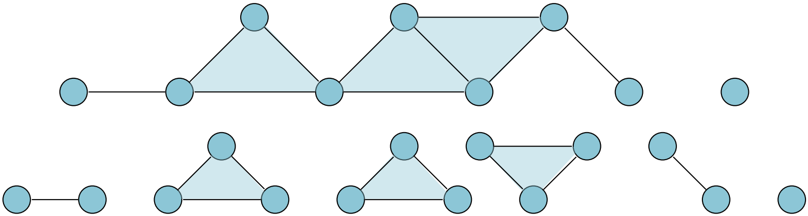
The Geometry of Discrete Decomposable Models

Danai Deligeorgaki



Introduction

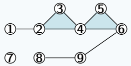
We study discrete decomposable models, a family of statistical models that lie in the class of **hierarchical models**. Decomposable models and their corresponding graphs are of wide use throughout **statistics** and data science. For instance, directed acyclic graphs (**DAGs**) can be approximated by decomposable graphs. The complexity of this approximation determines the complexity of probabilistic inference algorithms for DAG models such as **variable elimination**. Therefore, the combinatorics of the graphs defining the decomposable models carry important information in regard to **probabilistic inference**. The goal of this project is to explore at a deeper level the information encoded in combinatorial objects associated to **decomposable models**.

Definition

A **decomposable simplicial complex** Γ is a collection of simplices, i.e. nodes, edges, triangles, tetrahedra, etc., that are glued together (in a certain way). The simplices in Γ are called **faces** and the (non-trivial) inclusion-maximal faces are called **facets**.

For example, the graph on the right denotes a decomposable simplicial complex on 9 nodes. The edge {1,2}, the triangle {2,3,4} and the node {7} are some of its facets.

decomposable simplicial complex



Discrete decomposable models

Let $r_1, \ldots, r_m \in \mathbb{N}$ be the number outcomes of the discrete variables $X_1, X_2, \ldots, X_m$, respectively, and let $\mathcal{R} = r_1 \times \cdots \times r_m$ be the set of all possible outcomes. The joint distribution of $X_1, \ldots, X_m$ lies in the $(\#\mathcal{R} - 1)$-dimensional **probability simplex**

$$\Delta_{\#\mathcal{R}-1} = \{p \in \mathbb{R}^{\#\mathcal{R}} : p_i \geq 0, \text{ for all } i \in \mathcal{R} \text{ and } \sum_{i \in \mathcal{R}} p_i = 1\}.$$

The **decomposable model** associated with a decomposable simplicial complex Γ is

$$M_\Gamma = \{p \in \Delta_{\#\mathcal{R}-1} : p_i = \frac{1}{Z(\theta)} \prod_{F \in \text{facet}(\Gamma)} \theta_{i_F}^{(F)} \text{ for all } i \in \mathcal{R}\},$$

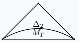
for $\theta_{i_F}^{(F)}$ positive parameters and $Z(\theta)$ normalizing constant.

From the model to the polytope

Apart from the graph Γ, there are other combinatorial objects linked to a decomposable model M_Γ. In fact, M_Γ can be written as the intersection of a toric variety V_{M_Γ} with the probability simplex $\Delta_{\#\mathcal{R}-1}$.

For example, for $\#\mathcal{R} = 3$,

$$M_\Gamma = V_{M_\Gamma} \cap \Delta_2$$



From the toric variety, which is an algebro-geometric object, we can pass to a polytope P_{M_Γ}, a geometric object. It is a property of toric varieties that the geometric properties of V_{M_Γ} are encoded in the **polytope** P_{M_Γ}.

In this project, we are investigating the structure of this polytope to see if it carries useful information in relation to probabilistic inference.

Getting to know the polytope

When investigating a polytope's combinatorics, there are several questions to be explored, such as

- What are the facets of the polytope P_{M_Γ}? Answered in [1].

The facets are given by $x_{i_F}^F \geq 0$ for $F \in$ facets(Γ) and $i_F \in \mathcal{R}_F$.

- Does P_{M_Γ} admit a regular unimodular triangulation? Answered in [2].

Yes!

What combinatorial information does this triangulation carry? Open.

- What are the enumerative properties of P_{M_Γ}? Our results.

To this end, we study the structure of an integer polynomial associated to P_{M_Γ}, called the $\boldsymbol{h^*}$**-polynomial**,

$$h^*(x) = h_0^* + h_1^* x + \cdots + h_{\#\mathcal{R}-1}^* x^{\#\mathcal{R}-1}.$$

This polynomial captures important information about the polytope, including its volume and whether or not the polytope P_{M_Γ}, and hence the model M_Γ, has the Gorenstein property. In fact, the decomposable model M_Γ is **Gorenstein** if and only if $h^*(x)$ is palindromic.

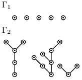
We characterize all Gorenstein discrete decomposable binary models for forests Γ.

Theorem 1

Let Γ be a forest on m nodes and $X_1, \ldots, X_m$ be binary variables.
Then M_Γ is Gorenstein if and only if all connected components of Γ have

- exactly one vertex,
- strictly more than one vertex.

Gorenstein examples



Future work

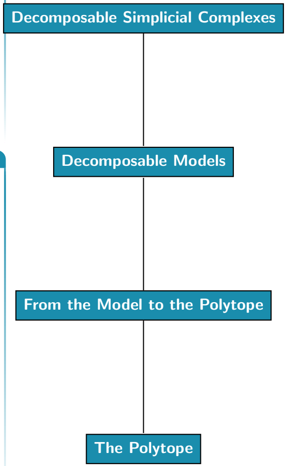
We will continue exploring the combinatorial properties of discrete decomposable models, and their interpretation in terms of statistics. Our current goals are to

1. Interpret the observations in Theorem 1 statistically.
2. Generalize Theorem 1 to characterize all discrete decomposable models. We already have a conjecture in this direction.
3. Analyze the information that the triangulation constructed in [2] carries.

References

[1] Markov bases of binary graph models
M. Develin, S. Sullivant
Annals of Combinatorics 7, 2003

[2] Gröbner bases and polyhedral geometry of reducible and cyclic models
S. Hoşten, S. Sullivant
Journal of Combinatorial Theory, Series A 100.2, 2002



Outline: Discrete Decomposable Models

Discrete Decomposable Models are defined through some fine geometric/combinatorial objects called *simplices*. They provide a way to do probabilistic inference, i.e. to compute the probability $P(X_{i_1}, X_{i_2}, ..., X_{i_k})$, given $P(X_1, X_2, ..., X_m)$ for discrete variables $X_1, X_2, ..., X_m$.

First, we will discuss

- ◇ simplices,
- ◇ simplicial complexes,
- ◇ decomposable simplicial complexes.

Simplex

A (geometric) **simplex** of dimension d can be thought of as a d-dimensional generalisation of a triangle: $d+1$ vertices, all connected to each other, forming a d-dimensional object.

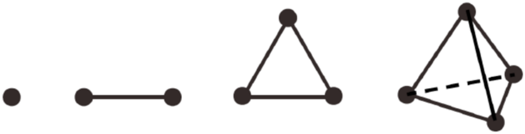
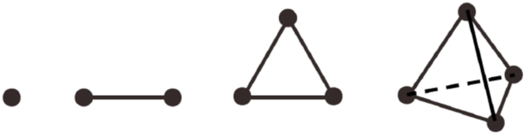


Figure appearing in 'Navigation and Stabilization of Swarms of Micro Aerial Vehicles in Complex Environment' by Petr Vsetecka and Martin Saska (2015)

Simplex

A (geometric) **simplex** of dimension d can be thought of as a d-dimensional generalisation of a triangle: $d+1$ vertices, all connected to each other, forming a d-dimensional object.

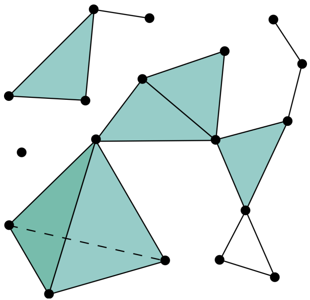


A simplex contains simplices of smaller dimension. We call
faces of a simplex: all simplices it contains and the empty set,
facets of a simplex of dimension d: all faces of dimension $d-1$.

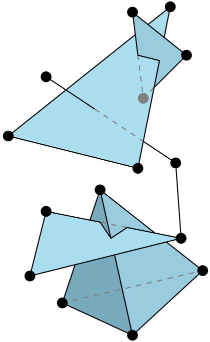
Simplicial Complex

A **simplicial complex** is a collection of simplices that may be glued together.
The simplices glued together should intersect in smaller simplices.

Example



Non-example

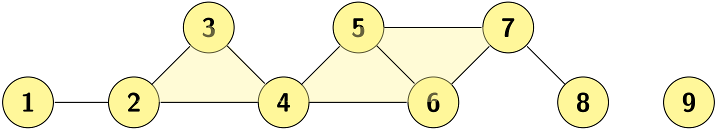


Figures taken from Wikipedia

Simplicial Complex

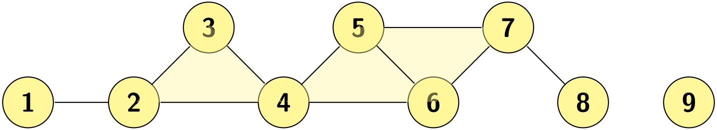
Given a simplicial complex Γ we call
faces of Γ: all simplices it contains and the empty set,
facets of Γ: all inclusion-wise maximal simplices it contains.

Example



Decomposable Simplicial Complexes

We are interested in simplicial complexes with even finer combinatorial/geometric properties.
Such examples are **decomposable simplicial complexes**.

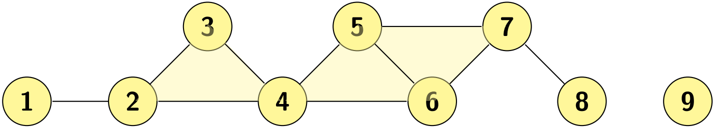


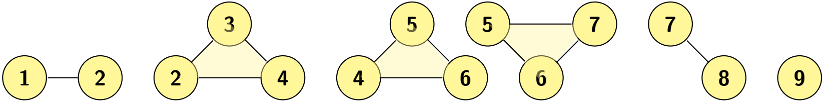
Decomposable Simplicial Complexes

We are interested in simplicial complexes with even finer combinatorial/geometric properties.

Such examples are **decomposable simplicial complexes**.

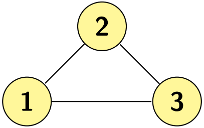
I think of decomposable simplicial complexes as simplicial complexes that intersect at a simplex and you can 'break' into two repeatedly and end up with the facets of the simplicial complex (see below).





Decomposable Simplicial Complexes

We are interested in simplicial complexes with even finer combinatorial/geometric properties.
Such examples are **decomposable simplicial complexes**.
I think of decomposable simplicial complexes as simplicial complexes that intersect at a simplex and you can 'break' into two repeatedly and end up with the facets of the simplicial complex (see below).
Counterexample:



⋄ We will see how to construct statistical models starting with a decomposable simplicial complex.

⋄ These models have nice combinatorial properties that allow us to apply probabilistic efficient inference techniques such as variable elimination.

Probability simplex

Let $X_1, \ldots, X_m$ be discrete random variables.
Let $r_1, \ldots, r_m \in \mathbb{N}$ be the number outcomes of $X_1, X_2, \ldots, X_m$, and let $\mathcal{R} = r_1 \times \cdots \times r_m$ be the set of all possible outcomes.

The joint distribution of $X_1, \ldots, X_m$ lies in the $(\#\mathcal{R} - 1)$-dimensional **probability simplex**

$$\Delta_{\#\mathcal{R}-1} = \{p \in \mathbb{R}^{\#\mathcal{R}} : p_i \geq 0, \text{ for all } i \in \mathcal{R} \text{ and } \sum_{i \in \mathcal{R}} p_i = 1\}.$$

for some choice of $r_1, \ldots, r_m$.

Probability simplex

Let $X_1, ..., X_m$ be discrete random variables.
Let $r_1, ..., r_m \in \mathbb{N}$ be the number outcomes of $X_1, X_2, ..., X_m$, and let $\mathcal{R} = r_1 \times \cdots \times r_m$ be the set of all possible outcomes.

The joint distribution of $X_1, ..., X_m$ lies in the $(\#\mathcal{R} - 1)$-dimensional **probability simplex**

$$\Delta_{\#\mathcal{R}-1} = \{p \in \mathbb{R}^{\#\mathcal{R}} : p_i \geq 0, \text{ for all } i \in \mathcal{R} \text{ and } \sum_{i \in \mathcal{R}} p_i = 1\}.$$

A (discrete) **statistical model** M is a subset of $\Delta_{\#\mathcal{R}-1}$ for some choice of $r_1, \ldots, r_m$.

Example: Independence Model

Consider variables X_1, X_2 such that $r_1 = r_2 = 2$ (binary).
There are 4 possible outcomes: $\{00, 01, 10, 11\}$.

Their joint probability distribution p satisfies

$$p_{00}, p_{01}, p_{10}, p_{11} \geq 0$$

and

$$p_{00} + p_{01} + p_{10} + p_{11} = 1.$$

We can think of the points $(p_{00}, p_{01}, p_{10}, p_{11})$ as a 3-dimensional probability simplex.

Example: Independence Model

Consider variables X_1, X_2 such that $r_1 = r_2 = 2$ (binary).
There are 4 possible outcomes: $\{00, 01, 10, 11\}$.

Their joint probability distribution p satisfies

$$p_{00}, p_{01}, p_{10}, p_{11} \geq 0$$

and

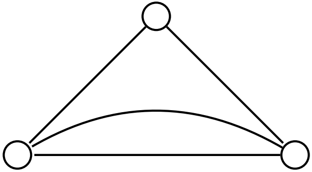
$$p_{00} + p_{01} + p_{10} + p_{11} = 1.$$

We can think of the points $(p_{00}, p_{01}, p_{10}, p_{11})$ as a 3-dimensional probability simplex.

A statistical model is a part of this simplex. It represents a set of candidates for the unknown distribution p.

We will talk about a family of statistical models that can be described in a nice geometric way:

as intersections of simplices with toric varieties.



Discrete Decomposable Models

Let $r_1, \ldots, r_m \in N$ be the number of outcomes of the discrete variables $X_1, X_2, \ldots, X_m$, respectively, and let $\mathcal{R} = r_1 \times \cdots \times r_m$ be the set of all possible outcomes.

Let Γ be a decomposable simplicial complex with m vertices.

The **decomposable model** M_Γ associated with Γ is

$$M_\Gamma = \{p \in \Delta_{\#\mathcal{R}-1} : p_i = \frac{1}{Z(\theta)} \prod_{F \in \text{facet}(\Gamma)} \theta_{i_F}^{(F)} \text{ for all } i \in \mathcal{R}\},$$

for $\theta_{i_F}^{(F)}$ positive parameters and $Z(\theta)$ normalizing constant.

Discrete Decomposable Models

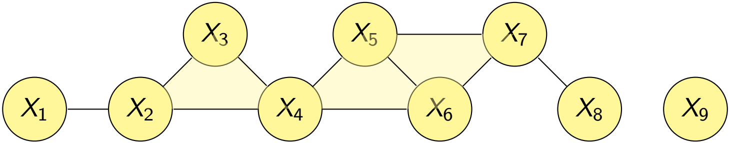
Let $r_1, \ldots, r_m \in N$ be the number of outcomes of the discrete variables $X_1, X_2, \ldots, X_m$, respectively, and let $\mathcal{R} = r_1 \times \cdots \times r_m$ be the set of all possible outcomes.

Let Γ be a decomposable simplicial complex with m vertices.

The **decomposable model** M_Γ associated with Γ is

$$M_\Gamma = \{p \in \Delta_{\#\mathcal{R}-1} : p_i = \frac{1}{Z(\theta)} \prod_{F \in \text{facet}(\Gamma)} \theta_{i_F}^{(F)} \text{ for all } i \in \mathcal{R}\},$$

for $\theta_{i_F}^{(F)}$ positive parameters and $Z(\theta)$ normalizing constant.



Example: Independence Model

Consider variables X_1, X_2 such that $r_1 = r_2 = 2$ (binary).
Let Γ be a simplicial complex consisting of 2 vertices.

Example: Independence Model

Consider variables X_1, X_2 such that $r_1 = r_2 = 2$ (binary).
Let Γ be a simplicial complex consisting of 2 vertices.

Γ

The decomposable model M_Γ associated with Γ is

$$M_\Gamma = \Big\{ p = (p_{00}, p_{01}, p_{10}, p_{11}) \in \Delta_3 :$$

$$p_{i_1 i_2} = \frac{\theta^{(1)}_{i_1}\theta^{(2)}_{i_2}}{\theta^{(1)}_0\theta^{(2)}_0 + \theta^{(1)}_0\theta^{(2)}_1 + \theta^{(1)}_1\theta^{(2)}_0 + \theta^{(1)}_1\theta^{(2)}_1},\ i_1 i_2 \in \{00, 01, 10, 11\} \Big\},$$

where $\theta^{(1)}_0, \theta^{(1)}_1, \theta^{(2)}_0, \theta^{(2)}_1$ are positive parameters.

Example: Independence Model

Consider variables X_1, X_2 such that $r_1 = r_2 = 2$ (binary).
Let Γ be a simplicial complex consisting of 2 vertices.

Γ

The decomposable model M_Γ associated with Γ is

$$M_\Gamma = \Big\{ p = (p_{00}, p_{01}, p_{10}, p_{11}) \in \Delta_3 :$$

$$p_{i_1 i_2} = \frac{\theta^{(1)}_{i_1} \theta^{(2)}_{i_2}}{\theta^{(1)}_0 \theta^{(2)}_0 + \theta^{(1)}_0 \theta^{(2)}_1 + \theta^{(1)}_1 \theta^{(2)}_0 + \theta^{(1)}_1 \theta^{(2)}_1}, \; i_1 i_2 \in \{00, 01, 10, 11\} \Big\},$$

where $\theta^{(1)}_0, \theta^{(1)}_1, \theta^{(2)}_0, \theta^{(2)}_1$ are positive parameters.

It turns out that M_Γ contains all positive distributions for which X_1, X_2 are **independent**.

Outline: From the Model to the Polytope

We discussed how graphs are associated to discrete decomposable models. We will briefly introduce other objects that relate to decomposable models:

- ◇ varieties
- ◇ polytopes

Varieties

- ◇ Algebraic varieties are algebro-geometric objects.
- ◇ You can think of an **algebraic variety** as the set of solutions of a system of polynomial equations over the real (or complex) numbers.
- ◇ For example, the **unit circle** is the set of real pairs (x, y) such that $x^2 + y^2 - 1 = 0$.

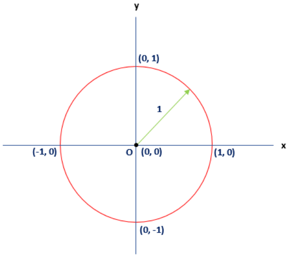
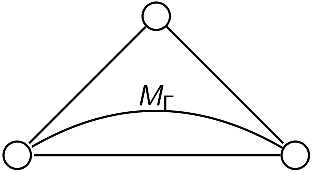


Figure taken from 'javaTpoint'

From the Model to the Variety

Every decomposable model M_Γ can be written as the intersection of a (toric) variety V_{M_Γ} with the probability simplex $\Delta_{\#\mathcal{R}-1}$,

$$M_\Gamma = V_{M_\Gamma} \cap \Delta_{\#\mathcal{R}-1}.$$



Example: Independence Model

◇ Let X_1, X_2 be two binary variables.
Let M_Γ be the model containing all possible joint distributions $p = (p_{00}, p_{01}, p_{10}, p_{11})$ for which X_1, X_2 are **independent**.

Example: Independence Model

◇ Let X_1, X_2 be two binary variables.
Let M_Γ be the model containing all possible joint distributions $p = (p_{00}, p_{01}, p_{10}, p_{11})$ for which X_1, X_2 are **independent**.

◇ As already discussed, all distributions p lie in Δ_3, i.e. they satisfy $p_{00}, p_{01}, p_{10}, p_{11} \geq 0$ and $p_{00} + p_{01} + p_{10} + p_{11} = 1$.

Example: Independence Model

$\diamond$ Let X_1, X_2 be two binary variables.
Let M_Γ be the model containing all possible joint distributions $p = (p_{00}, p_{01}, p_{10}, p_{11})$ for which X_1, X_2 are **independent**.

$\diamond$ As already discussed, all distributions p lie in Δ_3, i.e. they satisfy $p_{00}, p_{01}, p_{10}, p_{11} \geq 0$ and $p_{00} + p_{01} + p_{10} + p_{11} = 1$.

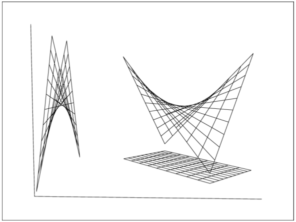
$\diamond$ It turns out that the distributions p also satisfy

$$p_{00}p_{11} - p_{01}p_{10} = 0$$

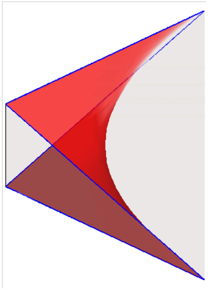
and these relations define the model.

Example: Independence Model

The solutions to the polynomial $p_{00}p_{11} - p_{01}p_{10} = 0$ define a (toric) variety V_{M_Γ} (see 3-dimensional surface on the right).



The variety V_{M_Γ} intersects with the simplex Δ_3 to give the independence model M_Γ (see the 3-dimensional surface on the right).



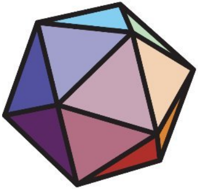
Figures taken from 'Real equivalence of complex matrix pencils and complex projections of real Segre varieties' by Adam Coffman, and 'Mixture decompositions of exponential families using a decomposition of their sample spaces' by Guido Montufar

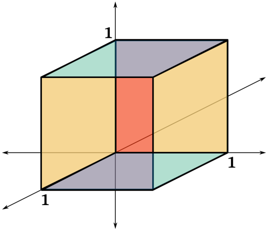
◇ We saw that a decomposable model is described in terms of a variety. Studying this variety can reveal valuable information about the model.

◇ We saw that a decomposable model is described in terms of a variety. Studying this variety can reveal valuable information about the model.

◇ The last step for today is to pass from a toric variety to a polytope. The polytopes corresponding to toric varieties capture their geometry!

Polytopes

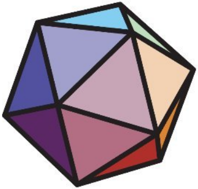


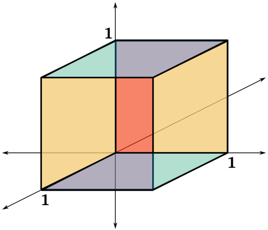


◇ A polytope is a bounded convex geometric object with "flat" sides.

Figures taken from Polytope@PolytopeSpace in Twitter, and Wikipedia

Polytopes

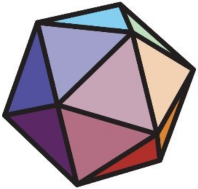


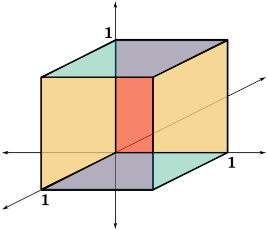


◇ A polytope is a bounded convex geometric object with "flat" sides.

◇ A way to construct a polytope is to start with some vertices in $\mathbb{R}^n$ and consider all points between them.

Polytopes





◇ A polytope is a bounded convex geometric object with "flat" sides.

◇ A way to construct a polytope is to start with some vertices in $\mathbb{R}^n$ and consider all points between them.

◇ For example, starting with the vertices $(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)$ in $\mathbb{R}^3$ we get the unit cube.

Polytopes

- **All simplices are polytopes.**

Polytopes

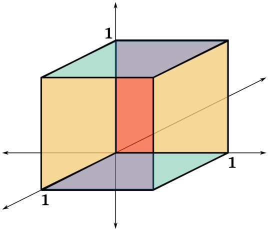
⋄ **All simplices are polytopes.**

⋄ The notions of face and facet generalise to all polytopes.
A polytope "contains" polytopes of smaller dimension. We call
faces of a polytope: all polytopes it contains and the empty set,
facets of a polytope of dimension d: all faces of dimension $d - 1$.

Polytopes

◇ **All simplices are polytopes.**

◇ The notions of face and facet generalise to all polytopes.
A polytope "contains" polytopes of smaller dimension. We call **faces** of a polytope: all polytopes it contains and the empty set, **facets** of a polytope of dimension d: all faces of dimension $d - 1$.



From the Variety to the Polytope

◇ There is a (toric) variety V_{M_Γ} associated to each discrete decomposable model M_Γ, and a polytope P_{M_Γ} associated to the variety.

From the Variety to the Polytope

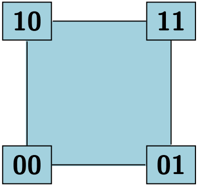
◇ There is a (toric) variety V_{M_Γ} associated to each discrete decomposable model M_Γ, and a polytope P_{M_Γ} associated to the variety.

◇ However, there is a way to construct the polytope directly from the model! Its vertices can be understood from the simplicial complex Γ and the number of outcomes of the variables that define it.

Example: Independence Model

Let X_1, X_2 be two binary variables.
Let M_Γ be the model containing all possible joint distributions $p = (p_{00}, p_{01}, p_{10}, p_{11})$ for which X_1, X_2 are **independent**.

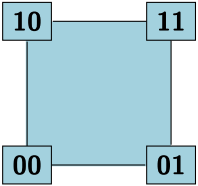
The polytope P_{M_Γ} that corresponds to M_Γ turns out to have vertices $00, 01, 10, 11$, similar to the outcomes of X_1, X_2.



Example: Independence Model

Let X_1, X_2 be two binary variables.
Let M_Γ be the model containing all possible joint distributions $p = (p_{00}, p_{01}, p_{10}, p_{11})$ for which X_1, X_2 are **independent**.

The polytope P_{M_Γ} that corresponds to M_Γ turns out to have vertices $00, 01, 10, 11$, similar to the outcomes of X_1, X_2.



We can subsequently study the structure/geometry of the square P_{M_Γ} to make conclusions about the model M_Γ.

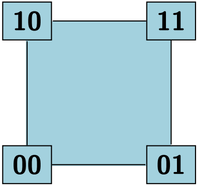
When investigating a polytope's combinatorics, there are several questions to be explored, such as

1) What are the facets of the polytope P_{M_Γ}? Answered

2) Does P_{M_Γ} admit a regular unimodular triangulation? Answered
What combinatorial information does this triangulation carry
Open.

3) What are the enumerative properties of P_{M_Γ} (such as the number of faces of each dimension and its volume)? First results.

P_{M_Γ}



Summary

Let X_1, X_2 be binary discrete variables.

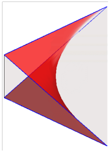
Starting with a decomposable simplicial complex Γ with m vertices

Γ

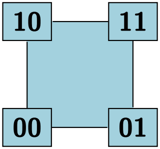
we can pass to a discrete decomposable model M_Γ (and vice versa)

$$\left\{ \text{positive part of the Independence Model} \right\}$$

The Independence model M'_Γ is linked to a (toric) variety V_{M_Γ} and a polytope P_{M_Γ}



V_{M_Γ}



P_{M_Γ}

Then we can investigate the structure of this polytope to see if it carries useful information in relation to probabilistic inference.

The end

Thanks!