Tobias Boege

The Gaussian conditional independence inference problem

Magdeburg, April 22, 2022

Institut für Algebra und Geometrie
Otto-von-Guericke-Universität Magdeburg

DFG-Graduiertenkolleg
MATHEMATISCHE
KOMPLEXITÄTSREDUKTION

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

"When does knowing Z make X irrelevant for Y?"

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

"When does knowing Z make X irrelevant for Y?"

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

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

"When does knowing Z make X irrelevant for Y?"

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

- $c_1 \perp\!\!\!\perp c_2$

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

"When does knowing Z make X irrelevant for Y?"

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

- $c_1 \perp\!\!\!\perp c_2$
- $\neg(c_1 \perp\!\!\!\perp c_2 \mid b)$ ...

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

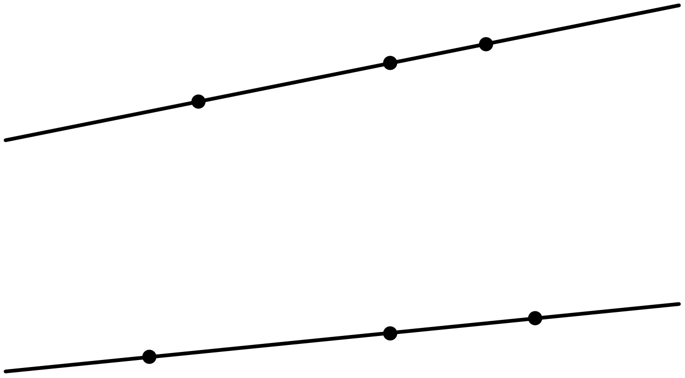
"When does knowing Z make X irrelevant for Y?"

Example: Two independent fair coins c_1 and c_2 are wired to a bell b which rings if and only if $c_1 = c_2$.

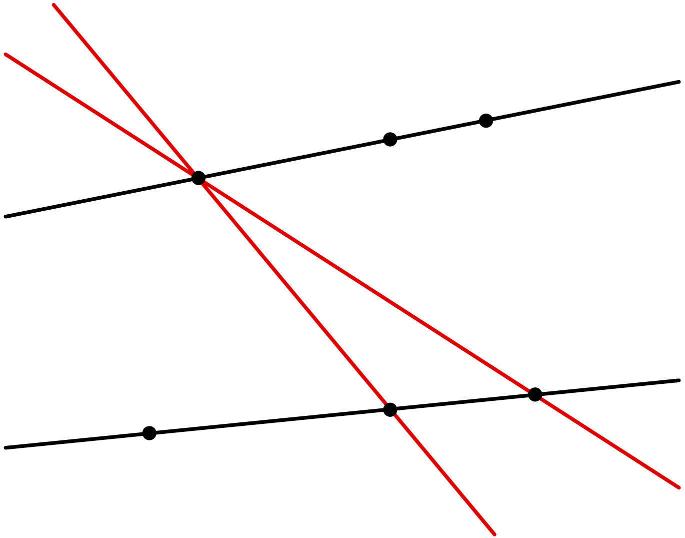
- $c_1 \perp\!\!\!\perp c_2$
- $\neg(c_1 \perp\!\!\!\perp c_2 \mid b)$ ...

Question: When can we conclude from some independences other independences?
E.g., is it possible that $c_1 \perp\!\!\!\perp b$?

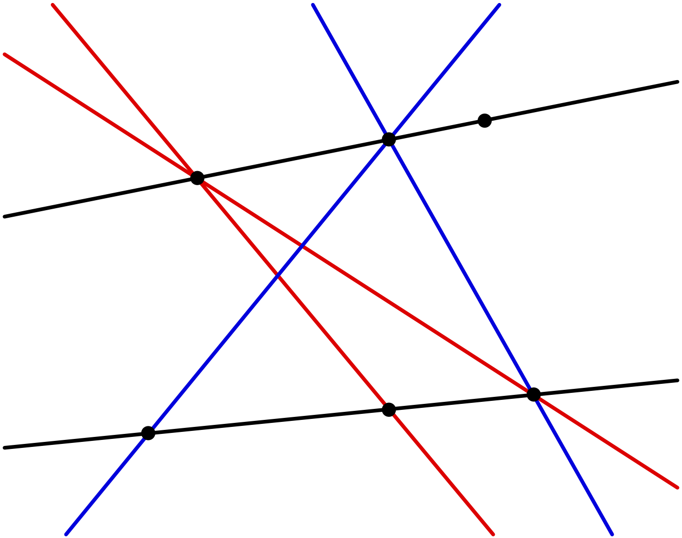
For ancient geometers: conditional independence ≈ collinearity



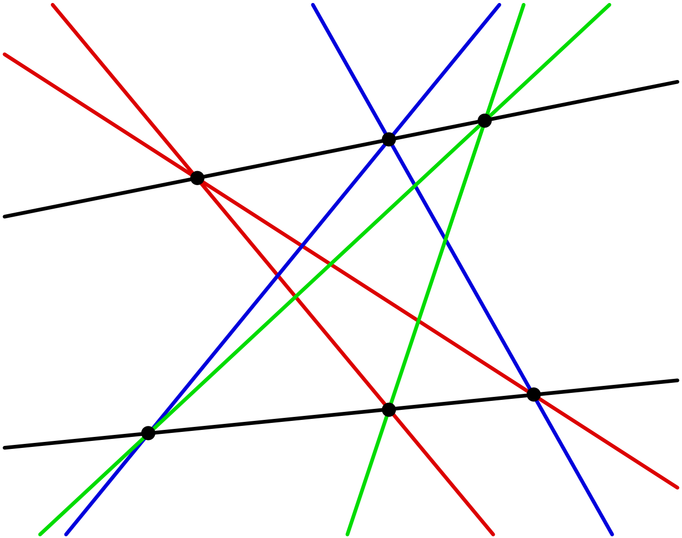
For ancient geometers: conditional independence ≈ collinearity



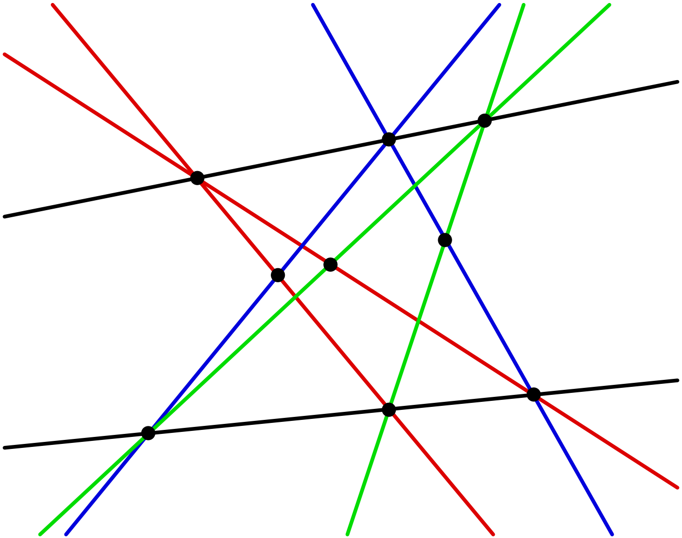
For ancient geometers: conditional independence ≈ collinearity



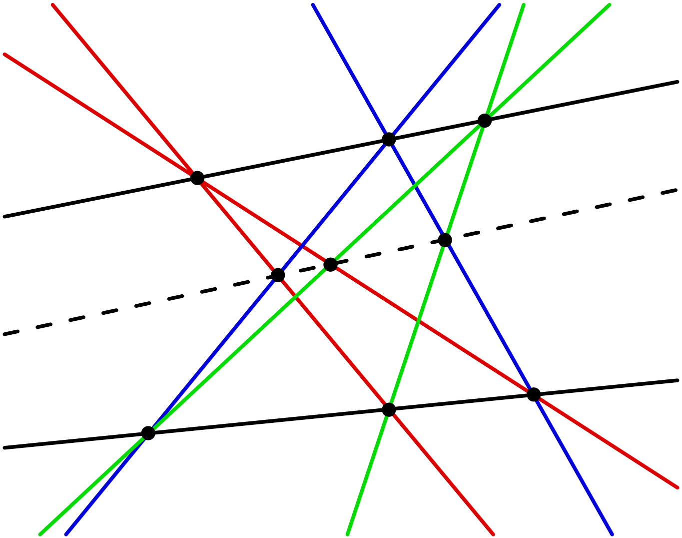
For ancient geometers: conditional independence ≈ collinearity



For ancient geometers: conditional independence ≈ collinearity



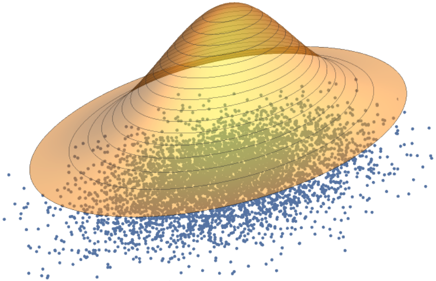
For ancient geometers: conditional independence ≈ collinearity



Gaussian distributions

The density of the multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ is

$$x \mapsto \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right).$$



Reasoning with graphical models

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR020412

Key Points:
- We develop a statistical graphical model to characterize the statewide California reservoir system
- We quantify the influence of external physical and economic factors (e.g., statewide PDSI and consumer price index) on the reservoir network
- Further analysis gives a system-wide health diagnosis as a function of PDSI, indicating when heavy management practices may be needed

Supporting Information:
- Supporting Information S1
- Supporting Information S2

A Statistical Graphical Model of the California Reservoir System

A. Taeb[1], J. T. Reager[2], M. Turmon[2], and V. Chandrasekaran[3]

[1]Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA, [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [3]Department of Computing and Mathematical Sciences and Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA

Abstract The recent California drought has highlighted the potential vulnerability of the state's water management infrastructure to multiyear dry intervals. Due to the high complexity of the network, dynamic storage changes in California reservoirs on a state-wide scale have previously been difficult to model using either traditional statistical or physical approaches. Indeed, although there is a significant line of research on exploring models for single (or a small number of) reservoirs, these approaches are not amenable to a system-wide modeling of the California reservoir network due to the spatial and hydrological heterogeneities of the system. In this work, we develop a state-wide statistical graphical model to characterize the dependencies among a collection of 55 major California reservoirs across the state; this model is defined with respect to a graph in which the nodes index reservoirs and the edges specify the

Gaussian conditional independence

Suppose a random vector $\xi = (\xi_i)_{i \in N}$ is normally distributed: $\xi \sim \mathcal{N}(\mu, \Sigma)$.

Definition

The polynomial $\Sigma[K] \coloneqq \det \Sigma_{K,K}$ is a *principal minor* of Σ and $\Sigma[ij \mid K] \coloneqq \det \Sigma_{iK,jK}$ is an *almost-principal minor*.

Gaussian conditional independence

Suppose a random vector $\xi = (\xi_i)_{i \in N}$ is normally distributed: $\xi \sim \mathcal{N}(\mu, \Sigma)$.

Definition

The polynomial $\Sigma[K] \coloneqq \det \Sigma_{K,K}$ is a *principal minor* of Σ and $\Sigma[ij \mid K] \coloneqq \det \Sigma_{iK,jK}$ is an *almost-principal minor*.

- Σ is positive-definite $\Leftrightarrow$ $\Sigma[K] > 0$.
- $[\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]$ holds $\Leftrightarrow$ $\Sigma[ij \mid K] = 0$.
- $\mathbb{E}[\xi] = \mu$ is irrelevant.

Models and inference

Definition

A *CI constraint* is a CI statement $[\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]$ or its negation $\neg[\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]$. They are *algebraic conditions* on the entries of Σ, equivalent to vanishing or non-vanishing of the almost-principal minors $\Sigma[ij \mid K]$.

Definition

The *model* of a set of CI constraints is the set of all positive-definite matrices which satisfy the constraints. The constraints are *feasible* if the model is non-empty.

Models and inference

Consider two sets of CI statements $\mathcal{L}$ and $\mathcal{M}$:

$$\bigwedge \mathcal{L} \Rightarrow \bigvee \mathcal{M}$$

Models and inference

Consider two sets of CI statements $\mathcal{L}$ and $\mathcal{M}$:

$$\bigwedge \mathcal{L} \Rightarrow \bigvee \mathcal{M} \text{ is not valid} \iff \mathcal{L} \cup \neg\mathcal{M} \text{ is feasible}$$

Models and inference

Consider two sets of CI statements $\mathcal{L}$ and $\mathcal{M}$:

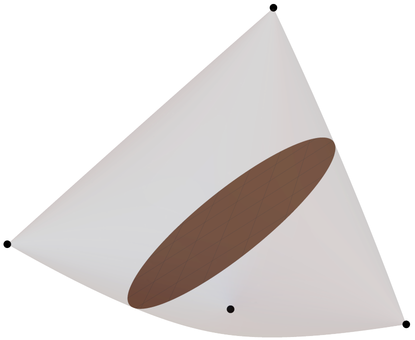
$$\bigwedge \mathcal{L} \Rightarrow \bigvee \mathcal{M} \text{ is not valid} \iff \mathcal{L} \cup \neg\mathcal{M} \text{ is feasible}$$

Inference rules are patterns of valid reasoning about relevance among jointly Gaussian random variables. This is equivalent to a purely semialgebraic problem.

Geometry of inference: Example

$$\Sigma = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}$$

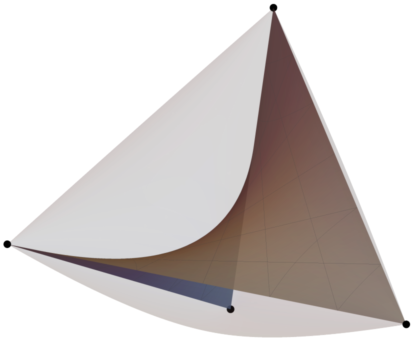
- If $\Sigma[12|] = a$



Geometry of inference: Example

$$\Sigma = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}$$

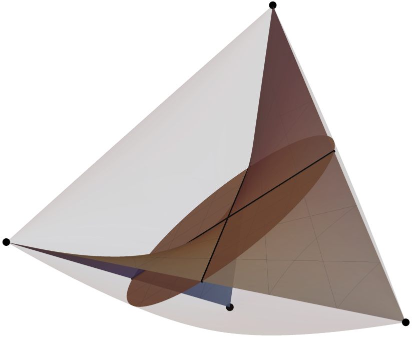
- If $\Sigma[12|] = a$ and $\Sigma[12|3] = a - bc$ vanish,



Geometry of inference: Example

$$\Sigma = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}$$

- If $\Sigma[12|] = a$ and $\Sigma[12|3] = a - bc$ vanish, then $bc = \Sigma[13|] \cdot \Sigma[23|]$ must vanish

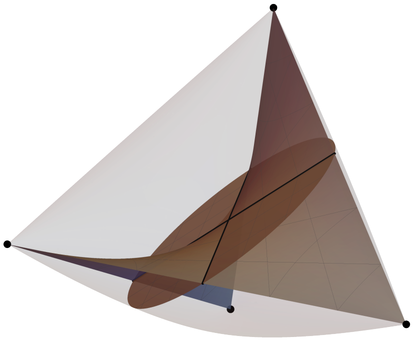


Geometry of inference: Example

$$\Sigma = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}$$

- If $\Sigma[12|] = a$ and $\Sigma[12|3] = a - bc$ vanish, then $bc = \Sigma[13|] \cdot \Sigma[23|]$ must vanish:

$$[12|] \wedge [12|3] \Rightarrow [13|] \vee [23|].$$



Computer algebra proves laws of probabilistic reasoning

$$[12|] \wedge [14|5] \wedge [23|5] \wedge [35|1] \wedge [45|2] \wedge [15|23] \wedge [34|12] \wedge [24|135] \;\Rightarrow\; [25|] \vee [34|].$$

Computer algebra proves laws of probabilistic reasoning

$$[12|] \wedge [14|5] \wedge [23|5] \wedge [35|1] \wedge [45|2] \wedge [15|23] \wedge [34|12] \wedge [24|135] \;\Rightarrow\; [25|] \vee [34|].$$

$$\Sigma[25|] \cdot \Sigma[34|] = \frac{1}{pqr(pt-d^2)} \cdot$$

$$\Big[\big(cd^2egr + bd^2fgr - ad^2grh - 2cd^2e^2i - 2bd^2efi - 2pdfgri + 2ad^2ehi + 2pdefi^2 - 2pdqhi^2 + 2pcqi^3 +$$

$$2pdqrij - 2pbqi^2j - pcegrt + pbfgrt + pagrht + 2pce^2it - 2pcqrit + 2pbqhit - 2paehit \big) \cdot \Sigma[12|] +$$

$$\big(pdqer + pbqgr - 2pbqei \big) \cdot \Sigma[14|5] - \big(pcdqr + p^2fgr - 2pbcqi + 2pb^2qj - 2p^2qrj \big) \cdot \Sigma[23|5] +$$

$$\big(cdqgr - 2cdqei + 2pqghi - 2pqfi^2 - pqgrj + 2pqeij - 2pe^2ft + 2pqfrt \big) \cdot \Sigma[35|1] +$$

$$\big(pd^2er - 2pbdei + p^2gri + 2pb^2et - 2p^2ert \big) \cdot \Sigma[45|2] - \big(2pdfi - 2pbft \big) \cdot \Sigma[15|23] -$$

$$\big(d^2gr - 2d^2ei - pgrt + 2peit \big) \cdot \Sigma[34|12] - 2pqi \cdot \Sigma[24|135] \Big].$$

The gaussoid axioms

Using algebraic tools Matúš derived basic inference rules for all regular Gaussian distributions, the gaussoid axioms:

$$
\begin{aligned}
[ij \mid L] \wedge [ik \mid jL] &\Rightarrow [ik \mid L] \wedge [ij \mid kL] \\
[ij \mid kL] \wedge [ik \mid jL] &\Rightarrow [ij \mid L] \wedge [ik \mid L] \\
[ij \mid L] \wedge [ik \mid L] &\Rightarrow [ij \mid kL] \wedge [ik \mid jL] \\
[ij \mid L] \wedge [ij \mid kL] &\Rightarrow [ik \mid L] \vee [jk \mid L]
\end{aligned}
$$

The gaussoid axioms

Using algebraic tools Matúš derived basic inference rules for all regular Gaussian distributions, the gaussoid axioms:

$$
\begin{aligned}
[ij \mid L] \wedge [ik \mid jL] &\Rightarrow [ik \mid L] \wedge [ij \mid kL] \\
[ij \mid kL] \wedge [ik \mid jL] &\Rightarrow [ij \mid L] \wedge [ik \mid L] \\
[ij \mid L] \wedge [ik \mid L] &\Rightarrow [ij \mid kL] \wedge [ik \mid jL] \\
[ij \mid L] \wedge [ij \mid kL] &\Rightarrow [ik \mid L] \vee [jk \mid L]
\end{aligned}
$$

Theorem (Sullivant / Šimeček)

There is no finite set of valid CI axioms which imply all valid inference rules for regular Gaussian distributions.

Software for computations with axioms

```
use Modern::Perl 2018;
use CInet::Base;
use CInet::Propositional;

propositional Gaussoids = cube(ijk|L) ::
    (ij|L)  & (ik|jL) => (ik|L)  & (ij|kL),
    (ij|kL) & (ik|jL) => (ij|L)  & (ik|L),
    (ij|L)  & (ik|L)  => (ij|kL) & (ik|jL),
    (ij|L)  & (ij|kL) => (ik|L)  | (jk|L);

say my $count = Gaussoids(4)->count;
say Gaussoids(4)->modulo(SymmetricGroup)->count;
say Gaussoids(4)->reduce(sub{ $a + $b->independences }, 0) / $count;
#= 679, 58, 3.958
```

Two-antecedental completeness

Theorem

Suppose $\varphi : \bigwedge \mathcal{L} \Rightarrow \bigvee \mathcal{M}$ *is valid for all regular Gaussian distributions.*

- *If* $|\mathcal{L}| \leq 1$, *then* $\mathcal{L} \subseteq \mathcal{M}$ *and* φ *is trivial.*
- *If* $|\mathcal{L}| = 2$, *then* φ *is implied by the gaussoid axioms.*

Two-antecedental completeness

Theorem

Suppose $\varphi : \bigwedge \mathcal{L} \Rightarrow \bigvee \mathcal{M}$ *is valid for all regular Gaussian distributions.*

- *If* $|\mathcal{L}| \leq 1$, *then* $\mathcal{L} \subseteq \mathcal{M}$ *and* φ *is trivial.*
- *If* $|\mathcal{L}| = 2$, *then* φ *is implied by the gaussoid axioms.*

i	$A^{(i)}$	i	$A^{(i)}$	i	$A^{(i)}$	i	$A^{(i)}$
1	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon & 0 & 1 & 0 \\ \varepsilon^2 & \varepsilon & 0 & 1 \end{pmatrix}$	2	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 1 & 0 & \varepsilon^2 \\ \varepsilon & 0 & 1 & 0 \\ \varepsilon & \varepsilon^2 & 0 & 1 \end{pmatrix}$	3	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & 1\text{-}\varepsilon^2 \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon & 0 & 1 & 0 \\ 1\text{-}\varepsilon^2 & \varepsilon & 0 & 1 \end{pmatrix}$	4	$\begin{pmatrix} 1 & 1\text{-}\varepsilon^2 & \varepsilon^2 & 0 \\ 1\text{-}\varepsilon^2 & 1 & 0 & \varepsilon \\ \varepsilon^2 & 0 & 1 & \text{-}\varepsilon \\ 0 & \varepsilon & \text{-}\varepsilon & 1 \end{pmatrix}$
6	$\begin{pmatrix} 1 & \varepsilon^2 & \varepsilon^2 & 0 \\ \varepsilon^2 & 1 & 0 & \varepsilon \\ \varepsilon^2 & 0 & 1 & \text{-}\varepsilon \\ 0 & \varepsilon & \text{-}\varepsilon & 1 \end{pmatrix}$	7	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & 0 \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon & 0 & 1 & \text{-}\varepsilon \\ 0 & \varepsilon & \text{-}\varepsilon & 1 \end{pmatrix}$	8	$\begin{pmatrix} 1 & \varepsilon & \varepsilon^2 & \varepsilon \\ \varepsilon & 1 & 0 & \varepsilon \\ \varepsilon^2 & 0 & 1 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 1 \end{pmatrix}$	9	$\begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 \\ \varepsilon & 1 & 0 & \varepsilon^2 \\ \varepsilon & 0 & 1 & \varepsilon \\ \varepsilon^2 & \varepsilon^2 & \varepsilon & 1 \end{pmatrix}$
	$\begin{pmatrix} 1 & \varepsilon & \varepsilon^2 & \varepsilon \end{pmatrix}$		$\begin{pmatrix} 1 & \varepsilon & \varepsilon^3 & \varepsilon^2 \end{pmatrix}$		$\begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 \end{pmatrix}$		$\begin{pmatrix} 1 & \text{-}\varepsilon & \varepsilon & \varepsilon \end{pmatrix}$

Universality theorems

How difficult is it to solve the inference problem for a given formula:

- algebraically?

Universality theorems

How difficult is it to solve the inference problem for a given formula:

- algebraically?
- computationally?

Universality theorems

How difficult is it to solve the inference problem for a given formula:

- algebraically?
- computationally?
- topologically?

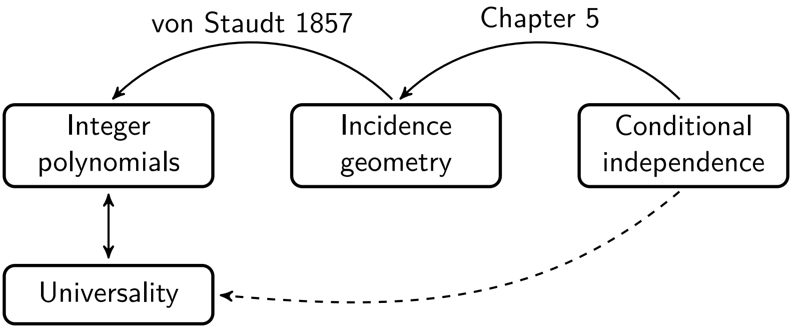
Universality theorems

How difficult is it to solve the inference problem for a given formula:

- algebraically?
- computationally?
- topologically?

Universality theorems: As hard as it could possibly be.

Universality theorems: Background



Theorem

To every polynomial system $\{f_i \bowtie 0\}$ there is a set of CI constraints which has a model over a field $\mathbb{K}/\mathbb{Q}$ if and only if the polynomial system has a solution in $\mathbb{K}$.

Universality theorems I: Algebraic numbers

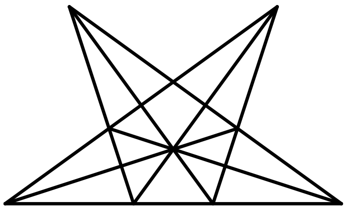
Šimeček's Question

Does every non-empty Gaussian CI model contain a rational point?

Universality theorems I: Algebraic numbers

Šimeček's Question

Does every non-empty Gaussian CI model contain a rational point?



The Perles configuration requires $\sqrt{5}$ to be realized.

Universality theorems I: Algebraic numbers

Šimeček's Question

Does every non-empty Gaussian CI model contain a rational point?

Theorem

For every finite real extension $\mathbb{K}/\mathbb{Q}$ there exists a Gaussian CI model $\mathcal{M}_{\mathbb{K}}$ such that: for every $\mathbb{L}/\mathbb{Q}$, $\mathcal{M}_{\mathbb{K}}$ has an $\mathbb{L}$-rational point if and only if $\mathbb{K} \subseteq \mathbb{L}$.

Universality theorems II: Complexity class

The complexity class $\exists\mathbb{R}$ contains all decision problems which can be reduced in polynomial time to the feasibility of a semialgebraic set:

Universality theorems II: Complexity class

The complexity class $\exists\mathbb{R}$ contains all decision problems which can be reduced in polynomial time to the feasibility of a semialgebraic set:

- polynomial optimization

Universality theorems II: Complexity class

The complexity class $\exists\mathbb{R}$ contains all decision problems which can be reduced in polynomial time to the feasibility of a semialgebraic set:

- polynomial optimization
- computational geometry

Universality theorems II: Complexity class

The complexity class $\exists\mathbb{R}$ contains all decision problems which can be reduced in polynomial time to the feasibility of a semialgebraic set:

- polynomial optimization
- computational geometry
- algebraic statistics ...

Universality theorems II: Complexity class

The complexity class $\exists\mathbb{R}$ contains all decision problems which can be reduced in polynomial time to the feasibility of a semialgebraic set:

- polynomial optimization
- computational geometry
- algebraic statistics ...

Theorem

The Gaussian CI inference problem is co-$\exists\mathbb{R}$-complete.

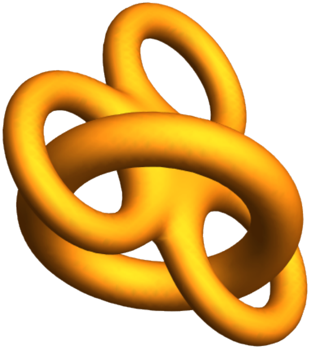
Universality theorems III: Homotopy type

In an oriented Gaussian CI model we incorporate the sign of almost-principal minors. Instead of just revealing dependence, this tells whether the variables are correlated positively or negatively.

Theorem

Oriented Gaussian CI models attain the homotopy types of all primary basic semialgebraic sets.

The space of counterexamples to an inference formula decomposes into oriented CI models →



Thank you for your attention