# Computational problems in probabilistic reasoning

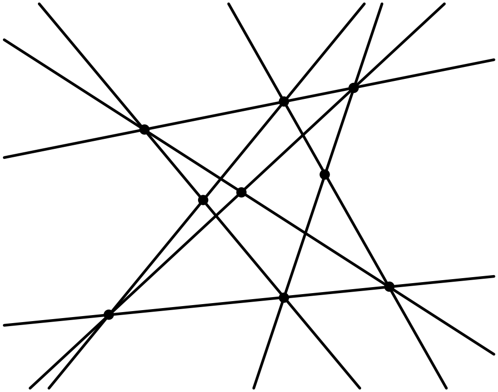
Tobias Boege
based on joint work with
Tabea Bacher and Ben Hollering

Aalto University
School of Science

SIAM AG 2023, TU Eindhoven,
Computational real algebraic geometry III,
13 July 2023

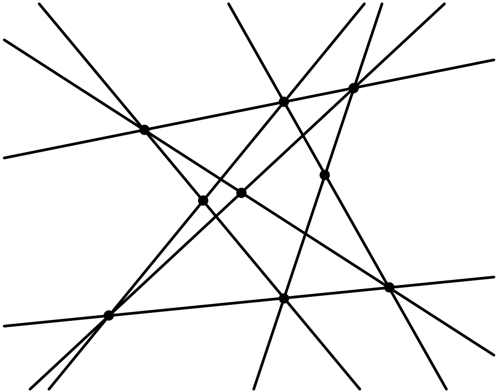
# Matroids and laws of geometry

- Matroids are combinatorial structures which model "special position" relations in geometry.



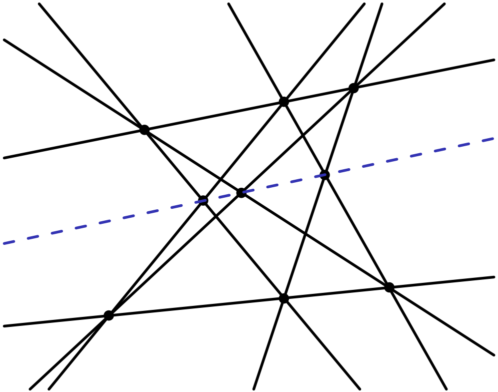
# Matroids and laws of geometry

- Matroids are combinatorial structures which model “special position” relations in geometry.
  - For example the matroid of a set of points in the projective plane records which triples of points lie on a line.



# Matroids and laws of geometry

- Matroids are combinatorial structures which model "special position" relations in geometry.
  - For example the matroid of a set of points in the projective plane records which triples of points lie on a line.
- Non-realizability of matroids captures the (non-obvious) laws of geometry.



# Conditional independence

Now think of $X, Y, Z$ as jointly distributed random variables instead of points in a common ambient space. The analogue of special position is:

# Conditional independence

Now think of $X, Y, Z$ as jointly distributed random variables instead of points in a common ambient space. The analogue of special position is:

## Conditional independence $X \perp\!\!\!\perp Y \mid Z$

"Does knowing $Z$ make $X$ irrelevant for $Y$?"

# Conditional independence

Now think of $X, Y, Z$ as jointly distributed random variables instead of points in a common ambient space. The analogue of special position is:

## Conditional independence $X \perp\!\!\!\perp Y \mid Z$

"Does knowing $Z$ make $X$ irrelevant for $Y$?"

**Laws of probabilistic reasoning**

*Let $X_1, \dots, X_n$ be jointly distributed random variables. Assume that $X_i \perp\!\!\!\perp X_j \mid X_K$ for some choices of $i, j \in [n]$ and $K \subseteq [n] \setminus \{i, j\}$. Which other CI statements $X_r \perp\!\!\!\perp X_s \mid X_T$ also hold?*

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector (“rank function”):

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector (“rank function”):

- $h(x)$: rank $\to$ entropy

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector (“rank function”):

- $h(x)$: rank → entropy
- $h(x, y, z) + h(z) = h(x, z) + h(y, z)$: modular pair → conditional independence

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector (“rank function”):

- $h(x)$: rank → entropy
- $h(x, y, z) + h(z) = h(x, z) + h(y, z)$: modular pair → conditional independence
- $h(x, y) = h(x) + h(y)$: independence → independence

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector (“rank function”):

- $h(x)$: rank → entropy
- $h(x, y, z) + h(z) = h(x, z) + h(y, z)$: modular pair → conditional independence
- $h(x, y) = h(x) + h(y)$: independence → independence
- $h(x, z) = h(z)$: closure operator → functional dependence

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector ("rank function"):

- $h(x)$: rank $\to$ entropy
- $h(x,y,z) + h(z) = h(x,z) + h(y,z)$: modular pair $\to$ conditional independence
- $h(x,y) = h(x) + h(y)$: independence $\to$ independence
- $h(x,z) = h(z)$: closure operator $\to$ functional dependence
- $h(x,z) = h(x) = h(z)$: parallel $\to$ functional equivalence

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector (“rank function”):

- $h(x)$: rank → entropy
- $h(x, y, z) + h(z) = h(x, z) + h(y, z)$: modular pair → conditional independence
- $h(x, y) = h(x) + h(y)$: independence → independence
- $h(x, z) = h(z)$: closure operator → functional dependence
- $h(x, z) = h(x) = h(z)$: parallel → functional equivalence
- $h(x) = 0$: loop → constant random variable

# Dictionary matroid theory — conditional independence

Special position properties of discrete random variables can be formulated in terms of linear functionals on the entropy vector (“rank function”):

- $h(x)$: rank → entropy
- $h(x,y,z) + h(z) = h(x,z) + h(y,z)$: modular pair → conditional independence
- $h(x,y) = h(x) + h(y)$: independence → independence
- $h(x,z) = h(z)$: closure operator → functional dependence
- $h(x,z) = h(x) = h(z)$: parallel → functional equivalence
- $h(x) = 0$: loop → constant random variable

Even though entropy is a transcendental function, all of these conditions are **polynomial** in the probabilities → algebraic statistics.

# Classification of binary CI models

**Goal:** **Create a database with all CI models of 4 binary random variables.**

# Classification of binary CI models

**Goal: Create a database with all CI models of 4 binary random variables.**

- The joint distribution of $n$ binary random variables $X_1, \dots, X_n$ is described by a $2 \times 2 \times \cdots \times 2$ tensor $p$ of non-negative real numbers which sum to 1:

$$p_{x_1 \dots x_n} = \Pr(X_i = x_i : i \in [n]).$$

# Classification of binary CI models

**Goal: Create a database with all CI models of 4 binary random variables.**

- The joint distribution of $n$ binary random variables $X_1, \ldots, X_n$ is described by a $2 \times 2 \times \cdots \times 2$ tensor $p$ of non-negative real numbers which sum to 1:

$$p_{x_1 \ldots x_n} = \Pr(X_i = x_i : i \in [n]).$$

- For $K \subseteq [n]$, the marginal $p^K$ has entries $p^K_{x_{i_1} \ldots x_{i_m}} = \Pr(X_{i_k} = x_{i_k} : k \in K)$.

# Classification of binary CI models

**Goal: Create a database with all CI models of 4 binary random variables.**

- The joint distribution of $n$ binary random variables $X_1, \dots, X_n$ is described by a $2 \times 2 \times \cdots \times 2$ tensor $p$ of non-negative real numbers which sum to 1:

$$p_{x_1 \dots x_n} = \Pr(X_i = x_i : i \in [n]).$$

- For $K \subseteq [n]$, the marginal $p^K$ has entries $p^K_{x_{i_1} \dots x_{i_m}} = \Pr(X_{i_k} = x_{i_k} : k \in K)$.
- The marginal CI statement $X_i \perp\!\!\!\perp X_j$ mandates the tensor decomposition:

$$p^{\{i,j\}} = p^i \otimes p^j.$$

# Classification of binary CI models

**Goal:** **Create a database with all CI models of 4 binary random variables.**

- The joint distribution of $n$ binary random variables $X_1, \ldots, X_n$ is described by a $2 \times 2 \times \cdots \times 2$ tensor $p$ of non-negative real numbers which sum to 1:

$$p_{x_1 \ldots x_n} = \Pr(X_i = x_i : i \in [n]).$$

- For $K \subseteq [n]$, the marginal $p^K$ has entries $p^K_{x_{i_1} \ldots x_{i_m}} = \Pr(X_{i_k} = x_{i_k} : k \in K)$.
- The marginal CI statement $X_i \perp\!\!\!\perp X_j$ mandates the tensor decomposition:

$$p^{\{i,j\}} = p^i \otimes p^j.$$

- The general CI statement $X_i \perp\!\!\!\perp X_j \mid X_K$ requires this decomposition for all $2^{|K|}$ slices of the marginal tensor $p^{\{i,j\} \cup K} \to$ **many** quadratic equations.

# Models and axioms

To every formula $\varphi = \bigwedge_p [X_{i_p} \perp\!\!\!\perp X_{j_p} \mid X_{K_p}] \Rightarrow \bigvee_q [X_{r_q} \perp\!\!\!\perp X_{s_q} \mid X_{T_q}]$ there is a semialgebraic set $K(\varphi)$ of counterexamples, i.e., real $2 \times 2 \times 2 \times 2$ tensors:

- $(\mathscr{P})$ with non-negative entries,
- $(\mathscr{I})$ satisfying all $X_{i_p} \perp\!\!\!\perp X_{j_p} \mid X_{K_p}$ but
- $(\mathscr{M})$ satisfying none of the $X_{r_q} \perp\!\!\!\perp X_{s_q} \mid X_{T_q}$.

# Models and axioms

To every formula $\varphi = \bigwedge_p [X_{i_p} \perp\!\!\!\perp X_{j_p} \mid X_{K_p}] \Rightarrow \bigvee_q [X_{r_q} \perp\!\!\!\perp X_{s_q} \mid X_{T_q}]$ there is a semialgebraic set $K(\varphi)$ of counterexamples, i.e., real $2 \times 2 \times 2 \times 2$ tensors:

($\mathscr{P}$) with non-negative entries,

($\mathscr{I}$) satisfying all $X_{i_p} \perp\!\!\!\perp X_{j_p} \mid X_{K_p}$ but

($\mathscr{M}$) satisfying none of the $X_{r_q} \perp\!\!\!\perp X_{s_q} \mid X_{T_q}$.

- $\varphi$ is valid (or an axiom) if and only if $K(\varphi) = \varnothing$.
- A set of CI statements implying nothing else is a model.

# Models and axioms

To every formula $\varphi = \bigwedge_p [X_{i_p} \perp\!\!\!\perp X_{j_p} \mid X_{K_p}] \Rightarrow \bigvee_q [X_{r_q} \perp\!\!\!\perp X_{s_q} \mid X_{T_q}]$ there is a semialgebraic set $K(\varphi)$ of counterexamples, i.e., real $2 \times 2 \times 2 \times 2$ tensors:

($\mathscr{P}$) with non-negative entries,

($\mathscr{I}$) satisfying all $X_{i_p} \perp\!\!\!\perp X_{j_p} \mid X_{K_p}$ but

($\mathscr{M}$) satisfying none of the $X_{r_q} \perp\!\!\!\perp X_{s_q} \mid X_{T_q}$.

- $\varphi$ is valid (or an axiom) if and only if $K(\varphi) = \emptyset$.
- A set of CI statements implying nothing else is a model.

## Conjecture

*The problem of deciding validity for binary distributions is $\forall\mathbb{R}$-complete. Moreover, all real algebraic numbers are necessary to certify invalidity.*

But in $n = 4$ we expect every model to be rationally realizable.

# Known laws I

**Theorem ([Mat18])**

*The following laws are valid and complete for 3 binary random variables* :*

$$[X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Z \mid Y] \Rightarrow [X \perp\!\!\!\perp Y \mid Z] \wedge [X \perp\!\!\!\perp Z] \tag{M1}$$

$$[X \perp\!\!\!\perp Y \mid Z] \wedge [X \perp\!\!\!\perp Z \mid Y] \Rightarrow [X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Z] \tag{M2}$$

$$[X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Y \mid Z] \Rightarrow [X \perp\!\!\!\perp Z] \vee [Y \perp\!\!\!\perp Z]. \tag{M3}$$

* If they satisfy no functional dependencies.

# Known laws I

**Theorem ([Mat18])**

*The following laws are valid and complete for 3 binary random variables* :*

$$[X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Z \mid Y] \Rightarrow [X \perp\!\!\!\perp Y \mid Z] \wedge [X \perp\!\!\!\perp Z] \tag{M1}$$

$$[X \perp\!\!\!\perp Y \mid Z] \wedge [X \perp\!\!\!\perp Z \mid Y] \Rightarrow [X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Z] \tag{M2}$$

$$[X \perp\!\!\!\perp Y] \wedge [X \perp\!\!\!\perp Y \mid Z] \Rightarrow [X \perp\!\!\!\perp Z] \vee [Y \perp\!\!\!\perp Z]. \tag{M3}$$

* If they satisfy no functional dependencies.

- SAT solvers can be used to derive more axioms logically implied by those above, to count or enumerate structures satisfying these axioms.

# Known laws II

**Theorem ([Šim07]*)**

*The following laws are valid for 4 binary random variables:*

$$[X \perp\!\!\!\perp Y \mid Z, W] \wedge [X \perp\!\!\!\perp Y \mid Z] \wedge [X \perp\!\!\!\perp W] \wedge [Z \perp\!\!\!\perp W] \;\Rightarrow\; [X \perp\!\!\!\perp W \mid Z] \vee [Y \perp\!\!\!\perp W] \tag{Š1}$$

$$[X \perp\!\!\!\perp Y \mid Z, W] \wedge [X \perp\!\!\!\perp Y \mid Z] \wedge [X \perp\!\!\!\perp W] \wedge [Y \perp\!\!\!\perp Z] \;\Rightarrow\; [X \perp\!\!\!\perp W \mid Y] \vee [Z \perp\!\!\!\perp W] \tag{Š2}$$

$$[X \perp\!\!\!\perp Y \mid Z, W] \wedge [X \perp\!\!\!\perp Y] \wedge [Y \perp\!\!\!\perp Z] \wedge [Z \perp\!\!\!\perp W] \wedge [X \perp\!\!\!\perp W] \;\Rightarrow\; [X \perp\!\!\!\perp W \mid Z] \vee [Z \perp\!\!\!\perp W \mid Y] \tag{Š3}$$

$$[X \perp\!\!\!\perp Y \mid Z, W] \wedge [X \perp\!\!\!\perp W \mid Y] \wedge [Z \perp\!\!\!\perp W \mid Y] \;\Rightarrow\; [Y \perp\!\!\!\perp Z \mid W] \vee [X \perp\!\!\!\perp Y \mid Z] \tag{Š4}$$

$$[X \perp\!\!\!\perp Y \mid Z, W] \wedge [X \perp\!\!\!\perp Y] \wedge [Y \perp\!\!\!\perp Z] \wedge [Y \perp\!\!\!\perp W] \wedge [X \perp\!\!\!\perp Y \mid Z] \wedge [Y \perp\!\!\!\perp Z \mid X] \;\Rightarrow\; [X \perp\!\!\!\perp W \mid Z] \vee [X \perp\!\!\!\perp Y \mid W]. \tag{Š5}$$

* (Š3) was incorrect in [Šim07].

# Audience participation

What is the vanishing ideal of the set of real non-negative $2 \times 2 \times 2 \times 2$ tensors $p$ which satisfy

$$\left.\begin{array}{ll} p_{0000}p_{1100} = p_{0100}p_{1000} & p_{0001}p_{1101} = p_{0101}p_{1001} \\ p_{0010}p_{1110} = p_{0110}p_{1010} & p_{0011}p_{1111} = p_{0111}p_{1011} \end{array}\right\} \quad [X \perp\!\!\!\perp Y \mid Z, W]$$

$$\left.\begin{array}{l} (p_{0000} + p_{0001})(p_{1010} + p_{1011}) = (p_{0010} + p_{0011})(p_{1000} + p_{1001}) \\ (p_{0100} + p_{0101})(p_{1110} + p_{1111}) = (p_{0110} + p_{0111})(p_{1100} + p_{1101}) \end{array}\right\} \quad [X \perp\!\!\!\perp Z \mid Y]$$

$$\left.\begin{array}{l} (p_{0000} + p_{0010})(p_{1001} + p_{1011}) = (p_{0001} + p_{0011})(p_{1000} + p_{1010}) \\ (p_{0100} + p_{0110})(p_{1101} + p_{1111}) = (p_{0101} + p_{0111})(p_{1100} + p_{1110}) \end{array}\right\} \quad [X \perp\!\!\!\perp W \mid Y]$$

# Classification of binary CI models

**Current state**:

# Classification of binary CI models

**Current state**:

- Matúš's axioms permit 178 models up to $S_4$ symmetry. SAT computation

# Classification of binary CI models

**Current state**:

- Matúš's axioms permit 178 models up to $S_4$ symmetry. SAT computation
- [Mat18] solves all cases which do not contain $[X \perp\!\!\!\perp Y \mid Z, W] \rightarrow 104$. linear program

# Classification of binary CI models

**Current state**:

- Matúš's axioms permit 178 models up to $S_4$ symmetry. SAT computation
- [Mat18] solves all cases which do not contain $[X \perp\!\!\!\perp Y \mid Z, W] \to 104$. linear program
- Šimeček's axioms handle some of those cases $\to 91$. SAT computation

# Classification of binary CI models

**Current state**:

- Matúš's axioms permit 178 models up to $S_4$ symmetry. SAT computation
- [Mat18] solves all cases which do not contain $[X \perp\!\!\!\perp Y \mid Z, W] \to 104$. linear program
- Šimeček's axioms handle some of those cases $\to$ 91. SAT computation
- Every model realizable by a regular Gaussian [LM07] on $n = 4$ is binary $\to$ 57.

# Classification of binary CI models

**Current state**:

- Matúš's axioms permit 178 models up to $S_4$ symmetry. SAT computation
- [Mat18] solves all cases which do not contain $[X \perp\!\!\!\perp Y \mid Z, W] \to 104$. linear program
- Šimeček's axioms handle some of those cases $\to$ 91. SAT computation
- Every model realizable by a regular Gaussian [LM07] on $n = 4$ is binary $\to$ 57.
- Sampling of $2 \times 2 \times 2 \times 2$ tensors $\to$ 39

# Classification of binary CI models

**Current state**:

- Matúš's axioms permit 178 models up to $S_4$ symmetry. SAT computation
- [Mat18] solves all cases which do not contain $[X \perp\!\!\!\perp Y \mid Z, W] \to 104$. linear program
- Šimeček's axioms handle some of those cases $\to$ 91. SAT computation
- Every model realizable by a regular Gaussian [LM07] on $n = 4$ is binary $\to$ 57.
- Sampling of $2 \times 2 \times 2 \times 2$ tensors $\to$ 39

Open e.g. $[X \perp\!\!\!\perp Y \mid Z, W] \wedge [X \perp\!\!\!\perp Z \mid Y] \wedge [X \perp\!\!\!\perp W \mid Y] \;\Rightarrow\; ?$

# Approach

- Change to moment coordinates!

# Approach

- Change to moment coordinates!
- Get conjectures for laws via numerical samples from CI varieties and try to prove them symbolically. (Analogous to numerical irreducible decomposition.)

# Approach

- Change to moment coordinates!
- Get conjectures for laws via numerical samples from CI varieties and try to prove them symbolically. (Analogous to numerical irreducible decomposition.)
- Certify numerically obtained counterexamples.

# Approach

- Change to moment coordinates!
- Get conjectures for laws via numerical samples from CI varieties and try to prove them symbolically. (Analogous to numerical irreducible decomposition.)
- Certify numerically obtained counterexamples.
- Factor out symmetry in polynomial systems.

# Approach

- Change to moment coordinates!
- Get conjectures for laws via numerical samples from CI varieties and try to prove them symbolically. (Analogous to numerical irreducible decomposition.)
- Certify numerically obtained counterexamples.
- Factor out symmetry in polynomial systems.
- Possibly CAD-based symbolic counterexamples in the future.

# Approach

- Change to moment coordinates!
- Get conjectures for laws via numerical samples from CI varieties and try to prove them symbolically. (Analogous to numerical irreducible decomposition.)
- Certify numerically obtained counterexamples.
- Factor out symmetry in polynomial systems.
- Possibly CAD-based symbolic counterexamples in the future.

**Theorem (Normal forms for proof and refutation)**

*The formula $\varphi$ is invalid if and only if $K(\varphi)$ contains a point $p \in \mathbb{R}^{2\times2\times2\times2}$ whose entries are algebraic over $\mathbb{Q}$. On the other hand, $\varphi$ is valid if and only if there are polynomials $f \in \mathscr{I}(\varphi), g \in \mathscr{P}(\varphi), h \in \mathscr{M}(\varphi)$ such that $f + g + h^2 = 0 \in \mathbb{Z}[p]$.*

# Approach

- Change to moment coordinates!
- Get conjectures for laws via numerical samples from CI varieties and try to prove them symbolically. (Analogous to numerical irreducible decomposition.)
- Certify numerically obtained counterexamples.
- Factor out symmetry in polynomial systems.
- Possibly CAD-based symbolic counterexamples in the future.

**Theorem (Normal forms for proof and refutation)**

*The formula $\varphi$ is invalid if and only if $K(\varphi)$ contains a point $p \in \mathbb{R}^{2\times2\times2\times2}$ whose entries are algebraic over $\mathbb{Q}$. On the other hand, $\varphi$ is valid if and only if there are polynomials $f \in \mathscr{I}(\varphi), g \in \mathscr{P}(\varphi), h \in \mathscr{M}(\varphi)$ such that $f + g + h^2 = 0 \in \mathbb{Z}[p]$.*

These certificates are not used in practice. Why?

# References


Radim Lněnička and František Matúš. “On Gaussian conditional independence structures”. eng. In: *Kybernetika* 43.3 (2007), pp. 327–342.

František Matúš. “On patterns of conditional independences and covariance signs among binary variables”. In: *Acta Math. Hung.* 154.2 (2018), pp. 511–524. ISSN: 0236-5294. DOI: 10.1007/s10474-018-0799-6.

Petr Šimeček. “Nezávislostní modely”. In Czech. Dissertation. Charles University, Prague, Czech Republic, 2007.