Tobias Boege

The Gaussian CI inference problem

Mini-Symposium (Decision Making Theory), UTIA Prague, 13 September 2021

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

DFG-Graduiertenkolleg
MATHEMATISCHE
KOMPLEXITÄTSREDUKTION

Gaussian conditional independence

Consider random variables $(\xi_i)_{i \in N} \sim \mathcal{N}(\mu, \Sigma)$. The *conditional independence (CI) statement* $\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K$ conveys, informally, that if ξ_K is known, then learning the value of ξ_i does not give any information about ξ_j.

Definition

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

If Σ is positive-definite, then $\Sigma[K] > 0$, and $\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K$ holds if and only if $\Sigma[ij|K] = 0$.

Almost-principal minors

$$
\begin{aligned}
\Sigma[ij|] &= x_{ij} \\
\Sigma[ij|k] &= x_{ij}x_{kk} - x_{ik}x_{jk} \\
\Sigma[ij|kl] &= x_{ij}x_{kk}x_{ll} - x_{il}x_{jl}x_{kk} + x_{il}x_{jk}x_{kl} + x_{ik}x_{jl}x_{kl} - x_{ij}x_{kl}^2 - x_{ik}x_{jk}x_{ll} \\
\Sigma[ij|klm] &= x_{ij}x_{kk}x_{ll}x_{mm} + x_{im}x_{jm}x_{kl}^2 - x_{im}x_{jl}x_{kl}x_{km} - x_{il}x_{jm}x_{kl}x_{km} + x_{il}x_{jl}x_{km}^2 \\
&\quad - x_{im}x_{jm}x_{kk}x_{ll} + x_{im}x_{jk}x_{km}x_{ll} + x_{ik}x_{jm}x_{km}x_{ll} - x_{ij}x_{km}^2x_{ll} \\
&\quad + x_{im}x_{jl}x_{kk}x_{lm} + x_{il}x_{jm}x_{kk}x_{lm} - x_{im}x_{jk}x_{kl}x_{lm} - x_{ik}x_{jm}x_{kl}x_{lm} \\
&\quad - x_{il}x_{jk}x_{km}x_{lm} - x_{ik}x_{jl}x_{km}x_{lm} + 2x_{ij}x_{kl}x_{km}x_{lm} + x_{ik}x_{jk}x_{lm}^2 \\
&\quad - x_{ij}x_{kk}x_{lm}^2 - x_{il}x_{jl}x_{kk}x_{mm} + x_{il}x_{jk}x_{kl}x_{mm} + x_{ik}x_{jl}x_{kl}x_{mm} \\
&\quad - x_{ij}x_{kl}^2x_{mm} - x_{ik}x_{jk}x_{ll}x_{mm} \\
&\vdots
\end{aligned}
$$

Gaussian CI models

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 $\Sigma[ij|K]$.

Definition

The *model* of a set of CI constraints is the set of all positive-definite matrices which satisfy the constraints.

Gaussian CI models

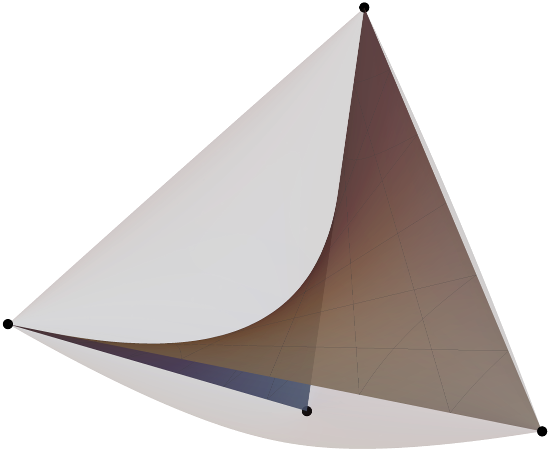


Figure: Gaussian model $\Sigma[12|3] = 0$ inside the elliptope.

Models and inference

Consider two sets of CI statements $\mathcal{P}$ and $\mathcal{Q}$:

$$\bigwedge \mathcal{P} \Rightarrow \bigvee \mathcal{Q}$$

Models and inference

Consider two sets of CI statements $\mathcal{P}$ and $\mathcal{Q}$:

$$\bigwedge \mathcal{P} \Rightarrow \bigvee \mathcal{Q} \text{ is not valid} \iff \mathcal{P} \cup \neg\mathcal{Q} \text{ has a model}$$

Models and inference

Consider two sets of CI statements $\mathcal{P}$ and $\mathcal{Q}$:

$$\underset{\text{is not valid}}{\bigwedge \mathcal{P} \Rightarrow \bigvee \mathcal{Q}} \iff \underset{\text{has a model}}{\mathcal{P} \cup \neg \mathcal{Q}}$$

Reasoning about relevance statements in normally distributed random variables is the same as reasoning about the vanishing of very special kinds of determinants on very special kinds of varieties inside the positive-definite matrices.

Examples of CI inference

Consider a general positive-definite 3×3 correlation matrix

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

- If $\Sigma[12|3] = a - bc$ and $\Sigma[13|] = b$ vanish, then $\Sigma[12|] = a$ and $\Sigma[13|2] = b - ac$ must vanish as well:

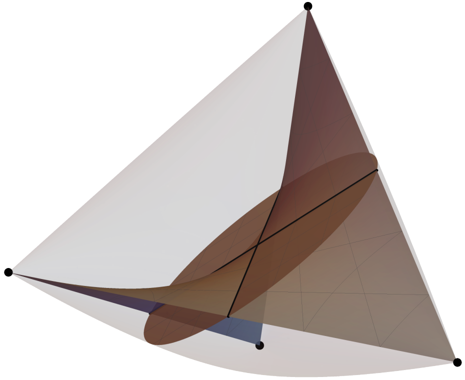
$$\begin{aligned} (12|3) \wedge (13|) &\Rightarrow (12|), \\ (12|3) \wedge (13|) &\Rightarrow (13|2). \end{aligned}$$

Examples of CI inference

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

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

$$(12|) \wedge (12|3) \;\Rightarrow\; (13|) \vee (23|).$$



No finite set of axioms

"There is no finite complete axiomatization of Gaussian CI":

Theorem (Sullivant 2009)

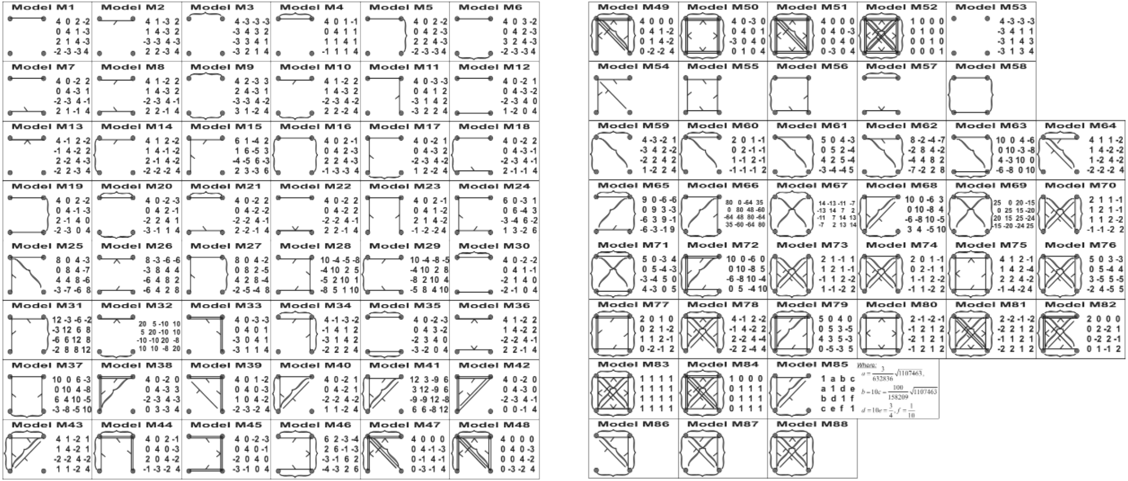
As the matrix size n grows, there exist valid inference rules for Gaussians which need arbitrarily many antecedents.

$$(12|3) \wedge (23|4) \wedge (34|1) \wedge (14|2) \;\Rightarrow\; (12|) \qquad (n = 4)$$

$$(12|3) \wedge (23|4) \wedge (34|5) \wedge (45|1) \wedge (15|2) \;\Rightarrow\; (12|) \qquad (n = 5)$$

$$(12|3) \wedge (23|4) \wedge (34|5) \wedge (45|6) \wedge (56|1) \wedge (16|2) \;\Rightarrow\; (12|) \qquad (n = 6)$$

$$\vdots$$

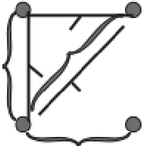


Petr Šimeček. Gaussian representation of independence models over four random variables.
In *COMPSTAT conference*, 2006.

Šimeček's Question

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

Model M85



$$\begin{pmatrix} 1 & a & b & c \\ a & 1 & d & e \\ b & d & 1 & f \\ c & e & f & 1 \end{pmatrix}$$

Where:

$$a = \frac{3}{632836}\sqrt{1107463},$$

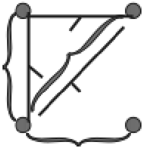
$$b = 10c = \frac{100}{158209}\sqrt{1107463}$$

$$d = 10e = \frac{3}{4}, f = \frac{1}{10}$$

Šimeček's Question

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

Model M85



$$\begin{matrix} 1 & a & b & c \\ a & 1 & d & e \\ b & d & 1 & f \\ c & e & f & 1 \end{matrix}$$

Where:

$$a = \frac{3}{632836}\sqrt{1107463},$$

$$b = 10c = \frac{100}{158209}\sqrt{1107463}$$

$$d = 10e = \frac{3}{4}, f = \frac{1}{10}$$

$$\begin{pmatrix} 357 & -21 & -343 & -147 \\ -21 & 357 & 119 & 51 \\ -343 & 119 & 357 & 153 \\ -147 & 51 & 153 & 357 \end{pmatrix}$$

Complexity bounds from real geometry

Theorem (Tarski's transfer principle)

If a polynomial system $\{f_i = 0, g_j > 0, h_k \neq 0\}$ has a solution over $\mathbb{R}$, then it has a solution in a finite real extension of $\mathbb{Q}$.

→ If $\bigwedge \mathcal{P} \Rightarrow \bigvee \mathcal{Q}$ is false, there exists a counterexample matrix Σ with algebraic entries.

Theorem (Positivstellensatz)

A polynomial F vanishes on the basic semialgebraic set $\{f_i = 0, g_j > 0, h_k \neq 0\}$ if and only if $0 \in \mathrm{ideal}(f_i) + \mathrm{cone}(g_j) + \mathrm{monoid}^2(F, g_j, h_k)$.

→ If $\bigwedge \mathcal{P} \Rightarrow \bigvee \mathcal{Q}$ is true, there exists an algebraic proof for it with rational coefficients.

Universality theorems

Theorem (B. 2021)

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 $\mathbb{K} \subseteq \mathbb{L}$.

→ The answer to Šimeček's question is **NO**.

Theorem (B. 2021)

The problem of deciding whether a CI inference formula is valid for all Gaussian distributions is polynomial-time equivalent to the existential theory of the reals.

→ Solving the inference problem can be used to check if arbitrary polynomial systems have a solution over $\mathbb{R}$.

A bit of the proof idea

$\Sigma[ij|] = x_{ij} \to$ impose $x_{kl} = x_{km} = x_{lm} = 0$ on a correlation matrix, then:

$$\begin{aligned}
\Sigma[ij|klm] = {} & x_{ij}x_{kk}x_{ll}x_{mm} + x_{im}x_{jm}\underline{x_{kl}^2} - x_{im}x_{jl}\underline{x_{kl}x_{km}} - x_{il}x_{jm}\underline{x_{kl}x_{km}} + x_{il}x_{jl}\underline{x_{km}^2} \\
& - x_{im}x_{jm}x_{kk}x_{ll} + x_{im}x_{jk}\underline{x_{km}}x_{ll} + x_{ik}x_{jm}\underline{x_{km}}x_{ll} - x_{ij}\underline{x_{km}^2}x_{ll} \\
& + x_{im}x_{jl}x_{kk}\underline{x_{lm}} + x_{il}x_{jm}x_{kk}\underline{x_{lm}} - x_{im}x_{jk}\underline{x_{kl}x_{lm}} - x_{ik}x_{jm}\underline{x_{kl}x_{lm}} \\
& - x_{il}x_{jk}\underline{x_{km}x_{lm}} - x_{ik}x_{jl}\underline{x_{km}x_{lm}} + 2x_{ij}\underline{x_{kl}x_{km}x_{lm}} + x_{ik}x_{jk}\underline{x_{lm}^2} \\
& - x_{ij}x_{kk}\underline{x_{lm}^2} - x_{il}x_{jl}x_{kk}x_{mm} + x_{il}x_{jk}\underline{x_{kl}}x_{mm} + x_{ik}x_{jl}\underline{x_{kl}}x_{mm} \\
& - x_{ij}\underline{x_{kl}^2}x_{mm} - x_{ik}x_{jk}x_{ll}x_{mm} \\
= {} & x_{ij} - \sum_{k=k,l,m} x_{ik}x_{jk} \quad = \quad x_{ij} - \left\langle \begin{pmatrix} x_{ik} \\ x_{il} \\ x_{im} \end{pmatrix}, \begin{pmatrix} x_{jk} \\ x_{jl} \\ x_{jm} \end{pmatrix} \right\rangle .
\end{aligned}$$

The rest is 19th century projective geometry. Keyword: *von Staudt constructions*.

Approximations to the inference problem

Approximations to the inference problem

Theorem (Matúš 2005)

The following relations hold for every symmetric matrix Σ:

$$\Sigma[ij|L]^2 = \Sigma[iL] \cdot \Sigma[jL] - \Sigma[L] \cdot \Sigma[ijL]$$
$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

Approximations to the inference problem

Theorem (Matúš 2005)

The following relations hold for every symmetric matrix Σ:

$$\Sigma[ij|L]^2 = \Sigma[iL] \cdot \Sigma[jL] - \Sigma[L] \cdot \Sigma[ijL] \quad \to \textit{semimatroids}$$

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L] \quad \to \textit{gaussoids}$$

These relations define essential geometric properties of symmetric matrices in principal and almost-principal minor coordinates. Study their combinatorics!

The Gaussian CI configuration space

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

The *Gaussian CI configuration space* $\mathscr{G} \subseteq \mathbb{R}^{2^n} \times \mathbb{R}^{\binom{n}{2}2^{n-2}}$ consists of all vectors of principal and almost-principal minors of $\Sigma \in \mathrm{PD}_n$.

$$\begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1, 1, 1, 1, 1 - a^2, 1 - b^2, 1 - c^2, \\ 1 + 2abc - a^2 - b^2 - c^2, \\ a, a - bc, b, b - ac, c, c - ab \end{pmatrix} \in \mathbb{R}^{8+6}.$$

The Gaussian CI configuration space

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

The *Gaussian CI configuration space* $\mathscr{G} \subseteq \mathbb{R}^{2^n} \times \mathbb{R}^{\binom{n}{2}2^{n-2}}$ consists of all vectors of principal and almost-principal minors of $\Sigma \in \mathrm{PD}_n$.

$$\begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1, 1, 1, 1, 1-a^2, 1-b^2, 1-c^2, \\ 1+2abc-a^2-b^2-c^2, \\ a, a-bc, b, b-ac, c, c-ab \end{pmatrix} \in \mathbb{R}^{8+6}.$$

Very wasteful encoding of a matrix, but this creates simple and useful relations on configuration vectors. The CI structure of Σ is encoded in the *zero pattern* of $c(\Sigma) \in \mathscr{G}$.

Combinatorial compatibility

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

Combinatorial compatibility means "fulfilling of relations under uncertainty":
What if we only knew that all $\Sigma[K] \neq 0$ and whether or not $\Sigma[ij|K] = 0$ for every $(ij|K)$?

Combinatorial compatibility

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

Combinatorial compatibility means "fulfilling of relations under uncertainty":
What if we only knew that all $\Sigma[K] \neq 0$ and whether or not $\Sigma[ij|K] = 0$ for every $(ij|K)$?

$$\begin{aligned} (ij|L) \wedge (ij|kL) &\Rightarrow (ik|L) \vee (jk|L) \\ (ik|L) \wedge (ij|kL) &\Rightarrow (ij|L) \\ &\vdots \end{aligned}$$

Combinatorial compatibility

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

Combinatorial compatibility means "fulfilling of relations under uncertainty":
What if we only knew that all $\Sigma[K] \neq 0$ and whether or not $\Sigma[ij|K] = 0$ for every $(ij|K)$?

$$\begin{aligned}
(ij|L) \wedge (ij|kL) &\Rightarrow (ik|L) \vee (jk|L) \\
(ik|L) \wedge (ij|kL) &\Rightarrow (ij|L) \wedge (ik|jL) \\
(ij|kL) \wedge (ik|jL) &\Rightarrow (ij|L) \wedge (ik|L) \\
(ij|L) \wedge (ik|L) &\Rightarrow (ij|kL) \wedge (ik|jL)
\end{aligned}$$

This yields the definition of *gaussoids*.

CI inference via SAT solvers

Since gaussoids have a finite axiomatization, a SAT solver like `CaDiCaL` can deduce implications under the gaussoid axioms:

$$
\begin{aligned}
&(12|3) \wedge (12|34) \wedge (24|1) \wedge (34|2) \\
&\Rightarrow\ (12|) \wedge (12|4) \wedge (24|) \wedge (24|3) \wedge (24|13) \wedge (34|)
\end{aligned}
$$

These conclusions are valid for all regular Gaussian distributions.

Oriented gaussoids

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

What if we only knew that all $\operatorname{sgn} \Sigma[K] = +1$ and the value of $\operatorname{sgn} \Sigma[ij|K]$ for every $(ij|K)$?

$$+(ij|L) \wedge -(ij|kL) \;\Rightarrow\; [+(ik|L) \wedge +(jk|L)] \vee [-(ik|L) \wedge -(jk|L)]$$

$\rightarrow$ *Oriented* and *orientable* gaussoids.

Oriented gaussoids

$$\Sigma[kL] \cdot \Sigma[ij|L] = \Sigma[L] \cdot \Sigma[ij|kL] + \Sigma[ik|L] \cdot \Sigma[jk|L]$$

What if we only knew that all $\operatorname{sgn} \Sigma[K] = +1$ and the value of $\operatorname{sgn} \Sigma[ij|K]$ for every $(ij|K)$?

$$+(ij|L) \wedge -(ij|kL) \;\Rightarrow\; [+(ik|L) \wedge +(jk|L)] \vee [-(ik|L) \wedge -(jk|L)]$$

→ *Oriented* and *orientable* gaussoids.

$$\begin{aligned}
(ij|L) \wedge (kl|L) \wedge (ik|jL) \wedge (jl|ikL) &\Rightarrow (ik|L) \\
(ij|L) \wedge (kl|iL) \wedge (kl|jL) \wedge (ij|klL) &\Rightarrow (kl|L) \\
(ij|L) \wedge (jl|kL) \wedge (kl|iL) \wedge (ik|jlL) &\Rightarrow (ik|L) \\
(ij|kL) \wedge (ik|lL) \wedge (il|jL) &\Rightarrow (ij|L) \\
(ij|kL) \wedge (ik|lL) \wedge (jl|iL) \wedge (kl|jL) &\Rightarrow (ij|L)
\end{aligned}$$

CI inference via SAT solvers II

Running the SAT solver `CaDiCaL` on the definition of oriented gaussoids confirms that on their supports

$$(12|) \wedge (13|4) \wedge (14|5) \wedge (15|23) \wedge (23|5) \wedge (24|135) \wedge (34|12) \wedge (35|1) \wedge (45|2)$$
$$\Rightarrow \text{ everything except } (25|K) \text{ for all } K.$$

Geometrically, the model is a Markov network.

Theorem (B. 2021+)

Orientable gaussoids have no finite complete axiomatization.

Selfadhesivity

Gaussian multiinformation functions satisfy a *selfadhesivity* property just like entropy vectors of discrete random variables (Matúš 2007):

Theorem (B. 2020+)

Let Σ be a positive-definite matrix on $\mathbb{R}^N$ and $K \subseteq N$. Let M be any set with $|M| = |N|$ and $M \cap N = K$. There exists a unique positive-definite Φ on $\mathbb{R}^{N \cup M}$ such that:

- $\Phi_N = \Sigma = \Phi_M$,
- $N \perp\!\!\!\perp M \mid K$ *holds for* Φ.

Structural selfadhesivity

Selfadhesivity can be formulated in CI terms and all necessary properties studied above can be strengthened by selfadhesivity. For example:

Structural selfadhesivity

Selfadhesivity can be formulated in CI terms and all necessary properties studied above can be strengthened by selfadhesivity. For example:

- $\mathcal{O}$ is a *selfadhesive* orientable gaussoid on N if for every $K \subseteq N$ and M as above there exists an orientable gaussoid $\overline{\mathcal{O}}$ on $N \cup M$ such that
  - $\overline{\mathcal{O}}|_N = \mathcal{O} = \overline{\mathcal{O}}|_M$,
  - $(N, M|K) \in \overline{\mathcal{O}}$.

  This property can be decided by 2^N calls to a program which decides orientability.

Every Gaussian CI structure is a selfadhesive orientable gaussoid, but not every orientable gaussoid is selfadhesive. Similarly for semimatroids.

The search for inference rules

Inference rules help characterize the *realizable* CI structures:

- 3-variate: 11 out of 64 by Matúš 2005.
- 4-variate: 629 out of 16 777 216 by Lněnička and Matúš 2007.
- 5-variate: *open!* (out of 1 208 925 819 614 629 174 706 176)
  - 254 826 gaussoids modulo symmetry,
  - 87 792 of which are orientable semimatroids,
  - 84 434 of which are *selfadhesive* orientable semimatroids.

The search for inference rules

Inference rules help characterize the *realizable* CI structures:

- 3-variate: 11 out of 64 by Matúš 2005.
- 4-variate: 629 out of 16 777 216 by Lněnička and Matúš 2007.
- 5-variate: *open!* (out of 1 208 925 819 614 629 174 706 176)
  - 254 826 gaussoids modulo symmetry,
  - 87 792 of which are orientable semimatroids,
  - 84 434 of which are *selfadhesive* orientable semimatroids.

Help wanted:

- Use linear programming and information inequalities.
- Non-linear information inequalities → Ahmadieh and Vinzant 2021.
- Tropical approximations and valuated gaussoids.
- Compute algebraic realization spaces.
- Find and certify real solutions to polynomial systems.

Tobias Boege, Alessio D'Alì, Thomas Kahle, and Bernd Sturmfels.
The Geometry of Gaussoids.
Found. Comput. Math., 19(4):775–812, 2019.

Tobias Boege.
Incidence geometry in the projective plane via almost-principal minors of symmetric matrices, 2021.
arXiv:2103.02589.

Jürgen Bokowski and Bernd Sturmfels.
Computational synthetic geometry, volume 1355 of *Lecture Notes in Mathematics*.
Springer, 1989.

Radim Lněnička and František Matúš.
On Gaussian conditional independence structures.
Kybernetika, 43(3):327–342, 2007.

František Matúš.
Conditional independence structures examined via minors.
Ann. Math. Artif. Intell., 21(1):99–30, 1997.

František Matúš.
Conditional independences in gaussian vectors and rings of polynomials.
In Gabriele Kern-Isberner, Wilhelm Rödder, and Friedhelm Kulmann, editors, *Conditionals, Information, and Inference*, pages 152–161. Springer, 2005.

František Matúš.
Adhesivity of polymatroids.
Discrete Math., 307(21):2464–2477, 2007.

Petr Šimeček.
Gaussian representation of independence models over four random variables.
In *COMPSTAT conference*, 2006.

Seth Sullivant.
Gaussian conditional independence relations have no finite complete characterization.
J. Pure Appl. Algebra, 213(8):1502–1506, 2009.