Identifiability of Gaussian Graphical Models

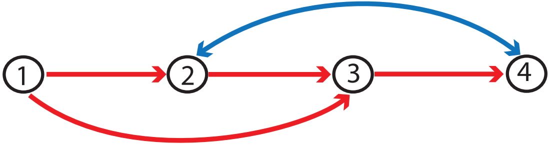
Seth Sullivant

North Carolina State University

March 1, 2010

Mixed Graphs

- $G = (V, B, D)$ graph with directed edges D ($i \to j$) and bidirected edges B ($i \leftrightarrow j$)
- Vertex set $V = [m] := \{1, 2, \dots m\}$
- G is acyclic: $i \to j \in D$ implies $i < j$.

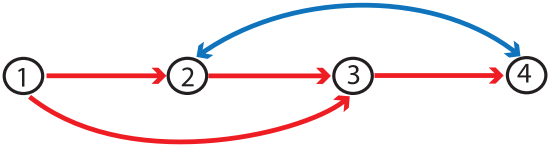


Gaussian graphical model is a statistical model that associates a family of normal distributions to the graph $G = (V, B, D)$.

- PD_m = cone of symmetric positive definite matrices
- Let $PD(B) := \{M \in PD_m : M_{ij} = 0 \text{ if } i \neq j \text{ and } i \leftrightarrow j \notin B\}$
- $\epsilon \in \mathbb{R}^m$, $\epsilon \sim \mathcal{N}(0, \Omega)$ with $\Omega \in PD(B)$
- For $i \to j \in D$, let $\lambda_{ij} \in \mathbb{R}$
- Define $X \in \mathbb{R}^m$ recursively by

$$X_j = \sum_{i : i \to j \in D} \lambda_{ij} X_i + \epsilon_j.$$

Example



$$\epsilon \sim \mathcal{N}(0, \Omega) \qquad \Omega = \begin{pmatrix} \omega_{11} & 0 & 0 & 0 \\ 0 & \omega_{22} & 0 & \omega_{24} \\ 0 & 0 & \omega_{33} & 0 \\ 0 & \omega_{42} & 0 & \omega_{44} \end{pmatrix}$$

$$X_1 = \epsilon_1, \ X_2 = \lambda_{12} X_1 + \epsilon_2, \ X_3 = \lambda_{13} X_1 + \lambda_{23} X_2 + \epsilon_3, \ X_4 = \lambda_{34} X_3 + \epsilon_4$$

Matrix Factorization

Let Λ $m \times m$ upper triangular matrix such that

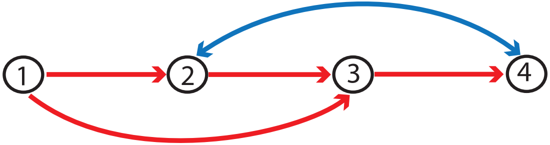
$$\Lambda_{ij} = \begin{cases} \lambda_{ij} & \text{if } i \rightarrow j \in D \\ 0 & \text{otherwise.} \end{cases}$$

Proposition

X from the graph $G = (V, B, D)$ is distributed $\mathcal{N}(0, \Sigma)$ where

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

Note that $(I - \Lambda)^{-1} = I + \Lambda + \Lambda^2 + \cdots + \Lambda^{m-1}$.



$$X_1 = \epsilon_1, \;\; X_2 = \lambda_{12}X_1 + \epsilon_2, \;\; X_3 = \lambda_{13}X_1 + \lambda_{23}X_2 + \epsilon_3, \;\; X_4 = \lambda_{34}X_3 + \epsilon_4$$

$$\Sigma = \begin{pmatrix} 1 & -\lambda_{12} & -\lambda_{13} & 0 \\ 0 & 1 & -\lambda_{23} & 0 \\ 0 & 0 & 1 & -\lambda_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-T} \begin{pmatrix} \omega_{11} & 0 & 0 & 0 \\ 0 & \omega_{22} & 0 & \omega_{24} \\ 0 & 0 & \omega_{33} & 0 \\ 0 & \omega_{24} & 0 & \omega_{44} \end{pmatrix} \begin{pmatrix} 1 & -\lambda_{12} & -\lambda_{13} & 0 \\ 0 & 1 & -\lambda_{23} & 0 \\ 0 & 0 & 1 & -\lambda_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1}$$

$$(I - \Lambda)^{-1} = \begin{pmatrix} 1 & \lambda_{12} & \lambda_{13} + \lambda_{12}\lambda_{23} & \lambda_{13}\lambda_{34} + \lambda_{12}\lambda_{23}\lambda_{34} \\ 0 & 1 & \lambda_{23} & \lambda_{23}\lambda_{34} \\ 0 & 0 & 1 & \lambda_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The Algebraic Perspective

- Let $\mathbb{R}^D = \{\Lambda \in \mathbb{R}^{m \times m} : \lambda_{ij} = 0 \text{ if } i \rightarrow j \notin D\}$
- $PD(B) := \{M \in PD_m : M_{ij} = 0 \text{ if } i \neq j \text{ and } i \leftrightarrow j \notin B\}$

Definition

The Gaussian graphical model $\mathcal{M}_G \subseteq PD_m$ consists of all covariance matrices Σ, that arise for some choice of $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$.

More algebraically, we have a polynomial map

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

$$\mathcal{M}_G = \text{im}\phi_G.$$

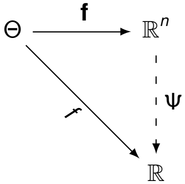
We would like to "understand" ϕ_G and $\mathcal{M}_G$.

Identification of Parameters: General Version

- Let $\Theta \subseteq \mathbb{R}^d$ d dimensional parameter space.
- Let $\mathbf{f} : \Theta \to \mathbb{R}^n$ function.
- Let $f : \Theta \to \mathbb{R}$ be another function (called a parameter).

Definition

- The parameter f is identifiable from $\mathbf{f}$ if there is a function $\Psi : \mathbf{f}(\Theta) \to \mathbb{R}$ such that $f(\theta) = \Psi \circ \mathbf{f}(\theta)$.
- The parameter f is is generically identifiable from $\mathbf{f}$ if there is a dense open subset $U \subseteq \Theta$ and a function $\Psi : \mathbf{f}(U) \to \mathbb{R}$ such that $f(\theta) = \Psi \circ \mathbf{f}(\theta)$ for all $\theta \in U$.



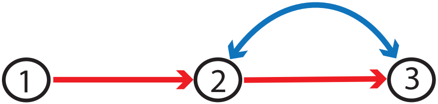
Identifiability of Models

Special interest is paid to the parameters $f(\theta) = \theta_i$. If all the parameters $f(\theta) = \theta_i$ are identifiable, then any parameter $f : \Theta \to \mathbb{R}$ is identifiable.

Definition

- The model $\mathbf{f}(\Theta)$ is (generically) identifiable if $f(\theta) = \theta_i$ is (generically) identifiable for all i.
- Equivalently, the model $\mathbf{f}(\Theta)$ is (generically) identifiable if and only if $\mathbf{f}$ is (generically) injective.

Instrumental Variable



$$\sigma_{11} = \omega_{11} \quad \sigma_{12} = \omega_{11}\lambda_{12} \quad \sigma_{13} = \omega_{11}\lambda_{12}\lambda_{23}$$

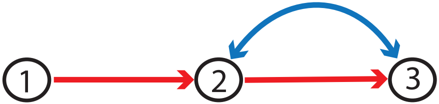
$$\sigma_{22} = \omega_{22} + \omega_{11}\lambda_{12}^2 \quad \sigma_{23} = \omega_{22}\lambda_{23} + \omega_{11}\lambda_{12}^2\lambda_{23} + \omega_{23}$$

$$\sigma_{33} = \omega_{33} + \omega_{22}\lambda_{23}^2 + \omega_{23}\lambda_{23} + \omega_{11}\lambda_{12}^2\lambda_{23}^2$$

Question

Is $f(\Lambda, \Omega) = \lambda_{12}$ identifiable? Is $f(\Lambda, \Omega) = \lambda_{23}$ identifiable?

Instrumental Variable



$$\sigma_{11} = \omega_{11} \quad \sigma_{12} = \omega_{11}\lambda_{12} \quad \sigma_{13} = \omega_{11}\lambda_{12}\lambda_{23}$$

$$\sigma_{22} = \omega_{22} + \omega_{11}\lambda_{12}^2 \quad \sigma_{23} = \omega_{22}\lambda_{23} + \omega_{11}\lambda_{12}^2\lambda_{23} + \omega_{23}$$

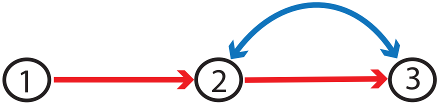
$$\sigma_{33} = \omega_{33} + \omega_{22}\lambda_{23}^2 + \omega_{23}\lambda_{23} + \omega_{11}\lambda_{12}^2\lambda_{23}^2$$

Question

Is $f(\Lambda, \Omega) = \lambda_{12}$ identifiable? Is $f(\Lambda, \Omega) = \lambda_{23}$ identifiable?

$\lambda_{12} = \frac{\sigma_{12}}{\sigma_{11}}$, and $\sigma_{11} > 0$
identifiable

Instrumental Variable



$$\sigma_{11} = \omega_{11} \quad \sigma_{12} = \omega_{11}\lambda_{12} \quad \sigma_{13} = \omega_{11}\lambda_{12}\lambda_{23}$$

$$\sigma_{22} = \omega_{22} + \omega_{11}\lambda_{12}^2 \quad \sigma_{23} = \omega_{22}\lambda_{23} + \omega_{11}\lambda_{12}^2\lambda_{23} + \omega_{23}$$

$$\sigma_{33} = \omega_{33} + \omega_{22}\lambda_{23}^2 + \omega_{23}\lambda_{23} + \omega_{11}\lambda_{12}^2\lambda_{23}^2$$

Question

Is $f(\Lambda, \Omega) = \lambda_{12}$ identifiable? Is $f(\Lambda, \Omega) = \lambda_{23}$ identifiable?

$\lambda_{12} = \frac{\sigma_{12}}{\sigma_{11}}$, and $\sigma_{11} > 0$
identifiable

$\lambda_{23} = \frac{\sigma_{13}}{\sigma_{12}}$, but $\sigma_{12} = 0$ if $\lambda_{12} = 0$
generically identifiable, not identifiable

Identification for Gaussian Graphical Models

Definition

Parameter $f(\Lambda, \Omega) = \lambda_{ij}$ is called the direct effect of X_i on X_j.

Let $\mathcal{P}(i,j)$ be the set of directed paths from i to j in G.

Definition

Parameter $f(\Lambda, \Omega) = ((I - \Lambda)^{-1})_{ij} = \sum_{P \in \mathcal{P}(i,j)} \prod_{j \to k \in P} \lambda_{jk}$ is called the total effect of X_i on X_j.

Proposition

If all the direct effects, or all the total effects, are (generically) identifiable for a graph G, then the model is (generically) identifiable.

Computational Algebra Approach to Identification

Definition

Let $S \subseteq \mathbb{R}^n$. Let $\mathcal{I}(S) \subseteq \mathbb{R}[\mathbf{p}] := \mathbb{R}[p_1, \ldots, p_n]$ be the vanishing ideal of S: $\mathcal{I}(S) := \{g \in \mathbb{R}[\mathbf{p}] : g(\mathbf{a}) = 0 \text{ for all } \mathbf{a} \in S\}$.

Let $\mathbf{f} : \Theta \to \mathbb{R}^n$ polynomial parametrization and f a parameter. Define the augmented map $\tilde{\mathbf{f}} : \Theta \to \mathbb{R}^{n+1}$ by

$$\tilde{\mathbf{f}}(\theta) = (f(\theta), \mathbf{f}(\theta)).$$

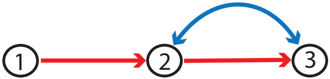
Let $\mathcal{I}(\tilde{\mathbf{f}}(\Theta)) \subseteq \mathbb{R}[q, \mathbf{p}] := \mathbb{R}[q, p_1, \ldots, p_n]$ augmented vanishing ideal.

Proposition

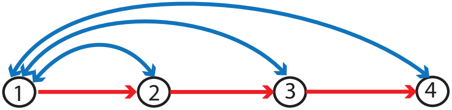
Suppose that $g(q, \mathbf{p}) \in \mathcal{I}(\tilde{\mathbf{f}}(\Theta))$ *is a polynomial such that* q *appears in this polynomial,* $g(q, \mathbf{p}) = \sum_{i=0}^{d} g_i(\mathbf{p}) q^i$ *and* $g_d(\mathbf{p})$ *does not belong to* $\mathcal{I}(\mathbf{f}(\Theta))$.

1. *If* g *is linear in* q, $g = g_1(\mathbf{p})q - g_0(\mathbf{p})$ *then* f *is generically identifiable by the formula* $f = \frac{g_0(\mathbf{p})}{g_1(\mathbf{p})}$. *If, in addition,* $g_1(\mathbf{p}) \neq 0$ *for* $\mathbf{p} \in \mathbf{f}(\Theta)$ *then* f *is identifiable.*
2. *If* g *has higher degree* d *in* q, *then* f *may or may not be generically identifiable. Generically, there are at most* d *possible choices for the parameter* $f(\theta)$ *given* $\mathbf{f}(\theta)$.
3. *If no such polynomial* g *exists then the parameter* f *is not generically identifiable.*

Example



If $f(\Lambda, \Omega) = \lambda_{23}$, then $\mathcal{I}(\tilde{\mathbf{f}}(\Theta)) = \langle \sigma_{12} q - \sigma_{13} \rangle \subset \mathbb{R}[q, \sigma_{11}, \ldots, \sigma_{33}]$ which implies that λ_{23} generically identifiable by $\lambda_{23} = \frac{\sigma_{13}}{\sigma_{12}}$.



If $f(\Lambda, \Omega) = \lambda_{23}$, then $\mathcal{I}(\tilde{\mathbf{f}}(\Theta)) =$

$$\left\langle \begin{array}{r} (\sigma_{14}\sigma_{22}\sigma_{23} - \sigma_{13}\sigma_{22}\sigma_{24}){\color{red}q^2} \\ (-\sigma_{14}\sigma_{23}^2 + \sigma_{13}\sigma_{23}\sigma_{24} - \sigma_{14}\sigma_{22}\sigma_{33} + \sigma_{12}\sigma_{24}\sigma_{33} + \sigma_{13}\sigma_{22}\sigma_{34} - \sigma_{12}\sigma_{23}\sigma_{34}){\color{red}q} \\ +(\sigma_{14}\sigma_{23}\sigma_{33} - \sigma_{13}\sigma_{24}\sigma_{33}) \end{array} \right\rangle$$

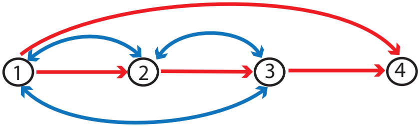
Gröbner Bases to Check for Certificate

Proposition

Let $\prec$ be an elimination term order with respect to the partition $\{q\} \cup \{p_1, \ldots, p_n\}$. Let $\mathcal{G}$ be a reduced Gröbner basis for $\mathcal{I}(\tilde{\mathbf{f}}(\Theta))$ with respect to the $\prec$.

1. *The Gröbner basis $\mathcal{G}$ contains polynomials of the lowest nonzero degree in q, of the form $g(q, \mathbf{p})$, if such a polynomial exists.*
2. *If no polynomial of $\mathcal{G}$ contains the indeterminate q, then f is not identifiable.*

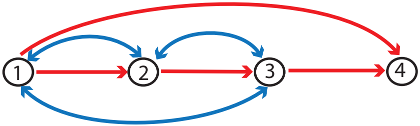
Luis Garcia and Sarah Spiegelvogel (Sam Houston State U.)



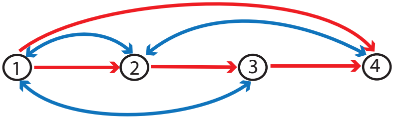
5 seconds

Computational Results

Luis Garcia and Sarah Spiegelvogel (Sam Houston State U.)



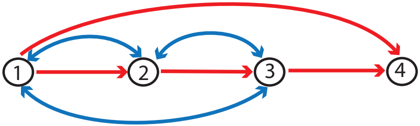
5 seconds



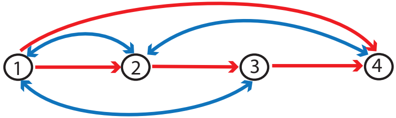
13 days

Computational Results

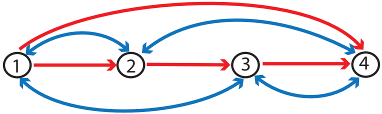
Luis Garcia and Sarah Spiegelvogel (Sam Houston State U.)



5 seconds



13 days



3.5 minutes

Theorems about Identifiability

Proposition

- ϕ_G *is not generically injective if* $\#B + \#D > \binom{n}{2}$
- ϕ_G *is not injective if there are two vertices* i, j *such that both* $i \leftrightarrow j \in B$ *and* $i \rightarrow j \in D$ *(this is called a bow).*
- ϕ_G *is injective if either B or D is empty.*

Theorem (Brito and Pearl)

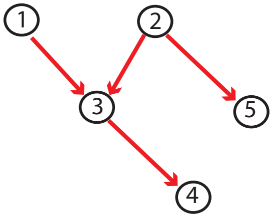
ϕ_G *is generically injective if G is bow-free.*

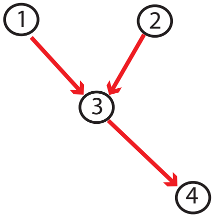
Lots of other techniques for identifying parameters: front door, back door, single door, IV pairs, accessory sets, G-criterion, ...

Converging Arborescence

Definition

- An arborescence is a directed graph whose underlying undirected graph is a tree.
- A converging arborescence is an arborescence with a unique sink.

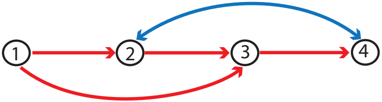




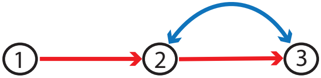
Identifiable Gaussian Graphical Models

Theorem (Drton-Foygel-S)

The Gaussian graphical model $\mathcal{M}_G$ is identifiable if and only if there is no subgraph $H = (V', B', D') \subseteq G$ such that (V', B') is connected and (V', D') is a converging arborescence.



Identifiable



Not Identifiable

Recursive Inversion Lemma

- For $i \in V$ let $P(i) = \{j \leq i : j \rightarrow i+1 \in D\}$ (parents)
- For $i \in V$ let $S(i) = \{j \leq i : j \leftrightarrow i+1 \in B\}$ (sisters)

Lemma

The parametrization ϕ_G is injective if and only if the rank condition

$$\operatorname{rank}\left(\Omega_{[i]\setminus S(i),[i]} (I - \Lambda)^{-1}_{[i],P(i)} \right) = |P(i)|$$

holds for all nodes $i = 1, \ldots, m-1$ and all pairs $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$.
IOW: ϕ_G is not injective if and only if there is an i, $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$ such that:

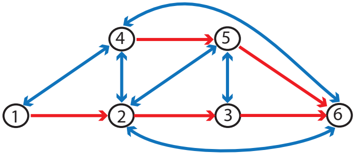
$$\ker \Omega_{[i]\setminus S(i),[i]} \cap \operatorname{im} (I - \Lambda)^{-1}_{[i],P(i)} \neq \{0\}.$$

Proof of Theorem: Necessity

Let G such that (V, D) is a converging arborescence and (V, B) is connected.

Lemma

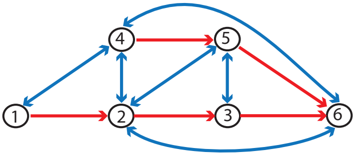
There is $\Lambda \in \mathbb{R}^D$ *such that* $\mathbf{1} = (1, \ldots, 1)^T \in \mathrm{im}(I - \Lambda)^{-1}_{[i],P(i)}$.



$$(I - \Lambda)^{-1}_{[5],P(5)} = \begin{pmatrix} \lambda_{12}\lambda_{23} & 0 \\ \lambda_{23} & 0 \\ 1 & 0 \\ 0 & \lambda_{45} \\ 0 & 1 \end{pmatrix}$$

Lemma

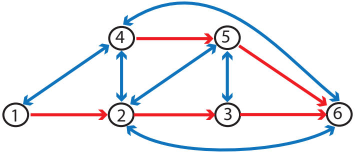
There is $\Omega \in PD(B)$ *such that* $\mathbf{1} = (1, \ldots, 1)^T \in \ker \Omega_{[i]\setminus S(i),[i]}$.



$$\Omega_{[5]\setminus S(5),[5]} = \begin{pmatrix} \omega_{11} & 0 & 0 & 0 & \omega_{14} \\ 0 & \omega_{33} & \omega_{35} & 0 & 0 \\ 0 & \omega_{35} & \omega_{55} & \omega_{25} & 0 \end{pmatrix}$$

Lemma

There is $\Omega \in PD(B)$ *such that* $\mathbf{1} = (1, \ldots, 1)^T \in \ker \Omega_{[i] \setminus S(i), [i]}$.

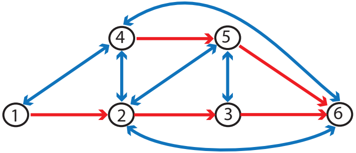


$$\Omega_{[5]\setminus S(5),[5]} = \begin{pmatrix} \omega_{11} & 0 & 0 & 0 & \omega_{14} \\ 0 & \omega_{33} & \omega_{35} & 0 & 0 \\ 0 & \omega_{35} & \omega_{55} & \omega_{25} & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 & \omega_{14} \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & \omega_{25} & 0 \end{pmatrix}$$

Lemma

There is $\Omega \in PD(B)$ *such that* $\mathbf{1} = (1, \ldots, 1)^T \in \ker \Omega_{[i] \setminus S(i), [i]}$.



$$\Omega_{[5]\setminus S(5),[5]} = \begin{pmatrix} \omega_{11} & 0 & 0 & 0 & \omega_{14} \\ 0 & \omega_{33} & \omega_{35} & 0 & 0 \\ 0 & \omega_{35} & \omega_{55} & \omega_{25} & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 & \omega_{14} \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & \omega_{25} & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \end{pmatrix}$$

Proof of Theorem: Sufficiency

Lemma

If ϕ_G is not injective and (V, D) is not a converging arborescence, or (V, B) is not connected, then there is a subgraph $H = (V', B', D') \subseteq (V, B, D)$ such that ϕ_H is not injective.

e.g.

$$\ker \Omega_{[i]\setminus S(i),[i]} \cap \operatorname{im}(I - \Lambda)^{-1}_{[i],P(i)} \neq \{0\}.$$

but (V, D) not converging arborescence implies that vectors in

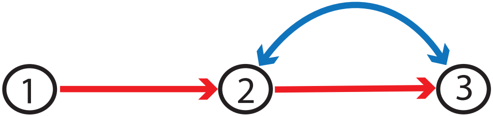
$$\ker \Omega_{[i]\setminus S(i),[i]} \cap \operatorname{im}(I - \Lambda)^{-1}_{[i],P(i)}$$

don't have full support.

Generic Identifiability: Wide Open

Problem

Characterize the mixed graphs G such that $\phi_G : \mathbb{R}^D \times PD(B)$ is generically injective.



Main difficulty: if ϕ_G is genericically injective and $H \subseteq G$, it does not imply ϕ_H is generically injective.