A Bayesian approach to causal discovery from soft interventions

Alessandro Mascaro[1], Federico Castelletti;[2]

[1]University of Milano-Bicocca, Milan, Italy • [2]Università Cattolica del Sacro Cuore, Milan, Italy

Background

Causal discovery

“We view the task of causal discovery as an induction game that scientists play against Nature. Nature possesses stable causal mechanisms that, on a detailed level of descriptions, are deterministic functional relationships between variables, some of which are unobservable. These mechanisms are organized in the form of an acyclic structure, which the scientist attempts to identify from the available observations” (Pearl, 2009)

- Structural Causal Models (SCMs): $X_j \leftarrow m(X_{\text{pa}_j}) + \varepsilon_j$
- Directed Acyclic Graphs (DAGs): $\mathcal{D} = (V, E)$. $i \rightarrow j$ iff i is a cause of j
- if Markovian, same factorization as Bayesian Networks:

$$f(X_1, \ldots, X_q | \boldsymbol{\theta}, \mathcal{D}) = \prod_{j=1}^{q} f(X_j | X_{\text{pa}_j}, \theta_j, \mathcal{D})$$

$\rightarrow$ Conditional independences are the link between SCMs and properties of data that we can test or detect, **using observational data alone**

However, different DAGs, may imply the same set of conditional independence statements: $\rightarrow$ we are able to distinguish up to a **Markov equivalence class** using observational data alone

Interventions

SCMs do not only describe a joint distribution, but also how this distribution changes as a result of *external interventions*

We can envisage two types of interventions:

Hard interventions:

- $X_i \leftarrow c$;
- $f(X_1, \ldots, X_q | \boldsymbol{\theta}, \mathcal{D}) =$
  $\prod_{j \neq i} f(X_j | X_{\text{pa}_j}, \theta_j, \mathcal{D})$

Soft interventions

- $X_i \leftarrow \tilde{m}(X_{\text{pa}_j}) + \varepsilon_i$;
- $f(X_1, \ldots, X_q | \boldsymbol{\theta}, \mathcal{D}) =$
  $\tilde{f}\left(X_i | X_{\text{pa}_i}, \tilde{\theta}_i, \mathcal{D}\right) \prod_{j \neq i} f(X_j | X_{\text{pa}_j}, \theta_j, \mathcal{D})$

$\rightarrow$ If interventional data is available, we can distinguish between DAGs inside the same Markov equivalence class by checking the local changes in the distributions implied by an intervention.

However we still have I-Markov equivalence classes of DAGs

Goal

1. To evaluate the posterior distribution of different causal structures $p(\mathcal{D}|\boldsymbol{\mathcal{X}}) \propto p(\boldsymbol{\mathcal{X}}|\mathcal{D})p(\mathcal{D})$ using a mixture of observational and (soft) interventional **Gaussian** data
2. To estimate the model parameters jointly using K datasets coming from different experimental settings;

A Bayesian approach

$$X_1, \ldots, X_q \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} \in \mathcal{C}_{\mathcal{D}}$$

Likelihood

$$f(\mathcal{X} | \boldsymbol{D}, \boldsymbol{L}, \tilde{\boldsymbol{D}}, \tilde{\boldsymbol{L}}, \mathcal{D}) = \prod_{j=1}^{q} d\mathcal{N}_{|A(j)|}\left(\boldsymbol{\mathcal{X}}_{\cdot j}^{A(j)} | - \boldsymbol{\mathcal{X}}_{\cdot \text{pa}_j} \boldsymbol{L}_{\prec j]}, \boldsymbol{D_{jj}} \boldsymbol{I_{|A(j)|}}\right) \cdot$$
$$\mathcal{N}_{n-|A(j)|}\left(\boldsymbol{\mathcal{X}}_{\cdot j}^{A(j)^c} | - \boldsymbol{\mathcal{X}}_{\cdot \text{pa}_j} \tilde{\boldsymbol{L}}_{\prec j]}, \tilde{\boldsymbol{D}}_{\boldsymbol{jj}} \boldsymbol{I_{n-|A(j)|}}\right)$$

- $A(j)$ is an index set of all those observation for which node j is not intervened upon;
- The parameters $(\boldsymbol{L}, \boldsymbol{D})$ come from a decomposition of $\boldsymbol{\Sigma}$: $\boldsymbol{\Sigma} = \boldsymbol{L}^{-T} \boldsymbol{D} \boldsymbol{L}^{-1}$;
- $(\tilde{\boldsymbol{L}}, \tilde{\boldsymbol{D}})$ are the new parameters induced by the soft intervention;

Prior specification

- Prior on causal structures:

$$p(\mathcal{D}) \propto p\left(\boldsymbol{S}^{\mathcal{D}}\right) = \pi^{|\boldsymbol{S}^{\mathcal{D}}|}(1-\pi)^{\frac{q(q-1)}{2} - |\boldsymbol{S}^{\mathcal{D}}|}$$

- Priors on parameters $(\boldsymbol{L}, \boldsymbol{D})$ (similarly for $(\tilde{\boldsymbol{L}}, \tilde{\boldsymbol{D}})$):

$$p(\boldsymbol{L}, \boldsymbol{D} | \mathcal{D}) = \prod_{j=1}^{q} p(\boldsymbol{L}_{\prec j]}, \boldsymbol{D}_{jj} | \mathcal{D})$$

$$\boldsymbol{D}_{jj} \sim I - Ga\left(a_j^{\mathcal{D}}/2, g/2\right)$$

$$\boldsymbol{L}_{\prec j]} \mid \boldsymbol{D}_{jj} \sim \mathcal{N}_{|\text{pa}_{\mathcal{D}}(j)|}\left(0, \boldsymbol{D}_{jj} \boldsymbol{I}_{|\text{pa}_{\mathcal{D}}(j)|}/g\right)$$

where $a_j^{\mathcal{D}} = a + |\text{pa}_j| - q + 1$.

$\rightarrow$ (conjugate) DAG-Wishart prior with hyperparameter chosen through Geiger-Heckerman (2002) prior construction method; ! Geiger-Heckerman (2002) prior construction method guarantees score equivalence of I-Markov equivalence DAGs

Posterior inference

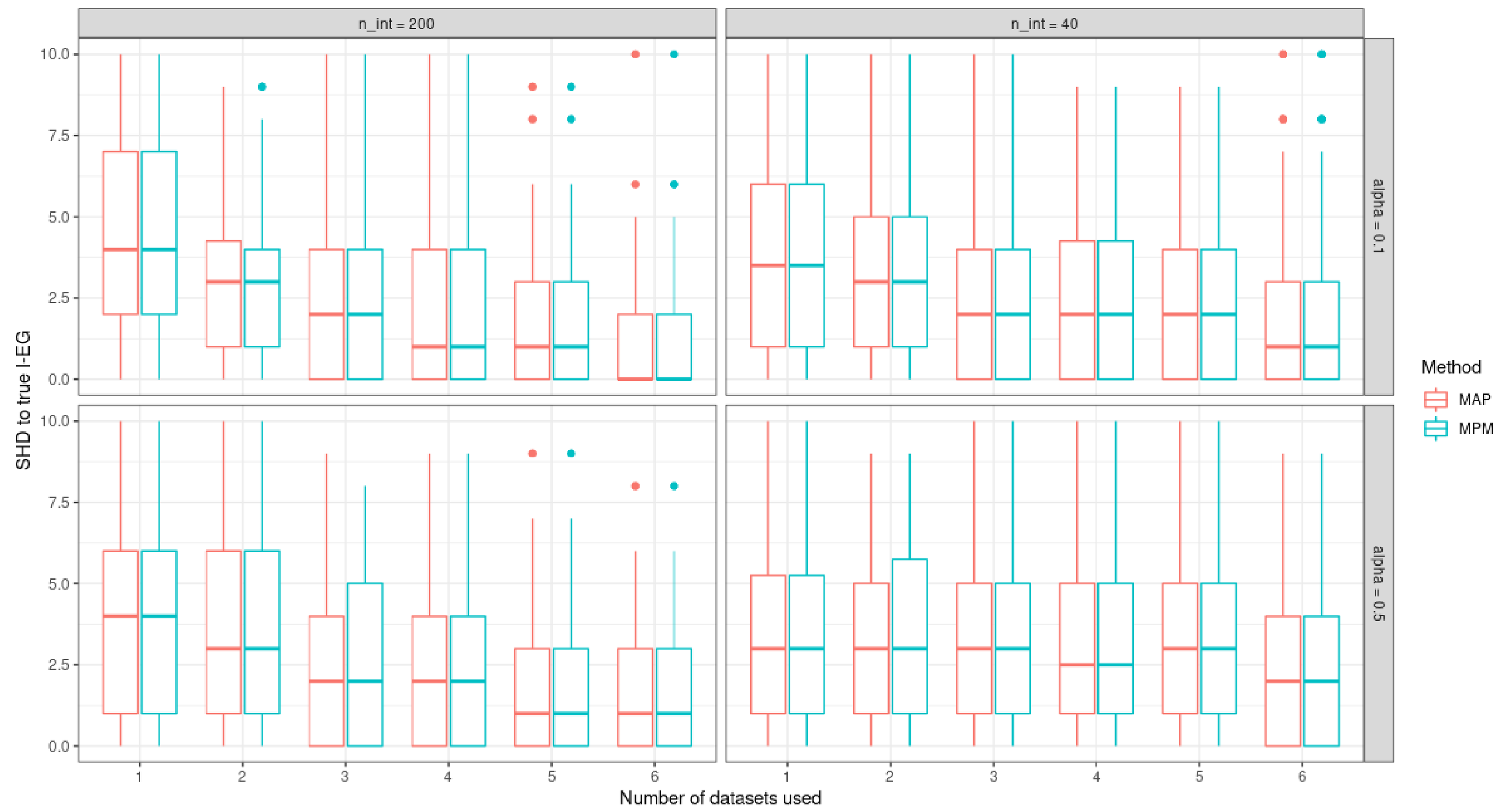
- Posterior of $\mathcal{D}$, obtained through a Partial Analytic Structure (PAS) algorithm (Godsill, 2012) which proceeds by iteratively updating the DAG and the parameters;

Simulation

Settings

- 50 DAGs with probability of edge inclusion $\pi = 0.5$ are generated;
- For each DAG:
  1. An observational dataset of size n_{obs} with $q = 10$ variables is generated from a linear SEM with weights uniformly drawn in the interval $[-2, -0.5] \cup [0.5, 2]$ and then standardized
  2. 5 intervention targets are sampled without replacement;
  3. For each intervention target, a dataset of size n_{int} with q variables is generated from a linear SEM and then standardized. The weights associated with the parents of the intervened node are then multiplied by a factor α;
- $\alpha \in \{0.1, 0.5\}$
- $(n_{\text{obs}}, n_{\text{int}}) \in \{(200, 200), (1000, 40)\}$

Results



Main references

- Castelletti, F., & Consonni, G. (2019). Objective Bayes model selection of Gaussian interventional essential graphs for the identification of signaling pathways. Annals of Applied Statistics, 13(4), 2289-2311.
- Geiger, D., & Heckerman, D. (2002). Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. The Annals of Statistics, 30(5), 1412-1440.
- Godsill, S.J. (2012) On the relationship between Markov chain Monte Carlo methods for model uncertainty. Journal of Computational and Graphical Statistics 10(2):230-248
- Pearl, J. (2009). Causality. Cambridge university press.

Contact information

Alessandro Mascaro, <a.mascaro3@campus.unimib.it>