Algebraic Geometry of Bayesian Networks

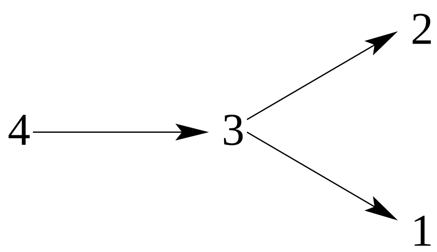
Luis David Garcia, Michael Stillman, and Bernd Sturmfels

lgarcia@math.vt.edu

Virginia Tech.

Bayesian Networks

- Let G be a directed acyclic graph with n nodes.
- The nodes represent random variables, denoted $X_1, \ldots, X_n$. The arrows represent causal dependencies among the variables.



- $p(u_1, u_2, u_3, u_4) = p(u_4)p(u_3|u_4)p(u_2|u_3, u_4)p(u_1|u_2, u_3, u_4)$
- The joint probability distribution is defined as:
  $p(X_1 = u_1, X_2 = u_2, \ldots, X_n = u_n) = \prod_{i=1}^{n} p(u_i|\,\mathrm{pa}(u_i))$
- $p(u_1, u_2, u_3, u_4) = p(u_4)p(u_3|u_4)p(u_2|u_3)p(u_1|u_3)$

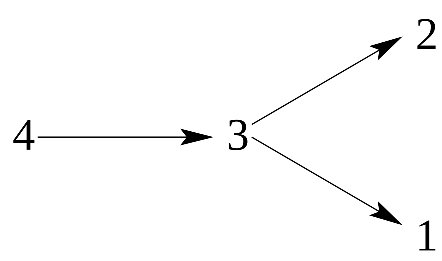
Find all distributions P that admit a recursive factorization according to G

Conditional Independence Relations

- The set of directed local Markov relations of G is the set of independence statements

$$\mathrm{local}(G) \quad = \quad \{X_i \perp\!\!\!\perp \mathrm{nd}(X_i) \mid \mathrm{pa}(X_i) \,:\; i = 1, 2, \ldots, n\},$$

- The set of directed global Markov relations, $\mathrm{global}(G)$, is the set of independent statements $A \perp\!\!\!\perp B \mid C$, for any triple A, B, C of disjoint subsets of vertices of G such that A and B are d-separated by C.



$$\mathrm{local}(G) = \{X_1 \perp\!\!\!\perp \{X_2, X_4\} \mid X_3,\; X_2 \perp\!\!\!\perp \{X_1, X_4\} \mid X_3\}$$

$$\mathrm{global}(G) = \mathrm{local}(G) \cup \{X_4 \perp\!\!\!\perp \{X_1, X_2\} \mid X_3\}$$

> Find all distributions P that satisfy a set of conditional independence relations obtained from G.

Ideals, Varieties, and Independent Statements

- Let $X_1, \dots, X_n$ be discrete random variables, where X_i takes values in $[d_i] = \{1, 2, \dots, d_i\}$.
- Let $\mathbb{R}^D$ denote the real vector space of n-dimensional tables of format $d_1 \times \cdots \times d_n$.
- Let $p_{u_1 u_2 \cdots u_n}$ be an indeterminate representing $p(X_1 = u_1,\ X_2 = u_2, \dots, X_n = u_n)$. Let $\mathbb{R}[D]$ be the ring generated by these indeterminates.

- Let X_1, X_2, X_3, X_4, X_5 be binary variables
- Let $I_{X_1 \perp\!\!\!\perp \{X_2, X_4\} | X_3}$ denote the ideal of $\mathbb{R}[D]$ generated by the 2×2-minors of the matrices.

$$\begin{pmatrix} p_{11k1+} & p_{11k2+} & p_{12k1+} & p_{12k2+} \\ p_{21k1+} & p_{21k2+} & p_{22k1+} & p_{22k2+} \end{pmatrix}$$

- $p_{u_1 u_2 u_3 u_4 +} = p_{u_1 u_2 u_3 u_4 1} + p_{u_1 u_2 u_3 u_4 2}$

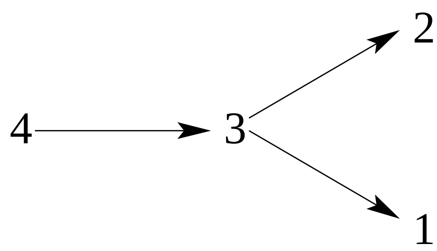
Ideals, Varieties, and Independent Statements (2)

- Let $\mathcal{M}$ be the independence model $\mathcal{M} = \{A^{(1)} \perp\!\!\!\perp B^{(1)} \mid C^{(1)}, \ldots, A^{(m)} \perp\!\!\!\perp B^{(m)} \mid C^{(m)}\}$. Then

$$I_{\mathcal{M}} = I_{A^{(1)} \perp\!\!\!\perp B^{(1)} | C^{(1)}} + \cdots + I_{A^{(m)} \perp\!\!\!\perp B^{(m)} | C^{(m)}}$$

- $V(I_{\mathcal{M}}) \subset \mathbb{C}^D$ is the set of all $d_1 \times \cdots \times d_n$-tables with complex number entries which satisfy the conditional independence statements in $\mathcal{M}$.
- $V_{\geq}(I_{\mathcal{M}} + \langle p - 1 \rangle)$ is the subset of the probability simplex specified by the model $\mathcal{M}$, where p denotes the sum of all unknowns.

Ideals, Varieties, and Independent Statements (3)



$$\text{local}(G) = \{X_1 \perp\!\!\!\perp \{X_2, X_4\} \mid X_3,\ X_2 \perp\!\!\!\perp \{X_1, X_4\} \mid X_3\}$$

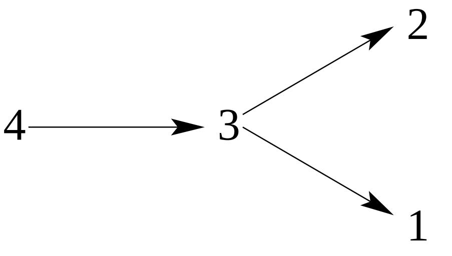
$$\text{global}(G) = \text{local}(G) \cup \{X_4 \perp\!\!\!\perp \{X_1, X_2\} \mid X_3\}$$

- Let $d_1 = d_2 = d_4 = 2$ and d_3 arbitrary
- The ideal $I_{\text{local}(G)} = I_{\text{global}(G)}$ is generated by the 2×2-minors of the following $2 \cdot d_3$ matrices

$$\begin{pmatrix} p_{11k1} & p_{11k2} & p_{12k1} & p_{12k2} \\ p_{21k1} & p_{21k2} & p_{22k1} & p_{22k2} \end{pmatrix} \qquad \begin{pmatrix} p_{11k1} & p_{11k2} & p_{21k1} & p_{21k2} \\ p_{12k1} & p_{12k2} & p_{22k1} & p_{22k2} \end{pmatrix}$$

- For each $k \in [d_3]$, the corresponding quadratic binomials define the Segre Variety $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \longrightarrow \mathbb{P}^7$
- $V(I_{\mathcal{M}})$ is the join of d_3 Segre varieties $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \longrightarrow \mathbb{P}^7$.

Homomorphisms and Recursive Factorization



- Recall $p(u_1, u_2, u_3, u_4) = p(u_1|u_3)p(u_2|u_3)p(u_3|u_4)p(u_4)$.
- Let x_{ij} be an indeterminate representing $p(X_1 = i|X_3 = j)$.
- Note that $\sum_{i=1}^{d_1} x_{ij} = 1$.
- Assume all the variables are binary, then $x_{21} = 1 - x_{11}$.
- Let $\mathbb{R}[E] = \mathbb{R}[x_{11}, x_{12}, y_{11}, y_{12}, z_{11}, z_{12}, w]$.
- The factorization of the joint probability distribution given by the graph G defines a map $\phi : \mathbb{R}^E \to \mathbb{R}^D$.

Homomorphisms and Recursive Factorization (2)

- $p(u_1, u_2, u_3, u_4) = p(u_1|u_3)p(u_2|u_3)p(u_3|u_4)p(u_4)$.
- This map is specified by the ring homomorphism $\Phi : \mathbb{R}[D] \rightarrow \mathbb{R}[E]$ which takes the unknowns

$$
\begin{aligned}
p_{1111} &\longrightarrow x_{11}y_{11}z_{11}w, \\
p_{1112} &\longrightarrow x_{11}y_{11}z_{12}(1-w) \\
&\vdots \\
p_{2222} &\longrightarrow (1-x_{12})(1-y_{12})(1-z_{12})(1-w)
\end{aligned}
$$

- The image of $\phi : \mathbb{R}^E \rightarrow \mathbb{R}^D$ lies in $V_{\mathrm{local}(G)}$
- $I_{\mathrm{local}(G)}$ is contained in the prime ideal $\mathrm{kernel}(\Phi)$.

Factorization Theorem

Theorem (Factorization Theorem). *Let G be a directed acyclic graph and P a probability distribution on $V(G)$. The following are equivalent:*

DF: P *admits a recursive factorization according to* G

DG: P *obeys the Directed Global Markov Property*

DL: P *obeys the Directed Local Markov Property*

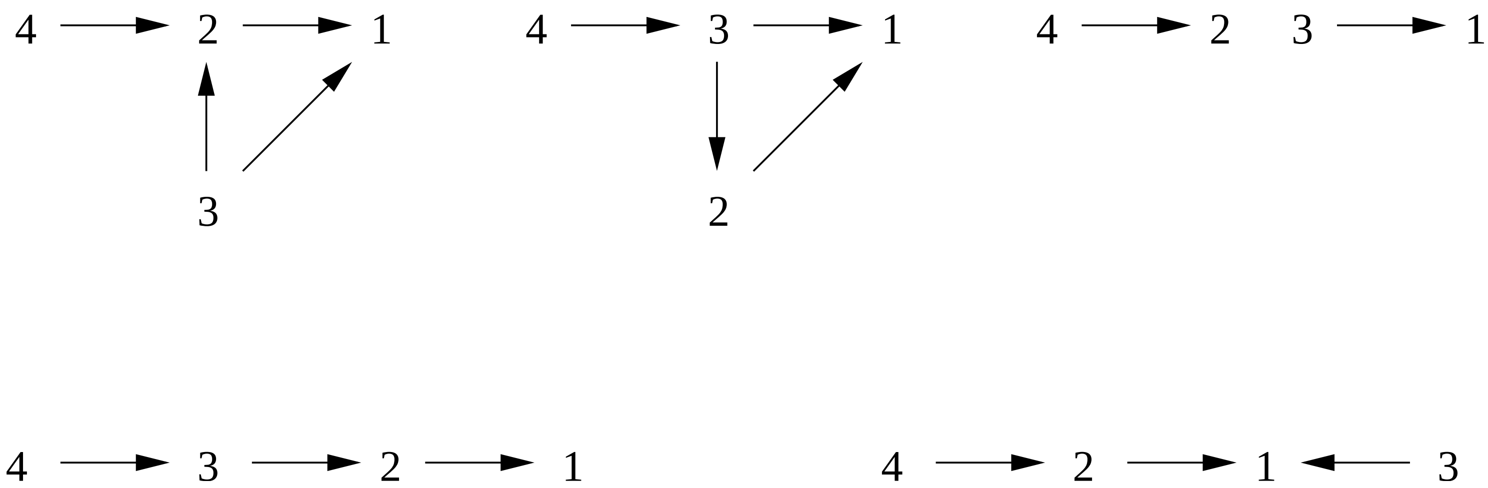
- Denote by $\mathbf{p}$ the product of all of the linear forms (marginals) $p_{++\cdots+u_{r+1}\cdots u_n}$.

Theorem. $\left(I_{\mathrm{local}(G)} : \mathbf{p}^{\infty}\right) \;=\; \left(I_{\mathrm{global}(G)} : \mathbf{p}^{\infty}\right) \;=\; \ker(\Phi).$

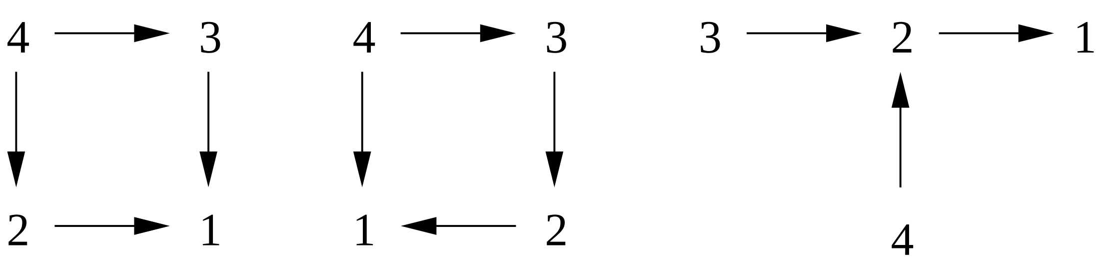
- The prime ideal $\ker(\Phi)$ is called the distinguished component.
- It is the set of all homogeneous polynomial functions on $\mathbb{R}^D$ which vanish on all probabilitity distributions that factor according to G.

Bayesian Networks on four random variables

Theorem. *Of the 30 local Markov ideals on four random variables, 22 are always* *prime, five are not prime but always radical*



and three are not radical.



A Prime example: $2 \longleftarrow 4 \longrightarrow 1 \longleftarrow 3$

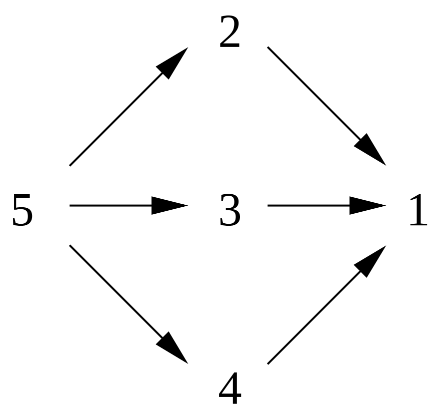
- $\mathrm{local}(G) = \{1 \perp\!\!\!\perp 2 \,|\, \{3,4\},\ 2 \perp\!\!\!\perp \{1,3\} \,|\, 4,\ 3 \perp\!\!\!\perp \{2,4\}\}$.
- $I_{\mathrm{local}(G)}$ is binomial in p_{ijkl} with $i \in \{+, 2, \ldots, d_1\}$.
- The generators are $p_{i_1 j_2 k_1 l} p_{i_2 j_1 k_2 l} - p_{i_1 j_1 k_1 l} p_{i_2 j_2 k_2 l}$, and $p_{+ j_1 k_2 l_1} p_{+ j_2 k_1 l_2} - p_{+ j_1 k_1 l_1} p_{+ j_2 k_2 l_2}$.
- The S-pairs within each group reduce to zero by the Gröbner basis property of the 2×2-minors of a generic matrix.
- The given set of irreducible quadrics is a reverse lexicographic Gröbner basis.
- The lowest variable is not a zero-divisor, and hence by symmetry none of the variables p_{ijkl} is a zero-divisor.
- Since $\left(I_{\mathrm{local}(G)} : \mathbf{p}^{\infty}\right) = \ker(\Phi)$, then $I_{\mathrm{local}(G)}$ coincides with the prime ideal $\ker(\Phi)$.

Global Markov Relations on Five nodes

Theorem. *Of the 301 global Markov ideals on five binary random variables, 220 are prime, 68 are radical but not prime, and 13 are not radical.*

# of components	1	3	5	7	17	25	29	33	39
# of ideals	220	8	41	3	9	1	2	3	1

- `http://math.cornell.edu/~mike/bayes/global5.html`.



- $I_{\mathrm{global}(G_{138})}$ has 207 minimal primes, and 37 embedded primes. Each of the 207 minimal primary components are prime.

Hidden Variables and Higher Secant Varieties

- Let G be a BN on n discrete random variables. The variables corresponding to the nodes $r+1, \ldots, n$ are hidden,
- The observable probabilities are $p_{i_1 i_2 \cdots i_r ++\cdots+} = \sum_{j_{r+1} \in [d_{r+1}]} \sum_{j_{r+2} \in [d_{r+2}]} \cdots \sum_{j_n \in [d_n]} p_{i_1 i_2 \cdots i_r j_{r+1} j_{r+2} \cdots j_n}$.
- Let $D' = [d_1] \times \cdots \times [d_r]$ and $\mathbb{R}[D'] \subset \mathbb{R}[D]$ generated by $p_{i_1 i_2 \cdots i_r ++\cdots+}$.
- Let $\pi : \mathbb{R}^D \to \mathbb{R}^{D'}$ denote the canonical linear epimorphism induced by the inclusion of $\mathbb{R}[D']$ in $\mathbb{R}[D]$.
- Let $P_G = \ker(\Phi)$ be its homogeneous prime ideal.
- $\pi(V_{\geq 0}(P_G)) \subset \pi(V(P_G))_{\geq 0} \subset \pi(V(P_G)) \subset \overline{\pi(V(P_G))} \subset \mathbb{R}^{D'}$.
- The set of all polynomial functions which vanish on the space $\pi(V_{\geq 0}(P_G))$ of observable probability distributions is the prime ideal

$$Q_G = P_G \cap \mathbb{R}[D'].$$

- Let G be a BN with $n+1$ random variables $F_1, \ldots, F_n, H$ and n directed edges (H, F_i), $i = 1, 2, \ldots, n$.
- H is the hidden variable, and its levels $1, 2, \ldots, d_{n+1} =: r$ are called the classes.
- The observed random variables $F_1, \ldots, F_n$ are the features of the model.
- $F_1 \perp\!\!\!\perp F_2 \perp\!\!\!\perp \cdots \perp\!\!\!\perp F_n \,|\, H.$
- $P_G = I_{\text{local}(G)}$. This is the ideal of the join of r copies of the *Segre variety*

$$X_{d_1,d_2,\ldots,d_n} \quad := \quad \mathbb{P}^{d_1-1} \times \mathbb{P}^{d_2-1} \times \cdots \times \mathbb{P}^{d_n-1} \quad \subset \quad \mathbb{P}^{d_1 d_2 \cdots d_n - 1}.$$

- The naive Bayes model with r classes and n features corresponds to the r-th secant variety of a Segre product of n projective spaces:

$$\overline{\pi(V(P_G))} \quad = \quad \mathrm{Sec}^r(X_{d_1,d_2,\ldots,d_n})$$