JOURNAL OF ALGEBRAIC STATISTICS
Vol. 7, No. 1, 2016, 1-13
ISSN 1309-3452 – www.jalgstat.com

Mode Poset Probability Polytopes

Guido Montúfar[1], Johannes Raugh[2]

[1] *Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany*
[2] *Department of Mathematics and Statistics, York University, Toronto, Canada*

Abstract. A mode of a probability distribution is an elementary event that has more probability mass than each of its direct neighbors, with respect to some vicinity structure on the set of elementary events. The mode inequalities cut out a polytope from the simplex of probability distributions. Related to this is the concept of strong modes. A strong mode is an elementary event that has more probability mass than all its direct neighbors together. The set of probability distributions with a given set of strong modes is again a polytope. We study the vertices, the facets, and the volume of such polytopes depending on the sets of (strong) modes and the vicinity structures.

2000 Mathematics Subject Classifications: 52B11, 06A07, 14P10

Key Words and Phrases: order polytope, partial order, implicitization, mode

1. Introduction

Many probability models used in practice are given in a parametric form. Sometimes it is useful to also have an implicit description in terms of properties that characterize the probability distributions that belong to the model. Such a description can be used to check whether a given probability distribution lies in the model or, otherwise, to estimate how far it lies from the model. For example, if a given model has a parametrization by polynomial functions, then one can show that it has a *semialgebraic description*; that is, an implicit description as the solution set of polynomial equations and polynomial inequalities. Finding this description is known as the *implicitization* problem, which in general is very hard to solve completely. See [4] for an overview and [3, 10, 1] for examples of implicit descriptions of probability models. Even if for a particular model it is in practice not possible to give a full implicit description, it may be possible to confine the model by simple polynomial equalities and inequalities. Here we are interested in simple confinements, in terms of natural classes of linear equalities and inequalities.

We consider polyhedral sets of discrete probability distributions defined by sets of modes. Given a vicinity structure in the set of elementary events, a mode is a local maximum point of the probability distribution. More precisely, an elementary event x is

Email addresses: `montufar@mis.mpg.de` (G. Montúfar), `jarauh@yorku.ca` (J. Rauh)

a (strict) mode of a probability distribution p if and only if $p_x > p_y$, for all neighbors y of x. The vicinity structure depends on the setting. For probability distributions on a set of fixed-length strings, it is natural to call two strings neighbors if and only if they have Hamming distance one from each other. For probability distributions on an integer interval, it is natural to call two integers neighbors if and only if they are consecutive. In general, a vicinity structure is just a graph with undirected edges.

Modes are important characteristics of probability distributions. In particular, the question whether a probability distribution underlying a statistical experiment has one or more modes is important in applications. Also, many statistical models consist of "nice" probability distributions that are "smooth" in some sense. Such probability distributions have only a limited number of modes. Another motivation for studying modes was given in [7], where it was observed that mode patterns are a practical way to differentiate between certain classes of parametric models.

Besides from modes, we are also interested in the related concept of strong modes introduced in [7]. An elementary event x is a (strict) strong mode of a probability distribution p if and only if $p_x > \sum_{y\in N(x)} p_y$, where $N(x)$ denotes the set of neighbors of x. Strong modes are special types of modes. They are easier to study than modes, since each of them is defined by a single inequality.

An observation that motivates our discussion is the following. Suppose that $p = \sum_{i=1}^k \lambda_i p^i$ is a mixture of k probability distributions. If p has a strict strong mode x, then x must be a strict mode of at least one of the distributions p^i, because if $p^i_x \leq p^i_{y_i}$ for some neighbor y_i of x for all i, then $p_x = \sum_{i=1}^k \lambda_i p^i_x \leq \sum_{i=1}^k \lambda_i p^i_{y_i} \leq \sum_{y\in N(x)} \sum_{i=1}^k \lambda_i p^i_y = \sum_{y\in N(x)} p_y$. In particular, a mixture of k uni-modal distributions has at most k strong modes. Surprisingly, the same statement is not true for modes. For instance, a mixture of k product distributions can have more than k modes [7]. The maximal possible number of modes of a mixture of k product distributions is not known in general.

Example 1. *Assume we want to know whether the following distribution of three binary variables is a mixture of k product distributions:*

$$p = \left[\begin{array}{cc|cc} p_{000} & p_{001} & p_{100} & p_{101} \\ p_{010} & p_{011} & p_{110} & p_{111} \end{array}\right] = \frac{1}{24}\left[\begin{array}{cc|cc} 1 & 5 & 5 & 2 \\ 4 & 1 & 1 & 5 \end{array}\right].$$

Product distributions have at most one strict mode. Since p has 4 strict strong modes, 001, 010, 100, 111, *we can rule out mixtures of less than 4 product distributions. On the other hand, every probability distribution of n binary variables is a mixture of at most 2^{n-1} product distributions [6]. We conclude that p is a mixture of 4 and not less product distributions.*

As this example illustrates, the pattern of (strong) modes of a probability distribution can provide sufficient information to decide whether or not the distribution belongs to a given probability model. The same general idea can be applied to more complex types of probability models, like the restricted Boltzmann machine [7].

Since (strong) modes are defined by linear inequalities, the set of probability distributions with a fixed pattern of (strong) modes is a polytope, which we call (strong) mode polytope. In this paper we describe the vertices, the facets, and the volume of these polytopes, depending on the vicinity structures and the considered patterns of (strong) modes. The number of facets tells us how many linear inequalities we need to verify in order to decide membership, and the volume tells us how likely it is to encounter a distribution from the polytope.

This paper is organized as follows: In Section 2 we study the polytopes of modes and discuss their relation to order and poset polytopes. In Section 3 we study the polytopes of strong modes. In Section 4 we summarize the results and discuss examples.

2. The Polytope of Modes

We consider a finite set of elementary events V and the set of probability distributions on this set, $\Delta(V)$, which is the standard $(|V|-1)$-simplex in $\mathbb{R}^V$. We endow V with a vicinity structure described by a graph. Let $G=(V,E)$ be a simple graph (i.e., no multiple edges and no loops). For any $x,y\in V$, if $(x,y)\in E$ is an edge in G, we write $x\sim y$. Since we assume that the graph is simple, $x\sim y$ implies $x\neq y$.

Definition 1. *A point $x\in V$ is a* mode *of a probability distribution $p\in\Delta(V)$ if $p_x\geq p_y$ for all $y\sim x$.*

Definition 2. *Consider a subset $\mathcal{C}\subseteq V$. The* polytope of $\mathcal{C}$-modes *in G is the set $\mathbf{M}(G,\mathcal{C})$ of all probability distributions $p\in\Delta(V)$ for which every $x\in\mathcal{C}$ is a mode.*

The set $\mathbf{M}(G,\mathcal{C})$ is always non-empty, since it contains the uniform distribution. It is a polytope, because it is defined as a bounded intersection of finitely many closed half-spaces. For a general overview on polytopes the reader is referred to [9].

Recall that a set of vertices of a graph is *independent* if it does not contain two adjacent elements. If $\mathcal{C}$ is not independent, then $\mathbf{M}(G,\mathcal{C})$ is not full-dimensional as a subset of $\Delta(V)$; that is, $\dim(\mathbf{M}(G,\mathcal{C}))<\dim(\Delta(V))=|V|-1$. For, if $x,y\in\mathcal{C}$ are neighbors, then the defining equations of $\mathbf{M}(G,\mathcal{C})$ imply that $p_x\geq p_y\geq p_x$ and hence that any $p\in\mathbf{M}(G,\mathcal{C})$ satisfies $p_x=p_y$. This degenerate case can be easily reduced to the independent case, as discussed in Appendix A. Therefore, in the following we assume that $\mathcal{C}$ is an independent set of vertices in G; that is, $x\not\sim y$ for all $x,y\in\mathcal{C}$.

In some applications, for example those mentioned in the introduction, it is more natural to study *strict modes*, which are points $x\in V$ with $p_x>p_y$ for all $y\sim x$. A description of the set of distributions with strict modes $\mathcal{C}$ is easy to obtain from a description of $\mathbf{M}(G,\mathcal{C})$. Moreover, it may also be of interest to require that there may be no further modes outside of $\mathcal{C}$. A good understanding of $\mathbf{M}(G,\mathcal{C})$ also allows to study this problem. We illustrate this below in Example 6.

Example 2. *Let G be a square with vertices $V=\{00,01,10,11\}$ and edges $E=\{(00,01),$ $(00,10),(01,11),(10,11)\}$, as illustrated in the left part of Figure 1. A probability distribution on V is a vector $p=[p_{00},p_{01},p_{10},p_{11}]^\top$ in $\mathbb{R}^V=\mathbb{R}^4$ with $p_{00},p_{01},p_{10},p_{11}\geq 0$*

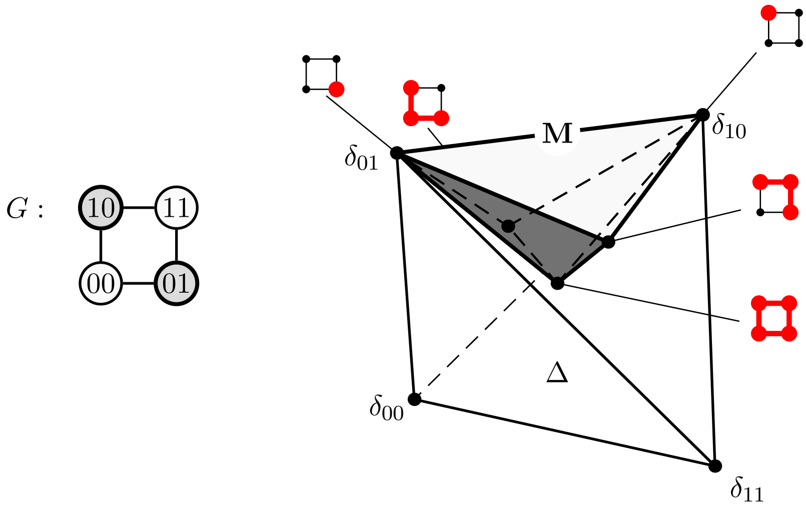

Figure 1: Illustration of Example 2. Left: The graph $G = (V, E)$, with $\mathcal{C} \subset V$ shown in dark gray. Right: The corresponding polytope $\mathbf{M}(G, \mathcal{C})$ of probability distributions with modes $\mathcal{C}$ in the 3-dimensional simplex $\Delta(V)$. Each vertex of this polytope is a uniform distribution supported on a subset of V. The corresponding support set is highlighted within G for each vertex. See Proposition 1.

and $p_{00} + p_{01} + p_{10} + p_{11} = 1$. The set $\Delta(V)$ of probability distributions on V is the 3-dimensional simplex with vertices

$$\delta_{00} = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \ \delta_{01} = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \ \delta_{10} = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \ \delta_{11} = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}.$$

Let $\mathcal{C} = \{01, 10\}$. The set $\mathbf{M}(G, \mathcal{C}) \subseteq \Delta(V)$ consists of all of probability distributions on V that satisfy $p_{01} \geq p_{00}, p_{11}$ and $p_{10} \geq p_{00}, p_{11}$. This is the solution set of the following system of linear inequalities (H-representation):

$$\left[\begin{array}{cccc} -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ \hline -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ \hline 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{array}\right] p \leq \left[\begin{array}{c} -1 \\ 1 \\ \hline 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right].$$

The first two inequalities mean that the entries of p add to one; the next two ensure the non-negativity of the entries p_x, $x \in V \setminus \mathcal{C}$; and the last 4 are mode inequalities. This description omits the inequalities $p_x \geq 0$, $x \in \mathcal{C}$, as they are redundant. The vertices of

$\mathbf{M}(G,\mathcal{C})$ are the columns of the following matrix (V-representation):

$$\begin{bmatrix} 0 & 0 & 1/3 & 0 & 1/4 \\ 1 & 0 & 1/3 & 1/3 & 1/4 \\ 0 & 1 & 1/3 & 1/3 & 1/4 \\ 0 & 0 & 0 & 1/3 & 1/4 \end{bmatrix}.$$

In particular, all vertices are uniform distributions supported on different subsets of V. The volume is $\operatorname{vol}(\mathbf{M}(G,\mathcal{C})) = \frac{1}{6}\operatorname{vol}(\Delta(V))$, as 6 congruent copies of the polytope build a perfect packing of the probability simplex. The situation is illustrated in Figure 1. We describe the general form of the vertices, facets, and volume in Propositions 1, 2, and 3.

Vertices

We have defined $\mathbf{M}(G,\mathcal{C})$ by linear inequalities (H-representation). Next we determine its vertices (V-representation). For any non-empty $W \subseteq V \setminus \mathcal{C}$ and $y \in V$ write $y \sim W$ if $y \sim x$ for some $x \in W$. Moreover, let $N_{\mathcal{C}}(W) = \{y \in \mathcal{C} : y \sim W\}$ and let $e_{\mathcal{C}}^{W}$ be the uniform distribution on $N_{\mathcal{C}}(W) \cup W$.

Proposition 1.

1. *For any $x \in \mathcal{C}$, the distribution δ_x is a vertex of $\mathbf{M}(G,\mathcal{C})$.*
2. *For any $W \subseteq V \setminus \mathcal{C}$, $W \neq \emptyset$, the distribution $e_{\mathcal{C}}^{W}$ is a vertex of $\mathbf{M}(G,\mathcal{C})$ if and only if for any $x, y \in W$, $x \neq y$, there is a path $x = x_0 \sim x_1 \sim \cdots \sim x_r = y$ in G with $x_0, x_2, \dots \in W$ and $x_1, x_3, \dots \in N_{\mathcal{C}}(W)$.*
3. *$\mathbf{M}(G,\mathcal{C})$ is the convex hull of $\{\delta_x : x \in \mathcal{C}\} \cup \{e_{\mathcal{C}}^{W} : \emptyset \neq W \subseteq V \setminus \mathcal{C}\}$.*

Proof.

1. Clearly, the vectors δ_x with $x \in \mathcal{C}$ belong to $\mathbf{M}(G,\mathcal{C})$ ($\mathcal{C}$ is independent). Since δ_x is a vertex of $\Delta(V)$, it is also a vertex of $\mathbf{M}(G,\mathcal{C})$.
2. Clearly, the vectors $e_{\mathcal{C}}^{W}$ with $\emptyset \neq W \subseteq V \setminus \mathcal{C}$ belong to $\mathbf{M}(G,\mathcal{C})$. Call a path such as in the statement of the proposition an *alternating path.* Suppose that there is no alternating path from x to y for some $x, y \in W$. Let $W_1 = \{z \in W :$ there is an alternating path from x to $z\}$ and let $W_2 = W \setminus W_1$. Then W_1, W_2 are non-empty, and $N_{\mathcal{C}}(W_1) \cap N_{\mathcal{C}}(W_2)$ is empty. Hence $e_{\mathcal{C}}^{W}$ is a convex combination of $e_{\mathcal{C}}^{W_1}$ and $e_{\mathcal{C}}^{W_2}$, and $e_{\mathcal{C}}^{W}$ is not a vertex.

   Let W be a non-empty subset of $V \setminus \mathcal{C}$ such that any pair of elements of W is connected by an alternating path. In view of item 3, to show that $e_{\mathcal{C}}^{W}$ is a vertex it suffices that for any different non-empty set $W' \subseteq V \setminus \mathcal{C}$, $W' \neq W$, we find a face of $\mathbf{M}(G,\mathcal{C})$ that contains $e_{\mathcal{C}}^{W}$ but not $e_{\mathcal{C}}^{W'}$. If there exists $x \in W' \setminus W$, then $e_{\mathcal{C}}^{W'}(x) > 0 = e_{\mathcal{C}}^{W}(x)$. Hence, $e_{\mathcal{C}}^{W}$ lies on the face of $\mathbf{M}(G,\mathcal{C})$ defined by $p_x \geq 0$, but $e_{\mathcal{C}}^{W'}$ does not. Otherwise, $W' \subsetneq W$. Let $x' \in W \setminus W'$ and $y' \in W' \neq \emptyset$. By assumption, there exists an alternating path from x' to y' in W. On this path, there exist $x \in W \setminus W'$ and $y \in \mathcal{C}$ with $y \sim x$ and $y \in N_{\mathcal{C}}(W')$. Therefore, $e_{\mathcal{C}}^{W'}(y) - e_{\mathcal{C}}^{W'}(x) > 0 = e_{\mathcal{C}}^{W}(y) - e_{\mathcal{C}}^{W}(x)$.

3. Next we show that each $p \in \mathbf{M}(G, \mathcal{C})$ can be written as a convex combination of $\{\delta_x : x \in \mathcal{C}\} \cup \{e_{\mathcal{C}}^{W} : \emptyset \neq W \subseteq V \setminus \mathcal{C}\}$. We do induction on the cardinality of $W := \operatorname{supp}(p) \setminus \mathcal{C}$. If $|W| = 0$, then $p \in \Delta(\mathcal{C})$ is a convex combination of $\{\delta_x : x \in \mathcal{C}\}$. Now assume $|W| > 0$. Let $\lambda = \min\{p_x : x \in W\}$. Then, $p - \lambda e_{\mathcal{C}}^{W} \geq 0$ (componentwise) and $\sum_x (p_x - \lambda e_{\mathcal{C}}^{W}(x)) = (1 - \lambda)$. Therefore,

$$p' := \frac{1}{1-\lambda}(p - \lambda e_{\mathcal{C}}^{W}) \in \Delta(V).$$

Moreover, one checks that $p' \in \mathbf{M}(G, \mathcal{C})$. By definition, $\operatorname{supp}(p') \setminus \mathcal{C} \subsetneq \operatorname{supp}(p) \setminus \mathcal{C}$. By induction, $\operatorname{supp}(p')$ is a convex combination of $\{\delta_x : x \in \mathcal{C}\} \cup \{e_{\mathcal{C}}^{W} : \emptyset \neq W \subseteq V \setminus \mathcal{C}\}$, and so the same is true for p. $\square$

Corollary 1. $\mathbf{M}(G, \mathcal{C})$ *is a full-dimensional sub-polytope of* $\Delta(V)$.

Proof. The convex hull of $\{\delta_x : x \in \mathcal{C}\} \cup \{e_{\mathcal{C}}^{\{y\}} : y \in V \setminus \mathcal{C}\}$ is a $(|V| - 1)$-simplex and a subset of $\mathbf{M}(G, \mathcal{C})$. Note that all vertices of this simplex are vertices of $\mathbf{M}(G, \mathcal{C})$. $\square$

Facets

The polytope $\mathbf{M}(G, \mathcal{C})$ is defined, as a subset of $\Delta(V)$, by the following inequalities:

$p_x \geq 0,$	for all $x \in V$,	(positivity inequalities)
$p_x \geq p_y,$	for all $x \in \mathcal{C}$ and $y \sim x$.	(mode inequalities)

Next we discuss which of these inequalities define facets.

Proposition 2.

1. *For any $x \in V \setminus \mathcal{C}$, the positivity inequality $p_x \geq 0$ defines a facet.*
2. *For $x \in \mathcal{C}$, the positivity inequality $p_x \geq 0$ defines a facet if and only if x is isolated in G.*
3. *For any $x \in \mathcal{C}$ and $y \sim x$, the mode inequality $p_x \geq p_y$ defines a facet.*

Proof.

1. For $x \in V \setminus \mathcal{C}$, the inequality $p_x \geq 0$ defines a facet of the sub-simplex from the proof of Corollary 1 and hence also of $\mathbf{M}(G, \mathcal{C})$.

2. If x is isolated, then x is a mode of any distribution. Therefore, $\mathbf{M}(G,\mathcal{C}) = \mathbf{M}(G,\mathcal{C} \setminus \{x\})$, and the statement follows from 1.

   Otherwise, suppose there exists $y \in V$ with $x \sim y$. Since $\mathcal{C}$ is independent, $y \notin \mathcal{C}$. Then $p_x = (p_x - p_y) + p_y$; that is, the inequality $p_x \geq 0$ is implied by the inequalities $p_x \geq p_y$ and $p_y \geq 0$. In fact, $p_x \geq 0$ defines a strict sub-face of the facet $p_y \geq 0$, since it does not contain δ_x. Therefore, $p_x \geq 0$ does not define a facet.

3. Let $W := \{z \in \mathcal{C} : z \sim y\} \setminus \{x\}$. The uniform distribution on $W \cup \{y\}$ satisfies all defining inequalities of $\mathbf{M}(G,\mathcal{C})$, except $p_x \geq p_y$. □

Triangulation and volume

The polytope $\mathbf{M}(G,\mathcal{C})$ has a natural triangulation that comes from one of $\Delta(V)$. Let $N = |V|$ be the cardinality of V. For any bijection $\sigma : \{1,\dots,N\} \to V$ let

$$\Delta_\sigma = \{p \in \Delta(V) : p_{\sigma(i)} \leq p_{\sigma(i+1)} \text{ for } i = 1,\dots,N-1\}.$$

Clearly, the sets Δ_σ form a triangulation of $\Delta(V)$. In particular, $\Delta(V) = \bigcup_\sigma \Delta_\sigma$ and $\mathrm{vol}(\Delta_\sigma \cup \Delta_{\sigma'}) = \mathrm{vol}(\Delta_\sigma) + \mathrm{vol}(\Delta_{\sigma'})$ whenever $\sigma \neq \sigma'$.

Lemma 1. *Let $\Sigma(G,\mathcal{C})$ be the set of all bijections $\sigma : \{1,\dots,N\} \to V$ that satisfy $\sigma^{-1}(x) < \sigma^{-1}(y)$ for all $y \in \mathcal{C}$ and $x \sim y$. Then $\mathbf{M}(G,\mathcal{C}) = \bigcup_{\sigma\in\Sigma(G,\mathcal{C})} \Delta_\sigma$.*

Proof. If $\sigma \in \Sigma(G,\mathcal{C})$ and $p \in \Delta_\sigma$, then $p \in \mathbf{M}(G,\mathcal{C})$ by definition. Conversely, let $p \in \mathbf{M}(G,\mathcal{C})$. Choose a bijection $\sigma : \{1,\dots,N\} \to V$ that satisfies the following:

1. $p_{\sigma(i+1)} \geq p_{\sigma(i)}$ for $i = 1,\dots,N-1$,
2. If $x \in \mathcal{C}$ and $y \sim x$, then $\sigma^{-1}(x) \leq \sigma^{-1}(y)$.

Clearly, $\sigma \in \Sigma$ and hence $p \in \Delta_\sigma$. □

Proposition 3. $\mathrm{vol}(\mathbf{M}(G,\mathcal{C})) = \frac{|\Sigma(G,\mathcal{C})|}{|V|!}\mathrm{vol}(\Delta(V))$.

Proof. All simplices Δ_σ have the same volume. Moreover, $\mathrm{vol}(\Delta_\sigma \cap \Delta_{\sigma'}) = 0$ for $\sigma \neq \sigma'$. In turn, $\mathrm{vol}(\mathbf{M}(G,\mathcal{C})) = |\Sigma(G,\mathcal{C})|\,\mathrm{vol}(\Delta_\sigma)$ and $\mathrm{vol}(\Delta(V)) = |V|!\,\mathrm{vol}(\Delta_\sigma)$. □

It remains to compute the cardinality of $\Sigma(G,\mathcal{C})$. It is not difficult to enumerate $\Sigma(G,\mathcal{C})$ by iterating over the set V. However, $\Sigma(G,\mathcal{C})$ may be very large and enumerating it can take a very long time. In fact, this is a special instance of the problem of counting the number of linear extensions of a partial order (see below); a problem which in many cases is known to be $\#P$-complete [2]. In our case, a simple lower bound is $|\Sigma(G,\mathcal{C})| \geq |\mathcal{C}|!|V \setminus \mathcal{C}|!$. Equality holds only when G is a complete bipartite graph and $\mathcal{C}$ is one of the maximal independent sets.

Relation to order polytopes

The results in this section can also be derived from results about order polytopes. To explain this, it is convenient to slightly generalize our settings. Instead of looking at a graph G and an independent subset $\mathcal{C}$ of nodes, consider a partial order $\succeq$ on V and let

$$\mathbf{M}(\succeq) := \{p \in \Delta(V) : p_x \geq p_y \text{ whenever } x \succeq y\}.$$

The polytope $\mathbf{M}(G, \mathcal{C})$ arises in the special case where $\succeq$ is defined by

$$x \succeq y \quad :\Longleftrightarrow \quad x \sim y \text{ and } x \in \mathcal{C}.$$

The relation $\succeq$ defined in this way from G and $\mathcal{C}$ is a partial order precisely when $\mathcal{C}$ is independent.

The *order polytope* of a partial order arises by looking at subsets of the unit hypercube instead of subsets of the probability simplex (see [8] and references):

$$\mathcal{O}(\succeq) := \left\{p \in [0,1]^V : p_x \geq p_y \text{ whenever } x \succeq y\right\}.$$

One can show that $\mathbf{M}(\succeq)$ is the vertex figure of $\mathcal{O}(\succeq)$ at the vertex 0. This observation allows to transfer the results from [8] to $\mathbf{M}(G, \mathcal{C})$ and, more generally, to $\mathbf{M}(\succeq)$.

In particular, our results about vertices, facets and volumes can be generalized to $\mathbf{M}(\succeq)$. Such generalizations appear interesting in their own right, but would go beyond the scope of this work. For more on order polytopes the reader is referred to [5].

3. The Polytope of Strong Modes

Definition 3. *A point $x \in V$ is a* strong mode *of a probability distribution $p \in \Delta(V)$ if $p_x \geq \sum_{y\sim x} p_y$.*

Definition 4. *Consider a subset $\mathcal{C} \subseteq V$. The* polytope of strong $\mathcal{C}$-modes *in G is the set $\mathbf{S}(G, \mathcal{C})$ of all probability distributions $p \in \Delta(V)$ for which every $x \in \mathcal{C}$ is a strong mode.*

Again, in applications one may be interested in *strict strong modes*, which are defined by strict inequalities of the form $p_x > \sum_{y\sim x} p_y$.

If $x \sim y$ for two strong modes of $p \in \Delta(V)$, then $p_x = p_y$ and $p_z = 0$ for all other neighbors z of x or y. In order to avoid such pathological cases, in the following we always assume that $\mathcal{C}$ is an independent subset of G.

Example 3. *Consider the graph $G = (V, E)$ from Example 2. Let $\mathcal{C} = \{01, 10\}$. The set $\mathbf{S}(G, \mathcal{C}) \subseteq \Delta(V)$ consists of all probability distributions on V that satisfy $p_{01} \geq p_{00} + p_{11}$ and $p_{10} \geq p_{00} + p_{11}$. This is the polytope with H-representation*

$$\left[\begin{array}{cccc} -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ \hline -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ \hline 1 & -1 & 0 & 1 \\ 1 & 0 & -1 & 1 \end{array}\right] p \leq \left[\begin{array}{c} -1 \\ 1 \\ \hline 0 \\ 0 \\ \hline 0 \\ 0 \end{array}\right].$$

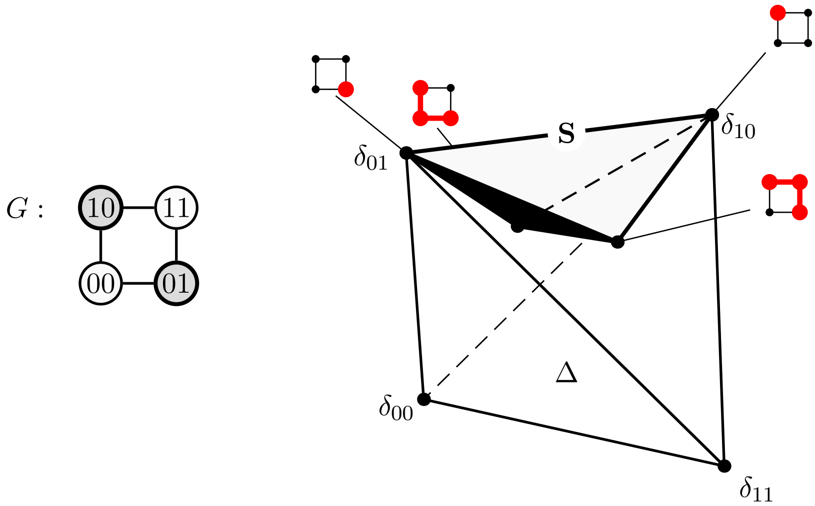

Figure 2: Illustration of Example 3. Left: The graph $G = (V, E)$, with $\mathcal{C} \subset V$ shown in dark gray. Right: The corresponding polytope $\mathbf{S}(G, \mathcal{C})$ of probability distributions with strong modes $\mathcal{C}$ in the 3-dimensional simplex $\Delta(V)$. Each vertex of this polytope is a uniform distribution supported on a subset of V. The corresponding support set is highlighted within G for each vertex. See Proposition 4.

The first two inequalities ensure that the entries of p add to one; the next two ensure the non-negativity of the entries p_x, $x \in V \setminus \mathcal{C}$; and the last two are strong mode inequalities. The V-representation is

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

In particular, all vertices are uniform distributions supported on different subsets of V. The volume is $\mathrm{vol}(\mathbf{S}(V, \mathcal{C})) = \frac{1}{9}\mathrm{vol}(\Delta(V))$, which can be computed using Proposition 6. The situation is illustrated in Figure 2. We describe the general form of the vertices, facets, and volume in Propositions 4, 5, and 6.

The next proposition describes the vertices of the polytope $\mathbf{S}(G, \mathcal{C})$. For any $x \in V$ let $N_{\mathcal{C}}(x) = \{y \in \mathcal{C} : y \sim x\}$ and let $f_{\mathcal{C}}^x$ be the uniform distribution on $N_{\mathcal{C}}(x) \cup \{x\}$.

Proposition 4. *$\mathbf{S}(G, \mathcal{C})$ is a $(|V| - 1)$-simplex with vertices $f_{\mathcal{C}}^x$, $x \in V$.*

Proof. The set of vectors $\{f_{\mathcal{C}}^x : x \in V\}$ is linearly independent. To see this, note that the matrix with columns $f_{\mathcal{C}}^x$ is in tridiagonal form when V is ordered such that the vertices in $\mathcal{C}$ come before the vertices in $V \setminus \mathcal{C}$. Therefore, the probability distributions $f_{\mathcal{C}}^x$ span a $(|V| - 1)$-dimensional simplex.

It is easy to check that $f_{\mathcal{C}}^x \in \mathbf{S}(G, \mathcal{C})$ for any $x \in V$. It remains to prove that any $p \in \mathbf{S}(G, \mathcal{C})$ lies in the convex hull of $\{f_{\mathcal{C}}^x : x \in V\}$. We do induction on the cardinality of $W := \mathrm{supp}(p) \setminus \mathcal{C}$. If $|W| = 0$, then $p \in \Delta(\mathcal{C})$ is a convex combination of $\{\delta_x : x \in \mathcal{C}\} =$

$\{f_{\mathcal{C}}^x : x \in \mathcal{C}\}$. Otherwise, let $x \in W$. Then

$$p' := \frac{1}{1-p_x}(p - p_x f_{\mathcal{C}}^x) \in \Delta(V),$$

since $p \in \mathbf{M}(G,\mathcal{C})$. Moreover, $p' \in \mathbf{M}(G,\mathcal{C})$. The statement now follows by induction, since $\operatorname{supp}(p') \setminus \mathcal{C} = W \setminus \{x\}$. □

Proposition 5. *The facets of* $\mathbf{S}(G,\mathcal{C})$ *are* $p_x \geq 0$, $x \in V \setminus \mathcal{C}$, *and* $p_x \geq \sum_{y\sim x} p_y$, $x \in \mathcal{C}$.

Proof. Each inequality defines a hyperplane that contains $|V|-1$ vertices. □

Proposition 6. $\operatorname{vol}(\mathbf{S}(G,\mathcal{C})) = \Big(\prod_{x\in V} \frac{1}{|N_{\mathcal{C}}(x)|+1}\Big) \operatorname{vol}(\Delta(V))$.

Proof. After rearrangement of columns, the matrix

$$(f_{\mathcal{C}}^x)_{x\in V} = \left((\delta_x)_{x\in\mathcal{C}}, \left(\tfrac{1}{|N_{\mathcal{C}}(x)|+1}\mathbb{1}_{N_{\mathcal{C}}(x)}\right)_{x\in V\setminus\mathcal{C}, x\sim\mathcal{C}}, (\delta_x)_{x\in V\setminus\mathcal{C}, x\not\sim\mathcal{C}}\right)$$

is in upper triangular from, with diagonal elements $\frac{1}{|N_{\mathcal{C}}(x)|+1}$, $x \in V$. The statement now follows from the next Lemma 2. □

Lemma 2. *Let* $\Delta = \operatorname{conv}\{e_0, \ldots, e_d\}$ *be the standard* d*-simplex in* $\mathbb{R}^{d+1}$ *and let* $s_0, \ldots, s_d \in \Delta$*. Then the* d*-volume of* $S = \operatorname{conv}\{s_0, \ldots, s_d\}$ *satisfies* $\operatorname{vol}(S) = |\det(s_0, \ldots, s_d)| \operatorname{vol}(\Delta)$.

Proof. The volume of the $(d+1)$-dimensional parallelepiped spanned by $s_0, \ldots, s_d \in \mathbb{R}^{d+1}$ is $|\det(s_0, \ldots, s_d)|$. The volume of an n-simplex with vertices $v_0, \ldots, v_n$ in $\mathbb{R}^n$ is $\frac{1}{n!}|\det(v_1 - v_0, \ldots, v_n - v_0)|$. Hence the volume of the $(d+1)$-simplex P with vertices $(0, s_0, \ldots, s_d)$ is $\operatorname{vol}(P) = \frac{1}{(d+1)!}|\det(s_0, \ldots, s_d)|$. Note that P is a pyramid over S of height $h = \frac{1}{\sqrt{d+1}}$. Thus $\operatorname{vol}(P) = \frac{h}{d+1}\operatorname{vol}(S)$. The volume of the regular d-simplex is $\operatorname{vol}(\Delta) = \frac{\sqrt{d+1}}{d!}$. The statement follows by combining these formulas. □

4. Summary and Examples

The description of the mode polytope given in Section 2 can be summarized as follows. There is one vertex for each mode and one for each non-empty set of non-modes that is connected in G by a path that alternates between non-modes and modes (Proposition 1). There is one facet for each non-mode and one for each edge connecting a mode with a non-mode (Proposition 2). The volume of the mode polytope, relative to the probability simplex, is equal to the number of linear extensions of the partial order defined by the mode inequalities, divided by the total number of linear orders of the set of elementary events (Proposition 3).

The description of the strong mode polytope given in Section 3 can be summarized as follows. This polytope is a full dimensional simplex (Proposition 4). The volume, relative to the ambient probability simplex, is equal to the inverse of the product of $|N_{\mathcal{C}}(x)+1|$, $x \in V$, where $N_{\mathcal{C}}(x)$ are the neighbors of x that are declared strong modes (Proposition 6).

Example 4. *Let G be the complete bipartite graph with $\mathcal{C}$ on one side and $V \setminus \mathcal{C}$ on the other. Let $m = |\mathcal{C}|$ and $n = |V \setminus \mathcal{C}|$.*

Then the polytope of modes $\mathbf{M}(G,\mathcal{C})$ has $m+2^n-1$ vertices, $n+mn$ facets, and volume $\mathrm{vol}(\mathbf{M}) = \frac{m!n!}{(m+n)!}\mathrm{vol}(\Delta)$. For the volume, note that in this example the number of linear extensions $|\Sigma(G,\mathcal{C})|$ is equal to the number of permutations of the modes times the number of permutations of the non-modes.

The polytope of strong modes $\mathbf{S}(G,\mathcal{C})$ is a simplex with $n+m$ vertices and $n+m$ facets. The volume is $\mathrm{vol}(\mathbf{S}) = \frac{1}{(m+1)^n}\mathrm{vol}(\Delta)$.

Example 5. *Generalizing Examples 2 and 3, let G be the edge graph of an n-cube, such that $V = \{0,1\}^n$ and two points are adjacent if and only if they are Hamming neighbors.*

- *Let $\mathcal{C}$ have cardinality $|\mathcal{C}| = k$ and minimum distance $\min_{x,y\in\mathcal{C}} |\{i \in \{1,\dots,n\}: x_i \neq y_i\}| \geq 3$. Then $\mathbf{M}(G,\mathcal{C})$ has $k+k(2^n-1)+2^n-k(n+1)$ vertices, 2^n-k+kn facets, and volume $\mathrm{vol}(\mathbf{M}) = \frac{|\Sigma|}{2^n!}\mathrm{vol}(\Delta) \geq \frac{k!(2^n-k)!}{2^n!}\mathrm{vol}(\Delta)$.*

  $\mathbf{S}(G,\mathcal{C})$ is a (2^n-1)-simplex with $\mathrm{vol}(\mathbf{S}) = 2^{-kn}\mathrm{vol}(\Delta)$.

- *Let $\mathcal{C}$ be the set of all even-parity strings, $\mathcal{C} = \{x \in V: \sum_{i=1}^n x_i = 0 \mod (2)\}$.*

  Then $\mathbf{M}(G,\mathcal{C})$ has $2^{n-1} + 2^{2^{n-1}} - 1$ vertices, $2^{n-1} + 2^{n-1}n$ facets, and volume $\mathrm{vol}(\mathbf{M}) = \frac{|\Sigma|}{2^n!}\mathrm{vol}(\Delta) \geq \frac{2^{n-1}!2^{n-1}!}{2^n!}\mathrm{vol}(\Delta)$. For $n = 2$ we have $|\Sigma| = 4$ and for $n = 3$ we have $|\Sigma| = 720$. The next open case is $n = 4$.

  $\mathbf{S}(G,\mathcal{C})$ is a (2^n-1)-simplex with $\mathrm{vol}(\mathbf{S}) = (n+1)^{-2^{n-1}}\mathrm{vol}(\Delta)$.

Example 6. *Suppose we are interested in probability distributions that have exactly one mode. First, suppose that G is the complete graph on V, i.e., any pair of nodes $(x,y) \in V \times V$ is an edge. Any distribution $p \in \Delta(V)$ that is generic, in the sense that $p(x) \neq p(y)$ for any $x \neq y$, has a unique mode. Thus, the set of distributions with exactly one mode arises from $\Delta(V)$ by removing hyperplanes, and so its volume is $\mathrm{vol}(\Delta(V))$. By symmetry, for any fixed $x \in V$, the set of distributions with exactly one mode $x \in V$ has volume $\frac{1}{|V|}\mathrm{vol}(\Delta(V))$. This result can also be derived from our Proposition 3, noting that there are $(|V|-1)!$ linear orders of V in which x is maximal.*

The situation is more complicated in other graphs, and one is led to inclusion-exclusion formulas. We illustrate this by means of the square $G = (V,E)$ from Figure 1 left. A distribution $p \in \Delta(V)$ has a unique mode 01 if and only if 01 is a mode and 10 is not a mode. Thus, using Proposition 3, for the vicinity structure G the set of distributions $p \in \Delta(V)$ with unique mode 01 has volume

$$\mathrm{vol}(\mathbf{M}(G,\{01\})) - \mathrm{vol}(\mathbf{M}(G,\{01,10\})) = \frac{8-4}{4!}\mathrm{vol}(\Delta(V)) = \frac{1}{6}\mathrm{vol}(\Delta(V)).$$

In total, by symmetry, the set of distributions with a unique mode has volume $\frac{4}{6}\mathrm{vol}(\Delta(V))$. This says that two thirds of all distributions have precisely one mode, and one third of all distributions have two modes.

A. Description of Degenerate Cases

When looking at $\mathbf{M}(G, \mathcal{C})$, the case that $\mathcal{C}$ is not independent can be reduced to the independent case: For any nodes $x, y \in \mathcal{C}$ with $x \sim y$, any $p \in \mathbf{M}(G, \mathcal{C})$ satisfies $p(x) = p(y)$. In this case we can contract the edge (x, y) and identify x and y. To be precise, we construct a new graph $G' = (V', E')$, where $V' = V \setminus \{y\}$ and $E' = (E \cap (V' \times V')) \cup \{(x, z) : z \in V' \setminus \{x\}, (z, y) \in E\}$, and we let $\mathcal{C}' = \mathcal{C} \setminus \{y\} = \mathcal{C} \cap V'$. Then the truncation $\psi : \Delta(V) \to \Delta(V')$ defined by

$$\psi(p)_z = \frac{p_z}{\sum_{z' \in V'} p_{z'}}, \quad \text{for all } z \in V',$$

restricts to a bijection $\mathbf{M}(G, \mathcal{C}) \cong \mathbf{M}(G', \mathcal{C}')$. Furthermore, as a projective map, it preserves the face structure of the polytopes. The inverse of the restricted map is given by $\psi^{-1}(q)_z = q_z/(q_x + \sum_{z' \in V'} q_{z'})$, if $z \in V'$, and $\psi^{-1}(q)_y = q_x/(q_x + \sum_{z' \in V'} q_{z'})$. Note that for any $z \in \mathcal{C}$ and $p \in \mathbf{M}(G, \mathcal{C})$, $z \in \mathcal{C}'$ is a mode of p if z is a mode of $\psi(p)$.

Acknowledgments

We are grateful to anonymous referees for helpful comments.

References

[1] Elizabeth S. Allman, John A. Rhodes, and Amelia Taylor. A semialgebraic description of the general Markov model on phylogenetic trees. *SIAM Journal on Discrete Mathematics*, 28:736–755, 2014. `http://dx.doi.org/10.1137/120901568`.

[2] Graham Brightwell and Peter Winkler. Counting linear extensions. *Order*, 8(3):225–242, 1991. `http://dx.doi.org/10.1007/BF00383444`.

[3] María Angélica Cueto, Enrique A. Tobis, and Josephine Yu. An implicitization challenge for binary factor analysis. *Journal of Symbolic Computation*, 45(12):1296–1315, 2010. `http://dx.doi.org/10.1016/j.jsc.2010.06.011`.

[4] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on algebraic statistics*. Oberwolfach Seminars, volume 39, Birkhäuser, Basel-Boston-Berlin, 2009. `http://dx.doi.org/10.1007/978-3-7643-8905-5`.

[5] Jiří Matoušek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002. `http://dx.doi.org/10.1007/978-1-4613-0039-7`.

[6] Guido Montúfar. Mixture decompositions of exponential families using a decomposition of their sample spaces. *Kybernetika*, 49(1):23–39, 2013. `http://www.kybernetika.cz/content/2013/1/23`.

[7] Guido Montúfar and Jason Morton. When does a mixture of products contain a product of mixtures? *SIAM Journal on Discrete Mathematics*, 29(1):321–347, 2015. `http://dx.doi.org/10.1137/140957081`.

[8] Richard P. Stanley. Two poset polytopes. *Discrete & Computational Geometry*, 1(1):9–23, 1986. `http://dx.doi.org/10.1007/BF02187680`.

[9] Günter M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer Verlag, 1995. `http://dx.doi.org/10.1007/978-1-4613-8431-1`.

[10] Piotr Zwiernik and Jim Q. Smith. Implicit inequality constraints in a binary tree model. *Electron. J. Statist.*, 5:1276–1312, 2011. `http://dx.doi.org/10.1214/11-EJS640`.