Learning the structure of graphical models by covariance queries

Piotr Zwiernik

Universitat Pompeu Fabra

(with Gábor Lugosi, Jakub Truszkowski, Vasiliki Velona)

TU Munich
23 October 2019

Gaussian graphical models

We consider random vectors $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_n})$ with covariance $\mathbf{\Sigma}$.

Concentration matrix $\mathbf{K} = \mathbf{\Sigma^{-1}}$.

A basic property of Gaussian vectors

$$\mathbf{K_{ij}} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{X_i} \perp\!\!\!\perp \mathbf{X_j} | \mathbf{X_{[n] \setminus \{i,j\}}}.$$

Concentration graph $\mathcal{G}(\mathbf{\Sigma}) = ([\mathbf{n}], \mathbf{E})$ has edges

$$\mathbf{E} = \{\mathbf{i} \sim \mathbf{j} | \mathbf{K_{ij}} \neq \mathbf{0}\}.$$

Global Markov property

Fix a graph $\mathbf{G}$ and $\mathbf{\Sigma}$ s.t. $\mathbf{G} = \mathcal{G}(\mathbf{\Sigma})$.

$\mathbf{S}$ separates $\mathbf{A}, \mathbf{B}$ if any path from $\mathbf{A}$ to $\mathbf{B}$ contains a vertex in $\mathbf{S}$.

Global Markov property:
If $\mathbf{S}$ separates $\mathbf{A}$ and $\mathbf{B}$ in $\mathbf{G}$ then $\mathbf{X_A} \perp\!\!\!\perp \mathbf{X_B} | \mathbf{X_S}$.

Generically $\mathbf{\Sigma}$ is faithful to G:
If $\mathbf{X_A} \perp\!\!\!\perp \mathbf{X_B} | \mathbf{X_S}$ then $\mathbf{S}$ separates $\mathbf{A}$ and $\mathbf{B}$ in $\mathbf{G}$.

Learning the structure

We address the problem of structure recovery.

Extensively studied in the high-dimensional setting.

Low sample and **time** complexity algorithms are in focus.

All algorithms start by computing the sample covariance matrix S.

Problem: When **n** is large even writing down S is expensive.

Question: Is it possible to recover the graph structure with less than quadratic time complexity?

The question only meaningful if the graph is sparse.

About the problem formulation

The trade off computational complexity vs statistical accuracy hard to study directly in this setting. We ask an identifiability question:

Can we recover $\mathbf{G} = \mathcal{G}(\boldsymbol{\Sigma})$ from $\mathbf{o}(\mathbf{n}^2)$ entries of $\boldsymbol{\Sigma}$?

(Can we consistently learn $\mathbf{G}$ with access to $\mathbf{o}(\mathbf{n}^2)$ entries of $\mathbf{S}$?)

We assume access to data though a **covariance oracle**.

The oracle takes a pair $\mathbf{i}, \mathbf{j} \in [\mathbf{n}]$ and outputs $\boldsymbol{\Sigma}_{\mathbf{ij}}$.

Note: We assume $\mathcal{G}(\boldsymbol{\Sigma})$ is connected.

$\boldsymbol{\Omega}(\mathbf{n}^2)$ queries needed to distinguish empty graph from a single edge.

Algebraic statistics connection

Separators and rank

$\mathbf{\Sigma}$ is consistent with $\mathbf{G}$ if $\mathbf{K_{ij}} = \mathbf{0}$ for all non-edges of $\mathbf{G}$.

Sullivant, Talaska, Draisma (2010): $\mathrm{rank}(\mathbf{\Sigma_{A,B}}) \leq \mathbf{r}$ for all $\mathbf{\Sigma}$ consistent with $\mathbf{G}$ if and only if $\exists \mathbf{S}$ with $|\mathbf{S}| \leq \mathbf{r}$ separating $\mathbf{A}$ from $\mathbf{B}$ in $\mathbf{G}$. Thus,

$$\mathrm{rank}(\mathbf{\Sigma_{A,B}}) \;\leq\; \min\{|\mathbf{S}| : \; \mathbf{S} \text{ separates } \mathbf{A} \text{ and } \mathbf{B}\}$$

with equality for generic $\mathbf{\Sigma}$ (in this case $\mathbf{G} = \mathcal{G}(\mathbf{\Sigma})$).

We always assume genericity getting:

1. $\mathrm{rank}(\mathbf{\Sigma_{A,B}})$ is the size of a minimal separator.
2. $\mathrm{rank}(\mathbf{\Sigma_{AC,BC}}) = \mathrm{rank}(\mathbf{\Sigma_{A,B}})$ if and only if $\mathbf{C}$ is a subset of a minimal separator.

$\text{rank}(\mathbf{\Sigma}_{\mathbf{AC},\mathbf{BC}}) = \text{rank}(\mathbf{\Sigma}_{\mathbf{A},\mathbf{B}})$ if and only if $\mathbf{C}$ is a subset of a minimal separator.

Proof: If $\text{rank}(\mathbf{\Sigma}_{\mathbf{AC},\mathbf{BC}}) = \mathbf{r}$ then any minimal separator $\mathbf{S}$ of $\mathbf{A} \cup \mathbf{C}$ and $\mathbf{B} \cup \mathbf{C}$ has $\mathbf{r}$ elements.

Naturally $\mathbf{S}$ is also a separator of $\mathbf{A}, \mathbf{B}$.

It is a minimal separator of $\mathbf{A}, \mathbf{B}$ because $\text{rank}(\mathbf{\Sigma}_{\mathbf{A},\mathbf{B}}) = \mathbf{r}$.

Necessarily $\mathbf{C} \subset \mathbf{S}$.

Trees

Graphical models on trees

When $\mathcal{G}(\mathbf{\Sigma})$ is a tree $\mathbf{T}$ then for every $\mathbf{i}, \mathbf{j} \in [\mathbf{n}]$

$$\rho_{\mathbf{ij}} = \prod_{\mathbf{uv} \in \overline{\mathbf{ij}}} \rho_{\mathbf{uv}},$$

where $\overline{\mathbf{ij}}$ is the unique path between $\mathbf{i}$ and $\mathbf{j}$ in $\mathbf{T}$.

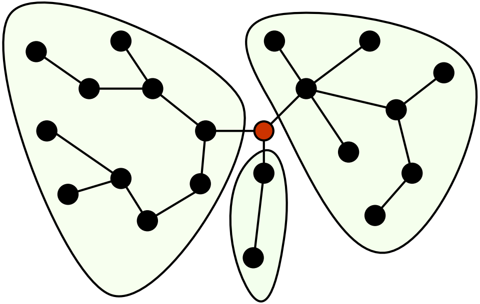
This implies that $\mathbf{T}$ is a maximum-weight spanning tree (with weights $|\rho_{ij}|$) that can be found simply (Chow-Liu, 1968)

To run the Chow-Liu algorithm all correlations must be known!

Recovering trees: divide and conquer

Randomized algorithm

1. Find a central vertex $\mathbf{v}$
2. Identify the connected components of $\mathbf{T} \setminus \{\mathbf{v}\}$.
3. Recurse.



Centrality

Centroid of $\mathbf{T}$ is the vertex that minimizes

$$\mathbf{c}(\mathbf{v}) = \frac{1}{|\mathbf{V}| - 1} \max_{\mathbf{C} \in \mathcal{C}^{\mathbf{v}}} |\mathbf{C}|$$

where $\mathcal{C}^{\mathbf{v}}$ is the set of connected components of $\mathbf{T} \setminus \{\mathbf{v}\}$.
(unclear how to find it using covariance queries)

A modified measure of centrality

$$\mathbf{c}(\mathbf{v}) \ = \ \frac{1}{(|\mathbf{V}| - 1)^2} \sum_{\mathbf{C} \in \mathcal{C}^{\mathbf{v}}} |\mathbf{C}|^2.$$

A modified measure of centrality

$$\mathbf{c}(\mathbf{v}) \; = \; \frac{1}{(|\mathbf{V}|-1)^2} \sum_{\mathbf{C}\in\mathcal{C}^{\mathbf{v}}} |\mathbf{C}|^2.$$

This can be approximated. For each $\mathbf{v} \in \mathbf{V}$:

1. Sample $\mathbf{k}$ pairs of vertices $(\mathbf{u_1}, \mathbf{v_1}), \ldots, (\mathbf{u_k}, \mathbf{v_k})$ with replacement.
2. Get $\mathbf{\Sigma_{A,B}}$ for $\mathbf{A} = \{\mathbf{u_i}, \mathbf{v}\}$, $\mathbf{B} = \{\mathbf{v_i}, \mathbf{v}\}$ and each $\mathbf{i}$.
3. Compute and minimize

$$\hat{\mathbf{c}}(\mathbf{v}) \; := \; \frac{1}{\mathbf{k}} \sum_{\mathbf{i}\leq\mathbf{k}} \mathbb{1}_{\{\mathbf{u_i},\mathbf{v_i} \text{ are not separated by } \mathbf{v}\}}$$

(separated iff $\det \Sigma_{A,B} = 0$)

Note: $\mathbf{k}\hat{\mathbf{c}}(\mathbf{v}) \; \sim \; \mathrm{Bin}(\mathbf{k}, \mathbf{c}(\mathbf{v}))$.

Note: $\mathbf{k}\hat{\mathbf{c}}(\mathbf{v}) \sim \mathrm{Bin}(\mathbf{k}, \mathbf{c}(\mathbf{v}))$.

Hoeffding inequality gives: $\mathbb{P}(|\hat{\mathbf{c}}(\mathbf{v}) - \mathbf{c}(\mathbf{v})| \geq \delta) \leq 2\exp(-2\delta^2\mathbf{k})$

The optimal vertex $\hat{\mathbf{v}}$ satisfies

$$\mathbb{P}\left(\mathbf{c}(\hat{\mathbf{v}}) > \sqrt{\frac{11}{12}}\right) \leq 2|\mathbf{V}|\exp(-\mathbf{k}/32).$$

This step only requires $\mathcal{O}(\mathbf{k}|\mathbf{V}|)$ queries.

Identifying components in $\mathcal{C}^{\hat{\mathbf{v}}}$ takes $\mathcal{O}(\mathbf{d}|\mathbf{V}|)$ queries.

With probability $\geq 1 - \epsilon$, the algorithm recovers the tree using

$$\mathcal{O}\left(\mathbf{n}\log(\mathbf{n})\max\left\{\log\left(\frac{\mathbf{n}}{\epsilon}, \mathbf{d}\right)\right\}\right)$$

queries.

General graphs

Graphs of small treewidth

A richer class of graphs is graphs with bounded treewidth.

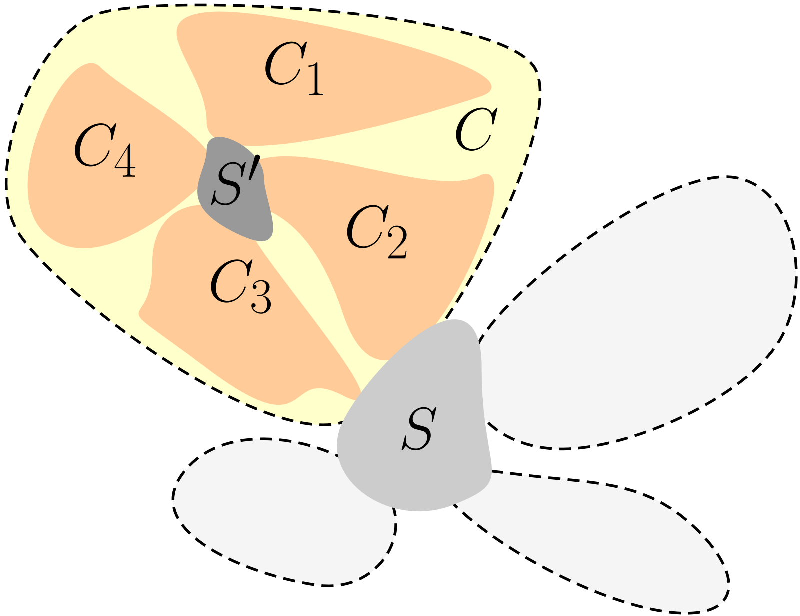
Such graphs have small balanced separators:
If treewidth $\leq \mathbf{k}$ then $\mathbf{G}$ has a separator $\mathbf{S} \subset \mathbf{V}$ with $|\mathbf{S}| \leq \mathbf{k}+\mathbf{1}$ and

$$\mathbf{c}(\mathbf{S}) \;\leq\; \frac{1}{2}\frac{|\mathbf{V}| - \mathbf{k}}{|\mathbf{V}| - (\mathbf{k}+1)}$$

where

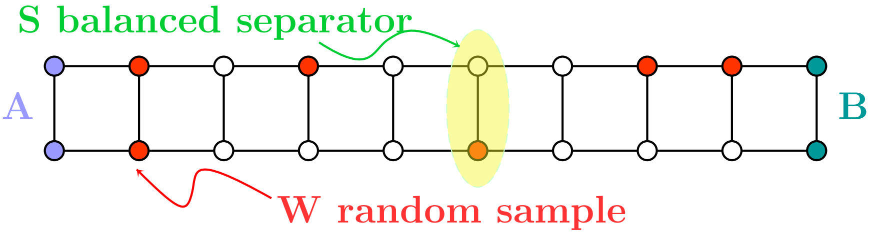
$$\mathbf{c}(\mathbf{S}) \;=\; \frac{1}{|\mathbf{V} \setminus \mathbf{S}|} \max_{\mathbf{C} \in \mathcal{C}^{\mathbf{S}}} |\mathbf{C}|.$$

Find a small balanced separator $\mathbf{S}$ and identify components of $\mathbf{V} \setminus \mathbf{S}$.


C_1
C
C_4
S'
C_2
C_3
S

Identifying small separators

1. Take a random sample **W** of **m** vertices of **G**. ($\mathbf{m} = \mathcal{O}(\mathbf{k})$)
2. Find $\mathbf{S} \subset \mathbf{V}$ with $|S| \leq k + 1$ that is a balanced separator of **W**.
3. Argue that **S** is a balanced separator of **G** as well.



A balanced separator of W

Recall: $\mathrm{rank}(\mathbf{\Sigma}_{\mathbf{AC},\mathbf{BC}}) = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{A},\mathbf{B}})$ if and only if **C** is a subset of a minimal separator.

Exhaustively search through $\mathbf{2}^{\mathcal{O}((\mathbf{k}))}$ approximately balanced partitions **A**, **B** of W, $\max(|\mathbf{A}|, |\mathbf{B}|) \leq (\mathbf{2/3})|\mathbf{W}|$.

Pick any leading to the minimal $\mathbf{r} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{A},\mathbf{B}}) \leq \mathbf{k} + \mathbf{1}$ and then:

1. Determine all $\mathbf{v} \in \mathbf{G}$ with $\mathrm{rank}(\mathbf{\Sigma}_{\mathbf{A},\mathbf{B}}) = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{Av},\mathbf{Bv}})$ (**v** lies in some minimal separator of **A** and **B**)
2. Starting from $\mathbf{S} = \emptyset$ gradually add vertices from 1. that satisfy $\mathrm{rank}(\mathbf{\Sigma}_{\mathbf{ASv},\mathbf{BSv}}) = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{A},\mathbf{B}})$

Next we argue that, with high probability, **S** is a balanced separator of **G** as well.

VC dimenion

The argument hinges on Vapnik-Chervonenkis theory.

The small sample $\mathbf{W}$ is a good representative of the entire graph.

Lemma (Feige and Mahdian, 2006) For any graph $\mathbf{G}$, let $\mathcal{F}_{\mathbf{S}}$ the set of all connected components of $\mathbf{G} \setminus \mathbf{S}$ and their complements in $\mathbf{V} \setminus \mathbf{S}$ and define

$$\mathcal{F}_{\mathbf{k}} := \bigcup_{\mathbf{S}:|\mathbf{S}| \leq \mathbf{k}} \mathcal{F}_{\mathbf{S}}.$$

Then the VC-dimension of $\mathcal{F}_{\mathbf{k}}$ is at most $\mathbf{11k}$.

The VC-inequality implies that for all sets $\mathbf{C} \in \mathcal{F}_{\mathbf{k}}$,

$$\frac{|\mathbf{C}|}{|\mathbf{V}|} - \delta \leq \frac{|\mathbf{W} \cap \mathbf{C}|}{|\mathbf{W}|} \leq \frac{|\mathbf{C}|}{|\mathbf{V}|} + \delta$$

whenever $|\mathbf{W}| = \tilde{\mathbf{\Omega}}(\mathbf{k}/\delta^{\mathbf{2}})$.

Connected components

Once a small balanced separator **S** is found, we determine $\mathcal{C}^{\mathbf{S}}$.

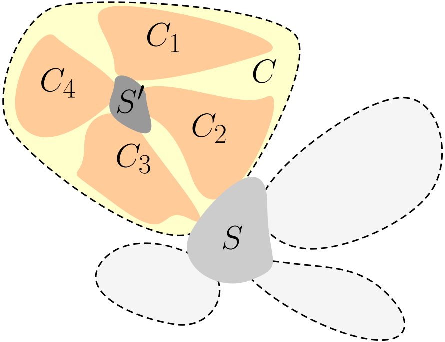
This can be done in linear time by checking ranks.

As for trees, we require small degree (although this can be relaxed).

Recursing in the connected components

Once the conn. components of $\mathbf{G} \setminus \mathbf{S}$ are known, we may recurse.

GMs are not closed under taking margins [Frydenberg, 1990].



Conditional covariances

A basic property of Gaussian graphical models (Lauritzen, 1996):

If $\mathbf{S}$ separates $\mathbf{C}$ from the rest of $\mathbf{G}$, then $\mathbf{K_C} = (\boldsymbol{\Sigma}_{\mathbf{C|S}})^{-1}$.

Note: As we recurse the conditioning set increases. However, recursion depth is $\mathcal{O}(\log(\mathbf{n}))$ and separators are small.

Main result

Let $\mathbf{\Sigma}$ be generic and let $\mathbf{G} = \mathcal{G}(\mathbf{\Sigma})$ be such that $\mathrm{tw}(\mathbf{G}) \leq \mathbf{k}$ and $\mathbf{\Delta}(\mathbf{G}) \leq \mathbf{d}$. Then, with probability at least $\mathbf{1} - \boldsymbol{\epsilon}$, the recursive algorithm reconstructs $\mathbf{G}$ with

$$\mathcal{O}\left((\mathbf{2}^{\mathcal{O}(\mathbf{k}\log\mathbf{k})} + \mathbf{dk}\log\mathbf{n}) \cdot \mathbf{k}^{\mathbf{2}}\mathbf{n} \cdot \log^{\mathbf{3}}\mathbf{n} \cdot \log\frac{\mathbf{1}}{\boldsymbol{\epsilon}}\right)$$

queries (or $\mathcal{O}_{\mathbf{k},\mathbf{d},\epsilon}(\mathbf{n}\log^{\mathbf{4}}(\mathbf{n}))$).

The computational complexity is $\mathcal{O}_{\mathbf{k},\mathbf{d},\epsilon}\left(\mathbf{n}\log^{\mathbf{5}}(\mathbf{n})\right)$.

We actually recover $\mathbf{K} = \mathbf{\Sigma}^{-\mathbf{1}}$, not just the concentration graph.

Further questions

- Imperfect oracle.
- Robustness.
- Non-gaussian models.
- Property testing.

Thank you!