Algebraic Aspects of Gaussian Bayesian Networks

Seth Sullivant

Harvard University → North Carolina State University

July 4, 2008

The Big Picture

Given a directed acyclic graph G, two ways to describe a Bayesian Network:

- Parametrically (recursive factorization of joint distribution)
- Conditional Independence Constraints

Theorem

A probability density function f factorizes according to G if and only if f satisfies the conditional independence statements implied by G.

Question

What happens when some of the random variables in the Bayes Net are hidden? What constraints replace conditional independence constraints?

Bayesian Networks

- G directed acyclic graph (DAG)
- $V(G) = [n] := \{1, 2, \dots, n\}$
- $i \to j \in E(G)$ must satisfy $i < j$.
- $\mathrm{pa}(i) = \{k \mid k \to i \in E(G)\}$
- Joint density $f(x)$ belongs to Bayes Net associated to G iff

$$f(x) = \prod_{i=1}^{n} f_i(x_i | x_{\mathrm{pa}(i)})$$

where $f_i(x_i | x_{\mathrm{pa}(i)})$ is the conditional density of X_i given its parents $X_{\mathrm{pa}(i)}$.

Gaussian Bayesian Networks

Proposition

For Gaussian random variables, the parametrization:

$$f(x) = \prod_{i=1}^{n} f_i(x_i | x_{\mathrm{pa}(i)})$$

is equivalent to the linear parametrization

$$X_i = \sum_{j \in \mathrm{pa}(i)} \lambda_{ji} X_j + Z_i$$

where $Z_i \sim \mathcal{N}(\nu_i, \psi_i^2)$ *and* $\lambda_{ji} \in \mathbb{R}$ *.*

The Trek Rule

- A trek from i to j is a simple path in G with no collider $k \to m$, $l \to m$.
- Every trek T has a topmost element $\mathrm{top}(T)$.
- $T(i,j)$ is set of all treks from i to j.
- For each $i \in [n]$ get variance parameter a_i.
- For each edge $k \to l$ in G get regression parameter λ_{kl}.

Proposition

$X \sim \mathcal{N}(\mu, \Sigma)$ in Bayes Net associated to G iff Σ satisfies:

$$\sigma_{ij} = \sum_{T \in T(i,j)} a_{\mathrm{top}(T)} \prod_{k \to l \in T} \lambda_{kl}$$

with $\lambda_{kl} \in \mathbb{R}$ and $a_i = \mathrm{Var}[X_i]$ is restricted.

The trek rules gives a polynomial parametrization

$$\phi_G : \mathbb{R}^{V(G)} \times \mathbb{R}^{E(G)} \longrightarrow \mathbb{R}^{\binom{n+1}{2}}$$

$$(a, \lambda) \mapsto \Sigma$$

Let

$$M_G \subseteq PD(n)$$

be the set of all covariance matrices that come from the Bayes Net associated to G (roughly, the image of ϕ_G).

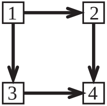
Definition

Let

$$I_G = \{p \in \mathbb{R}[\sigma_{ij} \mid 1 \leq i \leq j \leq n] \mid p(\Sigma) = 0 \ \forall \Sigma \in M_G\}$$

be the vanishing ideal of the Gaussian Bayesian network.

Example of the Trek Rule



$$X_1 = Z_1, \; X_2 = \lambda_{12}X_1 + Z_2, \; X_3 = \lambda_{13}X_1 + Z_3, \; X_4 = \lambda_{24}X_2 + \lambda_{34}X_3 + Z_4$$

$$\begin{array}{cccc}
\sigma_{11} = a_1 & \sigma_{12} = a_1\lambda_{12} & \sigma_{13} = a_1\lambda_{13} & \sigma_{14} = a_1\lambda_{12}\lambda_{24} + a_1\lambda_{13}\lambda_{34} \\
 & \sigma_{22} = a_2 & \sigma_{23} = a_1\lambda_{12}\lambda_{13} & \sigma_{24} = a_2\lambda_{24} + a_1\lambda_{12}\lambda_{13}\lambda_{34} \\
 & & \sigma_{33} = a_3 & \sigma_{34} = a_3\lambda_{34} + a_1\lambda_{13}\lambda_{12}\lambda_{24} \\
 & & & \sigma_{44} = a_4
\end{array}$$

I_G is the complete intersection of a quadric and a cubic:

$$I_G = \langle \sigma_{11}\sigma_{23} - \sigma_{13}\sigma_{21}, \sigma_{12}\sigma_{23}\sigma_{34} + \sigma_{13}\sigma_{24}\sigma_{23} + \cdots \rangle .$$

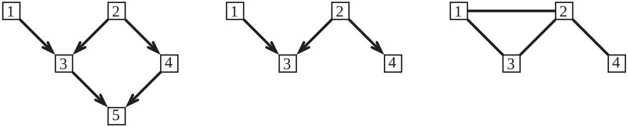
$$I_G = \langle \; |\Sigma_{12,13}|, \; |\Sigma_{123,234}| \; \rangle$$

Markov Properties of the DAG

Proposition (Moralization/*d*-separation)

$X_A \perp\!\!\!\perp X_B | X_C$ holds for Bayes Net associated to G if and only if C separates A and B in the moral graph $(G_{\mathrm{An}(A \cup B \cup C)})^m$.

Is $X_1 \perp\!\!\!\perp X_4 | X_3$?



Theorem

A probability density is in the Bayes Net model of G if and only if it satisfies all CI statements implied by G.

Conditional Independence is an Algebraic Condition

Proposition

If $X \sim \mathcal{N}(\mu, \Sigma)$ then $X_A \perp\!\!\!\perp X_B | X_C$ if and only if all $(\#C + 1) \times (\#C + 1)$ minors of $\Sigma_{A \cup C, B \cup C}$ are zero.

For each DAG G get a conditional independence ideal

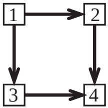
$$CI_G = \left\langle (\#C + 1) \text{ minors of } \Sigma_{A \cup C, B \cup C} \ : \ X_A \perp\!\!\!\perp X_B | X_C \text{ holds for } G \right\rangle .$$

Corollary

$$V(CI_G) \cap PD(n) = V(I_G) \cap PD(n) = M_G$$

Question

Is it always true that $CI_G = I_G$?

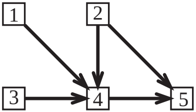


$$X_2 \perp\!\!\!\perp X_3 | X_1 \text{ and } X_1 \perp\!\!\!\perp X_4 | \{X_2, X_3\}$$

$$I_G = CI_G = \langle\ |\Sigma_{12,13}|, |\Sigma_{123,234}|\ \rangle$$

Theorem (S-, 2007)

If T is a tree then $I_T = CI_T$.



$$I_G = CI_G + \langle |\Sigma_{13,45}| \rangle$$

Question

Where do these extra determinantal constraints come from?

Question

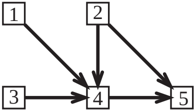
Why are they interesting?

Why Should We Care? Hidden Variables

- Partition $[n] = H \cup O$.
- H hidden variables, O observed variables.
- Density of observed variables is just $f_O(x_O)$.

Proposition

$$\begin{aligned} I_{G,O} &:= \{p \in \mathbb{R}[\sigma_{ij} \mid i,j \in O] \ : \ p(\Sigma_{O,O}) = 0 \ \forall \Sigma \in M_G\} \\ &= I_G \cap \mathbb{R}[\sigma_{ij} \ : \ i,j \in O] \end{aligned}$$

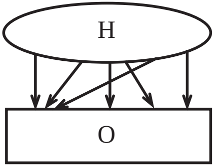


$$I_{G,1345} = \langle \sigma_{13}, |\Sigma_{13,45}| \rangle$$

A Special Grading

Definition

H is *upstream* from *O* if there are no edges $o \to h$ such that $o \in O$ and $H \in h$.



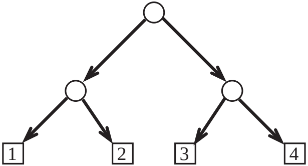
Grading: $\deg \sigma_{ij} = (1, \#(\{i\} \cap O) + \#(\{j\} \cap O))$.

Proposition (S-, 2007)

If H is upstream from O, I_G is homogenous with respect to the upstream grading. In particular, every homogeneous generating set of I_G contains a generating set of $I_{G,O}$.

Consequences for Trees

Let T be a directed tree (no colliders $i \rightarrow k, j \rightarrow k$) and suppose that O is the set of leaves of T. $J_T = I_{T,O}$ in this case.



Corollary

For a directed tree J_T is generated by tetrad constraints:

$$J_T = \left\langle \sigma_{ij}\sigma_{kl} - \sigma_{il}\sigma_{jk} \ : \ \{i,k\} \text{ splits from } \{j,l\} \right\rangle$$

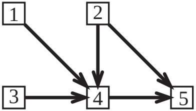
For tree above:

$$\sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23}$$

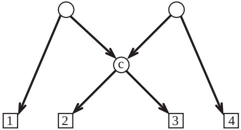
What Causes Extra Constraints? Tetrads and Beyond

Theorem (Spirtes, Glymour, Scheines)

A tetrad $|\Sigma_{ij,kl}| \in I_G$ (i.e. is zero for every covariance matrix in M_G) if and only if there is a choke point c between $\{i, j\}$ and $\{k, l\}$ in G.



4 is a choke point between $\{1, 3\}$ and $\{4, 5\}$.



c is NOT a choke point between $\{1, 2\}$ and $\{3, 4\}$

Definition

Let A, B, C, and D be four subsets of $V(G)$ (not necessarily disjoint). We say that (C, D) t-separates A from B if every trek from A to B passes through either a vertex in C on the A-side of the trek, or a vertex in D on the B-side of the trek.

Proposition

A set C d-separates A from B in G if and only if there is a partition $C = C_1 \cup C_2$ such that (C_1, C_2) t-separates $A \cup C$ from $B \cup C$.

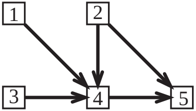
Theorem (S-Talaska)

The matrix $\Sigma_{A,B}$ has rank $\leq d$ if and only if there are $C, D \subset [n]$ with $\#C + \#D \leq d$ such that (C, D) t-separate A from B.

Proof.

- Extend the parametrization to treks with loops.
- $|\Sigma_{A,B}|$ is a determinant of path polynomials. Devise a variant of the Gessel-Viennot Theorem to expand $|\Sigma_{A,B}|$ combinatorially.
- Deduce that $|\Sigma_{A,B}| = 0$ if and only if every trek system has a sided crossing.
- Apply Max-Flow-Min-Cut theorem to deduce a blocking characterization.

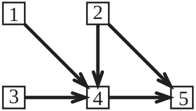
□



We have $|\Sigma_{13,45}| \in I_G$ because $(\emptyset, \{4\})$ t-separate $\{1,3\}$ from $\{4,5\}$.

Could also be deduced from CI statements $\{1,3\} \perp\!\!\!\perp 5 | \{2,4\}$ and $\{1,3\} \perp\!\!\!\perp 2$.

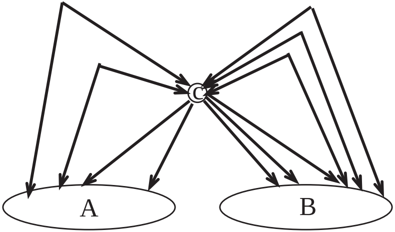
$$\begin{pmatrix} \sigma_{12} & \sigma_{14} & \sigma_{15} \\ \sigma_{22} & \sigma_{24} & \sigma_{25} \\ \sigma_{23} & \sigma_{34} & \sigma_{35} \\ \sigma_{24} & \sigma_{44} & \sigma_{45} \end{pmatrix} = \left(\begin{array}{ccc} 0 & \sigma_{14} & \sigma_{15} \\ >0 & \sigma_{24} & \sigma_{25} \\ 0 & \sigma_{34} & \sigma_{35} \\ \hline \sigma_{24} & \sigma_{44} & \sigma_{45} \end{array}\right)$$



We have $|\Sigma_{13,45}| \in I_G$ because $(\emptyset, \{4\})$ t-separate $\{1,3\}$ from $\{4,5\}$.
Could also be deduced from CI statements $\{1,3\} \perp\!\!\!\perp 5 | \{2,4\}$ and $\{1,3\} \perp\!\!\!\perp 2$.

$$\begin{pmatrix} \sigma_{12} & \sigma_{14} & \sigma_{15} \\ \sigma_{22} & \sigma_{24} & \sigma_{25} \\ \sigma_{23} & \sigma_{34} & \sigma_{35} \\ \sigma_{24} & \sigma_{44} & \sigma_{45} \end{pmatrix} = \begin{pmatrix} 0 & \sigma_{14} & \sigma_{15} \\ {\color{red}{>0}} & \sigma_{24} & \sigma_{25} \\ 0 & \sigma_{34} & \sigma_{35} \\ \hline \sigma_{24} & \sigma_{44} & \sigma_{45} \end{pmatrix}$$

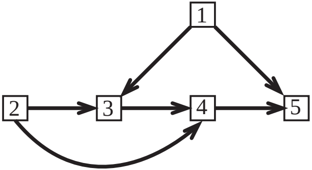
“Spiders”



$(\{c\}, \{c\})$ t-separates A from B.
$\Sigma_{A,B}$ has rank at most 2.

Questions and Open Problems

- Extend *t*-separation characterization of determinantal constraints to ancestral graphs and summary graphs.
- What does *t*-separation mean for general (non-Gaussian) Bayesian networks?
- How to determine general descriptions of other hidden variable constraints?



$$\begin{pmatrix} \sigma_{22} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{23} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ 0 & \sigma_{24} & \sigma_{34} & \sigma_{44} \\ 0 & \sigma_{25} & \sigma_{35} & \sigma_{45} \end{pmatrix}$$