Staged Tree Models with Toric Structure

Aida Maraj (University of Michigan)

Nonlinear Algebra and Statistics (NLASTATS)
Illinois Institute of Technology

based on joint work with
Christiane Görgen (MPI & U. Leipzig) and **Lisa Nicklasson** (U. of Genoa)

Görgen, M., Nicklasson. *Staged tree models with toric strucure*, J. of Symbolic Comput. (2022).

What is a toric ideal?

An ideal $I \subseteq \mathbb{R}[p_1, \ldots, p_n]$ is *toric* if one of the equivalent properties holds

I is a prime binomial ideal $\iff$ I is the kernel of a monomial map $\mathbb{R}[p_1, \ldots, p_n] \to \mathbb{R}[\theta_1^{\pm 1}, \ldots, \theta_m^{\pm 1}]$

$$\psi : \mathbb{R}[p_1, p_2, p_3, p_4] \to \mathbb{R}[\theta_1, \theta_2]$$

$$p_1 \mapsto \theta_1^3, \;\; p_2 \mapsto \theta_1^2\theta_2, \;\; p_3 \mapsto \theta_1\theta_2^2, \;\; p_4 \mapsto \theta_2^3$$

$$\ker \psi = \langle p_1 p_3 - p_2^2, \;\; p_2 p_4 - p_3^2, \;\; p_1 p_4 - p_2 p_3 \rangle$$

What is a staged tree model?

A **discrete probability distribution** is a point $p = (p_i) \in \mathbb{R}^n$ such that

- $p_1 + \cdots + p_n = 1$
- $0 < p_1, \ldots, p_n < 1$

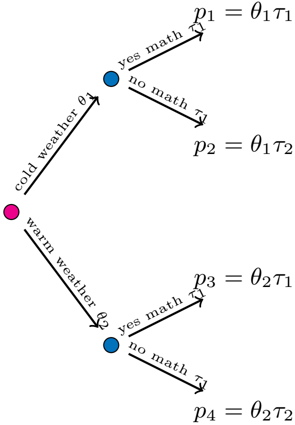
The collection
$\Delta_{n-1} = \{p \in \mathbb{R}^n \mid p_1 + \cdots + p_n = 1,\; p_1, \ldots, p_n \in (0,1)\}$ is the **probability simplex**.

Discrete statistical models are subsets $\mathcal{M}$ of Δ_{n-1}.

Staged Tree Models are parametrized discrete statistical models. They study the conditional dependency relations among events.

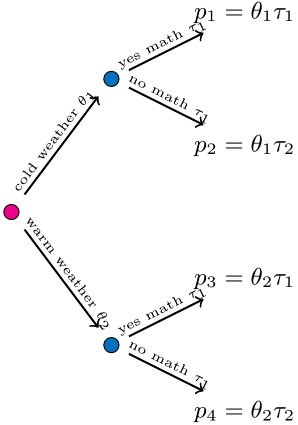
What is a staged tree model?

Staged trees are *rooted directed labeled trees with coloured vertices* that follow certain rules.



What is a staged tree model?

Staged trees are *rooted directed labeled trees with coloured vertices* that follow certain rules.

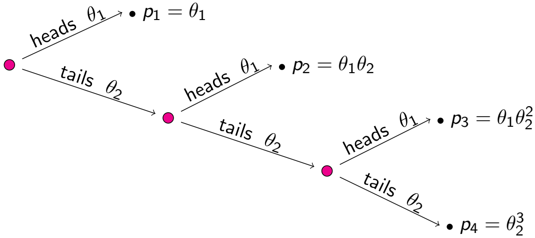


The **staged tree model** $\mathcal{M}_T$ is the collection of points $p = (p_i)$ that arrive from any possible labeling of the staged tree T.

$$\mathcal{M}_T = \{(p_1, p_2, p_3, p_4) = (\theta_1\tau_1, \theta_1\tau_2, \theta_2\tau_1, \theta_2\tau_2) \text{ such that }$$
$$\theta_1, \theta_2, \tau_1, \tau_2 \in (0,1), \theta_1 + \theta_2 = 1, \tau_1 + \tau_2 = 1\}$$

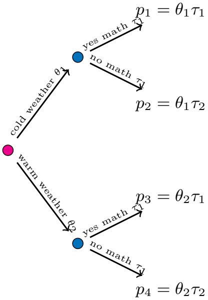
What is a staged tree model?

Staged trees are *rooted directed labeled trees with coloured vertices* that follow certain rules.



$$\mathcal{M}_T = \{(p_1, p_2, p_3, p_4) = (\theta_1, \theta_1\theta_2, \theta_1\theta_2^2, \theta_2^3) \mid \theta_1, \theta_2 \in (0,1) \text{ and } \theta_1 + \theta_2 = 1\}$$

The ideal of a staged tree model



$$\mathcal{M}_T = \{(p_1, p_2, p_3, p_4) = (\theta_1\tau_1, \theta_1\tau_2, \theta_2\tau_1, \theta_2\tau_2) \text{ such that } \theta_1, \theta_2, \tau_1, \tau_2 \in (0,1), \theta_1 + \theta_2 = 1, \tau_1 + \tau_2 = 1\}$$

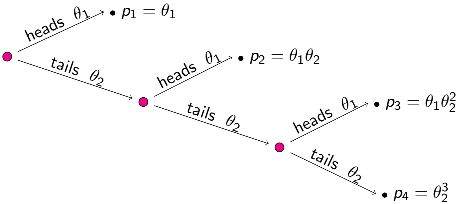
$$\varphi : \mathbb{R}[p_1, p_2, p_3, p_4] \to \mathbb{R}[\theta_1, \theta_2, \tau_1, \tau_2]/\langle \theta_1 + \theta_2 - 1, \tau_1 + \tau_2 - 1\rangle$$

$$p_1 \mapsto \theta_1\tau_1, \;\; p_2 \mapsto \theta_1\tau_2, \;\; p_3 \mapsto \theta_2\tau_1, \;\; p_4 \mapsto \theta_2\tau_2$$

$\ker\varphi = \langle \cancel{p_1 + p_2 + p_3 + p_4 - 1,}\; p_1p_4 - p_2p_3\rangle$ is the **ideal of the staged tree model**.

$$\mathcal{M}_T = V(\ker\varphi) \cap \Delta_3$$

The ideal of a staged tree model



$$\mathcal{M}_T = \{(p_1, p_2, p_3, p_4) = (\theta_1, \theta_1\theta_2, \theta_1\theta_2^2, \theta_2^3) \mid \theta_1, \theta_2 \in (0,1) \text{ and } \theta_1 + \theta_2 = 1\}$$

The parametrisation induces the map:

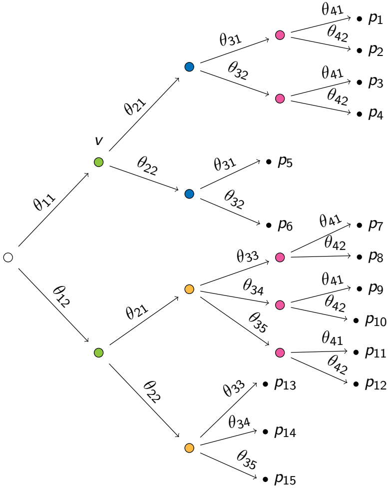
$$\varphi : \mathbb{R}[p_1, p_2, p_3, p_4] \to \mathbb{R}[\theta_1, \theta_2]/\langle \theta_1 + \theta_2 - 1 \rangle$$

$$p_1 \mapsto \theta_1, \;\; p_2 \mapsto \theta_1\theta_2, \;\; p_3 \mapsto \theta_1\theta_2^2, \;\; p_4 \mapsto \theta_2^3$$

The **ideal of the staged tree model** $\mathcal{M}_T$ is

$$\ker\varphi = \langle \cancel{p_1 + p_2 + p_3 + p_4 - 1,}$$
$$p_1(p_3 + p_4) - p_2(p_2 + p_3 + p_3), p_1p_4 - p_3(p_2 + p_3 + p_4), p_2p_4 - p_3(p_3 + p_4)\rangle$$

Balanced staged trees



For a vertex v, define

$$t(v) = \sum_{\lambda \in \{v\text{-to-leaf paths}\}} \text{prod. of the } \theta_{ij}\text{'s along } \lambda$$

as polynomial in $\mathbb{R}[\theta_{ij}]$. For example, in the figure

$$t(v) = \theta_{21}\theta_{31}\theta_{41} + \theta_{21}\theta_{31}\theta_{42} + \theta_{21}\theta_{32}\theta_{41} + \theta_{21}\theta_{32}\theta_{42} + \theta_{22}\theta_{31} + \theta_{22}\theta_{32}.$$

Let $v_1, \ldots, v_k$ denote the children of v.

A staged tree is **balanced** if for every pair of vertices v and w in the same stage

$$t(v_i)t(w_j) = t(v_j)t(w_i) \quad \text{for all } i, j.$$

$$\ker \varphi = \langle p_{12}p_{14} - p_{10}p_{15},\ p_{11}p_{14} - p_9p_{15},\ p_{12}p_{13} - p_8p_{15},\ p_{11}p_{13} - p_7p_{15},\ p_{10}p_{13} - p_8p_{14},\ p_9p_{13} - p_7p_{14},\ p_6p_{12} - p_4p_{15},\ p_5p_{12} - p_2p_{15}, \ldots$$

is a toric ideal generated by binomials of degree two! The staged tree is balanced.

Balanced staged trees

1. $\ker \varphi$ is toric in variables $p_1, \ldots, p_n$ if and only if the tree T is balanced.

E. Duarte and C. Görgen, *Equations defining probability tree models*, 2020.

2. The ideal $\ker \varphi$ of a balanced staged tree has a Gröbner basis with binomials of degree at most two. This ideal is Cohen-Macaulay, Kozul, and normal.

L. Ananiadi and E. Duarte, *Gröbner bases for staged trees*, 2021.

C. Görgen, AM, L. Nicklasson, *Staged tree models with toric structure, 2022.*

3. A Bayesian network is decomposable (a DAG is perfect) iff its staged tree presentation is balanced.

E. Duarte and L. Solus, *A new characterization of discrete decomposable models*, 2021.

Why toric models are wonderful?

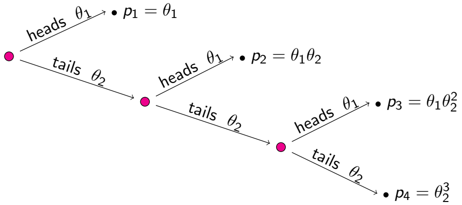
1. In a discrete toric model, the binomial generators of the toric ideal produce Markov bases, which contribute in hypothesis testing algorithms
   1. Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions." The Annals of statistics (1998).
   2. Sonja Petrović. *What is... a Markov basis?.* Notices of the American Mathematical Soc., 66(7), 2019.
   3. Seth Sullivant. *"Algebraic statistics".* American Mathematical Soc., (2018).

2. The polytope associated to a toric ideal is useful when studying the existence of maximum likelihood estimates
   1. Stephen Fienberg and Alessandro Rinaldo. *"Maximum likelihood estimation in log-linear models".* The Annals of Statistics, (2012).

3. The rich algebra, geometry, and combinatorics of a toric ideal facilitates computations on the maximum likelihood degree and estimate
   1. Steven Evans and Terrence Speed. *"Invariants of some probability models used in phylogenetic inference."* The Annals of Statistics (1993).
   2. Carlos Amendola, Dimitra Kosta, Kaie Kubjas. *"Maximum Likelihood Estimation of Toric Fano Varieties"* (2020): no.11.1, 15-30.
   3. Tobias Boege, Jane Ivy Coons, Chris Eur, Aida Maraj, Frank Rottger. "Reciprocal maximum likelihood degree of Brownian motion tree models" Le Matematiche 76.2 (2021)

Extending the class of Toric Staged Trees

Extending the class of Toric Staged Trees

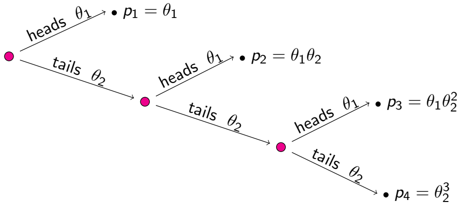


$$I_T = \langle p_1(p_3+p_4) - p_2(p_2+p_3+p_3),\ p_1p_4 - p_3(p_2+p_3+p_4),\ p_2p_4 - p_3(p_3+p_4)\rangle$$

The generators are the 2×2-minors of the matrix

$$M = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2+p_3+p_4 & p_3+p_4 & p_4 \end{bmatrix}$$

Extending the class of Toric Staged Trees



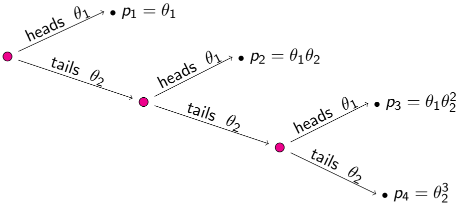
$$I_T = \langle p_1(p_3+p_4) - p_2(p_2+p_3+p_3),\ p_1p_4 - p_3(p_2+p_3+p_4),\ p_2p_4 - p_3(p_3+p_4)\rangle$$

The generators are the 2×2-minors of the matrix

$$M = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2+p_3+p_4 & p_3+p_4 & p_4 \end{bmatrix}$$

$$\xrightarrow{[\mathrm{R1}\to\mathrm{R1}+\mathrm{R2}]} \begin{bmatrix} \boldsymbol{p_1+p_2+p_3+p_4} & \boldsymbol{p_2+p_3+p_4} & \boldsymbol{p_3+p_4} \\ p_2+p_3+p_4 & p_3+p_4 & \boldsymbol{p_4} \end{bmatrix}$$

Extending the class of Toric Staged Trees

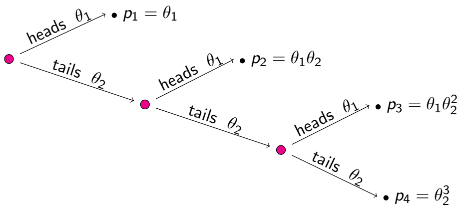


$$I_T = \langle p_1(p_3+p_4) - p_2(p_2+p_3+p_3),\ p_1p_4 - p_3(p_2+p_3+p_4),\ p_2p_4 - p_3(p_3+p_4)\rangle$$

The generators are the 2×2-minors of the matrix

$$\begin{aligned}
{\color{magenta}M} &= \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2+p_3+p_4 & p_3+p_4 & p_4 \end{bmatrix} \\
&\xrightarrow{[\text{R1}\to\text{R1}+\text{R2}]} \begin{bmatrix} \boldsymbol{p_1+p_2+p_3+p_4} & \boldsymbol{p_2+p_3+p_4} & \boldsymbol{p_3+p_4} \\ p_2+p_3+p_4 & p_3+p_4 & \boldsymbol{p_4} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{q_1} & \boldsymbol{q_2} & \boldsymbol{q_3} \\ q_2 & q_3 & \boldsymbol{q_4} \end{bmatrix}
\end{aligned}$$

Extending the class of Toric Staged Trees



$$I_T = \langle p_1(p_3+p_4) - p_2(p_2+p_3+p_3),\ p_1p_4 - p_3(p_2+p_3+p_4),\ p_2p_4 - p_3(p_3+p_4)\rangle$$

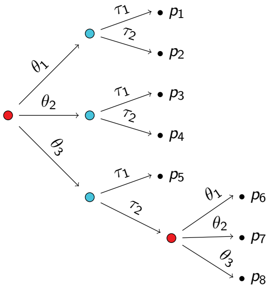
The generators are the 2×2-minors of the matrix

$$\begin{aligned} M &= \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2+p_3+p_4 & p_3+p_4 & p_4 \end{bmatrix} \\ &\xrightarrow{[\mathrm{R1}\to\mathrm{R1}+\mathrm{R2}]} \begin{bmatrix} \boldsymbol{p_1+p_2+p_3+p_4} & \boldsymbol{p_2+p_3+p_4} & \boldsymbol{p_3+p_4} \\ p_2+p_3+p_4 & p_3+p_4 & \boldsymbol{p_4} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{q_1} & \boldsymbol{q_2} & \boldsymbol{q_3} \\ q_2 & q_3 & \boldsymbol{q_4} \end{bmatrix} \end{aligned}$$

So,

$$\ker\varphi = \langle q_1q_3 - q_2^2,\ q_1q_4 - q_2q_3,\ q_2q_4 - q_3^2\rangle$$

The ideal of minors

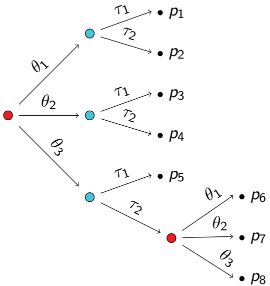


Let J be the ideal generated by the 2×2-minors of the two stage matrices.

$$M_1 = \begin{bmatrix} p_1 + p_2 & p_6 \\ p_3 + p_4 & p_7 \\ p_5 + \ldots + p_8 & p_8 \end{bmatrix} \quad M_2 = \begin{bmatrix} p_1 & p_3 & p_5 \\ p_2 & p_4 & p_6 + p_7 + p_8 \end{bmatrix}$$

$J \subseteq \ker \varphi$ and $\ker \varphi$ is a minimal prime of J.

The ideal of minors



Let J be the ideal generated by the 2×2-minors of the two stage matrices.

$$M_1 = \begin{bmatrix} p_1 + p_2 & p_6 \\ p_3 + p_4 & p_7 \\ p_5 + \ldots + p_8 & p_8 \end{bmatrix} \quad M_2 = \begin{bmatrix} p_1 & p_3 & p_5 \\ p_2 & p_4 & p_6 + p_7 + p_8 \end{bmatrix}$$

$J \subseteq \ker \varphi$ and $\ker \varphi$ is a minimal prime of J.

$$M_1 \xrightarrow{[\mathrm{R1}\to\mathrm{R1}+\mathrm{R2}+\mathrm{R3}]} \begin{bmatrix} p_1 + \ldots + p_8 & p_6 + p_7 + p_8 \\ p_3 + p_4 & p_7 \\ p_5 + \ldots + p_8 & p_8 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \\ q_5 & q_6 \end{bmatrix}$$

$$M_2 \xrightarrow{[\mathrm{R1}\to\mathrm{R1}+\mathrm{R2}]} \xrightarrow{[\mathrm{C1}\to\mathrm{C1}+\mathrm{C2}+\mathrm{C3}]} \begin{bmatrix} p_1 + \ldots + p_8 & p_3 + p_4 & p_5 + \ldots + p_8 \\ p_2 + p_4 + p_6 + p_7 + p_8 & p_4 & p_6 + p_7 + p_8 \end{bmatrix} = \begin{bmatrix} q_1 & q_3 & q_5 \\ q_7 & q_8 & q_2 \end{bmatrix}$$

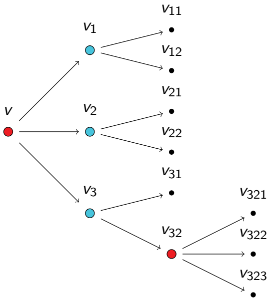
J is a binomial ideal in vars $q_1, \ldots, q_8 \rightsquigarrow \cdots \rightsquigarrow$ $\ker \varphi$ is toric in vars $q_1, \ldots, q_8$

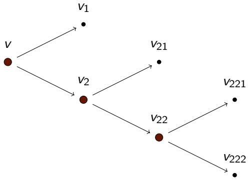
Theorem [Görgen, AM, Nicklasson, '22]: If the ideal of minors J is toric after a linear change of variables, then $\ker \varphi$ is toric in the new variables, also. (*details omitted*)

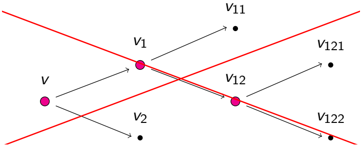
Applications of the ideal of minors; SIP trees

For a vertex u, let $u_1, u_2 \dots$ be its children ordered top-to-bottom in a staged tree $\mathcal{T}$, and $\mathcal{T}(u)$ be the subtree of $\mathcal{T}$ with root u.

A stage/colour c has the **Subtree Inclusion Property (SIP)** if there is an index i such that for every vertex u in stage c, each induced subtree $\mathcal{T}(u_j)$ contains a subtree identical to $\mathcal{T}(u_i)$ with root u_j.



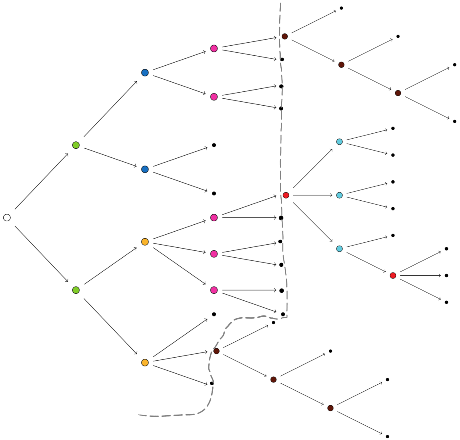




Theorem [Görgen, AM, Nicklasson, '22]: All SIP trees have toric structure.
The proof provides a linear change of variables that makes $\ker \varphi$ toric.

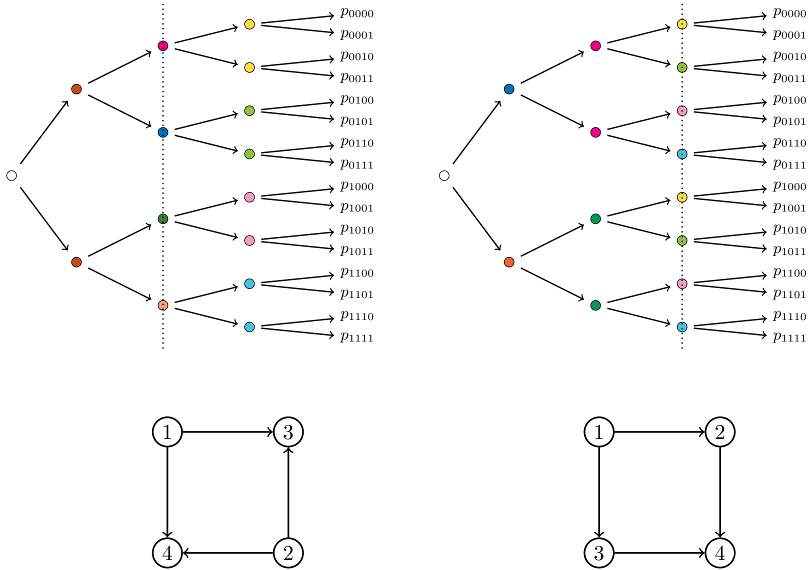
Applications of the ideal of minors; SIP trees

Theorem [Görgen, AM, Nicklasson, '22]: Staged trees with the Subtree Inclusion Property (SIP) have toric structure.



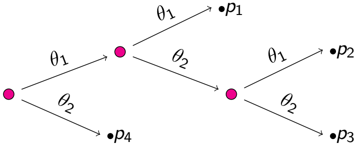
Theorem [Görgen, AM, Nicklasson, '22]: Glue SIP trees to leaves of a balanced tree with the condition that they use different colours from the colours of the balanced tree. The new staged tree has toric structure.

Applications to Bayesian Networks



One-Stage Trees

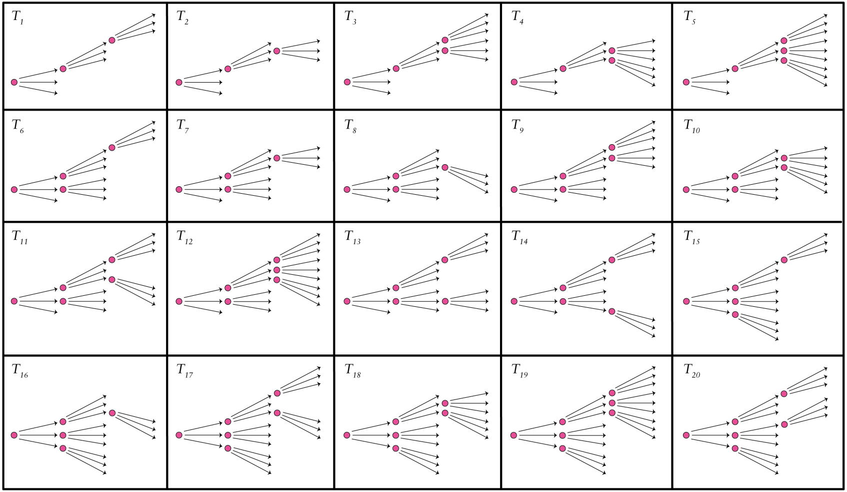
A *one-stage tree* is a tree where all internal vertices are in the same stage. They are not all covered by SIP.



Theorem [Görgen, AM, Nicklasson, '22]: All binary one-stage trees have toric structure with parametrisation given by Veronese algebras $\mathbb{R}[\theta_1^i \theta_2^{d-i} \mid 0 \leq i \leq d]$ where d is the depth of the tree.

One-Stage Trees

Theorem [Görgen, AM, Nicklasson, '22]: All one-stage trees of depth 3 in three parameters have toric structure.



Code in https://mathrepo.mis.mpg.de/StagedTreesWithToricStructures/

Are there non-toric Staged Tree Models?

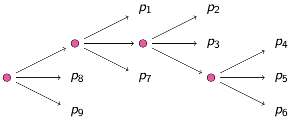
Conjecture: All one-stage trees have toric structure.

Are there non-toric Staged Tree Models?

Conjecture: All one-stage trees have toric structure. NOT*

Are there non-toric Staged Tree Models?

Conjecture: All one-stage trees have toric structure. NOT*



$$\begin{bmatrix} p_1 + \dots + p_7 & p_1 & p_3 & p_4 \\ p_8 & p_2 + \dots + p_6 & p_3 & p_5 \\ p_9 & p_7 & p_4 + p_5 + p_6 & p_6 \end{bmatrix}$$

*Ongoing work with **Arpan Pal** (Texas A&M)
Toric Structures in Statistical Models and Lie Algebras

A final word on toric models

In fact, the converse to this proposition also holds, with a slightly more inclusive definition of toric variety. Informally speaking, toric varieties are precisely those varieties that become linear spaces under taking logarithms.

8.3. The World is Toric

The occurrence of toric structures in an application can be either obvious or hidden. A typical example for the former is log-linear models in statistics. These are obviously toric, as seen around Example 8.26. In this section we discuss some scenarios where the toric structure is hidden, and it needs to be unearthed, often by a non-trivial chance of coordinates. Our style in this section is extremely informal. We briefly visit four fields where toric varieties arise. Under each header we focus on one concrete instance of a toric variety $X_A \subset \mathbb{P}^{p-1}$. The broader context is discussed alongside that example.

Mateusz Michałek, and Bernd Sturmfels. "Invitation to nonlinear algebra." American Mathematical Soc., (2021).

A final word on toric models

In fact, the converse to this proposition also holds, with a slightly more inclusive definition of toric variety. Informally speaking, toric varieties are precisely those varieties that become linear spaces under taking logarithms.

8.3. The World is Toric

The occurrence of toric structures in an application can be either obvious or hidden. A typical example for the former is log-linear models in statistics. These are obviously toric, as seen around Example 8.26. In this section we discuss some scenarios where the toric structure is hidden, and it needs to be unearthed, often by a non-trivial chance of coordinates. Our style in this section is extremely informal. We briefly visit four fields where toric varieties arise. Under each header we focus on one concrete instance of a toric variety $X_A \subset \mathbb{P}^{p-1}$. The broader context is discussed alongside that example.

Mateusz Michałek, and Bernd Sturmfels. "Invitation to nonlinear algebra." American Mathematical Soc., (2021).

Thank you for listening!

Extra Slides

Theorem (GMN 2021)

Let J be the ideal of 2×2-minors of the stage matrices of a staged tree. If there are linear forms $\ell_1, \ldots, \ell_n$ in $\mathbb{R}[p_1, ..., p_n]$ such that

1. *the ideal J is generated by binomials in the new variables $\ell_1, \ldots, \ell_n$,*
2. *$\ell_1, \ldots, \ell_n$ are linearly independent, and*
3. *each $\varphi(\ell_i)$ can be represented by a monomial in the θ_{ij}'s,*

then $\ker \varphi$ *is a toric ideal in the new variables $\ell_1, \ldots, \ell_n$.*

In the case $J = \ker \varphi$ we can drop 3.