Combinatorial Properties of Hierarchical Models

Daniel Irving Bernstein and Seth Sullivant

North Carolina State University

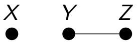
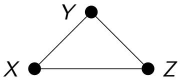
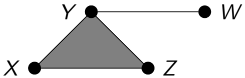
dibernst@ncsu.edu
http://www4.ncsu.edu/~dibernst/

http://arxiv.org/abs/1502.06131
http://arxiv.org/abs/1508.05461

Hierarchical Models

Definition (Hierarchical Model)

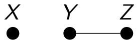
Let $X_1, \ldots, X_n$ be discrete random variables. A simplicial complex $\mathcal{C}$ on $X_1, \ldots, X_n$ specifies independence relations among the X_is. The collection of joint probability distributions on $X_1, \ldots, X_n$ that satisfy these relations is called a *hierarchical model*.

X is independent of Y and Z, but Y and Z are dependent	
There is no 3-way dependence	
X and Z are independent of W given Y	

Example

- Assume X has states x_1, x_2, x_3, Y has states y_1, y_2 and Z has states z_1, z_2.
- Observe (X, Y, Z) several times, record the counts in a $3 \times 2 \times 2$ array

$$\underset{z_1}{\begin{array}{c|cc} & y_1 & y_2 \\ \hline x_1 & 1 & 1 \\ x_2 & 2 & 0 \\ x_3 & 3 & 1 \end{array}} \qquad \underset{z_2}{\begin{array}{c|cc} & y_1 & y_2 \\ \hline x_1 & 2 & 1 \\ x_2 & 1 & 3 \\ x_3 & 0 & 2 \end{array}}$$

- Sufficient statistics for the model given by  are

$$\begin{array}{c|cc} & y_1 & y_2 \\ \hline z_2 & 3 & 6 \\ z_1 & 6 & 2 \end{array} \qquad \begin{array}{c|c} x_1 & 5 \\ x_2 & 6 \\ x_3 & 6 \end{array}$$

Design Matrix

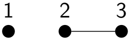
- Discrete random variables $X_1, \ldots, X_n$
- X_i has d_i states. Notation: $\mathbf{d} = (d_1, \ldots, d_n)$
- $\mathcal{C}$ denotes a simplicial complex on $[n]$
- The design matrix of the corresponding hierarchical model is

Definition

Let $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ be the matrix defined as follows:

- Columns are indexed by elements of $\bigoplus_{i=1}^{n}[d_i]$
- Rows are indexed by $\bigoplus_{F \in \text{facet}(\mathcal{C})} \bigoplus_{j \in F}[d_j]$
- Entry in row $(F, (j_1, \ldots, j_k))$ and column $(i_1, \ldots, i_n)$ is 1 if $i|_F = (j_1, \ldots, j_k)$
- All other entries are 0

Design Matrix - Example

- Let $n = 3$ with $d_1 = 3, d_2 = 2, d_3 = 2$
- Let $\mathcal{C}$ be the complex 

- Then $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is the following matrix:

	111	112	121	122	211	212	221	222	311	312	321	322
$\{1\}, 1$	1	1	1	1	0	0	0	0	0	0	0	0
$\{1\}, 2$	0	0	0	0	1	1	1	1	0	0	0	0
$\{1\}, 3$	0	0	0	0	0	0	0	0	1	1	1	1
$\{2,3\}, 11$	1	0	0	0	1	0	0	0	1	0	0	0
$\{2,3\}, 12$	0	1	0	0	0	1	0	0	0	1	0	0
$\{2,3\}, 21$	0	0	1	0	0	0	1	0	0	0	1	0
$\{2,3\}, 22$	0	0	0	1	0	0	0	1	0	0	0	1

Unimodularity

Definition (Unimodularity)

Assume $A \in \mathbb{Z}^{d \times n}$ has full row rank. We say that A is **unimodular** if all $d \times d$ submatrices have determinant $0, 1$, or -1.

Example

The matrix $\mathcal{A}$ is unimodular, whereas $\mathcal{B}$ is not

$$\mathcal{A} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathcal{B} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

Applications include:

- Integer programming over $\{x \in \mathbb{R}^n_{\geq 0} : Ax = b\}$
- Disclosure limitation
- Computing Markov basis and universal Gröbner basis of $\mathcal{I}_{\mathcal{A}}$

Unimodularity

Question

When is the design matrix $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ unimodular?

Observation

If $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is unimodular, then so is $\mathcal{A}_{\mathcal{C}} := \mathcal{A}_{\mathcal{C},(2,\ldots,2)}$.

- Terminology abuse "$\mathcal{C}$ is unimodular" means "$\mathcal{A}_{\mathcal{C}}$ is unimodular"

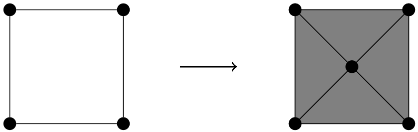
We have a complete classification of unimodular $\mathcal{C}$

Unimodularity-Preserving Operations

Definition (Adding a cone vertex)

If $\mathcal{C}$ is a simplicial complex on $[n]$, define cone($\mathcal{C}$) to be the complex on $[n+1]$ with facets

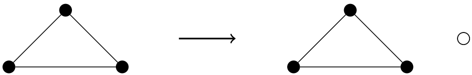
$$\text{facet}(\text{cone}(\mathcal{C})) = \{F \cup \{n+1\} : F \in \text{facet}(\mathcal{C})\}.$$



Unimodularity-Preserving Operations

Definition (Adding a ghost vertex)

If $\mathcal{C}$ is a simplicial complex on $[n]$, define $G(\mathcal{C})$ to be the simplicial complex on $[n+1]$ that has exactly the same faces as $\mathcal{C}$.

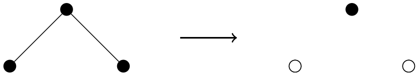


Unimodularity-Preserving Operations

Definition (Alexander Duality)

If $\mathcal{C}$ is a simplicial complex on $[n]$, then the *Alexander dual* complex $\mathcal{C}^*$ is the simplicial complex on $[n]$ with facets

$$\text{facet}(\mathcal{C}^*) = \{[n] \setminus S : S \text{ is a minimal non-face of } \mathcal{C}\}.$$



Unimodularity: Constructive Classification

Definition

We say that a simplicial complex $\mathcal{C}$ is *nuclear* if it satisfies one of the following:

1. $\mathcal{C} = \Delta_k$ for some $k \geq -2$ (i.e. a simplex)
2. $\mathcal{C} = \Delta_m \sqcup \Delta_n$ (i.e. a disjoint union of simplices)
3. $\mathcal{C} = \text{cone}(\mathcal{D})$ where $\mathcal{D}$ is nuclear
4. $\mathcal{C} = G(\mathcal{D})$ where $\mathcal{D}$ is nuclear
5. $\mathcal{C}$ is the Alexander dual of a nuclear complex.

Theorem (B.-Sullivant 2015)

The matrix $\mathcal{A}_\mathcal{C}$ is unimodular if and only if $\mathcal{C}$ is nuclear.

Simplicial Complex Minors

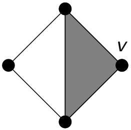
Definition (Deletion and Link)

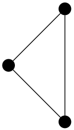
Let $\mathcal{C}$ be a simplicial complex on $[n]$. Let $v \in [n]$ be a vertex of $\mathcal{C}$. Then $\mathcal{C} \setminus v$ denotes the induced simplicial complex on $[n] \setminus \{v\}$, and $\text{link}_v(\mathcal{C})$ denotes the simplicial complex on $[n] \setminus \{v\}$ with facets

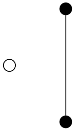
$$\text{facet}(\text{link}_v(\mathcal{C})) = \{F \setminus \{v\} : F \text{ is a facet of } \mathcal{C} \text{ with } v \in F\}.$$

Definition (Simplicial Complex Minor)

We say that $\mathcal{D}$ is a minor of $\mathcal{C}$ if $\mathcal{D}$ can be obtained from $\mathcal{C}$ via a series of deletion and link operations.

$\mathcal{C} =$ 

$\mathcal{C} \setminus v =$ 

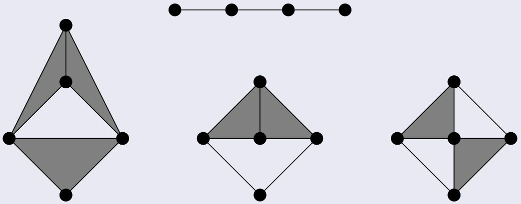
$\text{link}_v\mathcal{C} =$ 

Unimodularity: Excluded Minor Classification

Theorem (B.-Sullivant 2015)

The matrix $\mathcal{A}_{\mathcal{C}}$ is unimodular if and only if $\mathcal{C}$ has no simplicial complex minors isomorphic to any of the following

- $\partial\Delta_k \sqcup \{v\}$, *the disjoint union of the boundary of a simplex and an isolated vertex*
- O_6, *the boundary complex of an octahedron, or its Alexander dual O_6^**
- *The four simplicial complexes shown below*



Sketch of Proof

- $\mathcal{C}$ nuclear $\implies$ $\mathcal{C}$ unimodular
  - Simplices are unimodular
  - A disjoint union of two simplices is unimodular
  - Adding cone and ghost vertices and taking duals preserves unimodularity
- $\mathcal{C}$ unimodular $\implies$ $\mathcal{C}$ avoids forbidden minors
  - The forbidden minors are not unimodular
  - Taking minors preserves unimodularity
- $\mathcal{C}$ avoids forbidden minors $\implies$ $\mathcal{C}$ nuclear
  - If $\mathcal{C}$ avoids the forbidden minors but has a 4-cycle, then it must be an iterated cone over the 4-cycle. This is nuclear.
  - So focus on 4-cycle-free complexes. Then the 1-skeleton is either a complete graph, or two complete graphs glued along a clique.
  - Complex induction argument based on the link of a vertex of $\mathcal{C}$.

Next Steps - Unimodularity

Question

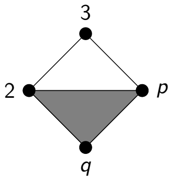
Given a simplicial complex $\mathcal{C}$ on $[n]$ and an integer vector $\mathbf{d} = (d_1, \ldots, d_n)$ with $d_i \geq 2$, is $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ unimodular?

Corollary (B.-Sullivant 2015)

If $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is unimodular, then $\mathcal{C}$ is nuclear.

Question

Let $\mathcal{C}$ and $\mathbf{d}$ be specified by the figure below. For which values of p and q is $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ unimodular?



Holes and Normality

Let $A \in \mathbb{N}^{d \times n}$. We define:

- $\mathbb{N}A := \{Ax : x \in \mathbb{N}^n\}$ (Semigroup generated by columns of A)
- $\mathbb{Z}A := \{Ax : x \in \mathbb{Z}^n\}$ (Lattice generated by columns of A)
- $\mathbb{R}_{\geq 0}A := \{Ax : x \in \mathbb{R}, x \geq 0\}$ (Cone generated by columns of A)

Definition (Normality)

We say that A is *normal* if

$$\mathbb{N}A = \mathbb{R}_{\geq 0}A \cap \mathbb{Z}A.$$

If A is not normal and

$$h \in \mathbb{R}_{\geq 0}A \cap \mathbb{Z}A \setminus \mathbb{N}A$$

then we say that h is a *hole* of $\mathbb{N}A$.

Normality: Non-example

The following matrix is *not* normal

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

because $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is a hole. Note:

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix}$$

so $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}_{\geq 0}A \cap \mathbb{Z}A$. However, $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \notin \mathbb{N}A$.

Normality

Question

When is $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ normal?

Observation

If $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is normal, then so is $\mathcal{A}_{\mathcal{C}} := \mathcal{A}_{\mathcal{C},(2,\ldots,2)}$.

- Terminology abuse "$\mathcal{C}$ is normal" means "$\mathcal{A}_{\mathcal{C}}$ is normal"

Applications include:

- Integer table feasibility problem
- Toric fiber products for constructing Markov bases work best with normal $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ (Rauh-Sullivant 2014)
- Sequential importance sampling works best with normal $\mathcal{A}_{\mathcal{C},\mathbf{d}}$

We have some partial results towards classification of normal $\mathcal{C}$

Known Classification Results - Normality

Theorem (Sullivant 2010)

If $\mathcal{C}$ is a graph, then $\mathcal{A}_{\mathcal{C}}$ is normal if and only if $\mathcal{C}$ is free of K_4-minors.

Theorem (Bruns, Hemmecke, Hibi, Ichim, Ohsugi, Köppe, Söger 2007-2011)

Let $\mathcal{C}$ be a complex whose facets are all $m-1$ element subsets of $[m]$. Then $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is normal in precisely the following situations up to symmetry:

1. *At most two of the d_v are greater than two*
2. $m = 3$ *and* $\mathbf{d} = (3,3,a)$ *for any* $a \in \mathbb{N}$
3. $m = 3$ *and* $\mathbf{d} = (3,4,4),(3,4,5)$ *or* $(3,5,5)$.

Theorem (Rauh-Sullivant 2014)

Let $\mathcal{C}$ be the four-cycle graph. Then $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is normal if $\mathbf{d} = (2,a,2,b)$ *or* $\mathbf{d} = (2,a,3,b)$ *with* $a,b,\in \mathbb{N}$.

Corollary of Unimodular Classification

Definition

Let $\mathcal{C}$ be a simplicial complex on $[n]$. We say a facet of $\mathcal{C}$ that has $n-1$ vertices is called a *big facet*.

Proposition

If $\mathcal{C}$ is a complex with a big facet, then $\mathcal{C}$ is normal if and only if unimodular.

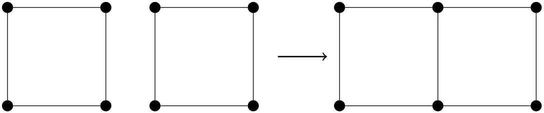
So our classification result on unimodular $\mathcal{C}$ immediately gives a classification of the normal $\mathcal{C}$ when $\mathcal{C}$ has a big facet.

Normality Preserving Operations

Theorem (Sullivant 2010)

Normality of $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is preserved under the following operations on the simplicial complex

1. *Deleting vertices*
2. *Contracting edges*
3. *Gluing two simplicial complexes along a common face*
4. *Adding or removing a cone or ghost vertex.*



Theorem (B.-Sullivant 2015)

Normality of $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is preserved when taking links of vertices of $\mathcal{C}$.

Minimally Non-Normal Simplicial Complexes

Question

Which simplicial complexes are minimally non-normal with respect to the operations of deleting vertices, contracting edges, gluing two complexes along a facet, removing cone and ghost vertices, and taking links of vertices?

Computational method:

- All simplicial complexes on 3 or fewer vertices are normal
- Choose two normal simplicial complexes $\mathcal{C}, \mathcal{D}$ on $n-1$ vertices. Create simplicial complex $\mathcal{C}'$ on n vertices by attaching a new vertex v to $\mathcal{C}$ such that $\text{link}_v(\mathcal{C}') = \mathcal{D}$
- See if (non)normality of $\mathcal{C}'$ can be certified by reducing to a smaller complex via our normality-preserving operations
- If not, check normality of $\mathcal{C}'$ using Normaliz. If non-normal, then minimally non-normal

Minimally Non-Normal Simplicial Complexes

We were able to use the computational method to determine normality on all complexes on up to 6 vertices

So far, we know that the set of minimally non-normal simplicial complexes contains:

- 20 sporadic complexes, obtained by computational method
- Two infinite families, obtained by theoretical means

Next Steps and Ongoing Work

- Develop new procedures for constructing normal $\mathcal{C}$
- Develop methods for constructing holes of $\mathbb{N}\mathcal{A}_{\mathcal{C}}$
- Classify normal complexes within certain families (e.g., surfaces)
- When does a non-normal $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ have finitely many holes?
- Find facet description of the cone $\mathbb{R}_{\geq 0}\mathcal{A}_{\mathcal{C},\mathbf{d}}$

References

- **Daniel Irving Bernstein and Seth Sullivant.**
  Unimodular Binary Hierarchical Models.
  ArXiv:1502.06131, **2015.**
- **Daniel Irving Bernstein and Seth Sullivant.**
  Normal Binary Hierarchical Models
  ArXiv:1508.05461, **2015**
- Winfried Bruns, Raymond Hemmecke, Bogdan Ichim, Matthias Köppe and Christof Söger.
  Challenging Computations of Hilbert Bases of Cones Associated with Algebraic Statistics
  Exp. Math. 20, 25 - 33 (2011)
- Mathias Drton, Bernd Sturmfels, Seth Sullivant.
  Lectures on Algebraic Statistics
  Birkhauser. Oberwolfach Seminars, Vol 39
- Takayuki HIbi and Hidefumi Ohsugi
  Toric Ideals Arising from Contingency Tables
  Commutative Algebra and Combinatorics
  Ramanujan Mathematical Society Lecture Notes Series, Number 4, Ramanujan Math. Soc., Mysore, 2007, pp. 91-115.
- Johannes Rauh and Seth Sullivant
  Lifting markov bases and higher codimension toric fiber products
  ArXiv:1404.6392 2014
- Bernd Sturmfels.
  Gröbner Bases and Convex Polytopes, volume 8 of *University Lecture Series*.
  American Mathematical Society, Providence, RI, 1996.
- Seth Sullivant.
  Normal binary graph models.
  Ann. Inst. Statist. Math., 62(4):717–726, 2010.