pmatrix} t_0^n & t_1^n & \ldots & t_n^n \\ t_0^{n-1} & t_1^{n-1} & \ldots & t_n^{n-1} \\ \vdots & & & \\ t_0 & t_1 & \ldots & t_n \\ 1 & 1 & \ldots & 1 \end{pmatrix}$$

which is a Vandermonde matrix and it is known to be nondegenerate.

Example 6.2 (mixture with identifiable supporting subspace)

Consider a case of 3 binary variables, and let mixing measure μ be concentrated at two points,

$$\beta^{(1)}=({}^{1}/_{6}, {}^{5}/_{6}; {}^{1}/_{8}, {}^{7}/_{8}; {}^{1}/_{6}, {}^{5}/_{6}), \quad \beta^{(2)}=({}^{5}/_{6}, {}^{1}/_{6}; {}^{7}/_{8}, {}^{1}/_{8}; {}^{5}/_{6}, {}^{1}/_{6}),$$

with weights ${}^{1}/_{2}$.

The observed distribution $\mathrm{P}=\mathrm{P}_{\mu}=\frac{1}{2}\mathrm{P}_{\beta^{(1)}}+\frac{1}{2}\mathrm{P}_{\beta^{(2)}}$ has the following elementary probabilities

$$\begin{array}{llll} p_{(1,1,1)}={}^{11}/_{36} & p_{(1,2,1)}={}^{1}/_{18} & p_{(2,1,1)}={}^{5}/_{72} & p_{(2,2,1)}={}^{5}/_{72} \\ p_{(1,1,2)}={}^{5}/_{72} & p_{(1,2,2)}={}^{5}/_{72} & p_{(2,1,2)}={}^{1}/_{18} & p_{(2,2,2)}={}^{11}/_{36} \end{array}$$

and its moment matrix is

$$\begin{pmatrix} {}^{1}/_{2} & ? & ? & {}^{3}/_{8} & {}^{1}/_{8} & {}^{13}/_{36} & {}^{5}/_{36} \\ {}^{1}/_{2} & ? & ? & {}^{1}/_{8} & {}^{3}/_{8} & {}^{5}/_{36} & {}^{13}/_{36} \\ {}^{1}/_{2} & {}^{3}/_{8} & {}^{1}/_{8} & ? & ? & {}^{3}/_{8} & {}^{1}/_{8} \\ {}^{1}/_{2} & {}^{1}/_{8} & {}^{3}/_{8} & ? & ? & {}^{1}/_{8} & {}^{3}/_{8} \\ {}^{1}/_{2} & {}^{13}/_{36} & {}^{5}/_{36} & {}^{3}/_{8} & {}^{1}/_{8} & ? & ? \\ {}^{1}/_{2} & {}^{5}/_{36} & {}^{13}/_{36} & {}^{1}/_{8} & {}^{3}/_{8} & ? & ? \end{pmatrix}.$$

The rank of the observed part of the moment matrix is 2 (the left bottom 2×2 minor is non-degenerate, and there is no non-degenerate minors of size 3×3). Thus, the rank of the mixing measure that produces the observed one is at least 2. One can also see that the distribution P is stable (*i.e.*, after removing any variable the rank of the visible part of the moment matrix is 2). Thus, according to Theorem 4.12, the supporting subspace of an essential mixing distribution is identifiable.

Let us forget for a moment that we know the fact that $\mathrm{P}=\frac{1}{2}\mathrm{P}_{\beta^{(1)}}+\frac{1}{2}\mathrm{P}_{\beta^{(2)}}$, and try to find the supporting subspace of an essential mixing measure using the method of Theorem 4.12.

First, we restrict our attention to the last two rows of moment matrix. Subcolumns 1 and 4 are linearly independent, and thus there exists a unique linear combination of them that produces the subcolumn 2:

$$({}^{13}/_{36}, {}^{5}/_{36})={}^{1}/_{18}\cdot({}^{1}/_{2}, {}^{1}/_{2})+{}^{8}/_{9}\cdot({}^{3}/_{8}, {}^{1}/_{8}).$$

Consequently, the value of the moment of height 1 at the top left corner is:

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

$$m_{(200000)} = {}^{1}\!/_{18} \cdot {}^{1}\!/_{2} + {}^{8}\!/_{9} \cdot {}^{3}\!/_{8} = {}^{13}\!/_{36}\,.$$

Continuing this process for other moments of height 1, we obtain the moment matrix:

$$\begin{pmatrix} {}^{1}\!/_{2} & \mathbf{{}^{13}\!/_{36}} & \mathbf{{}^{5}\!/_{36}} & {}^{3}\!/_{8} & {}^{1}\!/_{8} & {}^{13}\!/_{36} & {}^{5}\!/_{36} \\ {}^{1}\!/_{2} & \mathbf{{}^{5}\!/_{36}} & \mathbf{{}^{13}\!/_{36}} & {}^{1}\!/_{8} & {}^{3}\!/_{8} & {}^{5}\!/_{36} & {}^{13}\!/_{36} \\ {}^{1}\!/_{2} & {}^{3}\!/_{8} & {}^{1}\!/_{8} & \mathbf{{}^{25}\!/_{64}} & \mathbf{{}^{7}\!/_{64}} & {}^{3}\!/_{8} & {}^{1}\!/_{8} \\ {}^{1}\!/_{2} & {}^{1}\!/_{8} & {}^{3}\!/_{8} & \mathbf{{}^{7}\!/_{64}} & \mathbf{{}^{25}\!/_{64}} & {}^{1}\!/_{8} & {}^{3}\!/_{8} \\ {}^{1}\!/_{2} & {}^{13}\!/_{36} & {}^{5}\!/_{36} & {}^{3}\!/_{8} & {}^{1}\!/_{8} & \mathbf{{}^{13}\!/_{36}} & \mathbf{{}^{5}\!/_{36}} \\ {}^{1}\!/_{2} & {}^{5}\!/_{36} & {}^{13}\!/_{36} & {}^{1}\!/_{8} & {}^{3}\!/_{8} & \mathbf{{}^{5}\!/_{36}} & \mathbf{{}^{13}\!/_{36}} \end{pmatrix}.$$

Now, to obtain a basis of supporting subspace, we can take any two linearly independent columns of completed moment matrix (if we want to have vectors satisfying conditions 2.1, we have to normalize any column except the first). In particular, we can take the first and the second columns as basis vectors:

$$\lambda^{(1)} = ({}^{1}\!/_{2}, {}^{1}\!/_{2}; {}^{1}\!/_{2}, {}^{1}\!/_{2}; {}^{1}\!/_{2}, {}^{1}\!/_{2}) \quad \lambda^{(2)} = ({}^{13}\!/_{18}, {}^{5}\!/_{18}; {}^{3}\!/_{4}, {}^{1}\!/_{4}; {}^{13}\!/_{18}, {}^{5}\!/_{18}).$$

As it was expected, one can see that Lin($\lambda^{(1)}$, $\lambda^{(2)}$) = Lin($\beta^{(1)}$, $\beta^{(2)}$).

Example 6.3 (identifiable moments of a stable distribution)

Consider distribution P from Example 6.2. We can see that P is stable and $\mathcal{L}$(P) is 2-stable. Thus, according to Theorem 5.3, all its moments of order 3 are identifiable.

In Example 6.2, we have already recovered all moments of order 2. Here we show how the process of recovery of moments of order 3 works.

For the purpose of this example, consider columns of moment matrix indexed by (1, 0; 1, 0; 0, 0), (1, 0; 0, 1; 0, 0), and (1, 1; 0, 0; 0, 0):

$$\begin{pmatrix} \cdots & m_{(2,0;1,0;0,0)} & \cdots & m_{(2,0;0,1;0,0)} & \cdots & m_{(2,1;0,0;0,0)} & \cdots \\ \cdots & m_{(1,1;1,0;0,0)} & \cdots & m_{(1,1;0,1;0,0)} & \cdots & m_{(1,2;0,0;0,0)} & \cdots \\ \cdots & m_{(1,0;2,0;0,0)} & \cdots & m_{(1,0;1,1;0,0)} & \cdots & m_{(1,1;1,0;0,0)} & \cdots \\ \cdots & m_{(1,0;1,1;0,0)} & \cdots & m_{(1,0;0,2;0,0)} & \cdots & m_{(1,1;0,1;0,0)} & \cdots \\ \cdots & m_{(1,0;1,0;1,0)} & \cdots & m_{(1,0;0,1;1,0)} & \cdots & m_{(1,1;0,0;1,0)} & \cdots \\ \cdots & m_{(1,0;1,0;0,1)} & \cdots & m_{(1,0;0,1;0,1)} & \cdots & m_{(1,1;0,0;0,1)} & \cdots \end{pmatrix}.$$

At the beginning, we know only moments of height 0:

$$\begin{pmatrix} \cdots & m_{(2,0;1,0;0,0)} & \cdots & m_{(2,0;0,1;0,0)} & \cdots & m_{(2,1;0,0;0,0)} & \cdots \\ \cdots & m_{(1,1;1,0;0,0)} & \cdots & m_{(1,1;0,1;0,0)} & \cdots & m_{(1,2;0,0;0,0)} & \cdots \\ \cdots & m_{(1,0;2,0;0,0)} & \cdots & m_{(1,0;1,1;0,0)} & \cdots & m_{(1,1;1,0;0,0)} & \cdots \\ \cdots & m_{(1,0;1,1;0,0)} & \cdots & m_{(1,0;0,2;0,0)} & \cdots & m_{(1,1;0,1;0,0)} & \cdots \\ \cdots & {}^{11}\!/_{36} & \cdots & {}^{1}\!/_{18} & \cdots & m_{(1,1;0,0;1,0)} & \cdots \\ \cdots & {}^{5}\!/_{72} & \cdots & {}^{5}\!/_{72} & \cdots & m_{(1,1;0,0;0,1)} & \cdots \end{pmatrix}.$$

This allows us to find 2 linear combinations:

$$\begin{pmatrix} m_{(2,0;1,0;0,0)} \\ m_{(1,1;1,0;0,0)} \\ m_{(1,0;2,0;0,0)} \\ m_{(1,0;1,1;0,0)} \\ {}^{11}\!/_{36} \\ {}^{5}\!/_{72} \end{pmatrix} = -{}^{5}\!/_{32}\,\lambda^{(1)} + {}^{17}\!/_{32}\,\lambda^{(2)}, \quad \begin{pmatrix} m_{(2,0;0,1;0,0)} \\ m_{(1,1;0,1;0,0)} \\ m_{(1,0;1,1;0,0)} \\ m_{(1,0;0,2;0,0)} \\ {}^{1}\!/_{18} \\ {}^{5}\!/_{72} \end{pmatrix} = {}^{5}\!/_{32}\,\lambda^{(1)} - {}^{1}\!/_{32}\,\lambda^{(2)}.$$

This allows us to recover all moments in the first two columns – and, consequently, two moments in the third column:

$$\begin{pmatrix} \cdots & {}^{11}\!/_{36} & \cdots & {}^{1}\!/_{18} & \cdots & m_{(2,1;0,0;0,0)} & \cdots \\ \cdots & {}^{5}\!/_{72} & \cdots & {}^{5}\!/_{72} & \cdots & m_{(1,2;0,0;0,0)} & \cdots \\ \cdots & {}^{41}\!/_{128} & \cdots & {}^{7}\!/_{128} & \cdots & {}^{5}\!/_{72} & \cdots \\ \cdots & {}^{7}\!/_{128} & \cdots & {}^{9}\!/_{128} & \cdots & {}^{5}\!/_{72} & \cdots \\ \cdots & {}^{11}\!/_{36} & \cdots & {}^{1}\!/_{18} & \cdots & m_{(1,1;0,0;1,0)} & \cdots \\ \cdots & {}^{5}\!/_{72} & \cdots & {}^{5}\!/_{72} & \cdots & m_{(1,1;0,0;0,1)} & \cdots \end{pmatrix}.$$

Now, we can find one more linear combination:

$$\begin{pmatrix} m_{(2,1;0,0;0,0)} \\ m_{(1,2;0,0;0,0)} \\ {}^{5}\!/_{72} \\ {}^{5}\!/_{72} \\ m_{(1,1;0,0;1,0)} \\ m_{(1,1;0,0;0,1)} \end{pmatrix} = {}^{5}\!/_{36}\,\lambda^{(1)} + 0\lambda^{(2)}.$$

This allows us to recover remaining moments in the third column:

$$\begin{pmatrix} \cdots & {}^{11}\!/_{36} & \cdots & {}^{1}\!/_{18} & \cdots & {}^{5}\!/_{72} & \cdots \\ \cdots & {}^{5}\!/_{72} & \cdots & {}^{5}\!/_{72} & \cdots & {}^{5}\!/_{72} & \cdots \\ \cdots & {}^{41}\!/_{128} & \cdots & {}^{7}\!/_{128} & \cdots & {}^{5}\!/_{72} & \cdots \\ \cdots & {}^{7}\!/_{128} & \cdots & {}^{9}\!/_{128} & \cdots & {}^{5}\!/_{72} & \cdots \\ \cdots & {}^{11}\!/_{36} & \cdots & {}^{1}\!/_{18} & \cdots & {}^{5}\!/_{72} & \cdots \\ \cdots & {}^{5}\!/_{72} & \cdots & {}^{5}\!/_{72} & \cdots & {}^{5}\!/_{72} & \cdots \end{pmatrix}.$$

The remaining columns with moments of order 3 are handled in similar manner.

Example 6.4 (mixture with identifiable supporting subspace is not identifiable itself)

Now consider mixing measure $\bar\mu$ concentrated at 3 points

$$\bar\beta^{(1)}=({}^{1}\!/_{18},{}^{17}\!/_{18};0,1;{}^{1}\!/_{18},{}^{17}\!/_{18}),\quad \bar\beta^{(2)}=({}^{1}\!/_{2},{}^{1}\!/_{2};{}^{1}\!/_{2},{}^{1}\!/_{2};{}^{1}\!/_{2},{}^{1}\!/_{2}),$$
$$\bar\beta^{(3)}=({}^{17}\!/_{18},{}^{1}\!/_{18};1,0;{}^{17}\!/_{18},{}^{1}\!/_{18}).$$

Mixture $\bar{P} = P_{\bar\mu}$ coincides with mixture P from Example 6.2 (which one can see by ensuring that they have the same elementary probabilities).

As mixing measure $\bar\mu$ is essential (rank($\bar\beta^{(1)}$, $\bar\beta^{(2)}$, $\bar\beta^{(3)}$) = 2), its supporting subspace, according to Theorem 4.12, must coincide with the one of μ – and, in fact, one can see that $\mathcal{L}(\bar\mu)$ = Lin($\bar\beta^{(1)}$, $\bar\beta^{(2)}$, $\bar\beta^{(3)}$) = Lin($\beta^{(1)}$, $\beta^{(2)}$) = $\mathcal{L}(\mu)$.

Moreover, one can see that μ and $\bar\mu$ have equal moments of order 2 and 3.

This example illustrates that one can identify the supporting subspace of an essential mixing measure and some of its moments – but not the mixing measure itself.

Example 6.5 (visible rank of a distribution can be smaller its rank)

Consider a case of 3 binary variables and mixing measure μ, concentrated at points

$$\beta^{(1)}=(1,0;1,0;1,0),\quad \beta^{(2)}=(0,1;1,0;1,0),$$
$$\beta^{(3)}=(0,1;0,1;1,0),\quad \beta^{(4)}=(0,1;0,1;0,1)$$

with weights ${}^{1}/_{4}$.

The observed distribution P_μ has the elementary probabilities:

$$\begin{matrix} p_{(1,1,1)}={}^{1}\!/_{4} & p_{(1,2,1)}=0 & p_{(2,1,1)}={}^{1}\!/_{4} & p_{(2,2,1)}={}^{1}\!/_{4} \\ p_{(1,1,2)}=0 & p_{(1,2,2)}=0 & p_{(2,1,2)}=0 & p_{(2,2,2)}={}^{1}\!/_{4} \end{matrix}$$

and its moment matrix is

$$\begin{pmatrix} 1/4 & ? & ? & 1/4 & 0 & 1/4 & 0 \\ 3/4 & ? & ? & 1/4 & 1/2 & 1/2 & 1/4 \\ 1/2 & 1/4 & 1/4 & ? & ? & 1/2 & 0 \\ 1/2 & 0 & 1/2 & ? & ? & 1/4 & 1/4 \\ 3/4 & 1/4 & 1/2 & 1/2 & 1/4 & ? & ? \\ 1/4 & 0 & 1/4 & 0 & 1/4 & ? & ? \end{pmatrix}.$$

It is easy to see from the moment matrix that the visible rank of P_μ is 2. (Note that moments of order > 2 do not help to construct a minor of size 3, as any column containing such moments has at most 2 observable moments.)

As the visible rank is 2 and the mixing measure is concentrated at 4 points, we may conclude that $2 \le \text{rank}(P_\mu) \le 4$. We shall show that $\text{rank}(P_\mu) \ne 2$.

Assume, in contrary, that $\text{rank}(P_\mu) = 2$. Then P_μ is stable, and we can use Theorem 4.12 to find a basis of supporting subspace.

Using the same calculations from Example 6.2, we obtain completion of the moment matrix:

$$\begin{pmatrix} 1/4 & \mathbf{1/8} & \mathbf{1/8} & 1/4 & 0 & 1/4 & 0 \\ 3/4 & \mathbf{1/8} & \mathbf{5/8} & 1/4 & 1/2 & 1/2 & 1/4 \\ 1/2 & 1/4 & 1/4 & \mathbf{1/2} & \mathbf{0} & 1/2 & 0 \\ 1/2 & 0 & 1/2 & \mathbf{0} & \mathbf{1/2} & 1/4 & 1/4 \\ 3/4 & 1/4 & 1/2 & 1/2 & 1/4 & \mathbf{5/8} & \mathbf{1/8} \\ 1/4 & 0 & 1/4 & 0 & 1/4 & \mathbf{1/8} & \mathbf{1/8} \end{pmatrix}.$$

Any 2 linearly independent columns of the moment matrix can be taken as a basis of the supporting subspace. Let us take the second and the last columns, multiplied by 4, *i.e.*:

$$\lambda^{(1)} = (1/2, 1/2; 1, 0; 1, 0), \qquad \lambda^{(2)} = (0, 1; 0, 1; 1/2, 1/2).$$

Having this, we can parametrize the support of a mixing measure by parameter t:

$$\beta(t) = (t/2, 1 - t/2; t, 1 - t; 1 - t/2, t/2), \qquad t \in [0, 1]$$

(note that *supporting subspace* has dimensionality 2, but *support*, which is a (subset of a) linear manifold, has dimensionality 1).

If there exists a mixing measure $\bar{\mu}$ that is carried by this subspace and produces the observed distribution P_{μ}, it should satisfy:

$$\frac{1}{4}=p_{(1,1,1)}=\int_0^1 \frac{t}{2}\cdot t\cdot\left(1-\frac{t}{2}\right)\overline{\mu}(\mathrm{d}t)=\int_0^1\left(1-\frac{t}{2}\right)\cdot t\cdot\frac{t}{2}\,\overline{\mu}(\mathrm{d}t)=p_{(2,1,2)}=0$$

which gives us a contradiction.

Thus, rank(P_{μ}) ≥ 3, while the visible rank of P_{μ} is 2. This, in particular, means that P_{μ} is not even 0-stable.

Example 6.6 (essential and non-essential mixing measures may have very different supporting subspaces)

Now consider mixing measure $\tilde{\mu}$, concentrated at 4 points

$$\tilde{\beta}^{(1)}=(1,0;1,0;1,0),\qquad \tilde{\beta}^{(2)}=(0,1;0,1;0,1),$$
$$\tilde{\beta}^{(3)}=\left(\frac{1}{3},\frac{2}{3};\frac{1}{2},\frac{1}{2};\frac{2}{3},\frac{1}{3}\right),\qquad \tilde{\beta}^{(4)}=\left(\frac{2}{3},\frac{1}{3};\frac{1}{2},\frac{1}{2};\frac{1}{3},\frac{2}{3}\right)$$

with weights $^1/_4$ each.

Again, distribution $\tilde{P}=P_{\tilde{\mu}}$ coincides with distribution P from Example 6.2. However, rank($\tilde{\mu}$) = 3, *i.e.* $\tilde{\mu}$ is not essential.

Looking to the moment matrix of $\tilde{\mu}$:

$$\begin{pmatrix}
\frac{1}{2} & \mathbf{\frac{7}{18}} & \mathbf{\frac{1}{9}} & \frac{3}{8} & \frac{1}{8} & \frac{13}{36} & \frac{5}{36}\\
\frac{1}{2} & \mathbf{\frac{1}{9}} & \mathbf{\frac{7}{18}} & \frac{1}{8} & \frac{3}{8} & \frac{5}{36} & \frac{13}{36}\\
\frac{1}{2} & \frac{3}{8} & \frac{1}{8} & \mathbf{\frac{3}{8}} & \mathbf{\frac{1}{8}} & \frac{3}{8} & \frac{1}{8}\\
\frac{1}{2} & \frac{1}{8} & \frac{3}{8} & \mathbf{\frac{1}{8}} & \mathbf{\frac{3}{8}} & \frac{1}{8} & \frac{3}{8}\\
\frac{1}{2} & \frac{13}{36} & \frac{5}{36} & \frac{3}{8} & \frac{1}{8} & \mathbf{\frac{7}{18}} & \mathbf{\frac{1}{9}}\\
\frac{1}{2} & \frac{5}{36} & \frac{13}{36} & \frac{1}{8} & \frac{3}{8} & \mathbf{\frac{1}{9}} & \mathbf{\frac{7}{18}}
\end{pmatrix}$$

one can observe that it differs from the moment matrix of μ only in moments of height 1 (and higher).

One can also observe that $\mathcal{L}(\mu)\not\subseteq\mathcal{L}(\tilde{\mu})$. It is not by coincidence: it is impossible to construct a stable nonessential mixing measure $\tilde{\mu}'$ such that $\mathcal{L}(\mu)\subseteq\mathcal{L}(\tilde{\mu})$. This impossibility is a key point in the proof of Theorem 4.18.

Example 6.7 (nonstable distribution may have nonidentifiable supporting subspace)

Once more, consider a case of 3 binary variables, and 2 mixing measures: μ, concentrated at points

NIH-PA Author Manuscript
NIH-PA Author Manuscript

$$\beta^{(1)}=(1,0;{}^{1}\!/_{3},{}^{2}\!/_{3};1,0),\quad \beta^{(2)}=(0,1;{}^{2}\!/_{3},{}^{1}\!/_{3};1,0)$$

with weights ${}^{1}/_{2}$, and $\bar{\mu}$, concentrated at points

$$\bar{\beta}^{(1)}=({}^{1}\!/_{3},{}^{2}\!/_{3};1,0;1,0),\quad \bar{\beta}^{(2)}=({}^{2}\!/_{3},{}^{1}\!/_{3};0,1;1,0)$$

also with weights ${}^{1}/_{2}$.

The observed distributions P_{μ} and $P_{\bar{\mu}}$ are equal and have the following elementary probabilities

$$\begin{array}{llll} p_{(1,1,1)}={}^{1}\!/_{6} & p_{(1,2,1)}={}^{1}\!/_{3} & p_{(2,1,1)}={}^{1}\!/_{3} & p_{(2,2,1)}={}^{1}\!/_{6} \\ p_{(1,1,2)}=0 & p_{(1,2,2)}=0 & p_{(2,1,2)}=0 & p_{(2,2,2)}=0 \end{array}$$

and the observed part of their moment matrix is

$$\begin{pmatrix} {}^{1}\!/_{2} & ? & ? & {}^{1}\!/_{6} & {}^{1}\!/_{3} & {}^{1}\!/_{2} & 0 \\ {}^{1}\!/_{2} & ? & ? & {}^{1}\!/_{3} & {}^{1}\!/_{6} & {}^{1}\!/_{2} & 0 \\ {}^{1}\!/_{2} & {}^{1}\!/_{6} & {}^{1}\!/_{3} & ? & ? & {}^{1}\!/_{2} & 0 \\ {}^{1}\!/_{2} & {}^{1}\!/_{3} & {}^{1}\!/_{6} & ? & ? & {}^{1}\!/_{2} & 0 \\ 1 & {}^{1}\!/_{2} & {}^{1}\!/_{2} & {}^{1}\!/_{2} & {}^{1}\!/_{2} & ? & ? \\ 0 & 0 & 0 & 0 & 0 & ? & ? \end{pmatrix}.$$

The rank of the observer part of the moment matrix is 2, and rank(μ) = rank($\bar{\mu}$) = 2. Thus, both μ and $\bar{\mu}$ are essential.

However, the observed distribution is not stable: removing the first (or the second) random variable decreases the rank of the observed part of the moment matrix to 1. Consequently, Theorem 4.12 is not applicable here, and in fact one can see that $\mathcal{L}(\mu) \neq \mathcal{L}(\bar{\mu})$.

Example 6.8 (there may exist nonstable mixing measures of ranks prohibited by Thm. 4.18)

Consider a case of 4 binary variables, and 2 mixing measures: μ, concentrated at 2 points

$$\beta^{(1)}=({}^{5}\!/_{6},{}^{1}\!/_{6};1,0;1,0;1,0),\quad \beta^{(2)}=({}^{1}\!/_{6},{}^{5}\!/_{6};0,1;0,1;0,1)$$

with weights ${}^{1}/_{2}$, and $\tilde{\mu}$, concentrated at 4 points

$$\tilde{\beta}^{(1)}=(1,0;1,0;1,0;1,0), \qquad \tilde{\beta}^{(2)}=(0,1;0,1;0,1;0,1),$$
$$\tilde{\beta}^{(3)}=({}^{2}\!/_{3},{}^{1}\!/_{3};1,0;1,0;1,0), \quad \tilde{\beta}^{(4)}=({}^{1}\!/_{3},{}^{2}\!/_{3};0,1;0,1;0,1)$$

with weights $^{1}/_{4}$.

Elementary probabilities for P_μ and $P_{\tilde{\mu}}$ are equal; they are:

$$p_{(1,1,1,1)}=p_{(2,2,2,2)}={}^{5}\!/_{12}, \quad p_{(2,1,1,1)}=p_{(1,2,2,2)}={}^{1}\!/_{12}, \quad p_a=0 \quad \text{for all other } a.$$

We have the following facts:

- $P_\mu = P_{\tilde{\mu}}$;
- P_μ is stable;
- $\mathcal{L}(P_\mu)$ is 3-stable;
- $\text{rank}(P_\mu) = 2$;
- $\text{rank}(\tilde{\mu}) = 3$.

The only condition of Theorem 4.18 that is not satisfied is that $\tilde{\mu}$ is not stable—and this is enough to obtain $\text{rank}(P_\mu) < \text{rank}(\tilde{\mu}) \leq$ stability level of $\mathcal{L}(P_\mu)$.

Example 6.9 ((K − 1)-stable mixing measure of rank K may have a row selector with all minors of height 0 being degenerate)

Consider a case of 4 ternary random variables (*i.e.*, $J = 4$, $L_1 = L_2 = L_3 = L_4 = 3$). Let μ be a measure concentrated in 3 points $\beta^{(1)}$, $\beta^{(2)}$, $\beta^{(3)}$ with equal weights (*i.e.*, $\mu=\frac{1}{3}\delta(\beta^{(1)})+\frac{1}{3}\delta(\beta^{(2)})+\frac{1}{3}\delta(\beta^{(1)})$), where

$$\beta^{(1)}=(1,0,0;1,0,0;1,0,0;1,0,0)$$
$$\beta^{(2)}=(0,1,0;0,1,0;0,1,0;1,0,0)$$
$$\beta^{(3)}=(0,0,1;0,0,1;0,0,1;1,0,0).$$

As $\beta^{(1)}$, $\beta^{(2)}$, $\beta^{(3)}$ are linearly independent, they can be taken as a basis of $\mathcal{L}(\mu)$. The second-order moments of μ are:

$$\begin{pmatrix}
\mathbf{1/3} & \mathbf{0} & \mathbf{0} & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\
\mathbf{0} & \mathbf{1/3} & \mathbf{0} & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\
\mathbf{0} & \mathbf{0} & \mathbf{1/3} & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\
1/3 & 0 & 0 & \mathbf{1/3} & \mathbf{0} & \mathbf{0} & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\
0 & 1/3 & 0 & \mathbf{0} & \mathbf{1/3} & \mathbf{0} & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\
0 & 0 & 1/3 & \mathbf{0} & \mathbf{0} & \mathbf{1/3} & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\
1/3 & 0 & 0 & 1/3 & 0 & 0 & \mathbf{1/3} & \mathbf{0} & \mathbf{0} & 1/3 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 1/3 & 0 & \mathbf{0} & \mathbf{1/3} & \mathbf{0} & 1/3 & 0 & 0 \\
0 & 0 & 1/3 & 0 & 0 & 1/3 & \mathbf{0} & \mathbf{0} & \mathbf{1/3} & 1/3 & 0 & 0 \\
1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & \mathbf{1} & \mathbf{0} & \mathbf{0} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{pmatrix}.$$

Let r be a row selector that selects the boxed rows, $r = (v[1, 3], v[2, 2], v[3, 1])$. Then we have the following properties:

- rank(μ) = 3;
- P_μ is stable;
- μ is 2-stable but not 3-stable;
- $r(\beta^{(1)}, \beta^{(2)}, \beta^{(3)})$ is nondegenerate.

However, there is no column selector c such that height of (r, c) is 0 and $(r, c)(\mu)$ is nondegenerate (the only way for c to produce a minor selector of height 0 is to select the last 3 columns of the above matrix, which gives a minor of rank 1).

Example 6.10 (special case: joint distribution of 3 binary random variables with identifiable mixing distribution)

Consider a joint distribution of 3 binary random variables given by elementary probabilities:

$$\begin{matrix} p_{(1,1,1)}=a & p_{(1,2,1)}=0 & p_{(2,1,1)}=0 & p_{(2,2,1)}=d \\ p_{(1,1,2)}=0 & p_{(1,2,2)}=b & p_{(2,1,2)}=c & p_{(2,2,2)}=0 \end{matrix} \quad (6.1)$$

($a, b, c, d \geq 0$ and $a + b + c + d = 1$).

The moment matrix of this distribution is:

$$\begin{pmatrix} a+b & ? & ? & a & b & a & b \\ c+d & ? & ? & c & d & d & c \\ a+c & a & c & ? & ? & a & c \\ b+d & b & d & ? & ? & d & b \\ a+d & a & d & a & d & ? & ? \\ b+c & b & c & c & b & ? & ? \end{pmatrix}.$$

Note that if P is represented as a probabilistic mixture $\int \mathrm{P}_\beta \mu(\mathrm{d}\beta)$, then P($A$) = 0 implies $\mathrm{P}_\beta(A)$ = 0 for μ-almost all β.

We shall show that there are only 4 independent distributions that have form (6.1). The independence condition

$$\mathrm{P}(X_1{=}l_1 \wedge X_2{=}l_2 \wedge X_3{=}l_3){=}\mathrm{P}(X_1{=}l_1) \cdot \mathrm{P}(X_2{=}l_2) \cdot \mathrm{P}(X_3{=}l_3)$$

gives the equations:

$$\begin{array}{ll} (c{+}d)(a{+}c)(a{+}d){=}0 & (\text{for } l_1{=}2,\ l_2{=}1,\ l_3{=}1) \\ (a{+}b)(b{+}d)(a{+}d){=}0 & (\text{for } l_1{=}1,\ l_2{=}2,\ l_3{=}1) \\ (a{+}b)(a{+}c)(b{+}c){=}0 & (\text{for } l_1{=}1,\ l_2{=}1,\ l_3{=}2) \\ (c{+}d)(b{+}d)(b{+}c){=}0 & (\text{for } l_1{=}2,\ l_2{=}2,\ l_3{=}2). \end{array}$$

Assume that $a \neq 0$. Then from the first equation one obtains c+d = 0, and consequently c = d = 0; from the second equation one obtains b + d = 0, and consequently b = 0. Similarly, assuming $b \neq 0$ gives a = c = d = 0, etc. Thus, the only 4 independent distributions having form (6.1) are:

$$\begin{array}{llllll} (\mathrm{a}) & a{=}1, & b{=}c{=}d{=}0 & (\mathrm{b}) & b{=}1, & a{=}c{=}d{=}0 \\ (\mathrm{c}) & c{=}1, & a{=}b{=}d{=}0 & (\mathrm{d}) & d{=}1, & a{=}b{=}c{=}0 \end{array}$$

and the corresponding vectors in β-space are:

$$\begin{array}{ll} \beta^{(a)}{=}(1,0;1,0;1,0) & \beta^{(b)}{=}(1,0;0,1;0,1) \\ \beta^{(c)}{=}(0,1;1,0;0,1) & \beta^{(d)}{=}(0,1;0,1;1,0). \end{array}$$

Summarizing, one can conclude that any distribution in form (6.1) has a unique representation as a mixture of independent distributions:

$$\mathrm{P}{=}a \cdot \mathrm{P}_{\beta^{(a)}}{+}b \cdot \mathrm{P}_{\beta^{(b)}}{+}c \cdot \mathrm{P}_{\beta^{(c)}}{+}d \cdot \mathrm{P}_{\beta^{(d)}}$$

Example 6.11 (dimensionality of the supporting subspace may be smaller if nonprobabilistic mixtures are allowed)

Consider a distribution (denoted P) as in Example 6.10 with $a = 0.4$, $b = 0.3$, $c = 0.2$, and $d = 0.1$. The moment matrix of distribution P is:

$$\begin{pmatrix} 0.7 & ? & ? & 0.4 & 0.3 & 0.4 & 0.3 & \dots & ? & \dots \\ 0.3 & ? & ? & 0.2 & 0.1 & 0.1 & 0.2 & \dots & ? & \dots \\ 0.6 & 0.4 & 0.2 & ? & ? & 0.4 & 0.2 & \dots & 0 & \dots \\ 0.4 & 0.3 & 0.1 & ? & ? & 0.1 & 03 & \dots & 0.3 & \dots \\ 0.5 & 0.4 & 0.1 & 0.4 & 0.1 & ? & ? & \dots & ? & \dots \\ 0.5 & 0.3 & 0.2 & 0.2 & 0.3 & ? & ? & \dots & ? & \dots \end{pmatrix}$$

(the rightmost shown column has index (1, 0; 0, 0; 0, 1)).

As it follows from Example 6.10, rank(P) = 4 (it is easy to verify that vectors $\beta^{(a)}$, $\beta^{(b)}$, $\beta^{(c)}$, and $\beta^{(d)}$ are linearly independent). The moment matrix of distribution P, however, does not contain nondegenerate minors of size 4 (it even does not contain nondegenerate minors of size 3). Thus, according to Definition 4.10, distribution P is not even 0-stable.

On the other hand, the moment matrix satisfies other conditions of Definition 4.10 if one takes $K = 2$ and $k = 1$: for every $j = 1, 2, 3$ one can find a minor selector s such that s does not touch j, is of height 0 and is nondegenerate. For example, for $j = 2$ one can take $r = ((1, 0; 0, 0; 0, 0), (0, 1; 0, 0; 0, 0))$, $c = ((0, 0; 0, 0; 0, 0), (0, 0; 0, 0; 0, 1))$, $s = (r, c)$, which gives minor

$$s(\mathrm{P})=\begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.2 \end{pmatrix}.$$

This suggests that one can uniquely construct a matrix of rank 2 that coincides with the moment matrix of distribution P in places corresponding to moments of height 0. (Note, however, that at this point there are no reasons to think that such a matrix is a moment matrix of some mixing measure.) This matrix is:

$$M=\begin{pmatrix} 0.7 & 0.48 & 0.22 & 0.4 & 0.3 & 0.4 & 0.3 & \dots & 0.12 & \dots \\ 0.3 & 0.22 & 0.08 & 0.2 & 0.1 & 0.1 & 0.2 & \dots & 0.18 & \dots \\ 0.6 & 0.4 & 0.2 & 0.32 & 0.28 & 0.4 & 0.2 & \dots & 0 & \dots \\ 0.4 & 0.3 & 0.1 & 0.28 & 0.12 & 0.1 & 0.3 & \dots & 0.3 & \dots \\ 0.5 & 0.4 & 0.1 & 0.4 & 0.1 & 0 & 0.5 & \dots & 0.6 & \dots \\ 0.5 & 0.3 & 0.2 & 0.2 & 0.3 & 0.5 & 0 & \dots & -0.3 & \dots \end{pmatrix}.$$

Obviously, this matrix cannot be a moment matrix of a probabilistic mixture of independent probability measures: such a mixture must have only nonnegative moments, while this one has $m_{(1,0;0,0;0,2)} = -0.3$.

Moreover, if M is a moment matrix of some measure μ, its covariance matrix, according to equation (3.4), should be:

$$\begin{pmatrix} -0.01 & 0.01 & -0.08 & 0.02 & 0.05 & -0.05 \\ 0.01 & -0.01 & 0.02 & -0.02 & -0.05 & 0.05 \\ -0.08 & 0.02 & -0.04 & 0.04 & 0.1 & -0.1 \\ 0.02 & -0.02 & 0.04 & -0.04 & -0.1 & 0.1 \\ 0.05 & -0.05 & 0.1 & -0.1 & -0.25 & 0.25 \\ -0.05 & 0.05 & -0.1 & 0.1 & 0.25 & -0.25 \end{pmatrix}.$$

As this matrix is not positive-semidefinite, it cannot be a covariance matrix of any probability measure.

Summarizing, we may conclude that the distribution P cannot be represented as a mixture of independent distributions with the mixing measure carried by a 2-dimensional subspace of β-space. Such a representation does not exist even if we allow the support of a mixing measure to include vectors β that correspond to signed measures (*i.e.*, if we remove restrictions $0 \le \beta_{jl} \le 1$ from conditions (2.1)).

One can obtain distribution P as a mixture of independent distributions with the mixing distributions supported by a 2-dimensional subspace of β-space only if both the distributions being mixed and the mixing measure are allowed to be signed measures. One possible way to do this is to take the mixing measure μ being concentrated at two points,

$$\beta^{(1)} = \begin{pmatrix} 1.2+0.2\sqrt{6} \\ -0.2-0.2\sqrt{6} \\ 1.6+0.4\sqrt{6} \\ -0.6-0.4\sqrt{6} \\ -2-\sqrt{6} \\ 3+\sqrt{6} \end{pmatrix}, \quad \beta^{(2)} = \begin{pmatrix} 1.2-0.2\sqrt{6} \\ -0.2+0.2\sqrt{6} \\ 1.6-0.4\sqrt{6} \\ -0.6+0.4\sqrt{6} \\ -2+\sqrt{6} \\ 3-\sqrt{6} \end{pmatrix}$$

with weights $w_1 = 0.5 - 1.25/\sqrt{6}$ and $w_2 = 0.5 + 1.25/\sqrt{6}$. One can verify that the equality $P = w_1 P_{\beta^{(1)}} + w_2 P_{\beta^{(2)}}$ holds.

The columns of the moment matrix M span a 2-dimensional subspace of β-space. One possible basis of this subspace is:

$$\lambda^1 = (0.8,\ 0.2; 0.8,\ 0.2; 0,\ 1), \quad \lambda^2 = (0.6,\ 0.4; 0.4,\ 0.6; 1,\ 0).$$

As one would expect, both $\beta^{(1)}$ and $\beta^{(2)}$ belong to this subspace:

$$\beta^{(1)} = (3+\sqrt{6})\lambda^1 + (-2-\sqrt{6})\lambda^2, \quad \beta^{(2)} = (3-\sqrt{6})\lambda^1 + (-2+\sqrt{6})\lambda^2.$$

7. Discussion

The results of this article are very important for the foundation of Linear Latent Structure (LLS) analysis [8]. One domain of application of LLS analysis is the analysis of surveys, where individuals from a sample are asked multiple questions in order to obtain a description of a relatively simple (but directly unobservable) underlying phenomenon.

In this context, the mixing distribution can be thought of as a description of a *latent variable* that characterizes the underlying phenomenon. The dimensionality of a mixing distribution corresponds to the "complexity" of the underlying phenomenon.

Stability is a characteristic of how well a questionnaire is "balanced." A small (in comparison to the number of questions in a survey) level of stability means that a questionnaire is poorly balanced: many questions are devoted to discover one "side" of the underlying phenomenon, while only a few of them are devoted to discover another "side." From this perspective, stability of the LLS model can be considered as a mathematical measure of "quality" of a questionnaire.

On the other hand, application of LLS analysis (as of any statistical method) is an attempt to infer something from a number of imprecise evidences. One has to avoid inferences that are supported by a single (or very few) evidence(s). Thus, stability characterizes how reliable is our inference.

The above arguments suggest that it is very natural to restrict consideration to stable cases.

The other point that we want to stress is the importance of *k*-stability for reliability of LLS inference.

To give the reader a sense of numbers, let us assume that we have an observed distribution P of 100 binary random variables, which is stable, rank(P) = 5, and $\mathcal{L}$(P) is 90-stable. (This should be considered as a very realistic case. In [1], LLS analysis was applied to the National Long Term Care Survey, containing 49 binary variables and 8 quaternary variables. The estimate of rank is 3, while the estimate of stability level is 40[5]).

According to Theorem 4.12, there exists a 5-dimensional model for P. But one may ask a question: "OK, 5-dimensional model exists, it is fine. But maybe there exists a 6-dimensional model that is much-much better than the 5-dimensional one in a sense that was not taken into account so far?".

Theorem 4.18 gives quite a strong answer to this question: *there are no stable mixing measures of rank 6, 7, …, 90*. One may find a mixing measure of rank 91. It is hard to believe, however, that a 91-dimensional model would be better than a 5-dimensional one.

There are a few questions that were not answered in this article.

[5]The estimate of stability level was performed later and was not included in [1].

The first one is how *generic* are stable distributions. One would be much more convinced of the importance of our theorems (which are in many cases applicable only to stable distributions) if stability is a “common case”, and nonstable distributions are “rare”.

The answer to this question is not simple, primarily because there is no single notion of genericity. The two most important synonyms of “be generic” are “have Lebesgue measure 1” and “have Baire category 2” [9]. However, on the one hand, these notions do not coincide, and, on the other hand, in the infinite-dimensional case there is no good equivalent of Lebesgue measure (and the family of all mixing measures is infinite-dimensional even if finitely many random variables are considered).

Still we hope that a convincing notion of genericity can be formulated and proved for the case of stable distributions. Our belief is based on the simple fact that nondegenerate matrices are the “common case” and degenerate ones are “rare.” Consider, for example, matrix $\begin{pmatrix} a & b \\ b & x \end{pmatrix}$ with a, $b > 0$. The only value of x which makes this matrix degenerate is b^2/a, while all other values of x produce nondegenerate matrices.

Another question is raised by Example 6.11. This example shows that one can obtain a lower-dimensional model (2 instead of 4 in the example) if signed measures are allowed.

From the point of view of linear algebra considerations employed in the present article, using signed measures is very natural (vectors with negative coordinates are “first-class citizens” in linear algebra). But notions like “negative probability,” although sometimes discussed, are not a practical tool of an applied statistician.

It would be good to find out practical examples where the LLS model involves signed measures and investigate whether it is possible to give a reasonable interpetation to these measures.

One more question is raised by Example 6.5. As this example shows, the visible rank of a distribution can be smaller than its rank. But to prove the fact that the example distribution has rank bigger than 2, we used a method beyond the direct analysis of the moment matrix.

The question is whether the rank of a distribution can be derived from the visible part of its moment matrix using only linear algebra methods.

Note that if stability level is bigger than rank (which, as discussed above, is true in practically interesting cases) the visible rank of P is equal to its rank: the proof of Corollary 4.17 constructs a required minor. Thus, this question has rather theoretical than practical interest.

References

1. Akushevich I, Kovtun M, Manton KG, Yashin AI. Linear latent structure analysis and modeling of multiple categorical variables. Comput Math Methods Medicine. 2009; 10:203–218.
2. Barndorff-Nielsen O. Identifiability of mixtures of exponential families. J Math Anal Appl. 1965; 12:115–121.

3. Dellacherie, C.; Meyer, P-A. Probabilities and Potential. Vol. I. North-Holland Publishing Co; Amsterdam: 1978.

4. Dunford, N.; Schwartz, JT. Linear Operators. Vol. I. Interscience Publishers, Inc; New York: 1958.

5. Everitt, BS.; Hand, DJ. Monographs on Applied Probability and Statistics. Chapman and Hall; London: 1981. Finite Mixture Distributions.

6. Knill, O. Probability Theory and Stochastic Processes with Applications. Overseas Press; 2009.

7. Kolmogorov, AN.; Fomin, SV. Elements of the theory of functions and functional analysis. 3. Moscow, Russia: Science; 1972. In Russian

8. Kovtun M, Akushevich I, Manton KG, Tolley HD. Linear latent structure analysis: Mixture distribution models with linear constraints. Statist Methodology. 2007; 4:90–110.

9. Oxtoby, JC. Graduate Texts in Mathematics. 2. Springer-Verlag; New York: 1980. Measure and Category. Number 2.

10. Shiryaev, AN. Probability. 3. Moscow, Russia: MCCSE; 2004. In Russian

11. Tallis GM, Chesson P. Identifiability of mixtures. J Austral Math Soc Ser A. 1982; 32:339–348.

12. Teicher H. On the mixture of distributions. Ann Math Stat. 1960; 31:55–73.

13. Teicher H. Identifiability of mixtures. Ann Math Stat. 1961; 32:244–248.

14. Teicher H. Identifiability of finite mixtures. Ann Math Stat. 1963; 34:1265–1269.

$$\begin{pmatrix} m_{v^0+w^0}(\mu) = m_{v^0+w^0}(\mu') & m_{v^0+w^1} & \cdots & m_{v^0+w^K} \\ m_{v^1+w^0} & m_{v^1+w^1} & \cdots & m_{v^1+w^K} \\ \vdots & \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots\ldots \\ m_{v^K+w^0} & m_{v^K+w^1} & \cdots & m_{v^K+w^K} \end{pmatrix}$$

Figure 1.

Illustration to the proof of Theorem 4.12.

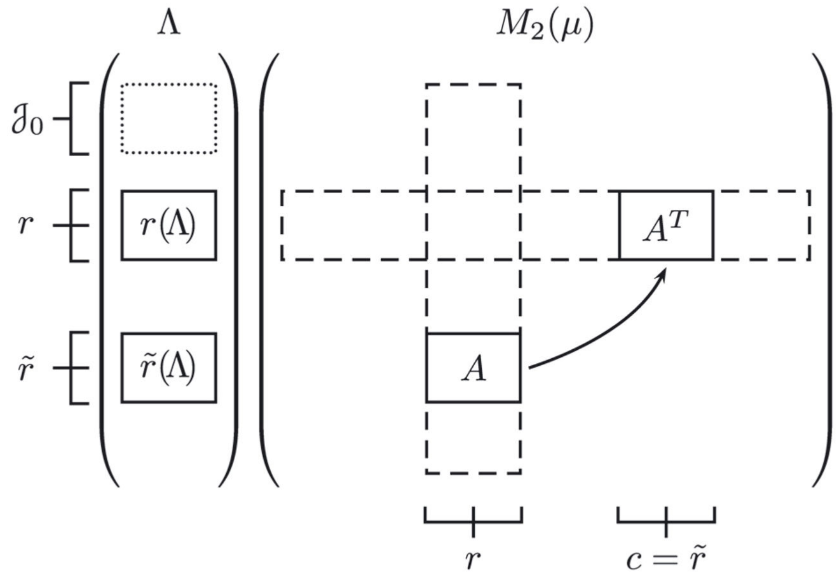

Figure 2.
Illustration to the proof of Proposition 4.16.

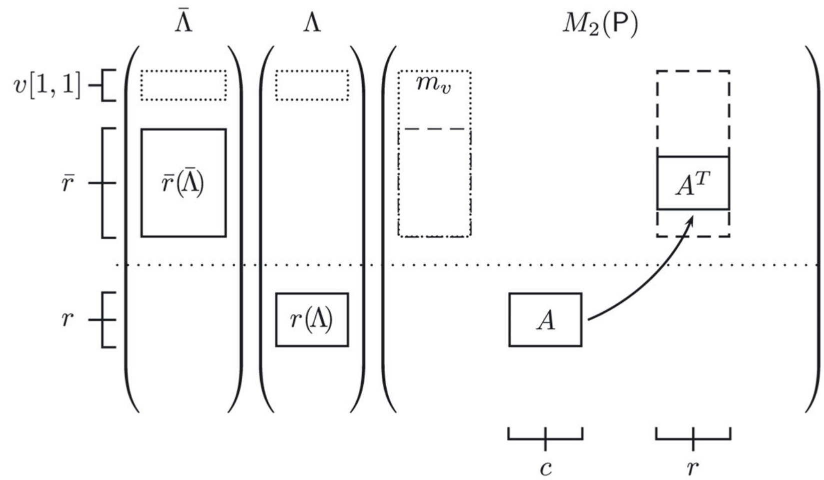

Figure 3.

Illustration to the proof of Theorem 4.18.