

Biological networks: Gaussian graphical models and the PC algorithm

M2 GENIOMHE – Session 2

Christophe Ambroise

M2 GENIOMHE – Université d'Évry

Session 1

- DAG factorisation: $p(\mathbf{x}) = \prod_t p(x_t \mid \mathbf{x}_{pa(t)})$
- Chain, fork: conditioning blocks; collider: conditioning opens
- d-separation $\Rightarrow$ conditional independence
- Markov blanket: parents, children, co-parents
- Markov equivalence: same skeleton + same v-structures $\rightarrow$ CPDAG

Prior knowledge

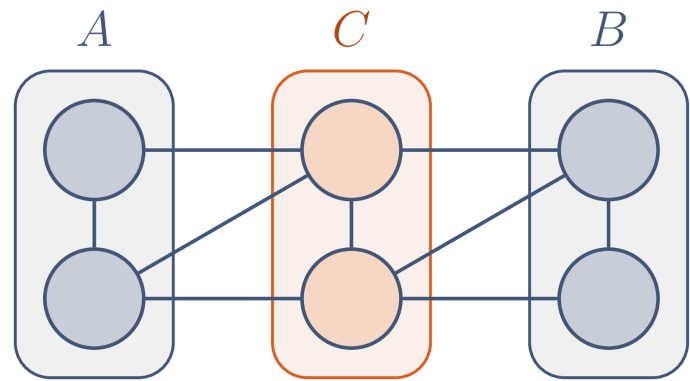
Markov random fields

Undirected graphs: motivation

- DAG: arrows imposed, often arbitrary (pixels, spins, co-expressed genes)
- Symmetric interactions → undirected edges
- Independence read by plain **graph separation**

Graph separation

C separates A from B when every path from A to B goes through C .



Markov properties of an undirected graph

Property	Statement
global	C separates A and $B \Rightarrow X_A \perp X_B \mid X_C$
local	$X_v \perp X_{V \setminus (\{v\} \cup N(v))} \mid X_{N(v)}$
pairwise	$u \approx v \Rightarrow X_u \perp X_v \mid X_{V \setminus \{u,v\}}$

- global $\Rightarrow$ local $\Rightarrow$ pairwise
- Equivalence for positive densities
- Markov blanket of v : its neighbours $N(v)$

Gibbs distribution

A Gibbs distribution for G is a product of non-negative potentials over the cliques of G , normalised by the partition function.

$$p(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}(G)} \psi_c(x_c), \quad Z = \sum_x \prod_{c \in \mathcal{C}(G)} \psi_c(x_c)$$

- Potential: compatibility, not a probability
- Z : sum over K^V configurations → main computational obstacle

Hammersley–Clifford theorem

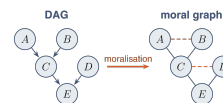
A strictly positive distribution satisfies the global Markov property of G if and only if it is a Gibbs distribution for G .

$$p > 0 : \quad p \text{ globally Markov w.r.t. } G \iff p(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}(G)} \psi_c(x_c)$$

Gaussian graphical models

Moralisation

Moral graph of a DAG: parents of each node married, then orientations dropped.



Proposition

A distribution factorising according to a DAG G is a Gibbs distribution for its moral graph G^m .

Price: lost independences (v-structures).

Multivariate Gaussian and precision matrix

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu)^\top K (x - \mu) \right), \quad K = \Sigma^{-1}$$

- K : precision (concentration) matrix
- Exponent: only pairwise terms $x_j K_{jk} x_k \rightarrow$ Gibbs distribution with cliques of size ≤ 2

Conditional Gaussian

For a Gaussian vector split into blocks A and B , the law of X_A given $X_B = b$ is Gaussian.

$$X_A | X_B = b \sim \mathcal{N}(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(b - \mu_B), \Sigma_{A|B})$$

$$\Sigma_{A|B} = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} = (K_{AA})^{-1}$$

Last equality: Schur complement.

Zeros of the precision matrix

In a Gaussian vector, two variables are independent given all the others if and only if the corresponding entry of K is zero.

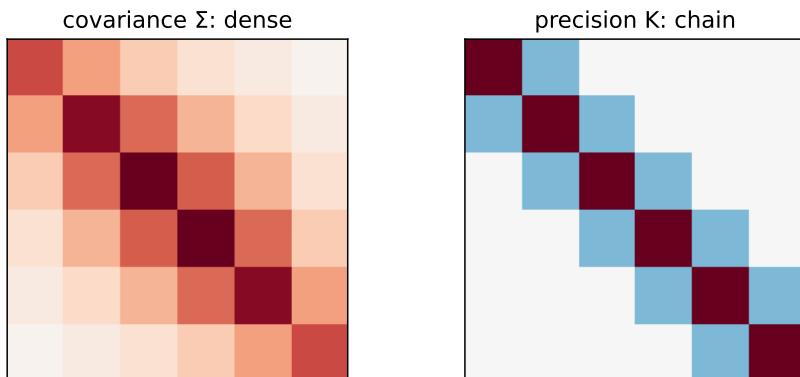
$$K_{ij} = 0 \iff X_i \perp X_j | X_{V \setminus \{i,j\}}$$

Partial correlation: $\rho_{ij \cdot V \setminus \{i,j\}} = -K_{ij} / \sqrt{K_{ii}K_{jj}}$.

Gaussian graphical model

Graph G with edge $i - j$ iff $K_{ij} \neq 0$: support of the precision matrix.

Covariance vs precision



Marginal correlation everywhere, conditional dependence only along the chain.

Covariance selection: known graph

Sample $x_1, \dots, x_N$, empirical covariance S . Log-likelihood up to constants:

$$\ell(K) = \frac{N}{2} (\log \det K - \text{tr}(SK)), \quad \nabla_K \ell \propto K^{-1} - S$$

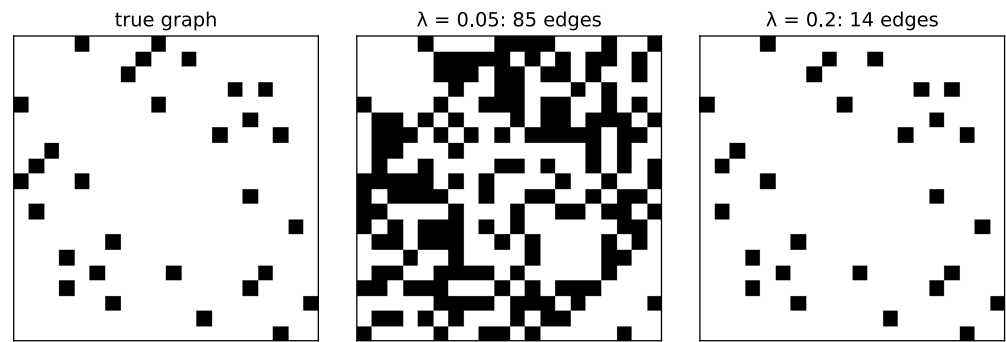
Graphical lasso: unknown graph

The graphical lasso penalises the ℓ_1 norm of the precision matrix to obtain a sparse estimate.

$$\hat{K}_\lambda = \arg \min_{K \succ 0} -\log \det K + \text{tr}(SK) + \lambda \sum_{s \neq t} |K_{st}|$$

- λ large $\rightarrow$ sparse graph
- Defined even for $p > N$
- Friedman, Hastie & Tibshirani (2008): block coordinate descent

Graphical lasso on simulated data



Choice of λ : cross-validation, StARS, BIC.

Graph learned: conditional dependences, **not** causal directions.

Structure learning

Structure learning: goals and obstacle

- Knowledge discovery: graph topology (gene networks, signalling pathways)
- Density estimation: fully specified model
- Obstacle: super-exponential number of DAGs

p	2	3	4	5	10
DAGs	3	25	543	29 281	4.2×10^{18}

Robinson (1973). Structure learning of DAGs: NP-hard (Chickering, 1996).

Three families of methods

Family	Principle	Output
constraint-based	conditional independence tests	CPDAG (PC, FCI)
score-based	score optimisation over graphs	DAG or CPDAG (hill climbing, GES)
functional	restricted SCM class	DAG (LiNGAM, ANM)

Assumptions

- **Causal Markov**: d-separation $\Rightarrow$ independence
- **Faithfulness**: independence $\Rightarrow$ d-separation
- **Causal sufficiency**: no unobserved common cause
- Reliable conditional independence tests

Key fact

i and j non-adjacent in G iff some set S makes them independent:
 $X_i \perp X_j \mid X_S$.

Constraint-based learning: the PC algorithm

Conditional independence tests

- Gaussian: partial correlation, Fisher z transform

$$z = \frac{1}{2} \log \frac{1 + \hat{\rho}_{ij \cdot S}}{1 - \hat{\rho}_{ij \cdot S}} \sqrt{N - |S| - 3} \approx \mathcal{N}(0, 1) \text{ under } H_0$$

- Discrete: χ^2 or G^2 test on contingency tables
- Non-linear: kernel tests (KCI)
- Level α : sparsity knob

PC step 1: skeleton

1. Start: complete undirected graph
2. For $\ell = 0, 1, 2, \dots$: for each edge $i - j$, test $X_i \perp X_j \mid X_S$ for $S \subseteq \text{adj}(i) \setminus \{j\}, |S| = \ell$
3. Independence accepted $\rightarrow$ edge removed, $\text{sepset}(i, j) = S$
4. Stop when no node has more than ℓ neighbours

Sparse graph $\rightarrow$ small conditioning sets $\rightarrow$ polynomial cost.

PC step 2: v-structures

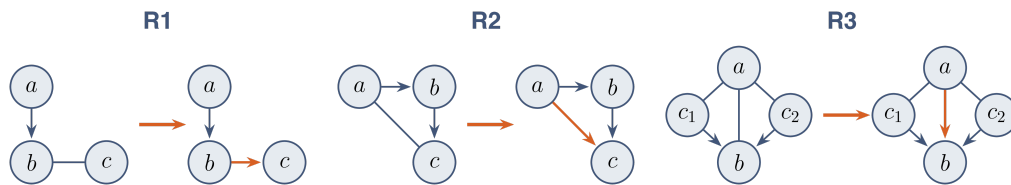
For each triple $i - k - j$ with i and j non-adjacent:

$$k \notin \text{sepset}(i, j) \implies i \rightarrow k \leftarrow j$$

- k in the separating set: chain or fork, no orientation
- k absent: collider, the only structure making i, j dependent given k

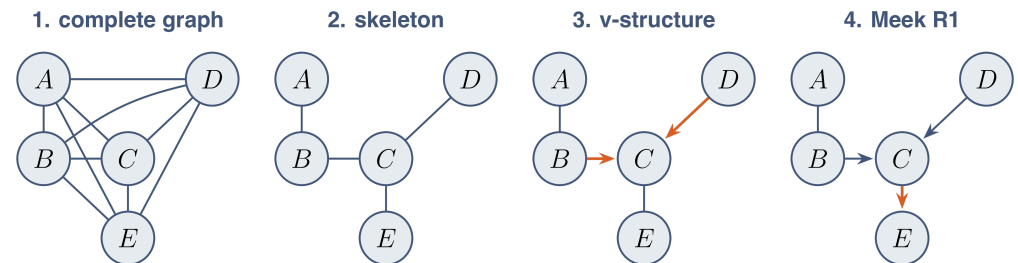
PC step 3: Meek rules

Orientation propagation, no new v-structure, no cycle:



- R1: $a \rightarrow b - c$, a, c non-adjacent $\rightarrow b \rightarrow c$
- R2: $a \rightarrow b \rightarrow c$ and $a - c \rightarrow a \rightarrow c$
- R3: two chains $a - c_1 \rightarrow b$, $a - c_2 \rightarrow b$, c_1, c_2 non-adjacent, $a - b \rightarrow a \rightarrow b$

PC on an example



- Removed: $A - C$ ($S = \{B\}$), $A - D$, $B - D$ ($S = \emptyset$), $A - E$, $B - E$, $D - E$ ($S = \{C\}$)
- $C \notin \text{sepset}(B, D) = \emptyset \rightarrow B \rightarrow C \leftarrow D$
- R1 $\rightarrow C \rightarrow E$; $A - B$ unoriented

PC with causal-learn

```
1 import numpy as np
2 from causallearn.search.ConstraintBased.PC import pc
3 rng = np.random.default_rng(0); N = 2000
4 A = rng.standard_normal(N); D = rng.standard_normal(N)
5 B = 0.8 * A + rng.standard_normal(N)
6 C = 0.7 * B - 0.6 * D + rng.standard_normal(N)
7 E = 0.9 * C + rng.standard_normal(N)
8 cg = pc(np.c_[A, B, C, D, E], alpha=0.01, indep_test="fisherz",
9         node_names=list("ABCDE"), show_progress=False)
10 print(cg.G)
```

Graph Nodes:
A;B;C;D;E

Graph Edges:
1. A --- B
2. B --> C
3. D --> C
4. C --> E

29

Evaluation of learned graphs

31

PC: properties and limits

- Consistent under the assumptions, oracle tests → exact CPDAG
- Test errors propagate; result depends on variable order → **PC-stable**
- Hidden confounders: FCI (partial ancestral graphs)
- Near-violations of faithfulness: unstable orientations

30

Structural Hamming distance

The structural Hamming distance counts the elementary edge operations turning the estimated graph into the true one.

$$\text{SHD}(\hat{G}, G) = \#\{\text{edge additions}\} + \#\{\text{deletions}\} + \#\{\text{reversals}\}$$

- Compared on CPDAGs when only the class is identifiable
- Complements: precision and recall on the skeleton, on arrowheads

32