Tutorial: Gaussian conditional independence and graphical models

Thomas Kahle
Otto-von-Guericke Universität Magdeburg

The central dogma of algebraic statistics

Statistical models are varieties

The central dogma of algebraic statistics

Statistical models are varieties

Today

Demonstrate algebraic approaches to conditional independence

- For Gaussian vectors $X = (X_1, \ldots, X_m)$ with values in $\mathbb{R}^m$.
- Source: Seth Sullivant's book manuscript "Algebraic Statistics".

The density

A random vector $X = (X_1, \ldots, X_m)$ has a Gaussian (or normal) distribution if its density with respect to the Lebesgue measure is

$$f(x) = \frac{1}{(2\pi)^{m/2} \det \Sigma^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

for some $\mu \in \mathbb{R}^m$ and $\Sigma \in \mathrm{PD}_m$ positive definite.

- Density wrt Lebesgue measure means

$$\mathrm{Prob}(X \in A) = \int_A f(x) dx \qquad A \subseteq \mathbb{R}^m$$

- μ is the mean
- Σ^{-1} the concentration matrix, and
- Σ the covariance matrix.

Marginals

Let $A \subseteq [m]$ and $X = (X_1, \ldots, X_m)$ a Gaussian random vector.

- The marginal density $f_A(x_A)$ of $X_A = (X_i)_{i \in A}$ is defined by

$$f_A(x_A) = \int_{\mathbb{R}^{[m]\setminus A}} f(x_A, x_{[m]\setminus A}) dx_{[m]\setminus A}$$

- The marginal X_A of a Gaussian X is itself Gaussian with mean $\mu_A = (\mu_i)_{i \in A}$ and covariance $\Sigma_{A \times A} = (\Sigma_{ij})_{i,j \in A}$.

Independence

Let $A, B \subseteq [m]$ be disjoint. X_A is independent of X_B ($A \perp\!\!\!\perp B$), if

$$f_{A \cup B}(x_A, x_B) = f_A(x_A) f_B(x_B)$$

This happens if and only if $\Sigma_{A \times B} = 0$.

Example Independence

$X_1 =$ delay of your flight to Atlanta,
$X_2 =$ delay of my flight to Atlanta.
With no further information, a reasonable first assumption: $X_1 \perp\!\!\!\perp X_2$.

Example Independence

$X_1 =$ delay of your flight to Atlanta,
$X_2 =$ delay of my flight to Atlanta.
With no further information, a reasonable first assumption: $X_1 \perp\!\!\!\perp X_2$.

or maybe not?

Assume our day of arrival sees a lot of rain (a variable X_3 takes high value).

- X_1 and X_2 show correlation (e.g. both more likely delayed)
- This correlation is explained by X_3
- Conditionally on X_3 being large, X_1 and X_2 are still independent.
- Capture this by dividing by the marginal density of X_3.

Conditionals

Let $A, B \subseteq [m]$ be disjoint.

- For each fixed $x_B \in \mathbb{R}^B$, the conditional density $f_{A|B}(x_A, x_B)$ of A given $X_B = x_B$ is defined by

$$f_{A|B}(x_A, x_B) = \frac{f_{A\cup B}(x_A, x_B)}{f_B(x_B)}$$

- The conditional density of a Gaussian is Gaussian with mean

$$\mu_A + \Sigma_{A\times B}\Sigma_{B\times B}^{-1}(x_B - \mu_B)$$

  and covariance

$$\Sigma_{A\times A} - \Sigma_{A\times B}\Sigma_{B\times B}^{-1}\Sigma_{B\times A}.$$

Definition

Let $A, B, C \subseteq [m]$ be pairwise disjoint and f be a Gaussian density. A is conditionally independent of B given C, written $A \perp\!\!\!\perp B \,|C$ if for all $x_A \in \mathbb{R}^A, x_B \in \mathbb{R}^B, x_C \in \mathbb{R}^C$

$$f_{AB|C}(x_A, x_B, x_C) = f_{A|C}(x_A, x_C) f_{B|C}(x_B, x_C).$$

Definition

Let $A, B, C \subseteq [m]$ be pairwise disjoint and f be a Gaussian density. A is conditionally independent of B given C, written $A \perp\!\!\!\perp B \,|C$ if for all $x_A \in \mathbb{R}^A, x_B \in \mathbb{R}^B, x_C \in \mathbb{R}^C$

$$f_{AB|C}(x_A, x_B, x_C) = f_{A|C}(x_A, x_C) f_{B|C}(x_B, x_C).$$

Convention: omit $\cup$, i.e. $AC = A \cup C$ and so on.

Definition

Let $A, B, C \subseteq [m]$ be pairwise disjoint and f be a Gaussian density. A is conditionally independent of B given C, written $A \perp\!\!\!\perp B \,|C$ if for all $x_A \in \mathbb{R}^A, x_B \in \mathbb{R}^B, x_C \in \mathbb{R}^C$

$$f_{AB|C}(x_A, x_B, x_C) = f_{A|C}(x_A, x_C) f_{B|C}(x_B, x_C).$$

Convention: omit $\cup$, i.e. $AC = A \cup C$ and so on.

Proposition

$A \perp\!\!\!\perp B \,|C$ if and only if $\operatorname{rk} \Sigma_{AC \times BC} = |C|$.

$A \perp\!\!\!\perp B \,|C$ if and only if $\operatorname{rk} \Sigma_{AC \times BC} = |C|$.

Proof

Conditional distribution of X_{AB} given $X_C = x_c$ has covariance

$$\Sigma_{AB \times AB} - \Sigma_{AB \times C} \Sigma_{C \times C}^{-1} \Sigma_{C \times AB}$$

Conditional independence happens if $A \times B$ submatrix vanishes:

$$S = \Sigma_{A \times B} - \Sigma_{A \times C} \Sigma_{C \times C}^{-1} \Sigma_{C \times B} = 0$$

This matrix is the Schur complement in

$$\Sigma_{AC \times BC} = \begin{pmatrix} \Sigma_{A \times B} & \Sigma_{A \times C} \\ \Sigma_{C \times B} & \Sigma_{C \times C}. \end{pmatrix} \longrightarrow \begin{pmatrix} S & \Sigma_{A \times C} \\ 0 & \Sigma_{C \times C}. \end{pmatrix}.$$

(subtract right column times $\Sigma_{C \times C}^{-1} \Sigma_{C \times B}$ from left column) □

If you don't like densities, this can be your starting point

Definition

Let $A, B, C \subseteq [m]$ be pw. disjoint. The corresponding conditional independence (CI) ideal is

$$I_{A \perp\!\!\!\perp B \,|\, C} = \langle (|C| + 1) - \text{minors of } \Sigma_{AC \times BC} \rangle$$

The conditional independence model is

$$\mathcal{M}_{A \perp\!\!\!\perp B \,|\, C} = V(I_{A \perp\!\!\!\perp B \,|\, C}) \cap \mathrm{PD}_m.$$

(note: this is a semi-algebraic set)

Our goal: Study Gaussian conditional independence using conditional independence ideals

Proposition ("CI Axioms")

1. $A \perp\!\!\!\perp B \,|C \;\Rightarrow\; B \perp\!\!\!\perp A \,|C$ (symmetry)
2. $A \perp\!\!\!\perp B \cup D \,|C \;\Rightarrow\; A \perp\!\!\!\perp B \,|C$ (decomposition)
3. $A \perp\!\!\!\perp B \cup D \,|C \;\Rightarrow\; A \perp\!\!\!\perp B \,|C \cup D$ (weak union)
4. $A \perp\!\!\!\perp B \,|C \cup D$ and $A \perp\!\!\!\perp D \,|C \;\Rightarrow\; A \perp\!\!\!\perp B \cup D \,|C$ (contraction)

Proof

- Proof for Gaussians is exercise in linear algebra.
- Can be proven for general (non-Gaussian) densities

Special properties of Gaussian conditional independence

- The "intersection axiom"

$$A \perp\!\!\!\perp B \,|\, C \cup D \text{ and } A \perp\!\!\!\perp C \,|\, B \cup D \Rightarrow A \perp\!\!\!\perp B \cup C \,|\, D$$

  holds for all strictly positive densities
- "Gaussoid axiom"

$$A \perp\!\!\!\perp B \,|\, \{c\} \cup D \text{ and } A \perp\!\!\!\perp B \,|\, D$$
$$\Rightarrow A \perp\!\!\!\perp B \cup \{c\} \,|\, D \text{ or } A \cup \{c\} \perp\!\!\!\perp B \,|\, D$$

  holds for Gaussians.

Special properties of Gaussian conditional independence

- The "intersection axiom"

$$A \perp\!\!\!\perp B \,|\, C \cup D \text{ and } A \perp\!\!\!\perp C \,|\, B \cup D \Rightarrow A \perp\!\!\!\perp B \cup C \,|\, D$$

  holds for all strictly positive densities
- "Gaussoid axiom"

$$A \perp\!\!\!\perp B \,|\, \{c\} \cup D \text{ and } A \perp\!\!\!\perp B \,|\, D \Rightarrow A \perp\!\!\!\perp B \cup \{c\} \,|\, D \text{ or } A \cup \{c\} \perp\!\!\!\perp B \,|\, D$$

  holds for Gaussians.

Why are these lemmas called axioms?

Q: Is there a finite axiomatization of Gaussian CI?

Conjunctions of CI statements

Want to answer questions like: Given a density satisfies a set

$$\mathcal{C} = \{ A_1 \perp\!\!\!\perp B_1 \,|C_1 \,, \ldots A_n \perp\!\!\!\perp B_n \,|C_n \}$$

of CI statements, what other properties does it have?

Algebraic approach

The covariances that satisfy $\mathcal{C}$:

$$\mathcal{M}_{\mathcal{C}} = \mathsf{PD}_m \cap V(I_{A_1 \perp\!\!\!\perp B_1 \,|C_1}) \cap \cdots \cap V(I_{A_n \perp\!\!\!\perp B_n \,|C_n})$$

Approach: Compute primary decomposition (or minimal primes) of

$$I_{\mathcal{C}} = I_{A_1 \perp\!\!\!\perp B_1 \,|C_1} + \cdots + I_{A_n \perp\!\!\!\perp B_n \,|C_n}$$

Example

Let's study the contraction property algebraically:

$$A \perp\!\!\!\perp B \,|C \cup D \;\text{ and }\; A \perp\!\!\!\perp D \,|C \;\Rightarrow\; A \perp\!\!\!\perp B \cup D \,|C$$

With $m = 3$, $A = \{1\}$, $B = \{2\}$, $C = \emptyset$, $D = \{3\}$ we get

$$\mathcal{C} = \{ 1 \perp\!\!\!\perp 2 \,|3 \,,\, 1 \perp\!\!\!\perp 3 \}$$

$\Rightarrow$ Macaulay2.

Example

Let's study the contraction property algebraically:

$$A \perp\!\!\!\perp B \,|\, C \cup D \ \text{ and } \ A \perp\!\!\!\perp D \,|\, C \ \Rightarrow A \perp\!\!\!\perp B \cup D \,|\, C$$

With $m = 3$, $A = \{1\}$, $B = \{2\}$, $C = \emptyset$, $D = \{3\}$ we get

$$\mathcal{C} = \{1 \perp\!\!\!\perp 2 \,|\, 3 \,,\, 1 \perp\!\!\!\perp 3\}$$

$\Rightarrow$ Macaulay2.
Primary decomposition has two components:

$$V(I_{\mathcal{C}}) = V(\Sigma_{12}, \Sigma_{13}) \cup V(\Sigma_{13}, \Sigma_{33}).$$

- The second component does not intersect PD_3
- The first component is the desired conclusion $1 \perp\!\!\!\perp \{2, 3\}$

Success story (Sullivant, 2009)

For $n \geq 4$, consider the cyclic set of CI statements

$$\mathcal{C} = \{ 1 \perp\!\!\!\perp 2 \,|3 \,, \ldots, n-1 \perp\!\!\!\perp n \,|1 \,, n \perp\!\!\!\perp 1 \,|2 \,\}$$

(Binomial) primary decomposition yields

- $I_{\mathcal{C}}$ has two minimal primes
  - $\langle \Sigma_{12}, \Sigma_{23}, \ldots, \Sigma_{n1} \rangle$ corresponding to $1 \perp\!\!\!\perp 2 \,, 2 \perp\!\!\!\perp 3 \,, \ldots, n \perp\!\!\!\perp 1$
  - The toric ideal $I_{\mathcal{C}} : \left(\prod_{ij} \Sigma_{ij} \right)^{\infty}$ whose variety does not contain positive definite matrices.
- No subset of $\mathcal{C}$ implies the marginal independencies.

Success story (Sullivant, 2009)

For $n \geq 4$, consider the cyclic set of CI statements

$$\mathcal{C} = \{ 1 \perp\!\!\!\perp 2 \,|3\,, \ldots, n-1 \perp\!\!\!\perp n\,|1\,, n \perp\!\!\!\perp 1\,|2\, \}$$

(Binomial) primary decomposition yields

- $I_{\mathcal{C}}$ has two minimal primes
  - $\langle \Sigma_{12}, \Sigma_{23}, \ldots, \Sigma_{n1} \rangle$ corresponding to $1 \perp\!\!\!\perp 2\,, 2 \perp\!\!\!\perp 3\,, \ldots, n \perp\!\!\!\perp 1$
  - The toric ideal $I_{\mathcal{C}} : \left(\prod_{ij} \Sigma_{ij}\right)^{\infty}$ whose variety does not contain positive definite matrices.
- No subset of $\mathcal{C}$ implies the marginal independencies.

$\Rightarrow$ Gaussian conditional independence cannot be finitely axiomatized.

A good source for CI ideals + problems: graphical models.

A good source for CI ideals + problems: graphical models.

Graphical models

Let G be a graph, either directed, undirected, or mixed, whose vertices are random variables, and edges represent dependency.

- A graphical model assigns a set of covariance matrices to G
  - Use separation in the graph to define conditional independence
  - Use connection in the graph to parametrize

Simplest example

As the simplest example, consider an undirected graph $G = (V, E)$.

- The pairwise Markov property of G postulates that $v \perp\!\!\!\perp w \,|V \setminus \{v, w\}$ for every non-edge $(v, w) \notin E$.
- The global Markov property of G postulates $A \perp\!\!\!\perp B \,|C$ whenever C separates A and B in G.

Theorem

Both Markov properties yield the same set of covariance matrices and this set is characterized by $\Sigma^{-1}_{ij} = 0$ whenever $(i, j) \notin E$ (which yields determinantal constraints on Σ by Kramer's rule).

For DAGs, there is are natural parametrization

Let D be DAG (acyclic directed graph) on $[m]$ (top. ordered).

- Postulate *structural equations*

$$X_j = \sum_{i \in \mathsf{pa}(j)} \lambda_{ij} X_i + \epsilon_j, \qquad j \in [m],$$

  where ϵ_j is Gaussian with variance ϕ_j, and $\lambda_{ij} \in \mathbb{R}$.
- Then $X = (X_1, \ldots, X_m)$ is Gaussian with covariance

$$\Sigma = (I - \Lambda)^{-T} \Phi (I - \Lambda)^{-1}$$

  where $\Phi = \operatorname{diag}(\phi_1, \ldots, \phi_m)$ consists of the variances of $X_1, \ldots, X_m$, and Λ is upper triangular with entries λ_{ij} and ones on the diagonal.
- The DAGical model consist of all such covariance matrices.

For DAGs, there is are natural parametrization

Let D be DAG (acyclic directed graph) on $[m]$ (top. ordered).

- Postulate *structural equations*

$$X_j = \sum_{i \in \mathsf{pa}(j)} \lambda_{ij} X_i + \epsilon_j, \qquad j \in [m],$$

  where ϵ_j is Gaussian with variance ϕ_j, and $\lambda_{ij} \in \mathbb{R}$.

- Then $X = (X_1, \ldots, X_m)$ is Gaussian with covariance

$$\Sigma = (I - \Lambda)^{-T} \Phi (I - \Lambda)^{-1}$$

  where $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_m)$ consists of the variances of $X_1, \ldots, X_m$, and Λ is upper triangular with entries λ_{ij} and ones on the diagonal.

- The DAGical model consist of all such covariance matrices.

$\Rightarrow$ Question: What is the vanishing ideal in $\mathbb{R}[\Sigma]$?

Separation gives valid conditional independence constraints

A,B are d-separated by C if every path from A to B either

- contains a "collider" $\cdots \to v \leftarrow \ldots$ where neither v nor any descendent of v are contained in C
- contains a blocked vertex $v \in C$ with $\cdots \to v \to \ldots$

Theorem

A CI Statement $A \perp\!\!\!\perp B \,|C$ is valid for all covariances in the model if and only if C d-separates A and B in G.

Separation gives valid conditional independence constraints

A,B are d-separated by C if every path from A to B either

- contains a "collider" $\cdots \to v \leftarrow \ldots$ where neither v nor any descendent of v are contained in C
- contains a blocked vertex $v \in C$ with $\cdots \to v \to \ldots$

Theorem

A CI Statement $A \perp\!\!\!\perp B \,|C$ is valid for all covariances in the model if and only if C d-separates A and B in G.

Surprise

There are more vanishing minors on the model, and all of these can be found using *trek separation* of Sullivant, Talaska, and Draisma.

Trek separation is still not all. In general there are non-determinantal constraints too ($\to$ exercise).

General problem

Characterization of graphs for which the vanishing ideal equals the global Markov ideal.

Holds for trees and all graphs on ≤ 4 vertices.

Functionality of the graphical models package

- Creation of appropriate rings for conditional independence and graphical models in the Gaussian and discrete case: `gaussianRing`, `markovRing`.
- Deal with undirected, directed, and mixed graphs.
- Enumeration of separation statements: pairwise, local, global, d-, trek.
- Creation of conditional independence ideals from a list of statements: `conditionalIndependenceIdeal`.
- Write out parametrizations of graphicalModels as rational maps and compute `vanishingIdeal`.
- Solve Gaussian identifiability problems with `identifyParameters`.