Mixtures of Discrete Decomposable Graphical Models

Yulia Alexandr (UCLA)

joint work with Jane Ivy Coons and Nils Sturma

Applied Algebra Seminar at UW-Madison
November 7, 2024

Graphical models

Graphical models encode relationships between random variables using a graph structure:

- Vertices $\rightarrow$ random variables
- Edges $\rightarrow$ conditional dependence relations

Any graphical model adopts a natural parametrization which can be read from the structure of the underlying graph.

Widely used in:

⋆ statistics (causal inference)
⋆ machine learning (Bayesian networks, generative models)
⋆ computational biology (protein interaction networks)
⋆ phylogenetics (gene trees)
⋆ economics (dependencies between financial entities)
⋆ computer vision (image structures and relationships within scenes)

Example

Three random variables:

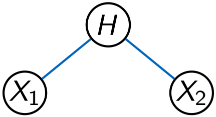
X_1 : length of a person's hair (bald, short, medium, and long).
X_2 : how often a person watches soccer (never, sometimes, and often).

Example

Three random variables:

X_1 : length of a person's hair (bald, short, medium, and long).
X_2 : how often a person watches soccer (never, sometimes, and often).

H : a person's gender!



The random variable G could be hidden or observed.

We write $X_1 \perp\!\!\!\perp X_2 | H$.

Parametric vs. implicit description

Given a model, parametrized by

$$\varphi : \theta = (\theta_1, \ldots, \theta_n) \mapsto \big(f_1(\theta), f_2(\theta), \ldots f_m(\theta)\big),$$

we are interested in describing the polynomials defining $\overline{\mathrm{image}}(\varphi)$. This process is called *implicitization*.

Parametric vs. implicit description

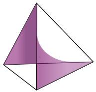
Given a model, parametrized by

$$\varphi : \theta = (\theta_1, \dots, \theta_n) \mapsto \big(f_1(\theta), f_2(\theta), \dots f_m(\theta)\big),$$

we are interested in describing the polynomials defining $\overline{\text{image}}(\varphi)$. This process is called *implicitization*.

Example: the independence model.
Parametrization:

$$(\theta_1, \theta_2) \mapsto \big(\underbrace{\theta_1\theta_2}_{p_1}, \underbrace{\theta_1(1-\theta_2)}_{p_2}, \underbrace{(1-\theta_1)\theta_2}_{p_3}, \underbrace{(1-\theta_1)(1-\theta_2)}_{p_4}\big).$$



Implicit ideal: $I = \langle p_1 p_4 - p_2 p_3, \;\; p_1 + p_2 + p_3 + p_4 - 1 \rangle$.

The generators of the ideal I are called *model invariants*.

Undirected Graphical Models

Setup: Random variables $(X_v)_{v \in V}$ and undirected graph $G = (V, E)$.

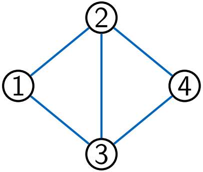
The graph G specifies dependencies between random variables.

Global Markov Property of G: all conditional independence statements

$$X_A \perp\!\!\!\perp X_B | X_C$$

for all disjoint sets A, B, and C such that C separates A and B in G.

Example:



$$X_1 \perp\!\!\!\perp X_4 | (X_2, X_3)$$

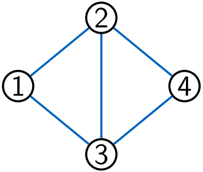
Parametrized Graphical Models

Factorization:

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \psi_C(x_C),$$

where $\mathcal{C}(G)$ is the collection of maximal cliques of G.

Example:



$$p(x_1, x_2, x_3, x_4) \propto \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$$

Theorem (Hammersley-Clifford)

A positive probability density satisfies the global Markov property on the graph G if and only if it factorizes according to G.

Discrete Undirected Graphical Models

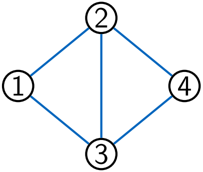
Finite state space $\mathcal{R} = \prod_{v \in V}[d_v]$. For $A \subset V$, let $\mathcal{R}_A = \prod_{v \in A}[d_v]$ and $d_A := \#\mathcal{R}_A = \prod_{v \in A} d_v$.

Definition

The discrete log-linear graphical model $\mathcal{M}_G$ consists of all probability distributions $p \in \Delta_{|\mathcal{R}|}$ such that

$$p_i = \frac{1}{Z(\theta)} \prod_{C \in \mathcal{C}(G)} \theta^{(C)}_{i_C}.$$

Example



$$p_{i_1 i_2 i_3 i_4} \propto \theta^{(C_1)}_{i_1 i_2 i_3} \cdot \theta^{(C_2)}_{i_2 i_3 i_4}$$

This is a log-linear model! It is parametrized by monomials and its Zariski closure is a toric variety.

Log-linear (toric) models

Every *log-linear model* is specified by an integer matrix $A \in \mathbb{Z}^{d\times n}$ with the vector of all ones in its rowspan.

Let $A = [A_1\ A_1\ \ \dots\ \ A_n]$ and $\theta^{A_j} := \theta_1^{a_{1j}} \dots \theta_d^{a_{dj}}$.

The log-linear model $\mathcal{M}_A$ is parametrized as

$$\theta \mapsto (\theta^{A_1}, \theta^{A_2}, \dots, \theta^{A_n}).$$

The implicit description of this model is given as

$$I_A = \langle p^u - p^v : u - v \in \ker_{\mathbb{Z}}(A)\rangle.$$

Example:

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

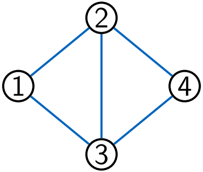
Parametrization: $(\theta_1, \theta_2) \mapsto (\theta_1^2, \quad \theta_1\theta_2, \quad \theta_2^2)$.

Ideal: $I_A = \langle p_1 p_3 - p_2^2 \rangle$.

A-matrix

Diamond graph, binary variables.

$$
\begin{array}{cc}
 & \begin{array}{cccccccccccccccc} 0000 & 0001 & 0010 & 0011 & 0100 & 0101 & 0110 & 0111 & 1000 & 1001 & 1010 & 1011 & 1100 & 1101 & 1110 & 1111 \end{array} \\
\begin{array}{c} 000\bullet \\ 001\bullet \\ 010\bullet \\ 011\bullet \\ 100\bullet \\ 101\bullet \\ 110\bullet \\ 111\bullet \\ \hline \bullet 000 \\ \bullet 001 \\ \bullet 010 \\ \bullet 011 \\ \bullet 100 \\ \bullet 101 \\ \bullet 110 \\ \bullet 111 \end{array} &
\left[\begin{array}{cccccccccccccccc}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]
\end{array}
$$



Mixture Models

We use the rth **mixture** to model a situation where the population is split into r subpopulations.

$$\mathrm{Mixt}^r(\mathcal{M}) = \{\pi_1 \boldsymbol{p}^1 + \ldots + \pi_r \boldsymbol{p}^r : \pi \in \Delta_r, \boldsymbol{p}^i \in \mathcal{M} \text{ for all } i \in [r]\}$$

Secant varieties: Given a variety W

$$\mathrm{Sec}^r(W) := \overline{\{\alpha_1 w^1 + \ldots + \alpha_r w^r : \sum \alpha_i = 1 \text{ and } w^i \in W \text{ for all } i \in [r]\}}$$

Parameterization of $\mathrm{Mixt}^r(\mathcal{M}_G)$:

$$p_i = \frac{1}{Z(\theta)} \sum_{j=1}^{r} \prod_{C \in \mathcal{C}} \theta_{i_C}^{(j,C)}$$

Mixture Models

We use the rth **mixture** to model a situation where the population is split into r subpopulations.

$$\mathrm{Mixt}^r(\mathcal{M}) = \{\pi_1 \boldsymbol{p}^1 + \ldots + \pi_r \boldsymbol{p}^r : \pi \in \Delta_r, \boldsymbol{p}^i \in \mathcal{M} \text{ for all } i \in [r]\}$$

Secant varieties: Given a variety W

$$\mathrm{Sec}^r(W) := \overline{\{\alpha_1 w^1 + \ldots + \alpha_r w^r : \sum \alpha_i = 1 \text{ and } w^i \in W \text{ for all } i \in [r]\}}$$

Questions: Dimension? Ideal $I_G^{(r)}$?

Expected dimension: $\min\{r \dim(\mathcal{M}_G) + (r-1), \quad \prod_{i \in V(G)} d_i - 1\}$.

Mixtures of the independence model

Independence model

$=$ graphical model with empty graph,

$=$ intersection of the probability simplex with the set of tensors of nonnegative rank at most 1.

Ideal of mixtures:

$r = 2$: Generated by all 3×3 minors of all flattenings. [Allman et al., 2015].

$r \geq 3$: Minors are not enough ("Salmon conjecture").

Dimension of mixtures:

- When the tensors are matrices, these are always defective.
- The dimension of the set of all rank r $m \times n$ matrices is $r(m+n-r) < r(m+n-1) + (r-1)$ when $r > 1$.
- Otherwise, "usually" of expected dimension, for details see [Landsberg, 2015, Section 5.5].

Sub-Ideals via Conditional Independence

Notation:

- For $S \subset V$, let $\mathcal{R}_S := \prod_{v \in S} [d_v]$ be the state space restricted to S
- For $i_S \in \mathcal{R}_S$, define the marginal

$$p_{i_S +} := \sum_{j \in \mathcal{R}_{V-S}} p_{i_S j}$$

- $I^{(r)}_{j_C; A \perp\!\!\!\perp B}$ = ideal of $(r+1) \times (r+1)$ minors of the matrix whose rows/columns are indexed by i_A / i_B and whose (i_A, i_B) entry is $p_{i_A i_B j_C +}$

Proposition (A.-Coons-Sturma, 2024)

Let $A, B, C \subset V$ be disjoint sets such that C separates A and B in G. Then for each $j_C \in \mathcal{R}_C$, $I^{(r)}_G$ contains $I^{(r)}_{j_C; A \perp\!\!\!\perp B}$.

$$\rightarrow \begin{bmatrix} p_{111} & p_{112} \\ p_{211} & p_{212} \end{bmatrix} \text{ and } \begin{bmatrix} p_{121} & p_{122} \\ p_{221} & p_{222} \end{bmatrix} \rightarrow \begin{matrix} p_{111}p_{212} - p_{112}p_{211} \\ p_{121}p_{222} - p_{122}p_{221} \end{matrix}.$$

Ideals

Question: Is $I_G^{(r)}$ the sum of these sub-ideals?

Ideals

Question: Is $I_G^{(r)}$ the sum of these sub-ideals? No!

Example (Second Mixture of the Binary 5-path)

By the proposition, the ideal $I_G^{(2)}$ contains 32 minimal cubic generators. However it also has 57 minimal quartic generators of the form:

$$\begin{aligned}
&p_{11222}p_{21112}p_{22121}p_{22211} - p_{11112}p_{21222}p_{22121}p_{22211} - p_{11221}p_{21112}p_{22122}p_{22211} + p_{11112}p_{21221}p_{22122}p_{22211} - \\
&p_{11222}p_{21111}p_{22121}p_{22212} + p_{11111}p_{21222}p_{22121}p_{22212} + p_{11221}p_{21111}p_{22122}p_{22212} - p_{11111}p_{21221}p_{22122}p_{22212} - \\
&p_{11212}p_{21122}p_{22111}p_{22221} + p_{11122}p_{21212}p_{22111}p_{22221} + p_{11211}p_{21122}p_{22112}p_{22221} - p_{11122}p_{21211}p_{22112}p_{22221} + \\
&p_{11212}p_{21121}p_{22111}p_{22222} - p_{11121}p_{21212}p_{22111}p_{22222} - p_{11211}p_{21121}p_{22112}p_{22222} + p_{11121}p_{21211}p_{22112}p_{22222}.
\end{aligned}$$

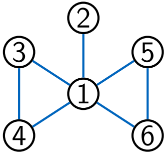
Shout-out: `MultigradedImplicitization.m2` by Joe Cummings and Ben Hollering

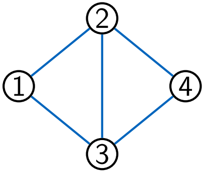
Clique-Stars

Definition

A graph G is a *clique star* if it is a union of cliques, $G = \cup_{i=1}^{k} \widetilde{C}_i$, and there is another clique S such that $\widetilde{C}_i \cap \widetilde{C}_j = S$ for all $i \neq j$.
Moreover, we write $C_i = \widetilde{C}_i \setminus S$.

Examples:

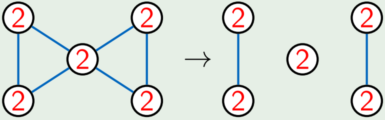




Clique-Stars: Ideal

Notation: $I^{(r)}_{j_S, d_{C_1} \times \cdots \times d_{C_k}}$ denotes the vanishing ideal of the rth mixture of the k-way independence model with the states $\prod_{i \in C} d_i$ for each clique C, with the fixed value $X_S = j_S \in \mathcal{R}_S$.

Example



$$\to \begin{bmatrix} p_{11111} & p_{12111} & p_{21111} & p_{22111} \\ p_{11112} & p_{12112} & p_{21112} & p_{22112} \\ p_{11121} & p_{12121} & p_{21121} & p_{22121} \\ p_{11122} & p_{12122} & p_{21122} & p_{22122} \end{bmatrix}$$

Theorem (A.-Coons-Sturma, 2024)

Let $G = (C_1 \cup \cdots \cup C_k \cup S, E)$ be a clique-star. Then

$$I^{(r)}_G = \sum_{j_S \in \mathcal{R}_S} I^{(r)}_{j_S, d_{C_1} \times \cdots \times d_{C_k}}.$$

Clique-Stars: Dimension

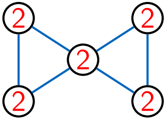
Theorem (A.-Coons-Sturma, 2024)

Let $G = (C_1 \cup \cdots \cup C_k \cup S, E)$ be a clique-star. Then

$$\dim(Sec^r(\overline{\mathcal{M}_G})) = \min\left\{ d_S \cdot \dim(\overline{\mathcal{T}^r_{d_{C_1}\times\cdots\times d_{C_k}}}) - 1, \prod_{v\in V} d_v - 1 \right\},$$

where $\mathcal{T}^r_{d_{C_1}\times\cdots\times d_{C_k}}$ is the set of $d_{C_1}\times\cdots\times d_{C_k}$ tensors of nonnegative rank at most r.

Example:



If $r = 2$ and all variables are binary, then

$$\dim(\mathrm{Sec}^2(\overline{\mathcal{M}_G})) = \min\{2\cdot 2\cdot(4+4-2)-1, 31\} = 23.$$

Expected dimension is 27 (similar for 3-path).

Proof: Restructure Jacobian of parametrization s.t. it is block-diagonal.

Dimensions

Let P_n denote the path with n vertices. We have seen that the secants of $\mathcal{M}_{P_3}$ are defective.

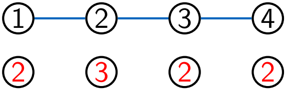
Question: Are the secants of $\mathcal{M}_{P_n}$ defective for $n > 3$?

Dimensions

Let P_n denote the path with n vertices. We have seen that the secants of $\mathcal{M}_{P_3}$ are defective.

Question: Are the secants of $\mathcal{M}_{P_n}$ defective for $n > 3$? **No!**

Surprising Example



The dimension of the toric model $\mathcal{M}_{P_4}$ with $d_1 = d_3 = d_4 = 2$ and $d_2 = 3$ is 10. Its second secant has dimension

$$21 = 2 \times 10 + (2 - 1),$$

which is the expected dimension.

Dimensions of Second Secants for Decomposable Graphs

Theorem (A.-Coons-Sturma 2024)

Let G be a decomposable graph that is not a clique star with $d_v \geq 2$ for all $v \in V$. Then

$$\dim(\mathrm{Mixt}^2 \mathcal{M}_G) = 2\dim(\mathcal{M}_G) + 1.$$

In particular, the secant variety has the expected dimension.

Why do we care?

- This means the parameters are "as identifiable as possible"
- In other words, they can be identified to the same extent as they can be for the log-linear model

Proof Strategy: Slicing Point Configurations

Theorem (Theorem 2.3, Draisma 2008)

- *Let* V_A *be the toric variety specified by integer matrix* $A \in \mathbb{Z}^{d \times n}$.
- *Let* $\mathbf{v} \in (\mathbb{R}^d)^*$.
- *Let* A_+ *denote the columns of* A *such that* $\mathbf{v} \cdot \mathbf{a} > 0$.
- *Similarly,* A_- *consists of the columns of* A *such that* $\mathbf{v} \cdot \mathbf{a} < 0$.

Then

$$\dim\big(\mathrm{Sec}^2(V_A)\big) \geq \mathrm{rank}(A_+) + \mathrm{rank}(A_-) - 1.$$

In particular, if we can separate the vertices of conv(A) with a hyperplane so that the columns on either side have full rank, then the secant has the expected dimension.

Proof Strategy

Graphs with three maximal cliques:

- we show that we can extend a hyperplane normal $\mathbf{v}$ for G to $\mathbf{v}'$ for G' when G' is obtained by:
  - adding a vertex without changing the clique structure, or
  - increasing d_v by 1 for some vertex v.
- Any such graph can be obtained from P_4 or $P_3 \sqcup P_1$ by a sequence of these operations, so we find hyperplanes for these two graphs.

Graphs with more than three maximal cliques:

- Find a separating hyperplane for a subgraph with three cliques
- Show that extending by zeros on the rest of the graph gives a separating hyperplane for all of G

Future Work

Conjecture (A.-Coons-Sturma, 2024)

If G is **any** *graph that is not a clique star with $d_v \geq 2$ for all $v \in V$, then its second mixture has the expected dimension.*

Question

Draisma's theorem can also be applied when

- we take r-mixtures for arbitrary r and/or
- we take mixtures of several different graphs (join varieties).

What happens then?

Question

Dimensions of mixtures of your favorite log-linear model?

Acknowledgements

Thanks to:

- Applied Algebra Seminar organizers,
- IMSI for supporting this research, and
- all of you for listening!

Part of this research was performed while the authors were visiting the Institute for Mathematical and Statistical Innovation (IMSI), which is supported by the National Science Foundation (Grant No. DMS-1929348).

References

E. Allman, J.A. Rhodes, B. Sturmfels and P. Zwiernik. "Tensors of non-negative rank two". *Linear Algebra and its Applications 473* (2015), Pages 37-53.

Y. Alexandr, J.I. Coons and N. Sturma. "Mixtures of discrete decomposable graphical models". To appear in *Algebraic Statistics* (2024).

J. Draisma. "A tropical approach to secant dimensions". *Journal of Pure and Applied Algebra 212*, 2 (2008). Pages 349-363.

J.M. Landsberg. *Tensors: Geometry and Applications*. Graduate Studies in Mathematics, 128 (2011).

S. Sullivant. *Algebraic Statistics*. Graduate Studies in Mathematics, 194 (2018).