Structure Learning for Cyclic Linear Causal Models

Carlos Améndola[1], Philipp Dettling[1], Mathias Drton[1],
Federica Onori[2], Jun Wu[1]

[1]Department of Mathematics,
Technische Universität München

[2]Department of Statistical Sciences,
Università La Sapienza di Roma

Conference on Uncertainty in Artificial Intelligence, August 2020

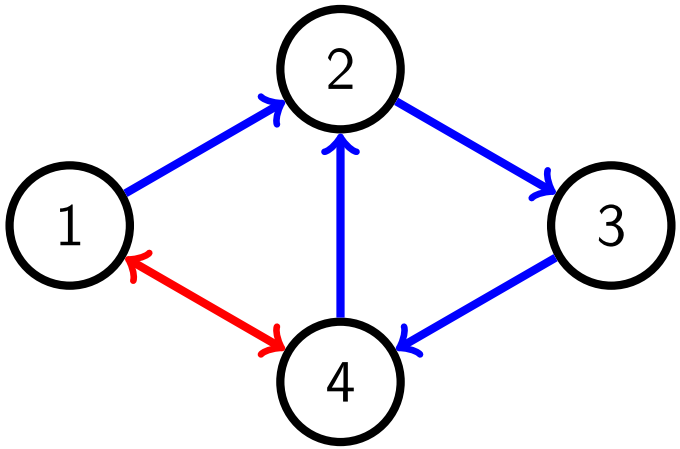
Introduction

- Simple mixed graph: $G = (V, D, B)$[1], allow cycles
- Random vector $X = (X_i : i \in V)$

$$X = \Lambda^T X + \varepsilon, \ \mathrm{Var}[\varepsilon] = \Omega$$

- Linear Gaussian Model of G:

$$\mathcal{M}_G = \left\{ (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} : \Lambda \in \mathbb{R}^D_{\mathrm{reg}}, \ \Omega \in PD(B) \right\}$$



[1] Christopher Nowzohour et al. "Distributional equivalence and structure learning for bow-free acyclic path diagrams". In: *Electron. J. Stat.* 11.2 (2017), pp. 5342–5374. DOI: 10.1214/17-EJS1372. URL: https://doi.org/10.1214/17-EJS1372.

Theoretical Results

Theorem 3.1 (Expected Dimension)

The linear Gaussian model given by a simple mixed graph G has dimension equaling the number of unknown parameters, i.e.

$$\dim(\mathcal{M}_G) = |V| + |D| + |B|.$$

Theorem 4.1 (Sufficient Condition)

Let G_1 and G_2 be two simple mixed graphs with same skeleton and collider triples. Then G_1 and G_2 are distributionally equivalent up to closure, i.e., $\overline{\mathcal{M}_{G_1}} = \overline{\mathcal{M}_{G_2}}$.

(collider triple: $i \to j \leftarrow k$, $i \leftrightarrow j \leftarrow k$, $i \to j \leftrightarrow k$, $i \leftrightarrow j \leftrightarrow k$)

Model Selection

- Score

$$s(G) = \frac{1}{n}\left(\max_{\Sigma \in \mathcal{M}_G} \ell(\Sigma; S) - \text{penalty}(p, k, n)\right)$$

  - notations: $p = |V|, k = |D| + |B|$
  - penalty: standard BIC / extended BIC
  - maximum log-likelihood: block coordinate-descent algorithm[2]
- Greedy Search
  - start: random graphs, by MCMC
  - local neighborhood: adding/removing/changing one edge
  - stop: no higher score / maximum iterations reached

[2] Mathias Drton et al. "The maximum likelihood threshold of a path diagram". In: *Ann. Statist.* 47.3 (2019), pp. 1536–1553. ISSN: 0090-5364. DOI: 10.1214/18-AOS1724. URL: https://doi.org/10.1214/18-AOS1724.

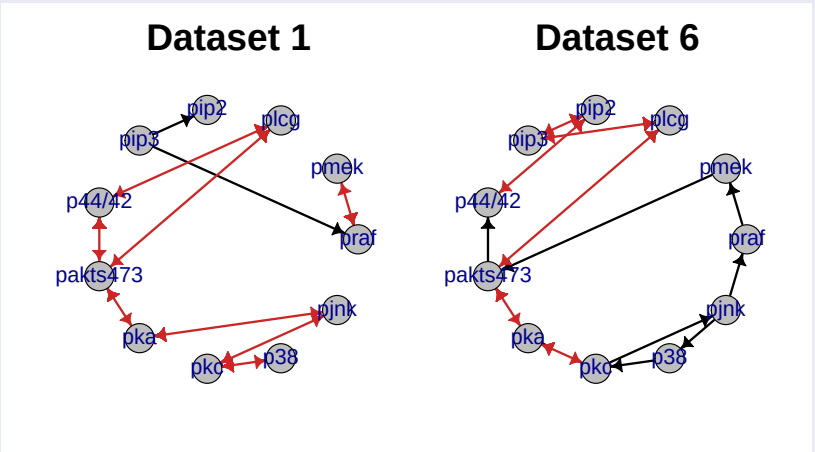
Numeric Experiments

Simulation Studies

- $\mathbf{p} \in \{\mathbf{5}, \mathbf{6}\}$, maximal degree $= 2$
- Evaluation:
  - lower and upper bounds for the frequency of getting equivalent models
  - minimal SHD: distance between equivalent classes

Protein Expression Data

- 14 datasets of $\mathbf{p} = \mathbf{11}$ proteins[a]
- Modeling strategy:
  - Gaussin copula model
  - Kendall's tau correlation matrix
  - extended BIC



[a] Karen Sachs et al. "Causal protein-signaling networks derived from multiparameter single-cell data". In: *Science* 308.5721 (2005), pp. 523–529.