Graphical Gaussian models and their groups

Piotr Zwiernik
TU Eindhoven
(*j.w. Jan Draisma, Sonja Kuhnt*)

Workshop on Graphical Models,
Fields Institute, Toronto, 16 Apr 2012

Outline and references

Outline:

1. Invariance of statistical models under group actions.
2. The maximal group that leaves the graphical Gaussian model invariant.
3. Finite sample breakdown points for the maximum likelihood estimator in the class of graphical Gaussian models.

This links to:

- Composite transformation (exponential) families; equivariant decision theory.
- Work of Gérard Letac and Hélène Massam on Wishart distributions; and related work of Steen Andersson (and collaborators) on Lie groups for Gaussian models.
- Hypergraphs, clique hypergraphs, and other related combinatorial concepts.
- “Groups and breakdowns” P. Laurie Davies, Ursula Gather.

Group invariance of models

- **Definition (Schervish):** Let $\mathcal{P}_0$ be a parametric family with parameter space Θ and sample space $(\mathcal{X}, \mathcal{B})$. Let G be a group of transformations on $\mathcal{X}$. We say that G *leaves* $\mathcal{P}_0$ *invariant* if for each $g \in G$ and each $\theta \in \Theta$ there exists $\theta^* \in \Theta$ such that $P_\theta(A) = P_{\theta^*}(gA)$ for every $A \in \mathcal{B}$.
- θ^* is unique and denoted by $\theta^* = g \cdot \theta$

- This induces the action of G on the model:

$$g \cdot P_\theta(A) \quad := \quad P_\theta(g^{-1}A) \quad = \quad P_{g \cdot \theta}(A)$$

- If there is a density $f(x;\theta)$ then

$$f(x;\theta) \quad = \quad f(g \cdot x;\, g \cdot \theta).$$

Gaussian Transformation models

- $\mathcal{X} = \mathbb{R}^m$, $G \subseteq \mathrm{GL}_m(\mathbb{R})$ acts on $\mathbb{R}^m$: $g \cdot x \;:=\; gx$
- density: $f(x;K) \;=\; (2\pi)^{-m/2}(\det K)^{1/2}\exp\{-\frac{1}{2}\,x^T K\,x\}$
- $\Theta \;=\; \mathcal{S}_m^+$ symmetric positive definite matrices
- $f(x;K) \;=\; f(gx; g^{-T}Kg^{-1})$ so G acts on $\mathcal{S}_m^+$: $g \cdot K \;:=\; g^{-T}Kg^{-1}$

- **Definition:** Let $\mathcal{K} \subset \mathcal{S}_m^+$. The *composite Gaussian transformation family* induced by G and $\mathcal{K}$:

$$M(G,\mathcal{K}) \quad = \quad \{f(x;\, g^T K\, g) : \; g \in G,\, K \in \mathcal{K}\}.$$

  - If $\mathcal{K} \;=\; \{K_0\}$ then *Gaussian transformation family* .

Gaussian graphical models

- undirected graph $\mathcal{G} = ([m], E)$, $[m] := \{1, \ldots, m\}$
- $f(x; K)$, K such that $K_{ij} = 0$ if $(i,j) \notin E$

- Parameter space: $\mathcal{S}_{\mathcal{G}}^{+} = \{K \in \mathcal{S}_m^{+} : K_{ij} = 0 \text{ if } (i,j) \notin E\}$.
- $\mathcal{S}_{\mathcal{G}}^{+} \subseteq \mathcal{S}_m^{+}$ so $G \subseteq \mathrm{GL}_m(\mathbb{R})$ acts on $\mathcal{S}_{\mathcal{G}}^{+}$: $g \cdot K := g^{-T} K g^{-1}$.

Definition of the group G

- **Definition:** $G \subseteq GL_m(\mathbb{R})$ such that $G \cdot \mathcal{S}_{\mathcal{G}}^{+} = \mathcal{S}_{\mathcal{G}}^{+}$
  - $K \in \mathcal{S}_{\mathcal{G}}^{+} \implies g^{-T} K g^{-1} \in \mathcal{S}_{\mathcal{G}}^{+}$ for all $g \in G$

Trivial example $m = 2$

- if $\mathcal{G} = 1 \bullet\!\!-\!\!\bullet 2$ then $G = GL_2(\mathbb{R})$
- if $\mathcal{G} = 1 \bullet \quad \bullet 2$ then $G = \left\{ \begin{bmatrix} * & 0 \\ 0 & * \end{bmatrix}, \begin{bmatrix} 0 & * \\ * & 0 \end{bmatrix} \right\}$

- full graph $\implies G = GL_m(\mathbb{R})$
- no edges $\implies G = T_m \cdot S_m$: $T_m =$ diagonal matrices, S_m=permutation matrices

Example: $\overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$

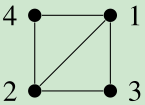
- $\mathcal{G}$: $\overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$
- $\mathcal{S}_{\mathcal{G}} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{bmatrix}, \quad G = \left\{ G^0 = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & 0 & * \end{bmatrix}, \begin{bmatrix} * & 0 & 0 \\ * & 0 & * \\ * & * & 0 \end{bmatrix} \right\}$
- if $K = [K_{ij}] \in \mathcal{S}_{\mathcal{G}}$ and $g^{-1} = [g_{ij}] \in G^0$ then $g^{-T} K g^{-1}$ equals

$$\begin{bmatrix} \text{too long} & g_{12}\, g_{22}\, K_{22} + g_{11}\, g_{22}\, K_{12} & g_{13}\, g_{33}\, K_{33} + g_{11}\, g_{33}\, K_{13} \\ g_{22}\, (g_{12}\, K_{22} + g_{11}\, K_{12}) & {g_{22}}^2\, K_{22} & 0 \\ g_{33}\, (g_{13}\, K_{33} + g_{11}\, K_{13}) & 0 & {g_{33}}^2\, K_{33} \end{bmatrix}$$

Preorder on the set of nodes

- the set of maximal cliques of $\mathcal{G}$: $\mathcal{C} \subseteq 2^{[m]}$
- define a preorder on $[m]$ by:

$$i \preccurlyeq j \quad \text{iff} \quad \forall_{C\in\mathcal{C}} \;\; j \in C \text{ implies } i \in C.$$



- $\mathcal{C} = \{\{1,2,3\},\{1,2,4\}\}$
- $1 \approx 2, \quad 1,2 \preccurlyeq 3, \quad 1,2 \preccurlyeq 4$

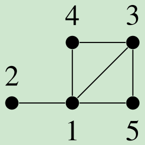
Equivalently: $i \preccurlyeq j$ iff $N(j) \cup j \subseteq N(i) \cup i.$

The poset $\mathbf{P}_{\mathcal{C}}$

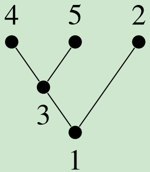
- equivalence relation on $[m]$: $i \sim j$ iff $i \preccurlyeq j$ and $j \preccurlyeq i$.
- the preorder $\preccurlyeq$ gives a partial order on $[m]/\sim$
- the resulting poset: $\mathbf{P}_{\mathcal{C}} = ([m]/\sim, \preccurlyeq)$

Recall: $i \preccurlyeq j$ iff $\forall_{C \in \mathcal{C}}$ $j \in C$ implies $i \in C$.

- $\mathcal{C} = \{\{1,2\},\{1,3,4\},\{1,3,5\}\}$



- the Hasse diagram



The connected component of the identity

- Recall: $\overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}, \quad 1 \preccurlyeq 2, 3$

$$\mathcal{S}_\mathcal{G} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{bmatrix}, \quad G = \left\{ G^0 = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & 0 & * \end{bmatrix}, \begin{bmatrix} * & 0 & 0 \\ * & 0 & * \\ * & * & 0 \end{bmatrix} \right\}$$

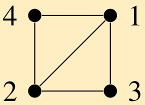
- **Proposition:** Let $G^0 \subseteq G$ be the closed connected component of the identity matrix I. Then $g \in G^0$ if and only if for every $i, j \in [m]$

$$g_{ij} \neq 0 \qquad \Longrightarrow \qquad j \preccurlyeq i.$$

The full group G

- Recall: $\overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$

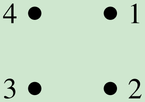
$$G = \left\{ \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & 0 & * \end{bmatrix}, \begin{bmatrix} * & 0 & 0 \\ * & 0 & * \\ * & * & 0 \end{bmatrix} \right\} \ni \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$



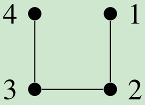
- Define $\widetilde{\mathcal{G}}$ as the induced graph on $[m]/\sim$.
- The coloring map $c : [m]/\sim \longrightarrow \mathbb{N}, \quad x \mapsto |x|$.
- For our example $\widetilde{\mathcal{G}} = \overset{3}{\bullet} - \overset{1,2}{\bullet} - \overset{4}{\bullet}$

- **Theorem:** For every $\mathcal{G}$ $\quad G = G^0 \rtimes \mathrm{Aut}(\widetilde{\mathcal{G}}, c).$

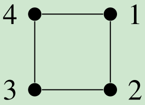
Some examples



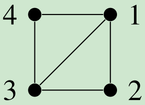
- $G^0 = T_4$
- $G = T_4 \rtimes S_4$



- $G^0 : g_{12}, g_{43}$ can be non-zero
- $G \simeq G^0 \rtimes Z_2$



- $G^0 = T_4$
- $G \simeq T_4 \rtimes D_8$



- $G^0 : g_{13}, g_{31}, g_{21}, g_{23}, g_{41}, g_{43}$ can be nonzero
- $G \simeq G^0 \rtimes Z_2$

Brief summary

We have:

- defined group invariance of models
  - (composite) transformation families
- identified the maximal group that leaves the graphical Gaussian model invariant

We will

- define finite sample breakdown points (fsbp)
- proceed with our algebraic and combinatorial analysis in order to analyse fsbp

High breakdown point

- a random sample: $\mathbf{x}_n = (x_1, \ldots, x_n)$.
- robustness of the sample mean: $x_1 \to \infty \implies \bar{x} \to \infty$.
  - The finite sample breakdown point is zero.
- robustness of the sample median: $x_1 \to \infty \;\not\Longrightarrow\; \mathrm{med}(x) \to \infty$.
  - The finite sample breakdown point is $\approx 1/2$.

- **Problem:** Compute the finite sample breakdown point for the maximum likelihood estimator for the graphical Gaussian model.
- Find bounds on the finite sample breakdown points

The maximum likelihood estimator

- sample covariance: $S_{\mathbf{x}_n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^T(x_i - \bar{x})$
- log-likelihood: $\ell(K;\mathbf{x}_n) = \frac{n}{2}\left(\log\det K - \langle S_{\mathbf{x}_n}, K\rangle\right)$,
  where $\langle S, K\rangle = \operatorname{trace}(SK)$
- $\ell(K; g\cdot\mathbf{x}_n) = \ell(g^{-1}\cdot K;\mathbf{x}_n)$ modulo $\log\det(g^T g)$, where
  $g\cdot\mathbf{x}_n = (gx_1,\ldots,gx_n)$
- $\widehat{K}_{\mathbf{x}_n}$ is G-equivariant: $\boxed{\widehat{K}_{g\cdot\mathbf{x}_n} = g\cdot\widehat{K}_{\mathbf{x}_n} = g^{-T}\,\widehat{K}_{\mathbf{x}_n}\,g^{-1}}$.

Finite sample breakdown points

- (unbounded) pseudometric $D : \mathcal{S}_{\mathcal{G}}^{+} \times \mathcal{S}_{\mathcal{G}}^{+} \to \mathbb{R}$
  - $D(K, L) \quad = \quad |\log\det(KL^{-1})|$
- $G_1 = \{g : \det(gg^T) \neq 1\}$
- $\mathbf{y}_{n,k}$ a random sample with k data points of $\mathbf{x}_n$ altered

- $\mathrm{fsbp}(\widehat{K}, \mathbf{x}_n, D) \quad := \quad \frac{1}{n} \min\{k : \sup_{\mathbf{y}_{n,k}} D(\widehat{K}_{\mathbf{x}_n}, \widehat{K}_{\mathbf{y}_{n,k}}) \; = \; \infty\}$

- **Theorem (Davies,Gather):** $\mathrm{fsbp}(\widehat{K}, \mathbf{x}_n, D) \quad \leq \quad \frac{1}{n}\left\lfloor \frac{n - n\Delta_{\mathbf{x}_n} + 1}{2} \right\rfloor$, where

$$\Delta_{\mathbf{x}_n} := \frac{1}{n} \max_k \{k : \exists A \subseteq [n], |A| = k \text{ and } g \in G_1 \text{ s.t. } gx_i = x_i \, \forall i \in A\}.$$

The link between $\Delta_{\mathbf{x}_n}$ and $\mathrm{fsbp}(\widehat{K}, \mathbf{x}_n, D)$

- Let $\Delta_{\mathbf{x}_n} = \frac{1}{n}(n-k)$. For $|A| = n-k$ let $g \in G_1$ be s.t. $gx_i = x_i$ for $i \in A$ and $gx_i \neq x_i$ if $i \in [n] \setminus A$.
- Then $g^l x_i = x_i$ for $l \geq 1$ and $i \in A$.
- But $D(\widehat{K}_{\mathbf{x}_n}, \widehat{K}_{g^l \cdot \mathbf{x}_n}) = D(\widehat{K}_{\mathbf{x}_n}, g^l \cdot \widehat{K}_{\mathbf{x}_n}) = |\log\det(\widehat{K}_{\mathbf{x}_n} g^l \widehat{K}_{\mathbf{x}_n}^{-1} (g^T)^l)| = l\,|\log\det(gg^T)| \longrightarrow \infty$ since $g \in G_1$.
- So $\mathrm{fsbp}(\widehat{K}, \mathbf{x}_n, D) \leq \frac{1}{n}k = 1 - \Delta_{\mathbf{x}_n}$.
- For details on the sharper bound see: Davies, Gather, “The breakdown point – Examples and Counterexamples”, 2007.

G^0-orbits in the sample space

- orbit of $x \in \mathbb{R}^m$: $\mathcal{O}_x := \{gx : g \in G\} \subseteq \mathbb{R}^m$
- G contains $T_m \implies x$ and $\mathrm{ind}(x)$ are in the same orbit
- $\mathcal{O}_I$: the subset of points in $\mathbb{R}^m$ with support given by $I \subseteq [m]$

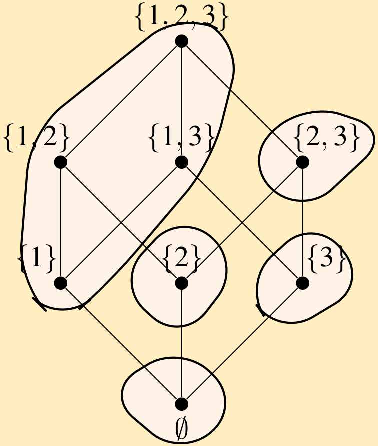
- **Lemma:** If $g \in G$ stabilizes every $x \in B \subseteq \mathbb{R}^m$ then there exists $g' \in G^0$ stabilizing all $x \in B$.

- **Theorem:** G^0-orbits are in $1-1$ correspondence with the up-sets of $\mathbf{P}_{\mathcal{C}}$:

$$\overline{\mathcal{O}}_I := \bigcup_{\uparrow J = I} \mathcal{O}_I \qquad \text{for all } I \in \mathbf{O}^{\uparrow}(\mathbf{P}_{\mathcal{C}}).$$

$\overline{\mathcal{O}}_{[m]}$ is the unique dense orbit.

Example: $\overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$



- For example

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

shows congruence of $\{1\}$ and $\{1, 2, 3\}$.

The generic case

- **Problem:** compute $\Delta_{\text{gen}} := \Delta_{\mathbf{x}_n}$ assuming that $\mathbf{x}_n$ is *generic*:
  - $x_1, \ldots, x_n$ lie in the dense G^0-orbit
  - we remove some other measure zero subsets

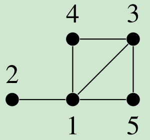
- **Example 1:** $\overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$
  - $x_i = g_i[1,0,0]^T$; w.l.o.g. assume $x_1 = [1,0,0]^T$
  - $G_{x_1} = \{g : g = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}\}$
  - if $ab \neq 1$ then $g \in G_1$ and no other point is stabilized; hence $\Delta_{\text{gen}} = 1/n$
- **Example 2:** the four-cycle
  - $G^0 = T_4$ and generic points are stabilized only by I; hence $\Delta_{\text{gen}} = 0$

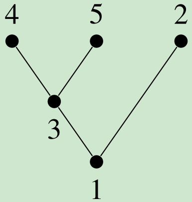
Bounds in the generic case

- **Theorem:** If in the Hasse diagram of $\mathbf{P}_{\mathcal{C}}$ every element covers at most one element (HD is a forest) then

$$\Delta_{\text{gen}} \quad = \quad \frac{1}{n}(\text{rank}(\mathbf{P}_{\mathcal{C}}) - 1).$$

- Consider the graph





- Here $\Delta_{\text{gen}} \; = \; 2/n$

Further problems

Non-generic data sets

- Is there a closed form formula?
- Is there an algorithmic procedure for arbitrary data sets?

G-orbits in $\mathcal{S}_{\mathcal{G}}^{+}$

- **Theorem (Letac, Massam)**: There is exactly one G-orbit in $\mathcal{S}_{\mathcal{G}}^{+}$ if and only if the Hasse diagram of the induced poset $\mathbf{P}_{\mathcal{C}}$ for each connected component of $\mathcal{G}$ is a tree (every element covers at most one element). This happens exactly when $\mathcal{G}$ contains no 4-cycle or 4-chain as an induced subgraph.
- Provide the description of G-orbits in the general case

Thank you!