Beiträge zur Algebra und Geometrie
Contributions to Algebra and Geometry
Volume 51 (2010), No. 1, 45-56.

Neighborliness of Marginal Polytopes

Thomas Kahle*

Max Planck Institute for Mathematics in the Sciences
Inselstrasse 22, D-04103 Leipzig, Germany
e-mail: kahle@mis.mpg.de

Abstract. A neighborliness property of marginal polytopes of hierarchical models, depending on the cardinality of the smallest non-face of the underlying simplicial complex, is shown. The case of binary variables is studied explicitly, then the general case is reduced to the binary case. A Markov basis for binary hierarchical models whose simplicial complexes is the complement of an interval is given.

MSC 2000: 52B05, 60C05
Keywords: marginal polytopes, Markov bases, hierarchical models

1. Introduction

The marginal polytope is an interesting combinatorial object that appears in statistics [9], coding theory [6, 12, 18] and, under a different name, in toric algebra [7]. It encodes in its face lattice the complete combinatorial information about the boundary of certain statistical models. To define it we have to take a very brief excursion to statistics, namely the theory of hierarchical models for contingency tables.

Consider a collection of n random variables taking values in finite sets $\mathcal{X}_i, i = 1, \dots, n$. We denote $N := \{1, \dots, n\}$, and its power set as $2^N := \{B : B \subseteq N\}$. For a subset $B \subseteq N$ of the variables, we denote its set of values as $\mathcal{X}_B := \prod_{i \in B} \mathcal{X}_i$, and abbreviate $\mathcal{X} := \mathcal{X}_N$. We have the natural projections

$$\begin{aligned} X_B &: \mathcal{X} \to \mathcal{X}_B \\ (x_i)_{i\in N} &\mapsto (x_i)_{i \in B} =: x_B. \end{aligned} \tag{1}$$

*The author is supported by the Volkswagen Foundation.

We slightly abuse notation and denote x_B the projection of $x = (x_i)_{i\in N}$, which is a function of x, and by the same symbol an arbitrary element $x_B \in \mathcal{X}_B$. A *contingency table* is a function $u : \mathcal{X} \to \mathbb{N}_0$. It is thereby a vector in the space $\mathbb{N}_0^{\mathcal{X}}$. For $B \subseteq N$ we define the *marginal table* $u_B \in \mathbb{N}_0^{\mathcal{X}_B}$ as the vector with components

$$u_B(x_B) := \sum_{y:X_B(y)=x_B} u(y). \tag{2}$$

A so called *hierarchical model* for contingency tables can be given by a simplicial complex Δ on the set N of variable indexes [4, 9]. The facets $\mathcal{F}$ of Δ are defined as the inclusion maximal faces. They determine the *marginal map*:

$$\begin{aligned} \pi_\Delta : \mathbb{R}^{\mathcal{X}} &\to \bigoplus_{F\in\mathcal{F}} \mathbb{R}^{\mathcal{X}_F} \\ u &\mapsto (u_F)_{F\in\mathcal{F}}. \end{aligned} \tag{3}$$

It is a linear map computing all marginal tables corresponding to facets. We define cylinder sets denoting for $B \subseteq N$, and $y_B \in \mathcal{X}_B$

$$\{X_B = y_B\} := \{x \in \mathcal{X} : X_B(x) = y_B\}. \tag{4}$$

With respect to the canonical basis, the matrix representing π_Δ is the $d \times |\mathcal{X}|$ matrix

$$A_\Delta := (A_{(B,y_B),x})_{(B,y_B),x} \text{ where } A_{(B,y_B),x} := \begin{cases} 1 & \text{if } X_B(x) = y_B, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The rows of this matrix are indexed by pairs (B, y_B), where $B \in \mathcal{F}$ is a facet of Δ and $y_B \in \mathcal{X}_B$ is a configuration on B. Then d is defined as the number of such pairs. If the simplicial complex is clear, we will sometimes omit the index Δ.

Definition 1. (Marginal polytope) *The marginal polytope is the convex hull of the columns of A_Δ:*

$$Q_\Delta := \operatorname{conv}\{A_x : x \in \mathcal{X}\} \subseteq \mathbb{R}^d. \tag{6}$$

Example 2. (Two independent binary variables) In the case of two binary variables, we have $\mathcal{X} = \{(00), (01), (10), (11)\}$. Let $\Delta = \{\{1\}, \{2\}\}$, then the matrix A_Δ is given as

$$A_\Delta = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \tag{7}$$

The columns are ordered as $(\{1\}, 0), (\{1\}, 1), (\{2\}, 0), (\{2\}, 1)$. If Δ was the whole power set, A_Δ would be the 4×4 identity matrix. The marginal polytopes are easily identified as a 2-dimensional square and a 3-dimensional simplex, respectively.

Our object of interest is the toric ideal:

$$\mathcal{I}_\Delta := \left\langle p^u - p^v : u, v \in \mathbb{N}_0^{\mathcal{X}}, \pi_\Delta(u) = \pi_\Delta(v) \right\rangle . \tag{8}$$

Here, we used the standard notation for monomials in the variables $\{p_x : x \in \mathcal{X}\}$, namely $p^u := \prod_{x\in\mathcal{X}} p_x^{u(x)}$. Throughout the whole paper we use the convention that $0^0 = 1$. The set of indexes with non-vanishing exponent will be called the *support* of the binomial $\mathrm{supp}(p^u - p^v) := \{x \in \mathcal{X} : u(x) + v(x) > 0\}$. The supports of u and v will also be called the *positive* respectively *negative support* of the binomial. The ideal $\mathcal{I}_\Delta$ is a homogeneous prime ideal in the polynomial ring $\mathbb{C}[p_x : x \in \mathcal{X}]$. In statistics the restriction of the corresponding variety to the non-negative real cone would be called the closure of an *exponential family.* This seminal observation is the cornerstone of what is now called algebraic statistics [5],[8],[15].

A first task is to find a suitable finite generating set of this ideal. Very useful is a Markov basis defined as follows:

Definition 3. *A finite set $M \subseteq \ker_{\mathbb{Z}} \pi_\Delta$ is called a Markov basis for the hierarchical model Δ if for each two contingency tables $u, v \in \mathbb{N}_0^{\mathcal{X}}$ with equal marginals $\pi_\Delta(u) = \pi_\Delta(v)$ there exists a sequence $m_i, i = 1, \ldots, l$ in $\pm M$ such that*

$$u = v + \sum_{i=1}^{l} m_i, \tag{9}$$

where

$$v + \sum_{i=1}^{k} m_i \in \mathbb{N}_0^{\mathcal{X}} \text{ for all } k = 1, \ldots, l. \tag{10}$$

The crucial property of a Markov basis is that any two tables, having the same marginals, can be connected without leaving the non-negative cone. A key theorem is that exactly a Markov basis gives the desired set of generators:

Theorem 4. [5] *A finite set M is a Markov basis if and only if*

$$\mathcal{I}_\Delta = \left\langle p^{m^+} - p^{m^-} : m \in M \right\rangle , \tag{11}$$

where $m^+(x) := \max\{0, m(x)\}$, $m^-(x) := \max\{0, -m(x)\}$, which allows the decomposition $m = m^+ - m^-$.

The elements in a Markov basis are referred to as *Markov moves.*

In the following section we will give our main theorem, which is a lower bound on the cardinality of the positive and negative support of any move. Then in Section 3 we will state and prove the neighborliness property of marginal polytopes. Finally, in Section 4 we discuss a case where the lower bound and an upper bound on the generators coincide and the Markov basis consists of very simple moves.

2. A lower degree bound

Theorem 5. *Let Δ be a simplicial complex on N and $\mathcal{I}_\Delta$ the corresponding toric ideal. Let g be the minimal cardinality of a non-face of Δ. Each generator of $\mathcal{I}_\Delta$ has degree at least 2^{g-1}. Moreover, the positive and negative supports of each generator both have cardinality bigger or equal to 2^{g-1}. The degree bound is realized only by square free binomials.*

Remark. Note that we give a lower bound on the – smallest – degree among the generators. Lower bounds on the largest degree have been considered for a measure of complexity of the model for instance in [8]. There it is shown that one finds a simplicial complex on $2n$ units, such that there exists a generator of degree 2^n. Furthermore, in [4] the authors study an algorithm which, for graph models, computes all generators of a given degree. Finally, in [13] the case of 2-margins of $(r, s, 3)$-tables is studied. It is shown that as r and s grow the support and degree of a maximal generator cannot be bounded. This has interesting implications for data disclosure.

Remark. (Graph models) A graph model is a hierarchical model for which $\dim \Delta \leq 1$ holds. If its graph is not complete, the bound reduces to the trivial bound $\deg m \geq 2$. On the other hand, for the complete graph, there are no quadratic generators.

Remark. (Type of generators) The vectors that achieve the bound (see Lemma 7) are natural generalizations of the quadratic Markov moves for the independence model [5].

We will prove Theorem 5 in two steps. First, the binary case is studied explicitly. Then the general case is reduced to the binary case.

2.1. The binary case

In this section we have $\mathcal{X} = \{0,1\}^N$. This will allow us to use a special orthogonal basis of $\ker_{\mathbb{Z}} A_\Delta$. Using this, we find that any element in the kernel has a lower bound for the cardinality of its support.

Put $\Delta^c := 2^N \setminus \Delta$ the set of non-faces of Δ. For elements $G \in \Delta^c$ we define the upper intervals

$$[G, N] := \{B \subseteq N : B \supseteq G\} \tag{12}$$

which are contained in Δ^c. Next, for each $B \subseteq N$ we define a vector $e_B \in \mathbb{R}^{\mathcal{X}}$ with components:

$$e_B(x) := (-1)^{E(B,x)} \tag{13}$$

where $E(B, x) := |\{i \in B : x_i = 1\}|$ is the number of entries equal to one that x has in B. Observe, that e_B depends on its argument only through x_B, the part in B. Therefore we will sometimes abuse notation and write $e_B(x_B)$ for the value of e_B at any configuration which projects to x_B. We have

Lemma 6. ([10]) *The set $\{e_B : B \subseteq N\}$ is an orthogonal basis of $\mathbb{R}^{\mathcal{X}}$ such that $\{e_B : B \in \Delta^c\}$ is a basis of $\ker_{\mathbb{Z}} A_\Delta$.*

Remark. (Characters) If we treat $\mathcal{X}$ as the additive group $(\mathbb{Z}/2\mathbb{Z})^n$ then the characters of this group form an orthonormal basis (with respect to the product induced by the Haar measure, which in this case is proportional to the standard product) of $\mathbb{C}^{\mathcal{X}}$. The characters are exactly given by the vectors $e_B, B \subseteq N$. In our case the characters are real functions and also a basis of $\mathbb{R}^{\mathcal{X}}$. See [10, 16] for details.

Lemma 7. *Let $G \in \Delta^c$ and $\mathcal{G} := [G, N]$. For $g := |G|$ it holds*

$$m^0_{\mathcal{G}}(x) := \sum_{B\in\mathcal{G}} e_B(x) = \begin{cases} 2^{n-g}\, e_G(x_G) & \text{if } x_{N\setminus G} = (0,\dots,0), \\ 0 & \text{otherwise }. \end{cases} \tag{14}$$

Furthermore, for any $x_C \in \mathcal{X}_C$ we have the identity

$$\sum_{x\in\{X_C=y_C\}} e_B(x) = \begin{cases} 2^{n-|C|} e_B(y_C) & \text{if } B \subseteq C, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Proof. For the second case in (14) assume we have $i \in N \setminus G$ such that $x_i = 1$. Since half of the sets in $[G, N]$ contain i, while the other half does not contain i, it follows that the sum equals zero if such an i exists. The first case is now clear. All the summands are equal to e_G in this case, and there are exactly 2^{n-g} terms. The identity (15) follows by the same argument. □

Remark. By choosing appropriate signs in the sum, one can achieve any of the cylinder sets $\{X_{N\setminus G} = x_{N\setminus G}\}$ instead of $\{X_{N\setminus G} = 0\}$ as the support. To be concrete, we have

$$\begin{aligned} m^{y_{N\setminus G}}_{\mathcal{G}}(x) &:= \sum_{B\in\mathcal{G}} (-1)^{E(B,y_{N\setminus G})} e_B(x) \\ &= \begin{cases} 2^{n-g}\, e_G(x_G) & \text{if } x_{N\setminus G} = y_{N\setminus G}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{16}$$

The vectors we have just constructed have minimal support. In the following we will deduce a technical, but elementary statement about large subsets of $\mathcal{X}$. In Lemma 9, it will follow that choosing G minimal in Δ^c, the value $2^n - 2^{|G|}$, as in Lemma 7, is the maximal number of zero components, which can be achieved by non-trivial linear combinations of the vectors $e_B, B \in \Delta^c$.

Lemma 8. *Let $g \in \{1,\dots,n\}$ be fixed. For $\mathcal{Y} \subseteq \mathcal{X}$ with $|\mathcal{Y}| > 2^n - 2^g$ the following statement holds:*

- *For each $B \subseteq N$ with $|B| \geq g$, $\mathcal{Y}$ contains one of the cylinder sets $\{X_B = x_B\}$. More formally: $\exists x_B \in \mathcal{X}_B$ such that $\{X_B = x_B\} \subseteq \mathcal{Y}$.*

Proof. The statement follows from a simple cardinality argument. Assume the contrary, let B be given, and $\forall x_B \in \mathcal{X}_B$, $\exists x \in \mathcal{X} \setminus \mathcal{Y}$ such that $x_B = X_B(x)$. These x are all distinct, since they differ on B. We find $|\mathcal{Y}| \leq 2^n - 2^g$. □

Lemma 9. *Let g denote the minimal cardinality among the sets in Δ^c. Then any non-zero linear combination of the vectors $e_B, B \in \Delta^c$ has at least 2^{g-1} positive and 2^{g-1} negative components.*

Proof. Assume we have a linear combination

$$m = \sum_{B\in\Delta^c} z^B e_B \in \ker \pi_\Delta \tag{17}$$

which has less then 2^{g-1} positive components. It has at least $2^n - 2^{g-1} + 1$ non-positive components. Let $\mathcal{Y}_\leq \subseteq \mathcal{X}$ denote the corresponding indexes. Let $G \in \Delta^c$ have cardinality g and choose $i \in G$ arbitrary. By Lemma 8 we find a cylinder set $\left\{X_{G\setminus\{i\}} = y_{G\setminus\{i\}}\right\}$ which is contained in $\mathcal{Y}_\leq$. We have

$$m(x) = \sum_{B\in\Delta^c} z^B e_B(x) \leq 0, \qquad x \in \mathcal{Y}_\leq. \tag{18}$$

Summing up these equations over the cylinder set $\left\{X_{G\setminus\{i\}} = y_{G\setminus\{i\}}\right\}$ yields

$$\sum_{x\in\left\{X_{G\setminus\{i\}}=y_{G\setminus\{i\}}\right\}} \sum_{B\in\Delta^c} z^B e_B(x) \leq 0. \tag{19}$$

Note that this summation is in fact the computation of the marginal $m_{G\setminus\{i\}}$ evaluated at the value $y_{G\setminus\{i\}}$. Since $m \in \ker \pi_\Delta$, and $G \setminus \{i\} \in \Delta$, equality must hold in (19). We find that every term in the sum was already zero:

$$\sum_{B\in\Delta^c} z^B e_B(x) = 0, \qquad x \in \left\{X_{G\setminus\{i\}} = y_{G\setminus\{i\}}\right\}. \tag{20}$$

We will now inductively show that $m = 0$. Contained in $\left\{X_{G\setminus\{i\}} = y_{G\setminus\{i\}}\right\}$ we have a smaller set $\{X_G = y_G\}$. Summing up the respective components of m for this set we find, using Lemma 7,

$$\begin{aligned} 0 &= \sum_{x\in\{X_G=y_G\}} \sum_{B\in\Delta^c} z^B e_B(x) \\ &= z^G 2^{n-g} e_G(x_G). \end{aligned} \tag{21}$$

It follows that $z^G = 0$. Applying the same argument, we can show that all coefficients z^H vanish for $|H| = g$. Inductively, we continue with sets of cardinality $g+1$. Finally, this argument yields that all coefficients vanish and m is zero. The whole procedure applies, mutatis mutandis, for the negative components as well. □

Lemma 9 completes the proof of Theorem 5 in the binary case. It shows that the degree of each Markov move is at least 2^{g-1}. Since in fact we have a lower bound for the support, the degree bound can only be realized by square free binomials.

2.2. The non-binary case

We now study the non-binary case. Let $\mathcal{X} = \prod_{i\in N} \mathcal{X}_i$ be some arbitrary, finite configuration space.

Definition 10. *Let $\phi_i : \mathcal{X}_i \to \{0,1\}, i \in N$ be surjective maps. For each $B \subseteq N$, the composed maps*

$$\begin{aligned} \phi_B : \mathcal{X}_B &\to \{0,1\}^B \\ x_B &\mapsto (\phi_i(x_i))_{i\in B} \end{aligned} \tag{22}$$

are called collapsing maps. Abbreviating, put $\phi := \phi_N$. We have an induced map on contingency tables:

$$\begin{aligned} \Phi : \mathbb{N}_0^{\mathcal{X}} &\to \mathbb{N}_0^{\{0,1\}^N} \\ (u(x))_{x\in\mathcal{X}} &\mapsto \left(\sum_{w\in\phi^{-1}(z)} u(w)\right)_{z\in\{0,1\}^N} . \end{aligned} \tag{23}$$

The key property of such a collapsing is that it commutes with marginalization.

Lemma 11. *Let $u \in \mathbb{N}_0^{\mathcal{X}}$. For $B \subseteq N, z_B \in \{0,1\}^B$ it holds:*

$$\sum_{x_B\in\phi_B^{-1}(z_B)} \sum_{w\in\{X_B=x_B\}} u(w) = \sum_{y\in\{X_B=z_B\}} \sum_{w\in\phi^{-1}(y)} u(w). \tag{24}$$

Note that for the cylinder set on the left hand side, $\{X_B = x_B\} \subseteq \mathcal{X}$, while on the right hand side $\{X_B = z_B\} \subseteq \{0,1\}^N$.

Proof. Since on each side, every w appears at most once, it suffices to show the set equality

$$\bigcup_{x_B\in\phi_B^{-1}(z_B)} \{X_B = x_B\} = \bigcup_{y\in\{X_B=z_B\}} \left\{\phi^{-1}(y)\right\}. \tag{25}$$

"⊆": Let w from the left hand side be given. One has $X_B(w) = x_B$ for some x_B with $\phi_B(x_B) = z_B$. Therefore $\phi(w) = y$ with $X_B(y) = z_B$ and w is contained in the right hand side.

"⊇": Let $w = \phi^{-1}(y), y \in \{X_B = z_B\}$ from the right hand side be given. We have $X_B(w) \in \phi^{-1}(z_B)$, so w is contained in the left hand side.

□

Lemma 12. *Let $u, v \in \mathbb{N}_0^{\mathcal{X}}$ be contingency tables. Denote the marginal map in the non-binary model as π_Δ, the corresponding binary one as ρ_Δ. In this case, $\pi_\Delta(u) = \pi_\Delta(v)$ implies $\rho_\Delta(\Phi(u)) = \rho_\Delta(\Phi(v))$.*

Proof. Let $B \in \Delta, z_B \in \{0,1\}^N$. We have to show that

$$\sum_{y\in\{X_B=z_B\}} \Phi(u)(y) = \sum_{y\in\{X_B=z_B\}} \Phi(v)(y). \tag{26}$$

By definition this equation is

$$\sum_{y\in\{X_B=z_B\}}\sum_{w\in\phi^{-1}(y)} u(w) = \sum_{y\in\{X_B=z_B\}}\sum_{w\in\phi^{-1}(y)} v(w). \tag{27}$$

Using Lemma 11 and the hypothesis, the statement follows. □

Proof of Theorem 5. Using the collapsing map, from generators of the non-binary model we can construct relations in the corresponding binary model as follows. Consider the polynomial rings $\mathfrak{R} := \mathbb{C}[p_x : x \in \mathcal{X}]$ and $\mathfrak{Q} := \mathbb{C}[q_z : z \in \{0,1\}^N]$. Given a simplicial complex Δ, denote $\mathcal{I}_\Delta \subseteq \mathfrak{R}$ the non-binary toric ideal and $\mathcal{J}_\Delta \subseteq \mathfrak{Q}$ the binary one.

To each binomial $p^{m^+} - p^{m^-} \in \mathfrak{R}$ associate the collapsed binomial $q^{\Phi(m^+)} - q^{\Phi(m^-)} \in \mathfrak{Q}$. By Lemma 12 it is clear that elements in the toric ideal $\mathcal{I}_\Delta$ are mapped to $\mathcal{J}_\Delta$. Furthermore, the supports of $q^{\Phi(m^+)}$ and $q^{\Phi(m^-)}$ will have smaller cardinality than the supports of p^{m^+} and p^{m^-}, respectively. Finally, if the non-binary model had a generator violating the statement of the theorem, then we can choose the maps $\phi_i : i \in N$ in such a way that this generator gets mapped to a non-zero binomial which violates the statement for the binary case. This contradiction concludes the proof. □

3. Neighborliness

Before stating the neighborliness property we will take another short excursion to statistics, introducing so called exponential families and their relation to marginal polytopes.

Let again Δ denote a simplicial complex. For each $x \in \mathcal{X}$ we have A_x the corresponding row of the marginal matrix A_Δ, as defined in (5). The exponential family associated to this complex is the parametrized family of probability measures

$$\mathbb{R}^{\mathcal{X}} \supseteq \mathcal{E}_\Delta := \left\{p_\theta(x) = Z(\theta)^{-1}\exp\left(\langle\theta, A_x\rangle\right) : \theta \in \mathbb{R}^d\right\}. \tag{28}$$

Here, $Z(\theta) := \sum_{x\in\mathcal{X}} \exp\left(\langle\theta, A_x\rangle\right)$ is a normalization, called the partition function. Like in Section 1, d is the number of rows of A_Δ. By construction, an exponential family is an open subset of the simplex of all probability measures on $\mathcal{X}$. In many applications one is interested in the closure $\overline{\mathcal{E}_\Delta}$, which is taken with respect to the usual topology of $\mathbb{R}^n$. The closure of $\mathcal{E}_\Delta$ equals the non-negative part of the toric variety $V(\mathcal{I}_\Delta)$ [8, Theorem 3.2]. By this fact, the Markov basis gives the implicit equations, cutting out the set $\overline{\mathcal{E}_\Delta}$. Before stating the main theorem we remind the reader of the following definition:

Definition 13. *A polytope P is called k-neighborly if the convex hull of any k or less of its vertices is a face of P.*

Note that, any l dimensional face, for $l < k$ of a k-neighborly polytope is a simplex. We can now state our main result in two equivalent formulations:

Theorem 14. *Let g be the minimal cardinality among the non-faces of Δ.*
Geometric Formulation: *The marginal polytope is $(2^{g-1}-1)$-neighborly.*
Probabilistic Formulation: *Every distribution p with $|\mathrm{supp}(p)| < 2^{g-1}$ is contained in $\overline{\mathcal{E}_\Delta}$.*

Proof. The probabilistic formulation is easy to see. Just observe that by Theorem 5 each monomial appearing in the set of generators $\left\{p^{m^+} - p^{m^-} : m \in M\right\}$ has cardinality of its support bounded from below by 2^{g-1}. Therefore a p with $|\mathrm{supp}(p)| < 2^{g-1}$ must fulfill the defining equations trivially.

Now, the geometric formulation is due to the well known fact that a set $\mathcal{Y} \subseteq \mathcal{X}$ is the support set of some $p \in \overline{\mathcal{E}_\Delta}$ if and only if $\mathrm{conv}\{A_y : y \in \mathcal{Y}\}$ is a face of the marginal polytope Q_Δ. This is a consequence of the fact that the marginals computed by A_Δ form a sufficient statistics for the exponential family $\mathcal{E}_\Delta$. $\square$

Remark. (The bound is sharp) On first sight one would maybe expect a better neighborliness property in the non-binary cases, for instance if every variable is ternary. However, one can easily see that the bound is sharp in the sense that already for the "no-three-way-interaction" model with ternary variables, given by $N = \{1,2,3\}, \mathcal{X}_i = \{0,1,2\}$ for $i = 1,2,3$ and $\Delta = \{B \subseteq \{1,2,3\} : |B| \leq 2\}$, one has square-free generators of degree 4. They can easily be computed with 4ti2 [1] or looked up in the Markov Bases Database [17]. Then a p supported exactly on the positive support is a counterexample for any improvement of Theorem 14.

Remark. (Maximizing multiinformation) The so called multiinformation is an entropic quantity which generalizes mutual information to more than two variables. Denoting $H(p) := -\sum_{x\in\mathcal{X}} p(x)\log p(x)$ the entropy of p, and $H_i(p) := -\sum_{x\in\mathcal{X}_i} p_{\{i\}}(x)\log p_{\{i\}}(x)$ the marginal entropy for $i \in N$, it is defined as

$$MI(p) := \sum_{i\in N} H_i(p) - H(p). \tag{29}$$

An interesting problem, considered in [3], is to maximize this function. There all global maximizer in the binary case are classified giving there support sets, and the question to construct a low dimensional family containing all maximizer is raised. It holds that any global maximizer is supported on two elements only. More generally, by [2, Theorem 3.2] any local maximizer p^* satisfies

$$|\mathrm{supp}(p^*)| \leq n + 1. \tag{30}$$

Let $\Delta_k := \{B \subseteq N : |B| \leq k\}$ denote the uniform simplicial complex of order k, then it is shown

Corollary 1. (Theorem 3.5 in [3]) *All global maximizer of MI are contained in $\overline{\mathcal{E}_{\Delta_2}}$.*

In view of the bound on the cardinality of the support, this now also follows from our Theorem 14, and more generally:

Corollary 2. *Any local maximizer of MI is contained in the closure of the uniform hierarchical model of order $k^* \geq \log_2(n+2)$.*

Considerations in this direction can be generalized even further when the multiinformation is replaced by the Kullback-Leibler divergence from a general exponential family. See [14] for details.

4. Markov bases of high dimensional models

Finally, in this last section, we will show an example where the moves $m_G^{y_{N\setminus G}}$ already constitute the full Markov Basis. Consider again the binary case $\mathcal{X} = \{0,1\}^n$. Let $G \subseteq N$. We denote

$$\Delta_{/G} := \{B \subseteq N : B \not\supseteq G\}, \tag{31}$$

the complex of all sets not containing G. We have seen that the toric ideal for this complex is generated in degree at least $2^{|G|-1}$. In this section we show, that if Δ has the structure (31) and the variables are binary, the Markov basis is given by the moves $m_G^{y_{N\setminus G}}$ as defined in (16), and therefore $I_{\Delta_{/G}}$ is generated in exactly degree $2^{|G|-1}$. As the no-three-way interaction model is of the form (31) it is also clear that the statement of the following theorem does not hold as soon as the variables are not binary.

Theorem 15. *Let $\mathcal{G} = [G, N]$. A Markov basis of the binary hierarchical model given by $\Delta_{/G}$ is*

$$M := \left\{ m_{\mathcal{G}}^{y_{N\setminus G}} : y_{N\setminus G} \in \mathcal{X}_{N\setminus G} \right\}. \tag{32}$$

Proof. We apply the standard technique [4] of reducing the degree of a given binomial via the moves in M. For convenience we introduce tableau notation [9] for monomials. In this notation, the monomial p^u is represented by listing each $x \in \mathcal{X}$, $u(x)$ times. For example $p_{000}p_{110}p_{111}^2$ will be written as the tableau

$$\begin{bmatrix} 000 \\ 110 \\ 111 \\ 111 \end{bmatrix}. \tag{33}$$

Assume $p^u - p^v \in \mathcal{I}_{\Delta_{/G}}$. Without loss of generality we assume that $G = \{l, \ldots, n\}$. We can assume that u and v have disjoint supports, otherwise we write $p^u - p^v = q(p^{u'} - p^{v'})$ and the following argument shows that $p^{u'} - p^{v'}$ can be expressed in terms of the Markov basis. Consider first the case $u(00\ldots0) \geq 1$. Since the marginals on the $n-1$ sets

$$\{1,2,\ldots,n-1\}, \{1,2,\ldots,n-2,n\}, \ldots, \{1,2,\ldots,l-1,l+1,\ldots,n\} \tag{34}$$

of u and v coincide, and $v(00\dots0) = 0$ we find that the given binomial has the form

$$\begin{bmatrix} 00\dots0\underline{000\dots0} \\ \dots \\ \vdots \\ \dots \\ \vdots \end{bmatrix} - \begin{bmatrix} 00\dots0\underline{100\dots0} \\ 00\dots0\underline{010\dots0} \\ \vdots \\ 00\dots0\underline{000\dots1} \\ \vdots \end{bmatrix} \tag{35}$$

where the set G is underlined. Applying the same argument in the other direction, namely that, since u has the same $n-1$ marginals on the sets (34) we find that $u(x) > 0$ for any x which has exactly two non-zero positions, both lying in G, formally $u(x) > 0$ for any x with $\operatorname{supp}(x) \subseteq G$ and $|\operatorname{supp}(x)| = 2$. We continue to find that $v(x) > 0$ for any x with $\operatorname{supp}(x) \subseteq G$ and $|\operatorname{supp}(x)| = 3$. Repeating this argument we find that p^u contains all configurations with zero outside G and an even number of ones in G. Conversely, p^v contains all configurations that are zero outside G and have an odd number of ones in G. All together, this is exactly the move $m_{\mathcal{G}}^{00\dots0}$. Obviously, in the general case, if in the beginning we would have started with some other configuration instead of $00\dots0$, say y, the same argument leads to the move $m_{\mathcal{G}}^{y_{N\setminus G}}$ instead. Abbreviating the specific move as m now, we write $p^u = Kp^{m^+}$ and $p^v = Lp^{m^-}$ with some monomials K, L and have

$$\begin{aligned} p^u - p^v &= Kp^{m^+} - Lp^{m^-} + Kp^{m^-} - Kp^{m^-} \\ &= K(p^{m^+} - p^{m^-}) - (L-K)p^{m^-}. \end{aligned} \tag{36}$$

The degree of $L-K$ is obviously smaller than the degree of $p^u - p^v$. Inductively it follows that $p^u - p^v$ can be written as a combination of the moves $m_{\mathcal{G}}^{y_{N\setminus G}}$. □

Acknowledgement. The author wishes to express his gratitude to Johannes Rauh, Bastian Steudel, and Nihat Ay for many valuable discussions.

References

[1] 4ti2 team: *4ti2 – a software package for algebraic, geometric and combinatorial problems on linear spaces.* Available at www.4ti2.de.

[2] Ay, N.: *An information-geometric approach to a theory of pragmatic structuring.* Ann. Probab. **30** (2002), 416–436. Zbl 1010.62007

[3] Ay, N.; Knauf, A.: *Maximizing multi-information.* Kybernetika **42** (2006), 517–538.

[4] Develin, M.; Sullivant, S.: *Markov bases of binary graph models.* Ann. Comb. **7** (2003), 441–466. Zbl 1037.05045

[5] Diaconis, P.; Sturmfels, B.: *Algebraic algorithms for sampling from conditional distributions.* Ann. Stat. **26** (1998), 363–397. Zbl 0952.62088

[6] Feldman, J.; Wainwright, M.; Karger, D. R.: *Using linear programming to decode linear codes.* IEEE Transactions on Information Theory **51**(3) (2005), 954–972.

[7] Fulton, W.: *Introduction to toric varieties.* Annals of Mathematical Studies, Princeton University Press, 1993. Zbl 0813.14039

[8] Geiger, D.; Meek, C.; Sturmfels, B.: *On the toric algebra of graphical models.* Ann. Stat. **34**(5) (2006), 1463–1492. Zbl 1104.60007

[9] Hoşten, S.; Sullivant, S.: *Gröbner bases and polyhedral geometry of reducible and cyclic models.* J. Comb. Theory, Series A **100**(2) (2002), 277–301. Zbl 1044.62065

[10] Kahle, T.; Ay, N.: *Support sets of distributions with given interaction structure.* Proceedings of the WUPES 2006, 2006.

[11] Kahle, T.; Olbrich, E.; Jost, J.; Ay, N.: *Complexity measures from interaction structures.* Phys. Rev. E **79** (2009), 026201.

[12] Kahle, T.; Wenzel, W.; Ay, N.: *Hierarchical models, marginal polytopes, and linear codes.* Kybernetika **45**(2) (2009), 189–208. Zbl 1167.94340

[13] De Loera, J. A.; Onn, S.: *Markov bases of three-way tables are arbitrarily complicated.* J. Symb. Comp. **41** (2006), 173–181. Zbl 1120.62043

[14] Matúš, F.; Ay, N.: *On maximization of the information divergence from an exponential family.* Proceedings of the WUPES 2003, University of Economics, Prague, 2003, 199–204.

[15] Pachter, L.; Sturmfels, B. (eds.): *Algebraic statistics for computational biology.* Cambridge University Press, 2005. Zbl 1108.62118

[16] Pontrjagin, L. S.: *Topological groups.* Second edition, Gordon and Breach, New York, London, Paris 1966. cf. Oxford Univ. Press, London 1939 Zbl 0022.17104

[17] Rauh, J.; Kahle, T.: *The Markov Bases Database.* http://mbdb.mis.mpg.de.

[18] Wainwright, M. J.; Jordan, M. I.: *Variational inference in graphical models: The view from the marginal polytope.* Allerton Conference on Communication, Control, and Computing, 2003.

Received September 8, 2008