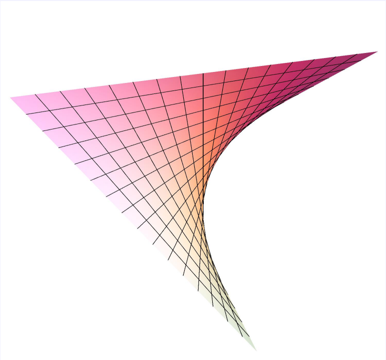
~~Some~~ What is algebraic statistics ?

Jan Draisma
Mathematical Institute
University of Bern
and Eindhoven University of Technology

Osnabrück, 30 November 2017

Constructible set in $\mathbb{C}^n$: defined by a finite, meaningful formula in the alphabet $\mathbb{C} \cup \{x_1, \ldots, x_n, \cdot, +, =, \neg, \vee, \wedge\}$.

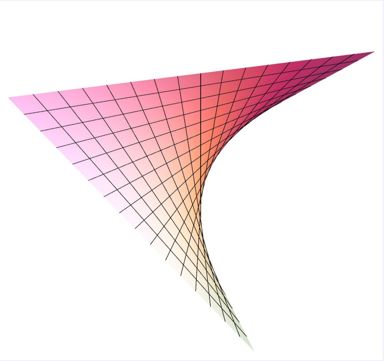


Example

$X = \{(x_1, x_2, x_3, x_4) \mid x_1 x_4 = x_2 x_3\}$

Constructible sets

Constructible set in $\mathbb{C}^n$: defined by a finite, meaningful formula in the alphabet $\mathbb{C} \cup \{x_1, \ldots, x_n, \cdot, +, =, \neg, \vee, \wedge\}$.

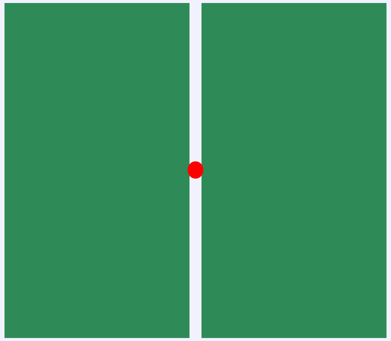


Example

$X = \{(x_1, x_2, x_3, x_4) \mid x_1 x_4 = x_2 x_3\}$

Preserved under union, intersection, complement and:

Theorem (Chevalley): The image of a constructible set under a polynomial map is constructible. (*And computable!*)

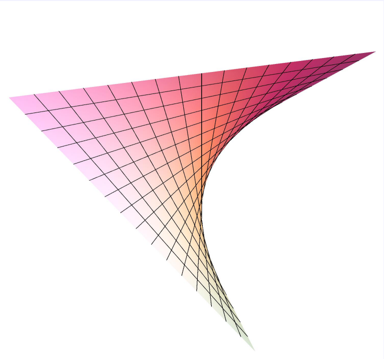


Example: $X = \mathbb{C}^2$, $\varphi(x, y) = (x, xy)$

$\varphi(X) = \{(u, v) \mid (\neg u = 0) \vee (u = 0 \wedge v = 0)\}$.

Constructible sets

Constructible set in $\mathbb{C}^n$: defined by a finite, meaningful formula in the alphabet $\mathbb{C} \cup \{x_1, \ldots, x_n, \cdot, +, =, \neg, \vee, \wedge\}$.

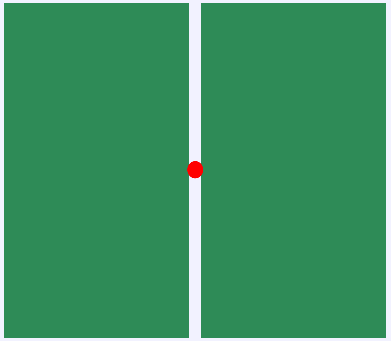


Example

$X = \{(x_1, x_2, x_3, x_4) \mid x_1 x_4 = x_2 x_3\}$

Preserved under union, intersection, complement and:

Theorem (Chevalley): The image of a constructible set under a polynomial map is constructible. (*And computable!*)



Example: $X = \mathbb{C}^2$, $\varphi(x, y) = (x, xy)$

$\varphi(X) = \{(u, v) \mid (\neg u = 0) \vee (u = 0 \wedge v = 0)\}$.

Example: $X = \mathbb{C}^{n \times k}$, $\varphi(A) = A \cdot A^T \in \mathbb{C}^{n \times n}$; $\varphi(X) = ?$

Semi-algebraic set in $\mathbb{R}^n$: defined by a finite, meaningful formula in the alphabet $\mathbb{R} \cup \{x_1, \ldots, x_n, \cdot, +, =, \neg, \vee, \wedge, \geq\}$.

Example: $\{x \in \mathbb{R}^4 \mid (x_1 x_4 = x_2 x_3) \wedge (x_1 + x_2 + x_3 + x_4 = 1) \wedge (x_1, \ldots, x_4 \geq 0)\}$—*probability distributions on* $\{1, 2\} \times \{1, 2\}$ *such that the first and second entry are independent.*

Semi-algebraic set in $\mathbb{R}^n$: defined by a finite, meaningful formula in the alphabet $\mathbb{R} \cup \{x_1, \ldots, x_n, \cdot, +, =, \neg, \vee, \wedge, \geq\}$.

Example: $\{x \in \mathbb{R}^4 \mid (x_1 x_4 = x_2 x_3) \wedge (x_1 + x_2 + x_3 + x_4 = 1) \wedge (x_1, \ldots, x_4 \geq 0)\}$—*probability distributions on* $\{1, 2\} \times \{1, 2\}$ *such that the first and second entry are independent.*

Theorem (Tarski): The image of a semialgebraic set under a polynomial map is semialgebraic. *(And computable!)*

Example: $X = \mathbb{R}^{n \times k}$, $\varphi(A) = A \cdot A^T$, $\varphi(X) = \{B \mid B = B^T$ and each principal $\ell \times \ell$-subdet of B is ≥ 0 for $\ell \leq k$ and $= 0$ for $\ell = k + 1\}$.

Semi-algebraic sets

Semi-algebraic set in $\mathbb{R}^n$: defined by a finite, meaningful formula in the alphabet $\mathbb{R} \cup \{x_1, \ldots, x_n, \cdot, +, =, \neg, \vee, \wedge, \geq\}$.

Example: $\{x \in \mathbb{R}^4 \mid (x_1 x_4 = x_2 x_3) \wedge (x_1 + x_2 + x_3 + x_4 = 1) \wedge (x_1, \ldots, x_4 \geq 0)\}$—*probability distributions on* $\{1,2\} \times \{1,2\}$ *such that the first and second entry are independent.*

Theorem (Tarski): The image of a semialgebraic set under a polynomial map is semialgebraic. *(And computable!)*

Example: $X = \mathbb{R}^{n \times k}$, $\varphi(A) = A \cdot A^T$, $\varphi(X) = \{B \mid B = B^T$ and each principal $\ell \times \ell$-subdet of B is ≥ 0 for $\ell \leq k$ and $= 0$ for $\ell = k+1\}$.

Example: $X = \mathbb{R}^{n \times k}_{\geq 0}$, $Y = \mathbb{R}^{k \times m}_{\geq 0}$, $\varphi(A, B) = A \cdot B$; $\varphi(X) = \{B \in \mathbb{R}_{\geq 0}$ of nonnegative rank $\leq k\}$—no “finite characterisation” for $k = 3$.

A word on algorithms

Symbolic (`4ti2`, `Macaulay2`, `Normaliz`, `Singular`, `...`)**:** manipulate polynomial equations (and inequalities), say with coefficients in $\mathbb{Q}$. *Typical application:* generators for all equations vanishing on all of $X \rightsquigarrow$ generators for all equations for $\varphi(X)$.

Example: Input $\{p_1 + p_2 - 1, q_1 - q_2 - 1\}$ and $\varphi(p,q) = (p_1q_1, p_1q_2, p_2q_1, p_2q_2)$. Output: $\{x_1x_4 - x_2x_3, x_1 + x_2 + x_3 + x_4 - 1\}$.

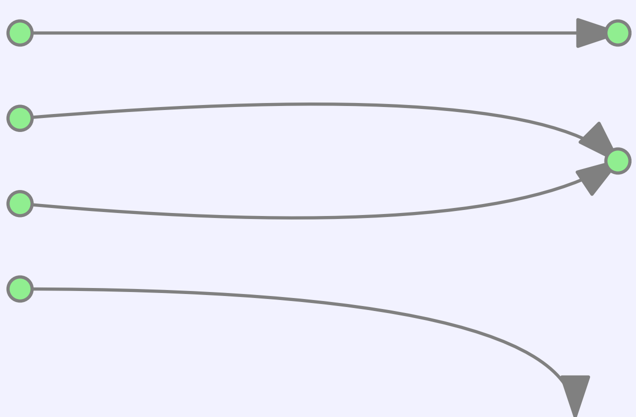
Symbolic (`4ti2, Macaulay2, Normaliz, Singular, ...`)**:** manipulate polynomial equations (and inequalities), say with coefficients in $\mathbb{Q}$. *Typical application:* generators for all equations vanishing on all of $X \rightsquigarrow$ generators for all equations for $\varphi(X)$.

Example: Input $\{p_1 + p_2 - 1, q_1 - q_2 - 1\}$ and $\varphi(p,q) = (p_1q_1, p_1q_2, p_2q_1, p_2q_2)$. Output: $\{x_1x_4 - x_2x_3, x_1 + x_2 + x_3 + x_4 - 1\}$.

Numerical (`Bertini, Macaulay, ...`)**:** solve square systems of equations using homotopy continuation.

$t = 1$
easy system

choice??



$t = 0$
original system

***Discrete* algebraic-statistical model:** semi-algebraic subset M of the *probabiliy simplex* $\Delta_{n-1} := \{x \in \mathbb{R}^n_{\geq 0} \mid \sum_i x_i = 1\}$. A point $x \in M$ is a probability distribution on $[n]$.

Discrete **algebraic-statistical model:** semi-algebraic subset M of the *probabiliy simplex* $\Delta_{n-1} := \{x \in \mathbb{R}^n_{\geq 0} \mid \sum_i x_i = 1\}$. A point $x \in M$ is a probability distribution on $[n]$.

Typical questions:

1. If M given as $\varphi(X)$, find a quantifier-free formula for M.
2. How does 1 vary with combinatorial parameters of M?
3. For $u \in \mathbb{Z}^n_{\geq 0}$ recording independent observations, compute $x \in M$ maximising the *likelihood* $L(u|x) = \binom{|u|}{u} x^u$ of u.

***Discrete* algebraic-statistical model:** semi-algebraic subset M of the *probabiliy simplex* $\Delta_{n-1} := \{x \in \mathbb{R}^n_{\geq 0} \mid \sum_i x_i = 1\}$. A point $x \in M$ is a probability distribution on $[n]$.

Typical questions:

1. If M given as $\varphi(X)$, find a quantifier-free formula for M.
2. How does 1 vary with combinatorial parameters of M?
3. For $u \in \mathbb{Z}^n_{\geq 0}$ recording independent observations, compute $x \in M$ maximising the *likelihood* $L(u|x) = \binom{|u|}{u} x^u$ of u.

Why 1?

• Model validation: if $u/(\sum_i u_i)$ almost satisfies the equations and inequalities, then M is a good model (without need to find x).

• Markov chains for Fisher's exact test for log-linear models.

Independence, two variables

Independence:

$M = \{(p_i q_j)_{i,j} \mid p_i \geq 0, q_j \geq 0, \sum_i p_i = \sum_j q_j = 1\} \subseteq \Delta_{mn-1}$.

Independence:

$M = \{(p_i q_j)_{i,j} \mid p_i \geq 0, q_j \geq 0, \sum_i p_i = \sum_j q_j = 1\} \subseteq \Delta_{mn-1}$.

1. $M = \{x \in \mathbb{R}^{m \times n} \mid \sum_{i,j} x_{ij} = 1, x_{ij} \geq 0, x_{ij}x_{kl} - x_{il}x_{kj} = 0\}$.
2. $x_{11} \geq 0$ and $x_{11}x_{22} - x_{12}x_{21} = 0$ + row and col permutations.
3. Observe $u \in \mathbb{Z}_{\geq 0}^{m \times n} \rightsquigarrow$ ML-estimate $x_{ij} := u_{i+}u_{+j}/u_{++}^2$.

Independence:
$M = \{(p_i q_j)_{i,j} \mid p_i \geq 0, q_j \geq 0, \sum_i p_i = \sum_j q_j = 1\} \subseteq \Delta_{mn-1}$.

1. $M = \{x \in \mathbb{R}^{m \times n} \mid \sum_{i,j} x_{ij} = 1, x_{ij} \geq 0, x_{ij}x_{kl} - x_{il}x_{kj} = 0\}$.
2. $x_{11} \geq 0$ and $x_{11}x_{22} - x_{12}x_{21} = 0$ + row and col permutations.
3. Observe $u \in \mathbb{Z}_{\geq 0}^{m \times n} \rightsquigarrow$ ML-estimate $x_{ij} := u_{i+}u_{+j}/u_{++}^2$.

Markov Chain for Fisher's exact test (Diaconis-Sturmfels): for some test function t, want to compute the probability of $t(v) \geq t(u)$ conditional on $v_{i+} = u_{i+}$ and $v_{+j} = u_{+j}$ and sampled from the same distribution x. Approximate by sampling such v using *Markov moves* from u:

$$\begin{matrix} +1 & -1 \\ -1 & +1 \end{matrix}$$

Mixture of M_1 and M_2:

$M_1 + M_2 := \{tx + (1-t)y \mid t \in [0,1], x \in M_1, y \in M_2\}$

Mixtures of independence

Mixture of M_1 and M_2:
$M_1 + M_2 := \{tx + (1-t)y \mid t \in [0,1], x \in M_1, y \in M_2\}$

Example: $M_1 = M_2 = M \subseteq \Delta_{mn-1}$ model of independence.
1. $M + M = \{x \mid \sum_{i,j} x_{ij} = 1, x_{ij} \geq 0, \mathrm{rk}(x) \leq 2\}$
(nonnegative rank 2 = nonnegative + rank 2)
2. one orbit of 3×3-dets, one orbit of inequalities
3. ML-estimate?

Mixtures of independence

Mixture of M_1 and M_2:
$M_1 + M_2 := \{tx + (1-t)y \mid t \in [0,1], x \in M_1, y \in M_2\}$

Example: $M_1 = M_2 = M \subseteq \Delta_{mn-1}$ model of independence.
1. $M + M = \{x \mid \sum_{i,j} x_{ij} = 1, x_{ij} \geq 0, \mathrm{rk}(x) \leq 2\}$
(nonnegative rank 2 = nonnegative + rank 2)
2. one orbit of 3×3-dets, one orbit of inequalities
3. ML-estimate?

Thm (Kubjas-Robeva-Sturmfels and Eggermont-Horobeţ-K):
$M + M + M = \{x$ of nonneg rank ≤ 3 and $\sum_{i,j} x_{ij} = 1\}$.
1. quant-free description; comps of the *algebraic boundary*
2. three orbits of boundary components
3. experiments: the EM algorithm often runs into the boundary

Zariski-closure: $\overline{M}$ of M in $\{\sum_i x_i = 1\} \subseteq \mathbb{C}^n$.

Zariski-closure: $\overline{M}$ of M in $\{\sum_i x_i = 1\} \subseteq \mathbb{C}^n$.

Critical points of the likelihood function:
$L(u|x) = C(u)x^u$, so $d_x L(u|.)(v) = L(u|x) \cdot \sum_i \frac{u_i}{x_i} v_i$
$\rightsquigarrow$ necessary for $x \in M^{\mathrm{reg}} \cap \mathbb{R}^n_{>0}$ to be the ML-estimate is that $x^{-1} \cdot T_x M \perp u$—*which makes sense for* $u \in \mathbb{C}^n$ *and* $x \in \overline{M} \cap (\mathbb{C}^*)^n$*!*

Zariski-closure: $\overline{M}$ of M in $\{\sum_i x_i = 1\} \subseteq \mathbb{C}^n$.

Critical points of the likelihood function:
$L(u|x) = C(u)x^u$, so $d_x L(u|.)(v) = L(u|x) \cdot \sum_i \frac{u_i}{x_i} v_i$
$\rightsquigarrow$ necessary for $x \in M^{\text{reg}} \cap \mathbb{R}^n_{>0}$ to be the ML-estimate is that $x^{-1} \cdot T_x M \perp u$—*which makes sense for* $u \in \mathbb{C}^n$ *and* $x \in \overline{M} \cap (\mathbb{C}^*)^n$*!*

ML-degree of M**:** the number of $x \in \overline{M}^{\text{reg}} \cap (\mathbb{C}^*)^n$ with $x^{-1} \cdot T_x \overline{M} \perp u$, for $u \in \mathbb{C}^n$ sufficiently general.

Zariski-closure: $\overline{M}$ of M in $\{\sum_i x_i = 1\} \subseteq \mathbb{C}^n$.

Critical points of the likelihood function:
$L(u|x) = C(u)x^u$, so $d_x L(u|.)(v) = L(u|x) \cdot \sum_i \frac{u_i}{x_i} v_i$
$\rightsquigarrow$ necessary for $x \in M^{\text{reg}} \cap \mathbb{R}^n_{>0}$ to be the ML-estimate is that $x^{-1} \cdot T_x M \perp u$—*which makes sense for* $u \in \mathbb{C}^n$ *and* $x \in \overline{M} \cap (\mathbb{C}^*)^n$*!*

ML-degree of M: the number of $x \in \overline{M}^{\text{reg}} \cap (\mathbb{C}^*)^n$ with $x^{-1} \cdot T_x \overline{M} \perp u$, for $u \in \mathbb{C}^n$ sufficiently general.

Example: ML-degree of independence is 1: $(u_{i+} u_{+j})/(u_{++}^2)$.

Theorem (Huh): all varieties with ML-degree 1 are image of a composition Ψ of a linear map $\mathbb{C}^n \to \mathbb{C}^r$ and a rational monomial map $\mathbb{C}^r \to (\mathbb{C}^*)^n$ with monomials of degree 0.

ML-degree for mixtures of independence

Theorem (Hauenstein-Rodriguez-Sturmfels):
$M \subseteq \Delta_{mn-1}$ independence. Values of ML-degree for $\overline{kM}$:

		(m,n)						
		(3,3)	(3,4)	(3,5)	(4,4)	(4,5)	(4,6)	(5,5)
	1	1	1	1	1	1	1	1
	2	10	26	58	191	843	3119	6776
k	3	1	1	1	191	843	3119	61326
	4				1	1	1	6776
	5							1

Bertini

ML-degree for mixtures of independence

Theorem (Hauenstein-Rodriguez-Sturmfels):
$M \subseteq \Delta_{mn-1}$ independence. Values of ML-degree for $\overline{kM}$:

		(m,n)						
		$(3,3)$	$(3,4)$	$(3,5)$	$(4,4)$	$(4,5)$	$(4,6)$	$(5,5)$
	1	1	1	1	1	1	1	1
	2	10	26	58	191	843	3119	6776
k	3	1	1	1	191	843	3119	61326
	4				1	1	1	6776
	5							1

Bertini

Theorem (D-Rodriguez)
The above symmetry really holds.

Theorem (Hauenstein-Rodriguez-Sturmfels):
$M \subseteq \Delta_{mn-1}$ independence. Values of ML-degree for $\overline{kM}$:

k \ (m,n)	$(3,3)$	$(3,4)$	$(3,5)$	$(4,4)$	$(4,5)$	$(4,6)$	$(5,5)$	
1	1	1	1	1	1	1	1	
2	10	26	58	191	843	3119	6776	
3	1	1	1	191	843	3119	61326	
4				1	1	1	6776	Bertini
5							1	

Theorem (D-Rodriguez)
The above symmetry really holds.

Theorem (Rodriguez-Wang)
The ML-degree for $m = 3, k = 2$ equals

Theorem (Hauenstein-Rodriguez-Sturmfels):
$M \subseteq \Delta_{mn-1}$ independence. Values of ML-degree for $\overline{kM}$:

k \ (m,n)	$(3,3)$	$(3,4)$	$(3,5)$	$(4,4)$	$(4,5)$	$(4,6)$	$(5,5)$	
1	1	1	1	1	1	1	1	
2	10	26	58	191	843	3119	6776	
3	1	1	1	191	843	3119	61326	
4				1	1	1	6776	Bertini
5							1	

Theorem (D-Rodriguez)
The above symmetry really holds.

Theorem (Rodriguez-Wang)
The ML-degree for $m = 3, k = 2$ equals $2^{n+1} - 6$.

Crazy but beautiful idea for ML-estimation

Idea (Hauenstein, Leykin, Rodriguez, Sottile, ...)

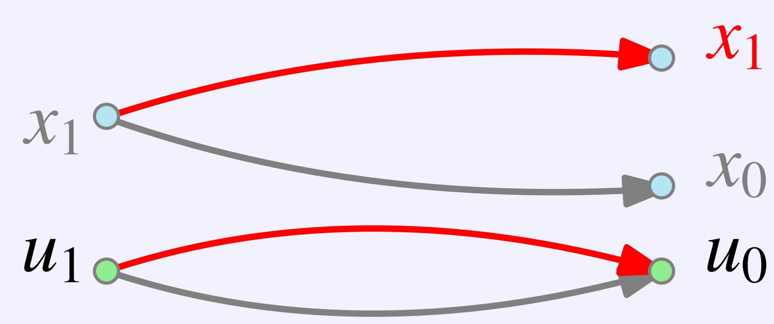
- construct a pair (x_1, u_1) with $x_1^{-1} T_{x_1} \overline{M} \perp u_1$ (easy: pick any $x_1 \in \overline{M}^{\text{reg}} \cap (\mathbb{C}^*)^n$, solve linear system for u_1).
- perform many homotopy continuations for the system $x_t \in \overline{M} \wedge x_t^{-1} T_{x_t} \overline{M} \perp u_t$ with u_0 the actual data.

Crazy but beautiful idea for ML-estimation

Idea (Hauenstein, Leykin, Rodriguez, Sottile, . . .)

- construct a pair (x_1, u_1) with $x_1^{-1} T_{x_1} \overline{M} \perp u_1$ (easy: pick any $x_1 \in \overline{M}^{\mathrm{reg}} \cap (\mathbb{C}^*)^n$, solve linear system for u_1).
- perform many homotopy continuations for the system $x_t \in \overline{M} \wedge x_t^{-1} T_{x_t} \overline{M} \perp u_t$ with u_0 the actual data.

x_0 depends on the chosen path u_t, and if $\overline{M} \cap (\mathbb{C}^*)^n$ is irreducible, all critical points for u_0 are obtained in this manner!



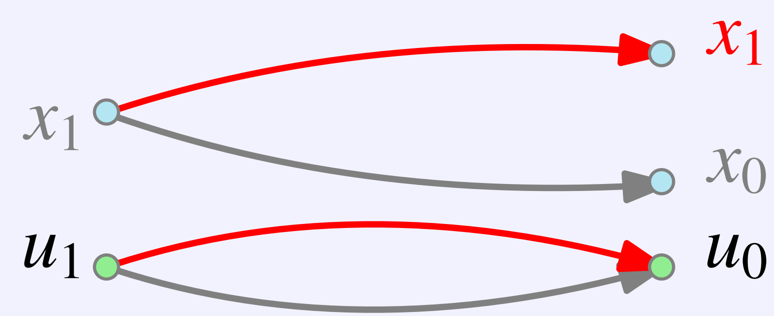
Crazy but beautiful idea for ML-estimation

Idea (Hauenstein, Leykin, Rodriguez, Sottile, . . .)

• construct a pair (x_1, u_1) with $x_1^{-1} T_{x_1} \overline{M} \perp u_1$ (easy: pick any $x_1 \in \overline{M}^{\text{reg}} \cap (\mathbb{C}^*)^n$, solve linear system for u_1).

• perform many homotopy continuations for the system $x_t \in \overline{M} \wedge x_t^{-1} T_{x_t} \overline{M} \perp u_t$ with u_0 the actual data.

x_0 depends on the chosen path u_t, and if $\overline{M} \cap (\mathbb{C}^*)^n$ is irreducible, all critical points for u_0 are obtained in this manner!



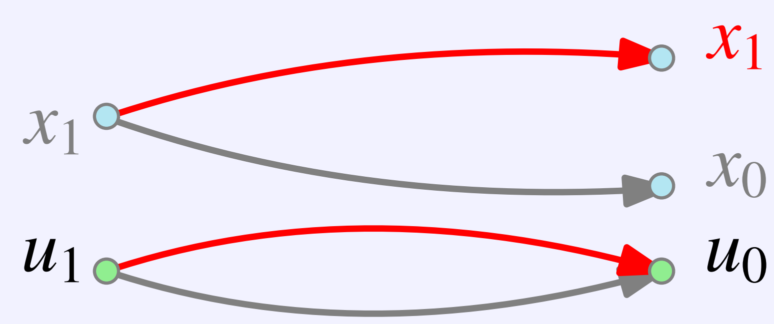
Question: When to stop?

Crazy but beautiful idea for ML-estimation

Idea (Hauenstein, Leykin, Rodriguez, Sottile, ...)

- construct a pair (x_1, u_1) with $x_1^{-1} T_{x_1} \overline{M} \perp u_1$ (easy: pick any $x_1 \in \overline{M}^{\text{reg}} \cap (\mathbb{C}^*)^n$, solve linear system for u_1).
- perform many homotopy continuations for the system $x_t \in \overline{M} \wedge x_t^{-1} T_{x_t} \overline{M} \perp u_t$ with u_0 the actual data.

x_0 depends on the chosen path u_t, and if $\overline{M} \cap (\mathbb{C}^*)^n$ is irreducible, all critical points for u_0 are obtained in this manner!



Question: When to stop?

Answer: when the trace test says you can!

Undirected graphical models

Setting

$G = (V, E)$ finite, simple undirected graph

$\Omega_i, i \in V$ finite sets

P a probability distribution on the state space $\Omega := \prod_{i \in V} \Omega_i$

$X_i : \Omega \to \Omega_i$ the ith coordinate function

$A \subseteq V \rightsquigarrow$ probability vector X_A taking values in Ω_A.

Setting

$G = (V, E)$ finite, simple undirected graph

$\Omega_i, i \in V$ finite sets

P a probability distribution on the state space $\Omega := \prod_{i \in V} \Omega_i$

$X_i : \Omega \to \Omega_i$ the ith coordinate function

$A \subseteq V \rightsquigarrow$ probability vector X_A taking values in Ω_A.

Conditional independence

$X_A \perp\!\!\!\perp X_B \mid X_C$ means: for each $x_C \in \Omega_C$ with $P(X_C = x_C) > 0$,

$P(X_A = x_A \wedge x_B = x_B \mid X_C = x_C) =$

$P(X_A = x_A \mid X_C = x_C) \cdot P(X_B = x_B \mid X_C = x_C)$.

Setting

$G = (V, E)$ finite, simple undirected graph

$\Omega_i, i \in V$ finite sets

P a probability distribution on the state space $\Omega := \prod_{i \in V} \Omega_i$

$X_i : \Omega \to \Omega_i$ the ith coordinate function

$A \subseteq V \rightsquigarrow$ probability vector X_A taking values in Ω_A.

Conditional independence

$X_A \perp\!\!\!\perp X_B \mid X_C$ means: for each $x_C \in \Omega_C$ with $P(X_C = x_C) > 0$,

$P(X_A = x_A \wedge x_B = x_B \mid X_C = x_C) =$

$P(X_A = x_A \mid X_C = x_C) \cdot P(X_B = x_B \mid X_C = x_C)$.

Pairwise Markov properties from G

$X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}$ for $i \neq j$ with $\{i, j\} \notin E$.

Example: Independence

$G =$

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

Undirected graphical models

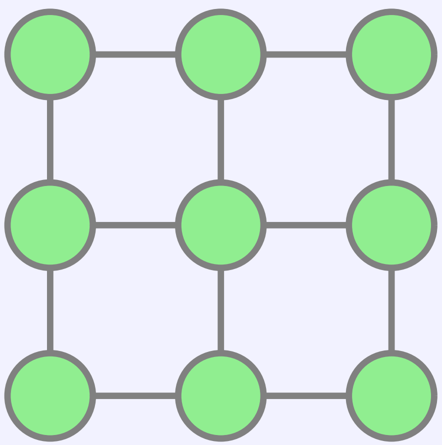
Example: Independence

$G =$

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

Example: Ising model

$G =$



$\Omega_i = \{-1, 1\}$ for all i

interaction parameters $c, d > 0$

$$P(x) = \frac{1}{Z} \cdot \left(\prod_{i \sim k, x_i = x_k} c\right) \cdot \left(\prod_{i \sim k, x_i \neq x_k} d\right)$$

Undirected graphical models

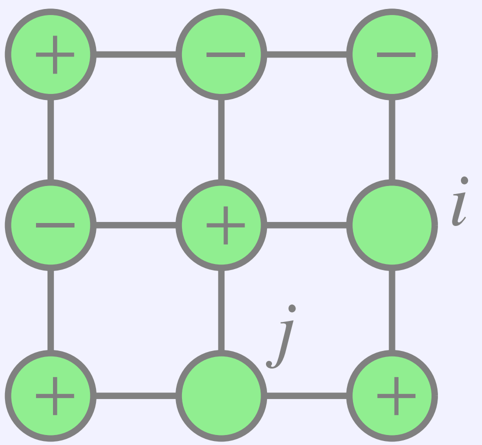
Example: Independence

$G =$

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

Example: Ising model

$G =$



$\Omega_i = \{-1, 1\}$ for all i

interaction parameters $c, d > 0$

$$P(x) = \frac{1}{Z} \cdot \left(\prod_{i \sim k, x_i = x_k} c\right) \cdot \left(\prod_{i \sim k, x_i \neq x_k} d\right)$$

$$P((X_i, X_j) = (1, 1) \mid \cdots) = \frac{(c^2 d)(c^3)}{(c^2 d + c d^2)(c^3 + d^3)} = P(X_i = 1 \mid \cdots) \cdot P(X_j = 1 \mid \cdots)$$

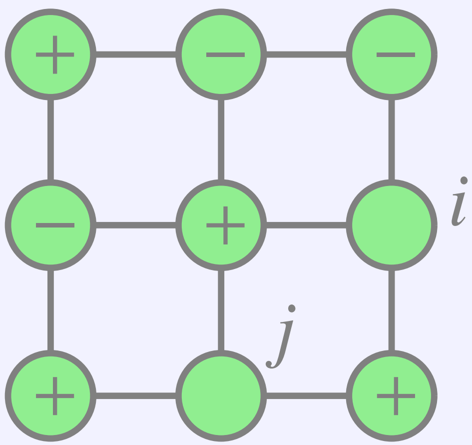
Undirected graphical models

Example: Independence

$G =$ 

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

Example: Ising model

$G =$ 

$\Omega_i = \{-1, 1\}$ for all i

interaction parameters $c, d > 0$

$$P(x) = \frac{1}{Z} \cdot \left(\prod_{i\sim k, x_i = x_k} c\right) \cdot \left(\prod_{i\sim k, x_i \neq x_k} d\right)$$

$$P((X_i, X_j) = (1, 1) \mid \cdots) = \frac{(c^2 d)(c^3)}{(c^2 d + cd^2)(c^3 + d^3)} = P(X_i = 1 \mid \cdots) \cdot P(X_j = 1 \mid \cdots)$$

$\rightsquigarrow$ P satisfies all the pairwise Markov properties for G.

Hammersley-Clifford Theorem
Assume $P > 0$ on all of Ω. Then P satisfies all the pairwise Markov properties $\Leftrightarrow \exists$ *interaction parameters* $\theta_C \in \mathbb{R}_{>0}^{\Omega_C}$, where C runs through the maximal cliques of G, such that $P(x) = \prod_C \theta_C(x_C)$.

Hammersley-Clifford Theorem
Assume $P > 0$ on all of Ω. Then P satisfies all the pairwise Markov properties $\Leftrightarrow \exists$ *interaction parameters* $\theta_C \in \mathbb{R}_{>0}^{\Omega_C}$, where C runs through the maximal cliques of G, such that $P(x) = \prod_C \theta_C(x_C)$.

Example
• *Independence*: maximal cliques=vertices $\rightsquigarrow P(x) = \prod_{i \in V} \theta_i(x_i)$.
• *Ising*: maximal cliques are edges $\{i, k\}$, and $c = \theta_{ik}(-1, -1) = \theta_{ik}(1, 1)$ and $d = \theta_{ik}(1, -1) = \theta_{ik}(-1, 1)$ (up to normalisation).

Hammersley-Clifford Theorem

Assume $P > 0$ on all of Ω. Then P satisfies all the pairwise Markov properties $\Leftrightarrow$ $\exists$ *interaction parameters* $\theta_C \in \mathbb{R}_{>0}^{\Omega_C}$, where C runs through the maximal cliques of G, such that $P(x) = \prod_C \theta_C(x_C)$.

Example

• *Independence*: maximal cliques=vertices $\rightsquigarrow$ $P(x) = \prod_{i \in V} \theta_i(x_i)$.
• *Ising*: maximal cliques are edges $\{i, k\}$, and $c = \theta_{ik}(-1,-1) = \theta_{ik}(1, 1)$ and $d = \theta_{ik}(1,-1) = \theta_{ik}(-1, 1)$ (up to normalisation).

Monomial parameterisation of $\widehat{M}$: $P(x) = \prod_C \theta_C(x_C)$.

Here we forget that the $P(x)$ must sum to 1 and must be positive. Hence the θ_C are unconstrained parameters. What *polynomial relations* among the $P(x)$ hold independently of the parameters θ_C?

Example: Independence

$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x_1', x_2', x_3) - P(x_1, x_2', x_3)P(x_1', x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$.

Example: Independence

$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x_1', x_2', x_3) - P(x_1, x_2', x_3)P(x_1', x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$.

Much more generally:

Example: Independence

$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x'_1, x'_2, x_3) - P(x_1, x'_2, x_3)P(x'_1, x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$.

Much more generally:

Independent Set Theorem (Hillar-Sullivant, 2012)

If $A \subseteq V$ is an independent set in G, then the ideal of $\widehat{M}$ is generated by boundedly many $\prod_{i \in A} \mathrm{Sym}(\Omega_i)$-orbits of binomials as $|\Omega_i| \to \infty$ for all $i \in A$.

Example: Independence
$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x'_1, x'_2, x_3) - P(x_1, x'_2, x_3)P(x'_1, x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\text{Sym}(\Omega_1) \times \text{Sym}(\Omega_2) \times \text{Sym}(\Omega_3)$.

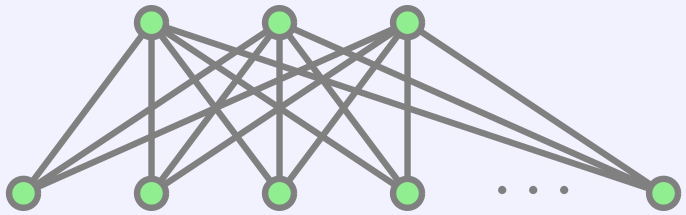
Much more generally:

Independent Set Theorem (Hillar-Sullivant, 2012)
If $A \subseteq V$ is an independent set in G, then the ideal of $\widehat{M}$ is generated by boundedly many $\prod_{i \in A} \text{Sym}(\Omega_i)$-orbits of binomials as $|\Omega_i| \to \infty$ for all $i \in A$.

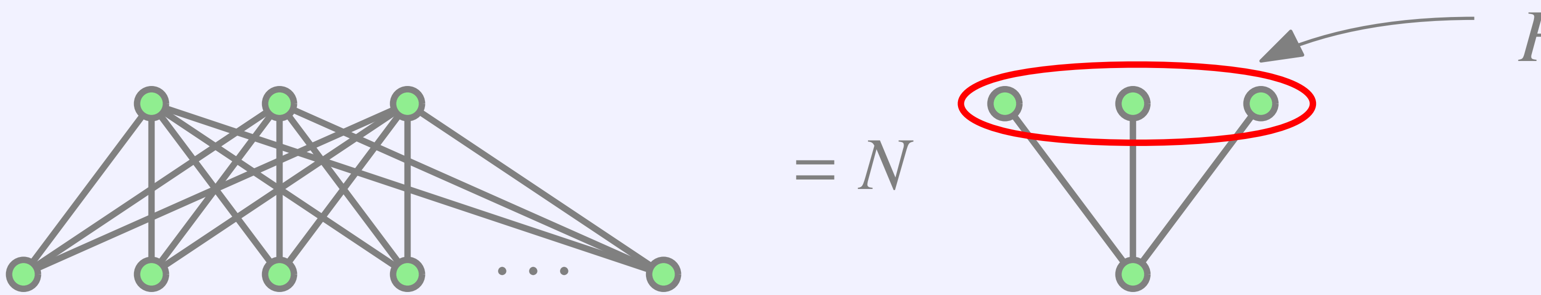
There are variants where G grows instead⇝Eggermont's talk!

Thank you!

Example (Rauh-Sullivant, 2014)
If G is $K_{3,N}$, and all state spaces are $\{0, 1\}$, then the ideal of $\widehat{M}$ is generated by binomials of degree ≤ 12, independently of N.

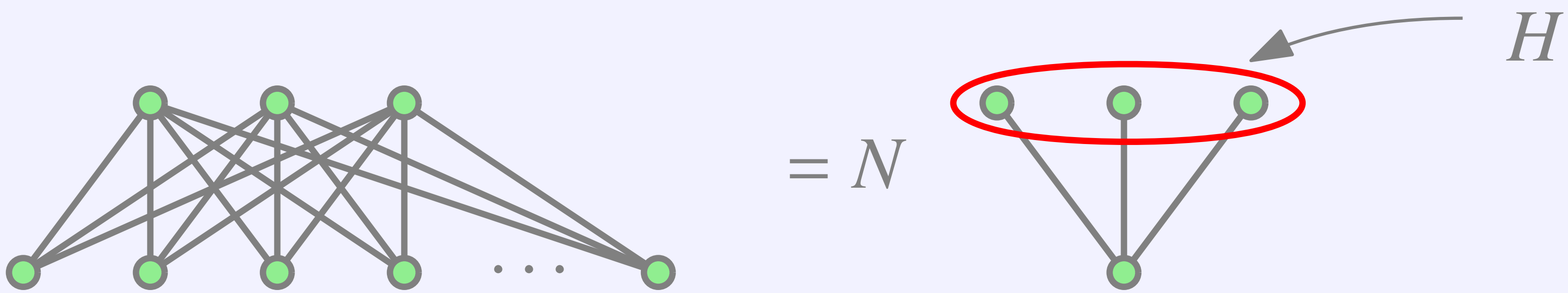


Example (Rauh-Sullivant, 2014)
If G is $K_{3,N}$, and all state spaces are $\{0,1\}$, then the ideal of $\widehat{M}$ is generated by binomials of degree ≤ 12, independently of N.



Construction: $G_1, \ldots, G_k$ finite graphs with a common induced subgraph $H \rightsquigarrow s_1 G_1 +_H \cdots +_H s_k G_k$ obtained from disjoint copies of the G_j by identifying their instances of H.

Example (Rauh-Sullivant, 2014)
If G is $K_{3,N}$, and all state spaces are $\{0,1\}$, then the ideal of $\widehat{M}$ is generated by binomials of degree ≤ 12, independently of N.



Construction: $G_1, \ldots, G_k$ finite graphs with a common induced subgraph $H \rightsquigarrow s_1 G_1 +_H \cdots +_H s_k G_k$ obtained from disjoint copies of the G_j by identifying their instances of H.

Theorem (D-Oosterhof, 2016)
Fixing state spaces for the vertices of each G_j, compatible with H, the ideal of $\widehat{M}(s_1 G_1 +_H \cdots +_H s_k G_k)$ is generated in bounded degree uniformly in the s_j.

Crucial fact: Suppose that G has vertex set $A \sqcup B$, where A is the vertex set of H; so G has state space $\Omega_A \times \Omega_B$.

Crucial fact: Suppose that G has vertex set $A \sqcup B$, where A is the vertex set of H; so G has state space $\Omega_A \times \Omega_B$.

Then sG has the vertex set $A \sqcup ([s] \times B)$ and state space $\Omega(s) := \Omega_A \times \Omega_B^s$. Any map $f : [s] \to [r]$ yields a map $\Omega(r) \to \Omega(s)$ and a linear map $\mathbb{R}^{\Omega(s)} \to \mathbb{R}^{\Omega(r)}$, which turns out to map $\widehat{M}(s)$ in the former space into $\widehat{M}(r)$ in the latter space.

Growing graphs

Crucial fact: Suppose that G has vertex set $A \sqcup B$, where A is the vertex set of H; so G has state space $\Omega_A \times \Omega_B$.

Then sG has the vertex set $A \sqcup ([s] \times B)$ and state space $\Omega(s) := \Omega_A \times \Omega_B^s$. Any map $f : [s] \to [r]$ yields a map $\Omega(r) \to \Omega(s)$ and a linear map $\mathbb{R}^{\Omega(s)} \to \mathbb{R}^{\Omega(r)}$, which turns out to map $\widehat{M}(s)$ in the former space into $\widehat{M}(r)$ in the latter space.

Thus $\widehat{M}$ is a *variety over the category* **Fin** *of finite sets*. We show that its ambient space is a Noetherian **Fin**-variety.

(The Independent Set Theorem concerns a $\mathbf{Fin}^{\mathrm{op}}$-variety!)

Open: What happens if both state spaces and graphs grow?

Definition

Given two models M_1 and M_2 in R^{Ω}, their *mixture* is the set $\{\lambda P + (1 - \lambda)Q \mid P \in M_1, Q \in M_2, \lambda \in [0, 1]\}$.

Mixtures

Definition
Given two models M_1 and M_2 in R^Ω, their *mixture* is the set $\{\lambda P + (1-\lambda)Q \mid P \in M_1, Q \in M_2, \lambda \in [0,1]\}$.

Example (Independence)
The mixture of two copies of independence is the set of all $|\Omega_1| \times |\Omega_2| \times |\Omega_3|$-tensors of *nonnegative rank* at most two whose entries add up to one.

Definition
Given two models M_1 and M_2 in R^{Ω}, their *mixture* is the set $\{\lambda P + (1-\lambda)Q \mid P \in M_1, Q \in M_2, \lambda \in [0,1]\}$.

Example (Independence)
The mixture of two copies of independence is the set of all $|\Omega_1| \times |\Omega_2| \times |\Omega_3|$-tensors of *nonnegative rank* at most two whose entries add up to one.

Theorem (Allman-Rhodes-Sturmfels-Zwiernik, 2013)
An $m_1 \times m_2 \times m_3$-tensor P with entries in $\mathbb{R}_{\geq 0}$ has nonnegative rank at most two if and only if P has rank at most two and is moreover *(log-)supermodular*: $P(x_1, x_2, x_3)P(y_1, y_2, y_3) \leq P(u_1, u_2, u_3)P(z_1, z_2, z_3)$ if $\{x_r, y_r\} = \{u_r, z_r\}$ and $u_r \leq z_r$ for all r, or in the $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$-orbit of such a tensor.

Boundedness of equations of some mixtures

Forget again about inequalities and summing up to 1.

Easy fact:
If G is a disjoint union of cliques, then $\widehat{M}$ is not only stable under $\prod_{i\in V}\mathrm{Sym}(\Omega_i)$, but even under $\prod_{i\in V}\mathrm{GL}_{\Omega_i}$. Hence the same holds for mixtures $\widehat{M}_1+\widehat{M}_2=\overline{\{P+Q \mid P\in\widehat{M}_1, Q\in\widehat{M}_2\}}$ coming from graphs G_1, G_2 that are unions of cliques.

Boundedness of equations of some mixtures

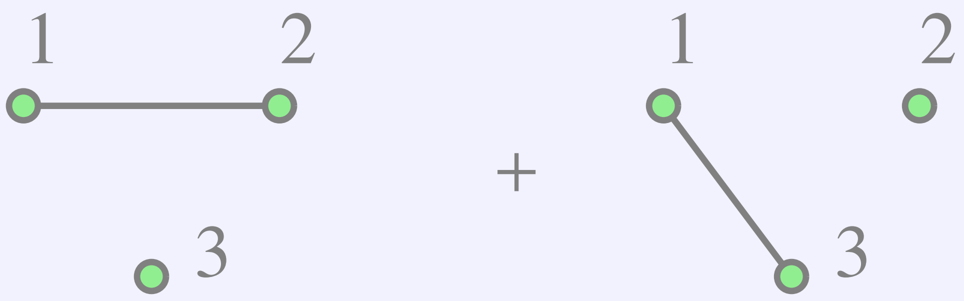
Forget again about inequalities and summing up to 1.

Easy fact:
If G is a disjoint union of cliques, then $\widehat{M}$ is not only stable under $\prod_{i \in V} \mathrm{Sym}(\Omega_i)$, but even under $\prod_{i \in V} \mathrm{GL}_{\Omega_i}$. Hence the same holds for mixtures $\widehat{M}_1 + \widehat{M}_2 = \overline{\{P + Q \mid P \in \widehat{M}_1, Q \in \widehat{M}_2\}}$ coming from graphs G_1, G_2 that are unions of cliques.

Theorem (D, 2017)
For any fixed k, a closed subvariety in a tensor product $W_1 \otimes \cdots \otimes W_k$ of vector spaces that depends functorially on $W_1, \ldots, W_k$ is defined by finitely many equations up to $\prod_i \mathrm{GL}(W_i)$, independently of the dimensions of the W_i.

A mixture challenge

A mixture challenge



$$M = \{P(x_1, x_2, x_3) = \lambda \frac{1}{Z(c,d)} c_{x_1,x_2} d_{x_3} + (1-\lambda) \frac{1}{Z'(e,f)} e_{x_1,x_3} f_{x_2}\}$$

Here: $\lambda \in [0, 1]$, $c \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_2}$, $d \in \mathbb{R}_{>0}^{\Omega_3}$, $e \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_3}$, $f \in \mathbb{R}_{>0}^{\Omega_2}$

A mixture challenge

$$M = \{P(x_1, x_2, x_3) = \lambda \frac{1}{Z(c,d)} c_{x_1,x_2} d_{x_3} + (1-\lambda) \frac{1}{Z'(e,f)} e_{x_1,x_3} f_{x_2}\}$$

Here: $\lambda \in [0,1]$, $c \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_2}$, $d \in \mathbb{R}_{>0}^{\Omega_3}$, $e \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_3}$, $f \in \mathbb{R}_{>0}^{\Omega_2}$

Challenge: Find a quantifier-free description of M!

Oosterhof found polynomial equations cutting out $\widehat{M}$ of degrees 3 and 6: certain 2×2-determinants of 3×3-determinants.

There's a beautiful relation with matrix spaces of rank two!

Many other topics

• equations for phylogenetic tree models (Casanellas et al, Allman-Rhodes, Sturmfels-Sullivant, Michałek et al, . . .)
• determinantal equations for Gaussian graphical models (Sullivant-Talaska-D)
• identifiability of Gaussian graphical models (Foygel-Drton-D)
• (non-)singularity of hypersurfaces defined by conditional independence statements for Gaussian graphical models (Lin-Uhler-Sturmfels-Bühlmann)
• twisted commutative algebras (Sam-Snowden)

$\vdots$

Many other topics

• equations for phylogenetic tree models (Casanellas et al, Allman-Rhodes, Sturmfels-Sullivant, Michałek et al, . . .)
• determinantal equations for Gaussian graphical models (Sullivant-Talaska-D)
• identifiability of Gaussian graphical models (Foygel-Drton-D)
• (non-)singularity of hypersurfaces defined by conditional independence statements for Gaussian graphical models (Lin-Uhler-Sturmfels-Bühlmann)
• twisted commutative algebras (Sam-Snowden)

⋮

Thank you!