Graphical models from an algebraic perspective

Elina Robeva
MIT

ICERM Nonlinear Algebra Bootcamp

September 11, 2018

Overview

- Undirected graphical models
  - Definition and parametric description
  - Markov properties and implicit description
  - Discrete and Gaussian
- Directed graphical models
  - Definition and parametric description
  - Markov properties, *d*-separation, and implicit description
  - Discrete and Gaussian
  - model equivalence
- Mixed graphical models

Undirected graphical models

Let $G = (V, E)$ be an undirected graph and $\mathcal{C}(G)$ the set of *maximal cliques* of G.

Let $(X_v : v \in V) \in \mathcal{X} := \prod_{v \in V} \mathcal{X}_v$ be a random vector.

<u>Notation:</u> $\mathcal{X}_A = \prod_{v \in A} \mathcal{X}_v$, $X_A = (X_v : v \in A)$, $x_A = (x_v : v \in A)$.

Undirected graphical models

Let $G = (V, E)$ be an undirected graph and $\mathcal{C}(G)$ the set of *maximal cliques* of G.

Let $(X_v : v \in V) \in \mathcal{X} := \prod_{v \in V} \mathcal{X}_v$ be a random vector.

<u>Notation:</u> $\mathcal{X}_A = \prod_{v \in A} \mathcal{X}_v$, $X_A = (X_v : v \in A)$, $x_A = (x_v : v \in A)$.

For each $C \in \mathcal{C}(G)$ let

$$\phi_C : \mathcal{X}_C \to \mathbb{R}_{\geq 0}$$

be a continuous function called a *clique potential*.

The *undirected graphical model* (or *markov random field*) corresponding to G and $\mathcal{X}$ is the set of all probability density functions on $\mathcal{X}$ of the form

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)$$

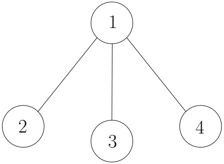
where

$$Z = \int_{\mathcal{X}} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C) d\mu(x)$$

is the normalizing constant.

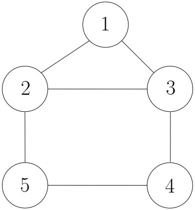
Undirected graphical models

Example



$$p(x_1, x_2, x_3, x_4) = \frac{1}{Z}\phi_{12}(x_1, x_2)\phi_{13}(x_1, x_3)\phi_{14}(x_1, x_4).$$

Example



$$p(x_1, x_2, x_3, x_4, x_5) = \frac{1}{Z}\phi_{123}(x_1, x_2, x_3)\phi_{25}(x_2, x_5)\phi_{34}(x_3, x_4)\phi_{45}(x_4, x_5).$$

Discrete undirected graphical models

Suppose that $\mathcal{X}_v = [r_v]$, $r_v \in \mathbb{N}$. Then, $X \in \mathcal{X} = \prod_{v \in V}[r_v]$. We use parameters

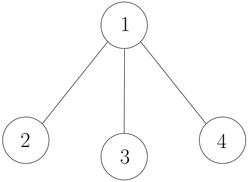
$$\theta^C_{x_C} := \phi_C(x_C), \quad C \in \mathcal{C}(G), x_r \in [r_v].$$

Then, we get the *rational parametrization*

$$p_x = \frac{1}{Z(\theta)} \prod_{C \in \mathcal{C}(G)} \theta^C_{x_C}.$$

The graphical model corresponding to G consists of all discrete distributions $p = (p_x : x \in \mathcal{X})$ that factor in this way.

Example



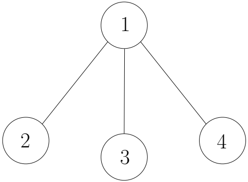
Let $r_1 = r_2 = r_3 = r_4 = 2$. The parametrization has the form

$$p_{x_1x_2x_3x_4} = \frac{1}{Z(\theta)} \theta^{(12)}_{x_1x_2} \theta^{(13)}_{x_1x_3} \theta^{(14)}_{x_1x_4}.$$

The ideal I_G is the ideal of the image of this parametrization.

Discrete undirected graphical models

Example



Let $r_1 = r_2 = r_3 = r_4 = 2$. The parametrization has the form

$$p_{x_1x_2x_3x_4} = \frac{1}{Z(\theta)}\theta^{(12)}_{x_1x_2}\theta^{(13)}_{x_1x_3}\theta^{(14)}_{x_1x_4}.$$

The ideal I_G is the ideal of the image of this parametrization.

```
S = QQ[a_(1,1)..a_(2,2), b_(1,1)..b_(2,2), c_(1,1)..c_(2,2)]
R = QQ[p_(1,1,1,1)..p_(2,2,2,2)]
L = {}
for i from 0 to 15 do (
s = last baseName (vars R)_(0,i);
L = append(L, a_(s_0,s_1)*b_(s_0,s_2)*c_(s_0,s_3))
)
phi = map(S, R, L)
I = ker phi
```

Output:

$$I_G = \langle \text{2-minors of } M_1\rangle + \langle \text{2-minors of } M_2\rangle + \langle \text{2-minors of } M_3\rangle + \langle \text{2-minors of } M_4\rangle$$

where

$$M_1 = \begin{pmatrix} p_{0000} & p_{0001} & p_{0010} & p_{0011} \\ p_{0100} & p_{0101} & p_{0110} & p_{0111} \end{pmatrix}, \quad M_2 = \begin{pmatrix} p_{1000} & p_{1001} & p_{1010} & p_{1011} \\ p_{1100} & p_{1101} & p_{1110} & p_{1111} \end{pmatrix}$$

$$M_3 = \begin{pmatrix} p_{0000} & p_{0001} & p_{0100} & p_{0101} \\ p_{0010} & p_{0011} & p_{0110} & p_{0111} \end{pmatrix}, \quad M_4 = \begin{pmatrix} p_{1000} & p_{1001} & p_{1100} & p_{1101} \\ p_{1010} & p_{1011} & p_{1110} & p_{1111} \end{pmatrix}.$$

Gaussian undirected graphical models

$X = (X_v : v \in V) \sim \mathcal{N}(\mu, \Sigma)$ Gaussian random vector, $K = \Sigma^{-1}$. The density of X is

$$p(x) = \frac{1}{Z} \exp\left(-\frac{1}{2}(x-\mu)^T K (x-\mu) \right)$$

When does it factorize according to $G = (V, E)$, i.e. $p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)$?

Gaussian undirected graphical models

$X = (X_v : v \in V) \sim \mathcal{N}(\mu, \Sigma)$ Gaussian random vector, $K = \Sigma^{-1}$. The density of X is

$$p(x) = \frac{1}{Z} \exp\left(-\frac{1}{2}(x-\mu)^T K (x-\mu)\right)$$

When does it factorize according to $G = (V, E)$, i.e. $p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)$?

$$p(x) = \frac{1}{Z} \prod_{v \in V} \exp\left(-\frac{1}{2}(x_v - \mu_v)^2 K_{vv}\right) \prod_{v \neq u} \exp\left(-\frac{1}{2}(x_v - \mu_v)(x_u - \mu_u) K_{vu}\right).$$

Gaussian undirected graphical models

$X = (X_v : v \in V) \sim \mathcal{N}(\mu, \Sigma)$ Gaussian random vector, $K = \Sigma^{-1}$. The density of X is

$$p(x) = \frac{1}{Z} \exp\left(-\frac{1}{2}(x-\mu)^T K (x-\mu)\right)$$

When does it factorize according to $G = (V, E)$, i.e. $p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)$?

$$p(x) = \frac{1}{Z} \prod_{v \in V} \exp\left(-\frac{1}{2}(x_v - \mu_v)^2 K_{vv}\right) \prod_{v \neq u} \exp\left(-\frac{1}{2}(x_v - \mu_v)(x_u - \mu_u) K_{vu}\right).$$

The density factorizes according to $G = (V, E)$ if and only if

$$K_{uv} = 0 \text{ for all } (u, v) \notin E.$$

Gaussian undirected graphical models

$X = (X_v : v \in V) \sim \mathcal{N}(\mu, \Sigma)$ Gaussian random vector, $K = \Sigma^{-1}$. The density of X is

$$p(x) = \frac{1}{Z} \exp\left(-\frac{1}{2}(x-\mu)^T K (x-\mu)\right)$$

When does it factorize according to $G = (V, E)$, i.e. $p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)$?

$$p(x) = \frac{1}{Z} \prod_{v \in V} \exp\left(-\frac{1}{2}(x_v - \mu_v)^2 K_{vv}\right) \prod_{v \neq u} \exp\left(-\frac{1}{2}(x_v - \mu_v)(x_u - \mu_u) K_{vu}\right).$$

The density factorizes according to $G = (V, E)$ if and only if

$$K_{uv} = 0 \text{ for all } (u, v) \notin E.$$

The *parametric description* of the Gaussian graphical model with respect to $G = (V, E)$ is

$$\mathcal{M}_G = \{\Sigma = K^{-1} : K \succ 0 \text{ and } K_{uv} = 0 \text{ for all } (u, v) \notin E\}.$$

The ideal of the model I_G is the ideal of the image of this parametrization.

Markov properties and conditional independence for undirected graphical models

A different way to define undirected graphical models is via conditional independence statements.

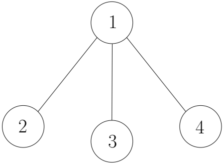
Markov properties and conditional independence for undirected graphical models

A different way to define undirected graphical models is via conditional independence statements.

Let $G = (V, E)$.
For $A, B, C \subseteq V$, say that A and B are **separated** by C if every path between $a \in A$ and $b \in B$ goes through a vertex in C.

The **global Markov property** associated to G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B | X_C$ for all disjoint sets A, B, C such that C separates A and B.

Example



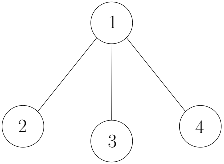
Markov properties and conditional independence for undirected graphical models

A different way to define undirected graphical models is via conditional independence statements.

Let $G = (V, E)$.
For $A, B, C \subseteq V$, say that A and B are **separated** by C if every path between $a \in A$ and $b \in B$ goes through a vertex in C.

The **global Markov property** associated to G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B | X_C$ for all disjoint sets A, B, C such that C separates A and B.

Example



Global Markov property:

$$X_2 \perp\!\!\!\perp X_3 | X_1$$
$$X_2 \perp\!\!\!\perp X_4 | X_1$$
$$X_3 \perp\!\!\!\perp X_4 | X_1$$

Conditional independence for discrete distributions

For discrete random variables conditional independence yields polynomial equations in

$$(p_x : x \in \mathcal{X}).$$

How?

Conditional independence for discrete distributions

For discrete random variables conditional independence yields polynomial equations in

$$(p_x : x \in \mathcal{X}).$$

How?

Example

If $V = \{1, 2\}$ and $\mathcal{X} = [m_1] \times [m_2]$, then $X_1 \perp\!\!\!\perp X_2$ is the same as

$$p_{ij} = p_{i+}p_{+j} \quad \text{for all } i \in [m_1], j \in [m_2].$$

Equivalently, the matrix

$$P = (p_{ij}) = \begin{pmatrix} p_{1+} \\ \vdots \\ p_{m_1+} \end{pmatrix} \begin{pmatrix} p_{+1} & \cdots & p_{+m_2} \end{pmatrix},$$

has rank 1. So, equivalently its 2×2 minors vanish, i.e. $p_{ij}p_{k\ell} - p_{i\ell}p_{kj} = 0$ for all $i, k \in [m_1], j, \ell \in [m_2]$.

Conditional independence for discrete distributions

Proposition

Let X be a discrete random vector with sample space $\mathcal{X} = \prod_{i=1}^{n}[m_i]$. Then for disjoint sets $A, B, C \subset [n]$, we have that $X_A \perp\!\!\!\perp X_B | X_C$ if and only if

$$p_{i_A i_B i_C +} p_{j_A j_B i_C +} - p_{i_A j_B i_C +} p_{j_A i_B i_C +} = 0 \quad \textit{for all} \;\; i_A \neq j_A \in \mathcal{X}_A, i_B \neq j_B \in \mathcal{X}_B, i_C \in \mathcal{X}_C.$$

Conditional independence for discrete distributions

Recall: the **global Markov property** w.r.t. G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B | X_C$ for all disjoint A, B, C s.t. C separates A and B.

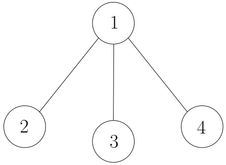
The global Markov properteis define an ideal $I_{\text{global}(G)} \subseteq \mathbb{R}[p_x : x \in \mathcal{X}]$.

Conditional independence for discrete distributions

Recall: the **global Markov property** w.r.t. G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B | X_C$ for all disjoint A, B, C s.t. C separates A and B.

The global Markov properteis define an ideal $I_{\text{global}(G)} \subseteq \mathbb{R}[p_x : x \in \mathcal{X}]$.

Example



Let $X_1, X_2, X_3, X_4 \in \{1, 2\}$. Global Markov property:

$$X_2 \perp\!\!\!\perp X_3, X_4 | X_1$$
$$X_3 \perp\!\!\!\perp X_2, X_4 | X_1$$
$$X_4 \perp\!\!\!\perp X_2, X_3 | X_1$$

Ideal associated to the global Markov property is

$$I_{\text{global}(G)} = \langle \text{2-minors of } M_1 \rangle + \langle \text{2-minors of } M_2 \rangle + \langle \text{2-minors of } M_3 \rangle + \langle \text{2-minors of } M_4 \rangle = I_G$$

where

$$M_1 = \begin{pmatrix} p_{0000} & p_{0001} & p_{0010} & p_{0011} \\ p_{0100} & p_{0101} & p_{0110} & p_{0111} \end{pmatrix}, \ M_2 = \begin{pmatrix} p_{1000} & p_{1001} & p_{1010} & p_{1011} \\ p_{1100} & p_{1101} & p_{1110} & p_{1111} \end{pmatrix}$$

$$M_3 = \begin{pmatrix} p_{0000} & p_{0001} & p_{0100} & p_{0101} \\ p_{0010} & p_{0011} & p_{0110} & p_{0111} \end{pmatrix}, \ M_4 = \begin{pmatrix} p_{1000} & p_{1001} & p_{1100} & p_{1101} \\ p_{1010} & p_{1011} & p_{1110} & p_{1111} \end{pmatrix}.$$

Conditional independence for Gaussian distributions

For Gaussian random variables $X = (X_v : v \in V) \sim \mathcal{N}(\mu, \Sigma)$, conditional independence statements yield polynomial equations in the entries of Σ!

Conditional independence for Gaussian distributions

For Gaussian random variables $X = (X_v : v \in V) \sim \mathcal{N}(\mu, \Sigma)$, conditional independence statements yield polynomial equations in the entries of Σ!

- **Independence** in a Gaussian distribution $X \sim \mathcal{N}(\mu, \Sigma)$ is equivalent to entries of Σ vanishing:
$$X_a \perp\!\!\!\perp X_b \iff \Sigma_{a,b} = 0.$$
- **Conditional independence** in a Gaussian distribution $X \sim \mathcal{N}(\mu, \Sigma)$ is equivalent to a rank condition:
$$X_A \perp\!\!\!\perp X_B | X_C \iff \text{rank}(\Sigma_{A\cup C, B\cup C}) \leq |C|.$$

Proof.

Exercise. □

Markov properties for undirected Gaussian graphical models

Proposition

The set of of Gaussian covariance matrices compatible with the global Markov properties for G is precisely

$$\mathcal{M}_G = \{\Sigma \succ 0 : \mathit{rank}(\Sigma_{A\cup C, B\cup C}) \leq |C| \text{ for all } A, B, C \subseteq V \text{ s.t. } C \text{ separates } A \text{ and } B\}.$$

The ideal $I_{\text{global}(G)} \subseteq \mathbb{R}[\Sigma]$ corresponding to the global Markov property for G is

$$I_{\text{global}(G)} = \langle (|C|+1)\text{-minors of } \Sigma_{A\cup C, B\cup C} : A, B, C \subseteq V \text{ s.t. } C \text{ separates } A \text{ and } B\rangle.$$

Markov properties for undirected Gaussian graphical models

Proposition

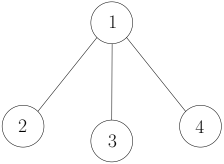
The set of of Gaussian covariance matrices compatible with the global Markov properties for G is precisely

$$\mathcal{M}_G = \{\Sigma \succ 0 : rank(\Sigma_{A\cup C, B\cup C}) \leq |C| \text{ for all } A, B, C \subseteq V \text{ s.t. } C \text{ separates } A \text{ and } B\}.$$

The ideal $I_{\text{global}(G)} \subseteq \mathbb{R}[\Sigma]$ corresponding to the global Markov property for G is

$$I_{\text{global}(G)} = \langle (|C|+1)\text{-minors of } \Sigma_{A\cup C, B\cup C} : A, B, C \subseteq V \text{ s.t. } C \text{ separates } A \text{ and } B \rangle.$$

Example



Global Markov property:

$$X_2 \perp\!\!\!\perp X_3, X_4 | X_1$$
$$X_2 \perp\!\!\!\perp X_3, X_4 | X_1$$
$$X_3 \perp\!\!\!\perp X_2, X_4 | X_1$$

The global Markov property yields the ideal

$$I_{\text{global}(G)} = \langle \det \Sigma_{12,13}, \det \Sigma_{12,14}, \det \Sigma_{13,14}, \det \Sigma_{12,34}, \det \Sigma_{13,24}, \det \Sigma_{14,23} \rangle.$$

Equivalence of parametric and implicit descriptions

Theorem (Hammersley-Clifford)

A continuous positive distribution P on X factorizes according to G if and only if it satisfies the global Markov property for the graph G.

- For discrete distributions:

$$\mathcal{V}(I_G) \cap \Delta_{(|\mathcal{X}|-1),+} = \mathcal{V}(I_{\text{global}(G)}) \cap \Delta_{(|\mathcal{X}|-1),+}.$$

- For Gaussian distributions

$$\mathcal{V}(I_G) \cap \{\Sigma \succ 0\} = \mathcal{V}(I_{\text{global}(G)}) \cap \{\Sigma \succ 0\}.$$

Directed acyclic graphical models

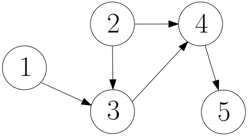
Let $G = (V, E)$ be a *directed acyclic graph* (or *DAG*). For each node $v \in V$, let $\text{pa}(v)$ be the parents of v. Let $X \in \prod_{v \in V} \mathcal{X}_v$ be our random variable.

The distribution $p(x)$ *factors according to the graph* G if

$$p(x) = \prod_{v \in V} p(x_v | x_{\text{pa}(v)}).$$

for all $x \in \mathcal{X}$.

Example



The distribution $p(x)$ factors according to this graph if

$$p(x) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_2, x_3)p(x_5|x_4)$$

for all $x \in \mathcal{X}$.

The *directed acyclic graphical model* (or *Bayesian network*) corresponding to a DAG G and a state space $\mathcal{X}$ is the set of all densities that factorize in according to G.

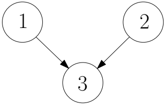
Discrete directed graphical models

The factorization gives a parametric description of discrete graphical models.

Discrete directed graphical models

The factorization gives a parametric description of discrete graphical models.

Example



Assume that variables are binary: $X_1, X_2, X_3 \in \{1, 2\}$. We have

$$p_{x_1,x_2,x_3} = p(x_1)p(x_2)p(x_3|x_1,x_2) = \theta^{(1)}_{x_1}\theta^{(2)}_{x_2}\theta^{(3)}_{x_3|x_1,x_2}.$$

Note that

$$1 = \theta^{(1)}_1 + \theta^{(1)}_2 = \theta^{(2)}_1 + \theta^{(2)}_2 = \theta^{(3)}_{1|x_1,x_2} + \theta^{(3)}_{2|x_1,x_2}$$

for all values $x_1, x_2 \in \{1, 2\}$. Using Macaulay2, we can compute the vanishing ideal I_G for this model:

```
S = QQ[a,b,c11,c12,c21,c22];
R = QQ[p111,p112,p121,p122,p211,p212,p221,p222];
f = map(S,R, { a*b*c11, a*b*(1-c11), a*(1-b)*c12, a*(1-b)*(1-c12),
(1-a)*b*c21, (1-a)*b*(1-c21), (1-a)*(1-b)*c22, (1-a)*(1-b)*(1-c22)});
I = kernel f
```

The output is:

$$I_G = \langle p_{11+}p_{22+} - p_{12+}p_{21+} \rangle = I_{1 \perp\!\!\!\perp 2}.$$

Gaussian directed graphical models

The factorization of a Gaussian DAG model also gives a parametrization of the model! How?

Gaussian directed graphical models

The factorization of a Gaussian DAG model also gives a parametrization of the model! How?

Theorem

Let $X \sim \mathcal{N}(\mu, \Sigma)$ be a Gaussian random vector. The density of X factors according to the DAG G if and only if we can write

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

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \sim \mathcal{N}(\nu, \Omega = diag(\omega_1, \ldots, \omega_n))$, i.e. the ϵ_i are independent of each other.

Proof.

Exercise. □

Gaussian directed graphical models

The factorization of a Gaussian DAG model also gives a parametrization of the model! How?

Theorem

Let $X \sim \mathcal{N}(\mu, \Sigma)$ be a Gaussian random vector. The density of X factors according to the DAG G if and only if we can write

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

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \sim \mathcal{N}(\nu, \Omega = diag(\omega_1, \ldots, \omega_n))$, i.e. the ϵ_i are independent of each other.

Proof.

Exercise. □

Equivalently,

$$X = \Lambda^T X + \epsilon, \text{ where } \Lambda_{ij} = \begin{cases} \lambda_{ij} & \text{if } i \to j \in E \\ 0 & \text{otherwise.} \end{cases}.$$

Gaussian directed graphical models

Note that

$$X = \Lambda^T X + \epsilon \quad \iff \quad X = (I - \Lambda)^{-T} \epsilon.$$

Therefore, the covariance matrix of X is

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

Gaussian directed graphical models

Note that

$$X = \Lambda^T X + \epsilon \quad \Longleftrightarrow \quad X = (I - \Lambda)^{-T} \epsilon.$$

Therefore, the covariance matrix of X is

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

Corollary

The Gaussian graphical model associated to the DAG $G = (V, E)$ is

$$\mathcal{M}_G = \{\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} : \ \Lambda \in \mathbb{R}^E \text{ and } \Omega \succ 0 \text{ is diagonal}\}.$$

Gaussian directed graphical models

Note that

$$X = \Lambda^T X + \epsilon \quad \Longleftrightarrow \quad X = (I - \Lambda)^{-T} \epsilon.$$

Therefore, the covariance matrix of X is

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

Corollary

The Gaussian graphical model associated to the DAG $G = (V, E)$ is

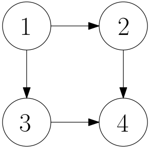
$$\mathcal{M}_G = \{\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} : \ \Lambda \in \mathbb{R}^E \textit{ and } \Omega \succ 0 \textit{ is diagonal}\}.$$

Definition

The Gaussian vanishing ideal for a given DAG G is the ideal $I_G \subseteq \mathbb{R}[\Sigma]$ of the image of this parametrization.

Gaussian directed graphical models

Example



$$\Lambda = \begin{pmatrix} 0 & \lambda_{12} & \lambda_{13} & 0 \\ 0 & 0 & 0 & \lambda_{24} \\ 0 & 0 & 0 & \lambda_{34} \end{pmatrix}, \quad (I-\Lambda)^{-1} = \begin{pmatrix} 1 & \lambda_{12} & \lambda_{13} & \lambda_{12}\lambda_{24}+\lambda_{13}\lambda_{34} \\ 0 & 1 & 0 & \lambda_{24} \\ 0 & 0 & 1 & \lambda_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Sigma = (I-\Lambda)^{-T} \begin{pmatrix} \omega_1 & & & \\ & \omega_2 & & \\ & & \omega_3 & \\ & & & \omega_4 \end{pmatrix} (I-\Lambda)^{-1}$$

$$= \begin{pmatrix} \omega_1 & \omega_1\lambda_{12} & \omega_1\lambda_{13} & \omega_1\lambda_{12}\lambda_{24}+\omega_1\lambda_{13}\lambda_{34} \\ \omega_1\lambda_{12} & \omega_2+\omega_1\lambda_{12}^2 & \omega_1\lambda_{12}\lambda_{13} & \omega_2\lambda_{24}+\omega_1\lambda_{12}^2\lambda_{24}+\omega_1\lambda_{12}\lambda_{13}\lambda_{34} \\ \dots & & & \\ \dots & & & \end{pmatrix}.$$

The ideal of the parametrization is $I_G = \langle |\Sigma_{12,13}|, |\Sigma_{123,234}| \rangle = I_{2 \perp\!\!\!\perp 3|1,\ 1 \perp\!\!\!\perp 4|2,3}$.

Markov properties for directed acyclic graphical models

Let $G = (V, E)$ be a DAG.

The **directed global Markov property** associated to G consists of all conditional independence statements $X_A \perp\!\!\!\perp X_B | X_C$ for all disjoint sets A, B, C such that C d-**separates** A and B.

d-separation

An *undirected path* in a DAG G is a sequence of nodes $u_0, \ldots, u_k$ such that either $u_i \leftarrow u_{i+1}$ or $u_i \rightarrow u_{i+1}$ for all $i \geq 0$.

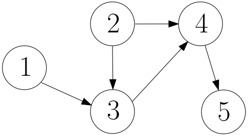
The vertex u_i is a **collider** in an undirected path if $u_{i-1} \rightarrow u_i \leftarrow u_{i+1}$.

Definition

Two nodes $u, v \in V$ in a DAG G are d-**separated** given $C \subseteq V \setminus \{u, v\}$ if for *every* undirected path π between u and v

- either $\exists$ a non-collider in C
- or $\exists$ a collider not in $C \cup \text{an}(C)$.

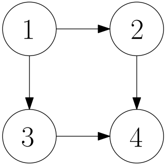
Example



d-separation:	Global Markov properties:
$1 \perp_d 2$	$X_1 \perp\!\!\!\perp X_2$
$1 \perp_d 4 \mid 2, 3$	$X_1 \perp\!\!\!\perp X_4 \mid X_2, X_3$
$1 \perp_d 5 \mid 4$	$X_1 \perp\!\!\!\perp X_5 \mid X_4$
$1 \not\perp_d 2 \mid 5$	

Markov properties for DAG models

Example



d-separation:

$$2 \perp_d 3|1$$
$$1 \perp_d 4|2,3$$

Global Markov properties:

$$X_2 \perp\!\!\!\perp X_3 | X_1$$
$$X_1 \perp\!\!\!\perp X_4 | X_2, X_3$$

- Discrete: let $X_1, X_2, X_3, X_4 \in \{1,2\}$. Then

$$\begin{aligned} I_{\text{global}(G)} = \langle & p_{111+}p_{122+} - p_{112+}p_{121+}, p_{211+}p_{222+} - p_{212+}p_{221+}, \\ & p_{1111}p_{2112} - p_{1112}p_{2111}, p_{1121}p_{2122} - p_{1122}p_{2121}, \\ & p_{1211}p_{2212} - p_{1212}p_{2211}, p_{1221}p_{2222} - p_{1222}p_{2221}\rangle. \end{aligned}$$

- Gaussian:

$$I_{\text{global}(G)} = \langle \det \Sigma_{12,13}, \det \Sigma_{123,234}\rangle = I_G.$$

Hammersley-Clifford Theorem for directed acyclic graphical models

Theorem

A probability density factorizes according to a DAG G if and only if it satisfies the global Markov property with respect to G.

Hammersley-Clifford Theorem for directed acyclic graphical models

Theorem

A probability density factorizes according to a DAG G if and only if it satisfies the global Markov property with respect to G.

For Gaussian directed acyclic graphical models:

$$\mathcal{M}_G = \{\Sigma \succ 0\} \cap \mathcal{V}(I_G) = \{\Sigma \succ 0\} \cap \mathcal{V}(I_{\text{global}(G)}).$$

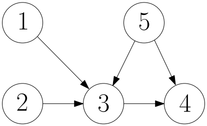
Note that

$$I_{\text{global}(G)} \subseteq I_G,$$

but equality doesn't always hold.

Gaussian directed graphical models in Macaulay2

Example



There is a Macaulay2 package called "GraphicalModels" specifically designed for working with parametrizations and conditional independence ideals in graphical models.

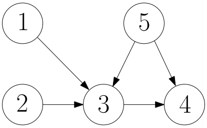
```
loadPackage "GraphicalModels"
G = digraph{{1,{3}},{2,{3}},{3,{4}},{5,{3,4}}}
R = gaussianRing G
I = conditionalIndependenceIdeal(R,globalMarkov(G))
J = gaussianVanishingIdeal(R)
I == J
```

Output: `false`

Reason: $|\Sigma_{12,34}| \in I_G$ but $|\Sigma_{12,34}| \notin I_{\text{global}(G)}$.

Gaussian directed graphical models in Macaulay2

Example



There is a Macaulay2 package called "GraphicalModels" specifically designed for working with parametrizations and conditional independence ideals in graphical models.

```
loadPackage "GraphicalModels"
G = digraph{{1,{3}},{2,{3}},{3,{4}},{5,{3,4}}}
R = gaussianRing G
I = conditionalIndependenceIdeal(R,globalMarkov(G))
J = gaussianVanishingIdeal(R)
I == J
```

Output: `false`

Reason: $|\Sigma_{12,34}| \in I_G$ but $|\Sigma_{12,34}| \notin I_{\text{global}(G)}$.

Theorem

For a Gaussian DAG model the following relationship holds between I_G and $I_{global(G)}$:

$$I_G = I_{global(G)} : \left(\prod_{A \subseteq V} \det(\Sigma_{A,A}) \right)^{\infty}.$$

Markov equivalence for directed acyclic graphical models

Undirected graphical models:

- unique set of global Markov statements,
- unique family of probability distributions.

Not true for directed graphical models!

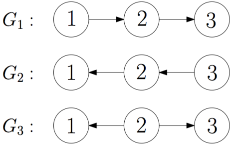
Markov equivalence for directed acyclic graphical models

Undirected graphical models:

- unique set of global Markov statements,
- unique family of probability distributions.

Not true for directed graphical models!

Example



All three of these DAGS have the global Markov property consisting of one statement:

$$X_1 \perp\!\!\!\perp X_3 | X_2.$$

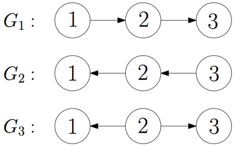
Markov equivalence for directed acyclic graphical models

Undirected graphical models:

- unique set of global Markov statements,
- unique family of probability distributions.

Not true for directed graphical models!

Example



All three of these DAGS have the global Markov property consisting of one statement:

$$X_1 \perp\!\!\!\perp X_3 | X_2.$$

Definition

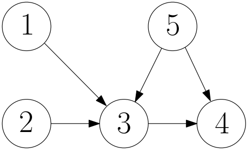
Two DAGs are **Markov equivalent** if they yield the same set of global Markov statements, i.e. they have the same *d*-separation.

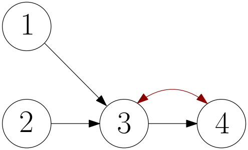
Theorem

Two DAGS G_1 and G_2 are Markov equivalent if and only if

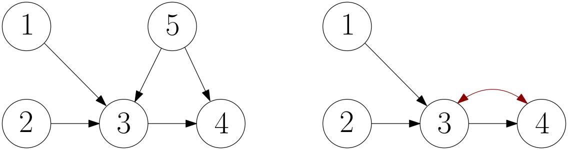
1. *G_1 and G_2 have the same underlying undirected graph,*
2. *G_1 and G_2 have the same* unshielded colliders, *i.e. triples of vertices u, v, w which induce the subgraph $u \to v \leftarrow w$.*

Linear Structural Equation Models





Linear Structural Equation Models



Definition

A *mixed graph* is a triple $G = (V, D, B)$ where

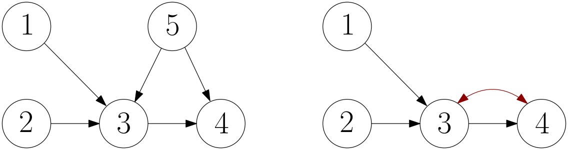
- D is the set of *directed edges* $i \to j$, and
- B is the set of *bidirected edges* $i \leftrightarrow j$.

Gaussian random vectors $X = (X_v : v \in V), \epsilon = (\epsilon_v : v \in V)$ such that

$$X = \Lambda^T X + \epsilon,$$

where $\Lambda \in \mathbb{R}^D$, and $\mathrm{Var}(\epsilon) = \Omega$, where $\Omega_{uv} = 0$ for $(u, v) \notin B$.

Linear Structural Equation Models



Definition

A *mixed graph* is a triple $G = (V, D, B)$ where

- D is the set of *directed edges* $i \to j$, and
- B is the set of *bidirected edges* $i \leftrightarrow j$.

Gaussian random vectors $X = (X_v : v \in V), \epsilon = (\epsilon_v : v \in V)$ such that

$$X = \Lambda^T X + \epsilon,$$

where $\Lambda \in \mathbb{R}^D$, and $\mathrm{Var}(\epsilon) = \Omega$, where $\Omega_{uv} = 0$ for $(u, v) \notin B$.

Example

$$\Lambda = \begin{pmatrix} 0 & 0 & \lambda_{13} & 0 \\ 0 & 0 & \lambda_{23} & 0 \\ 0 & 0 & 0 & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \omega_{11} & & & \\ & \omega_{22} & & \\ & & \omega_{33} & \omega_{34} \\ & & \omega_{34} & \omega_{44} \end{pmatrix}.$$

Linear Structural Equation Models

$$X = \Lambda^T X + \epsilon \quad \Longleftrightarrow \quad X = (I - \Lambda)^{-T}\epsilon.$$

Thus, if $\Sigma = \mathrm{Var}(X)$, then

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

Definition

The **linear structural equation model** associated to a mixed graph $G = (V, D, B)$ is

$$\mathcal{M}_G = \{(I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1} : \Lambda \in \mathbb{R}^D, \Omega \in PD(B)\}.$$

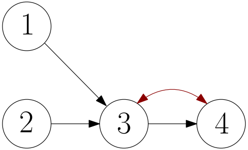
The **parametrization map** of this model is

$$\phi_G : \mathbb{R}^D \times PD(B) \to PD_V, \qquad (\Lambda, \Omega) \mapsto (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}.$$

What is the ideal of the image of ϕ_G? A complete characterization of generators isn't known, Markov properties aren't enough.

Linear Structural Equation Models

Example

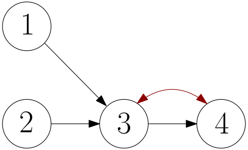


$$I_G = \langle |\Sigma_{12,45}| \rangle.$$

Not a conditional independence ideal!
Corresponds to **trek separation**.

Linear Structural Equation Models

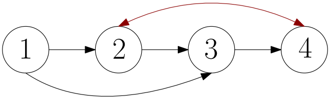
Example



$$I_G = \langle |\Sigma_{12,45}| \rangle.$$

Not a conditional independence ideal!
Corresponds to **trek separation**.

Example (Verma Graph)



$$I_G = \langle \sigma_{11}\sigma_{13}\sigma_{22}\sigma_{34} - \sigma_{11}\sigma_{13}\sigma_{23}\sigma_{24}$$
$$-\sigma_{11}\sigma_{14}\sigma_{22}\sigma_{33} + \sigma_{11}\sigma_{14}\sigma_{23}^2 - \sigma_{12}^2\sigma_{13}\sigma_{34}$$
$$+\sigma_{12}^2\sigma_{14}\sigma_{33} + \sigma_{12}\sigma_{13}^2\sigma_{24} - \sigma_{12}\sigma_{13}\sigma_{14}\sigma_{23} \rangle.$$

Not determinantal. It turns out that

$$I_G = \left\langle \begin{vmatrix} |\Sigma_{123,123}| & |\Sigma_{123,124}| \\ \Sigma_{1,3} & \Sigma_{1,4} \end{vmatrix} \right\rangle .$$

Linear Structural Equation Models

Open problems:

- Parameter identifiability: is ϕ_G (generically) injective?
- What is the dimension of the model $\mathcal{M}_G$?
- Covariance equivalence: what are the equivalence classes of mixed graphs?
- What are the generators of I_G?
- Maximum likelihood estimation: when does the MLE exist, what is the ML-degree?

  . . .

[1] S. Sullivant. *Algebraic Statistics* (2018)
[2] M. Drton. *Algebraic Problems in Linear Structural Equation Modeling* (2016)

Linear Structural Equation Models

Open problems:

- Parameter identifiability: is ϕ_G (generically) injective?
- What is the dimension of the model $\mathcal{M}_G$?
- Covariance equivalence: what are the equivalence classes of mixed graphs?
- What are the generators of I_G?
- Maximum likelihood estimation: when does the MLE exist, what is the ML-degree?

  ...

[1] S. Sullivant. *Algebraic Statistics* (2018)
[2] M. Drton. *Algebraic Problems in Linear Structural Equation Modeling* (2016)

Thank you!