The laws of Gaussian conditional independence

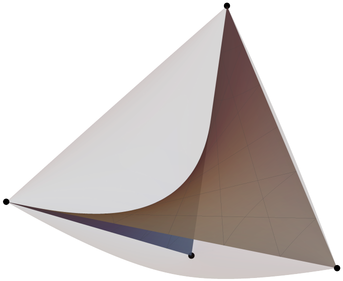
Tobias Boege

MPI-MiS Leipzig

Tagung der Fachgruppe Computeralgebra
München, 10 March 2022

The mantra of algebraic statistics

Statistical models are semialgebraic sets*



The set of all centered, standardized Gaussian distributions parametrized by their correlation matrices

$$\Sigma = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix} \in \mathrm{PD}_3$$

which satisfy the conditional independence statement $\xi_1 \perp\!\!\!\perp \xi_2 \mid \xi_3$.

*sometimes

Gaussian conditional independence

Consider random variables $(\xi_i)_{i \in N}$. 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.

Gaussian conditional independence

Consider random variables $(\xi_i)_{i \in N}$. 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*.

- Σ is PD if and only if $\Sigma[K] > 0$ for all $K \subseteq N$.
- $\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K$ holds if and only if $\Sigma[ij|\,K] = 0$.

Very special polynomials

$$\Sigma[i] = x_{ii}$$

$$\Sigma[ij] = x_{ii}x_{jj} - x_{ij}^2$$

$$\Sigma[ijk] = x_{ii}x_{jj}x_{kk} - x_{ik}^2x_{jj} + 2x_{ij}x_{ik}x_{jk} - x_{ii}x_{jk}^2 - x_{ij}^2x_{kk}$$

$$\Sigma[ijkl] = x_{ii}x_{jj}x_{kk}x_{ll} + x_{il}^2x_{jk}^2 - 2x_{ik}x_{il}x_{jk}x_{jl} + x_{ik}^2x_{jl}^2 - x_{il}^2x_{jj}x_{kk} + 2x_{ij}x_{il}x_{jl}x_{kk} - x_{ii}x_{jl}^2x_{kk} + 2x_{ik}x_{il}x_{jj}x_{kl} - 2x_{ij}x_{il}x_{jk}x_{kl} - 2x_{ij}x_{ik}x_{jl}x_{kl} + 2x_{ii}x_{jk}x_{jl}x_{kl} + x_{ij}^2x_{kl}^2 - x_{ii}x_{jj}x_{kl}^2 - x_{ik}^2x_{jj}x_{ll} + 2x_{ij}x_{ik}x_{jk}x_{ll} - x_{ii}x_{jk}^2x_{ll} - x_{ij}^2x_{kk}x_{ll}$$

$$\vdots$$

$$\Sigma[ij|\,] = x_{ij}$$

$$\Sigma[ij|k] = x_{ij}x_{kk} - x_{ik}x_{jk}$$

$$\Sigma[ij|kl] = x_{ij}x_{kk}x_{ll} - x_{il}x_{jl}x_{kk} + x_{il}x_{jk}x_{kl} + x_{ik}x_{jl}x_{kl} - x_{ij}x_{kl}^2 - x_{ik}x_{jk}x_{ll}$$

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

$$\vdots$$

Gaussian CI models

Definition

A *CI constraint* is a CI statement $\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K$ or its negation $\neg(\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K)$.
The *model* of a set of CI constraints is the set of all PD matrices which satisfy them.

Gaussian CI models

Definition

A *CI constraint* is a CI statement $\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K$ or its negation $\neg(\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K)$.
The *model* of a set of CI constraints is the set of all PD matrices which satisfy them.

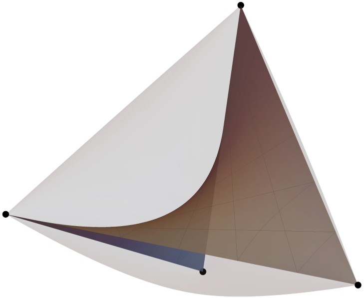


Figure: Model of $\Sigma[12 \mid 3] = a - bc = 0$ in the space of 3×3 correlation matrices.

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 $\mathcal{Q}$:

$$\bigwedge \mathcal{P} \Rightarrow \bigvee \mathcal{Q} \quad \text{is not valid} \qquad \Longleftrightarrow \qquad \mathcal{P} \cup \neg\mathcal{Q} \quad \text{has a point}$$

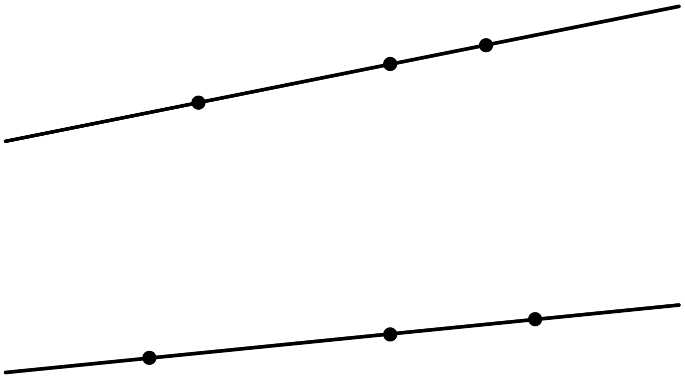
Models and inference

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

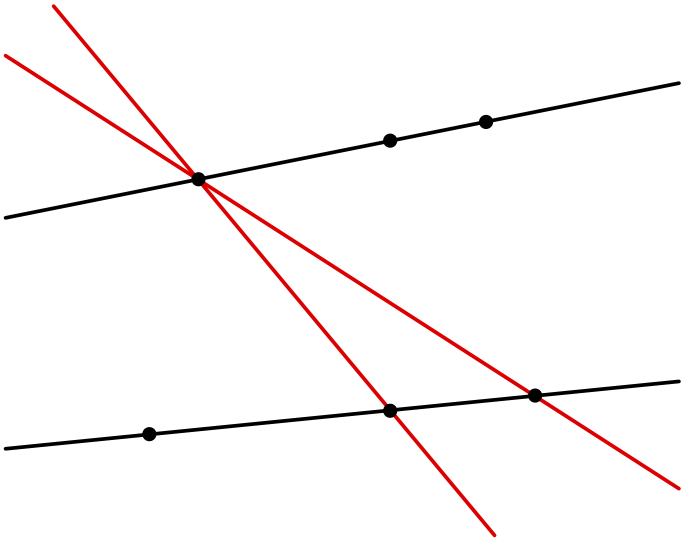
$$\bigwedge \mathcal{P} \Rightarrow \bigvee \mathcal{Q} \text{ is not valid} \quad \Longleftrightarrow \quad \mathcal{P} \cup \neg\mathcal{Q} \text{ has a point}$$

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.

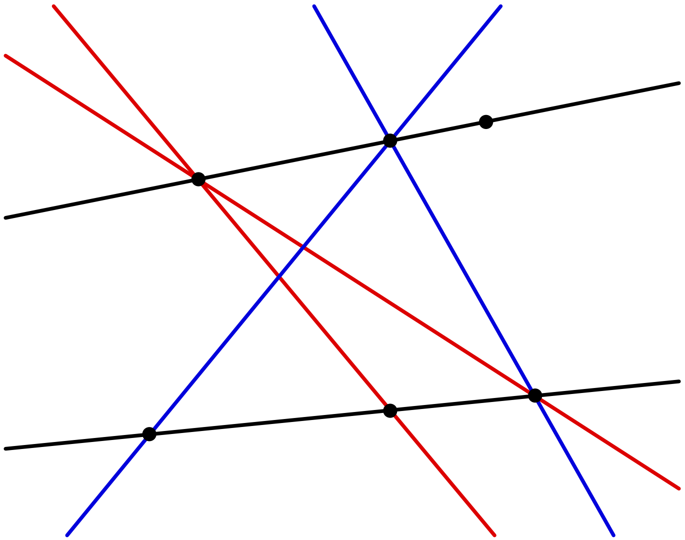
For geometers: conditional independence ≈ collinearity



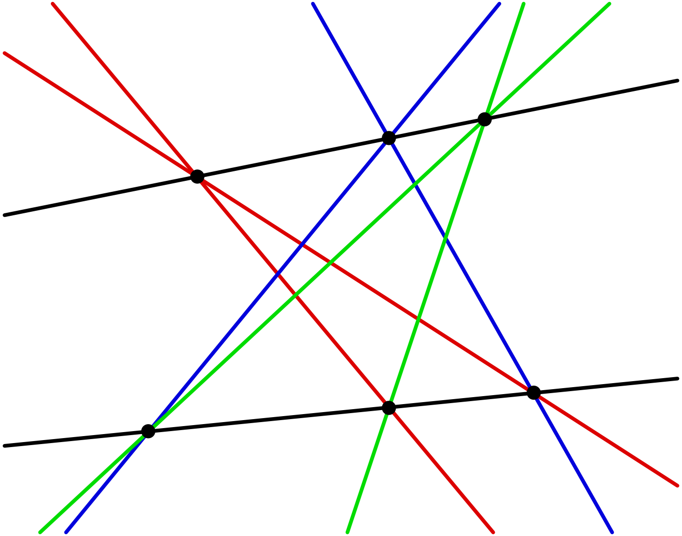
For geometers: conditional independence ≈ collinearity



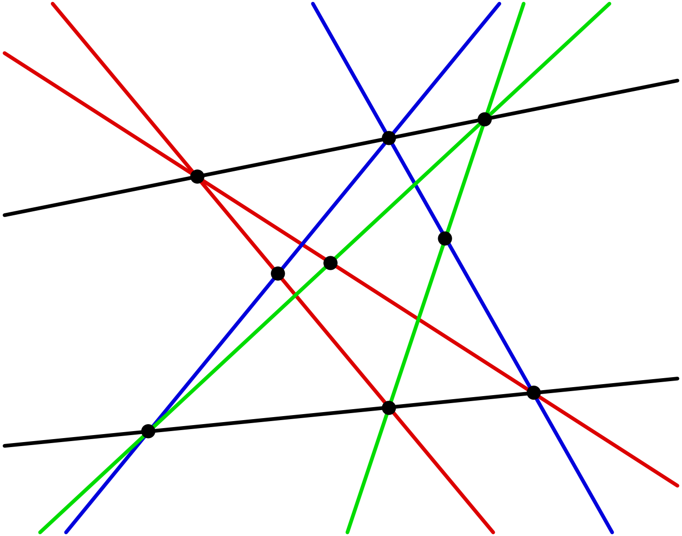
For geometers: conditional independence ≈ collinearity



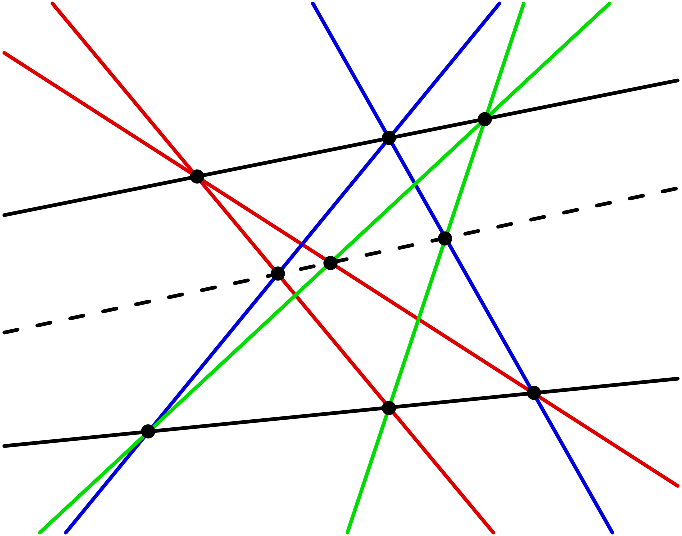
For geometers: conditional independence ≈ collinearity



For geometers: conditional independence ≈ collinearity



For geometers: conditional independence ≈ collinearity

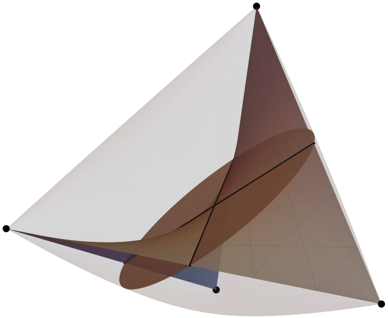


Example of CI inference

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

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

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



Normal form for proofs and refutations

Let $f_i \in \mathbb{Z}[t_1, \ldots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Tarski's transfer principle)

If a polynomial system $\{f_i \bowtie_i 0\}$, *where* $\bowtie_i \in \{=, \neq, <, \leq, \geq, >\}$, *has a solution over* $\mathbb{R}$, *then it has a solution in a finite real extension of* $\mathbb{Q}$.

Normal form for proofs and refutations

Let $f_i \in \mathbb{Z}[t_1, \ldots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Tarski's transfer principle)

If a polynomial system $\{f_i \bowtie_i 0\}$, where $\bowtie_i \in \{=, \neq, <, \leq, \geq, >\}$, 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.

$[12||] \wedge [12|3] \Rightarrow [13||]$ is false and a counterexample is

$$\begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1 \end{pmatrix}.$$

Normal form for proofs and refutations

Let $f_i, g_j, h_k \in \mathbb{Z}[t_1, \ldots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Positivstellensatz)

A polynomial system $\{f_i = 0, g_j \geq 0, h_k \neq 0\}$ *is infeasible if and only if there exist* $f \in \mathrm{ideal}(f_i)$, $g \in \mathrm{cone}(g_j)$ *and* $h \in \mathrm{monoid}(h_k)$ *such that* $g + h^2 = f$.

Normal form for proofs and refutations

Let $f_i, g_j, h_k \in \mathbb{Z}[t_1, \ldots, t_k]$ be integer polynomials in finitely many variables.

Theorem (Positivstellensatz)

A polynomial system $\{f_i = 0, g_j \geq 0, h_k \neq 0\}$ *is infeasible if and only if there exist* $f \in \text{ideal}(f_i)$, $g \in \text{cone}(g_j)$ *and* $h \in \text{monoid}(h_k)$ *such that* $g + h^2 = f$.

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

$[12|\,] \wedge [12|3] \Rightarrow [13|\,] \vee [23|\,]$ is true and a proof is the final polynomial

$$\Sigma[13|\,] \cdot \Sigma[23|\,] = \Sigma[3] \cdot \Sigma[12|\,] - \Sigma[12|3].$$

A 5×5 final polynomial

The following inference rule is valid for all positive-definite 5×5 matrices:

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

A 5×5 final polynomial

The following inference rule is valid for all positive-definite 5×5 matrices:

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

$$[25\,|\,][34\,|\,]\cdot[1][2][3][15] =$$

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

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

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

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

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

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

A 5 × 5 final polynomial

```
R = QQ[p,a,b,c,d, q,e,f,g, r,h,i, s,j, t];
X = genericSymmetricMatrix(R,p,5);
I = ideal(
  det X_{0}^{1}, det X_{0,3}^{2,3}, det X_{0,4}^{3,4},
  det X_{1,4}^{2,4}, det X_{2,0}^{4,0}, det X_{3,1}^{4,1},
  det X_{0,1,2}^{4,1,2}, det X_{2,0,1}^{3,0,1},
  det X_{1,0,2,4}^{3,0,2,4}
);
U = g*h*p*q*r*(p*t-d^2); -- [25||][34||] · [1][2][3][15] ∈ monoid(𝒱)
U % I --> 0, meaning monoid(𝒱) ∩ ideal(𝒱) ≠ ∅ in ℚ[X]
-- Get a proof that U is in I:
G = gens I; -- the equations generating ideal(𝒱)
H = U // G; -- linear combinators for U from G
U == G*H --> true
```

Difficult inference rules for PD_4

$$[12|3] \wedge [13|4] \wedge [14|2] \Rightarrow [12||]$$
$$\Leftarrow \quad [12||](e^2f^2[23] + qr^2s[24] + qe^2r[34]) \in \text{ideal}$$

$$[12|3] \wedge [13|4] \wedge [24|1] \wedge [34|2] \Rightarrow [12||]$$
$$\Leftarrow \quad [12||](qr[14] + c^2[23]) \in \text{ideal}$$

$$[23||] \wedge [14|2] \wedge [14|3] \wedge [23|14] \Rightarrow [34||]$$
$$\Leftarrow \quad [34||]^2(e^2[123] + b^2q^2s) \in \text{ideal}$$

$$[13||] \wedge [14|2] \wedge [24|3] \wedge [23|14] \Rightarrow [23||]$$
$$\Leftarrow \quad [23||](pq^2r[34] + a^2f^2[23]) \in \text{ideal}$$

$$[13||] \wedge [24||] \wedge [14|23] \wedge [23|14] \Rightarrow [13||] \wedge [23||]$$
$$\Leftarrow \quad ([13||]^2qr + [23||]^2ps)[1234] + qs(acr + dfp)^2 + pr(cfq + ads)^2 \in \text{ideal}$$

The gaussoid axioms

Theorem (Matúš 2005)

The following relations hold for every symmetric matrix Σ:

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

The gaussoid axioms

Theorem (Matúš 2005)

The following relations hold for every symmetric matrix Σ:

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

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

$$[ij|L] \wedge [ij|kL] \Rightarrow [ik|L] \vee [jk|L]$$

$$[ik|L] \wedge [ij|kL] \Rightarrow [ij|L]$$

$$\vdots$$

The gaussoid axioms

Theorem (Matúš 2005)

The following relations hold for every symmetric matrix Σ:

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

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

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

This yields the definition of *gaussoids*.

Good and bad news

Theorem (Two-antecedental completeness)

Every inference rule $[ij|K] \wedge [lm|N] \Rightarrow \bigvee \mathcal{Q}$ *with (at most) two antecedents which is valid for all positive-definite matrices is a consequence of the gaussoid axioms.*

Good and bad news

Theorem (Two-antecedental completeness)

Every inference rule $[ij|K] \wedge [lm|N] \Rightarrow \bigvee \mathcal{Q}$ with (at most) two antecedents which is valid for all positive-definite matrices is a consequence of the gaussoid axioms.

Theorem (Universality)

The problem of deciding whether a general CI inference formula is valid for all Gaussian distributions is polynomial-time equivalent to the existential theory of the reals. (And counterexamples can have arbitrarily high extension degrees over $\mathbb{Q}$.)

The search for inference rules (since at least 2008!)

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
  - 84 434 open cases after applying further combinatorial and polyhedral-geometric techniques.

Other open questions from *Geometry of gaussoids*

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

Let $\mathcal{J}_n$ be the ideal of homogeneous relations among principal and almost-principal minors of a generic symmetric $n \times n$ matrix.

Conjecture

Using all quadrics in $\mathcal{J}_n$ as final polynomials still only yields the gaussoid axioms.

Conjecture

$\mathcal{J}_n$ is generated by its quadrics.

References

Tobias Boege. "Gaussoids are two-antecedental approximations of Gaussian conditional independence structures". In: *Ann. Math. Artif. Intell.* (2021). DOI: 10.1007/s10472-021-09780-0.

Tobias Boege. *Incidence geometry in the projective plane via almost-principal minors of symmetric matrices*. 2021. arXiv: 2103.02589 [math.ST].

Tobias Boege, Alessio D'Alì, Thomas Kahle, and Bernd Sturmfels. "The Geometry of Gaussoids". In: *Found. Comput. Math.* 19.4 (2019), pp. 775–812. ISSN: 1615-3383. DOI: 10.1007/s10208-018-9396-x.

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