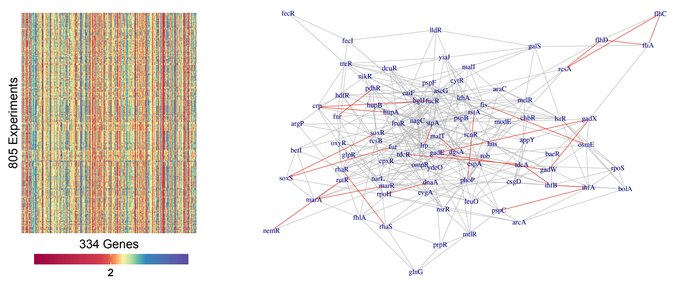
Munich Center for Machine Learning (MCML)
Research group Mathematical Statistics
Technical University of Munich

Confidence in Causal Discovery with Linear Causal Models

David Strieder[1], Mathias Drton[1]

Motivation

Inferring gene regulatory networks with graphical/causal models:



Gene expression data for *E. coli* and estimated (undirected) network (from Drton & Maathuis. *Structure Learning in Graphical Modeling*, 2017)

Structure learning methods may estimate

- the absence/presence of interactions in complex systems, and also
- causal directions (does gene 'X' regulate gene 'Y' or vice versa?).

Given a causal structure, standard statistical methods quantify the size of causal effects and provide an uncertainty assessment.

However, **'double-dipping' problem** occurs when both causal structure and the effect are being estimated from one data set.

We propose a new framework to construct **confidence sets for causal effects that capture both sources of uncertainty** (causal structure and numerical size of effect).

$\longrightarrow$ First results/paper at UAI 2021 (37th Conference on Uncertainty in AI)

Setup

Model assumptions that ensure identifiability:

- Observational data of $(X_1, .., X_d)$ follows **linear structural equation model**, represented by a **directed acyclic graph**, that is,

$$X_j = \sum_{k \neq j} \beta_{jk} X_k + \epsilon_j, \quad j = 1, ..., d,$$

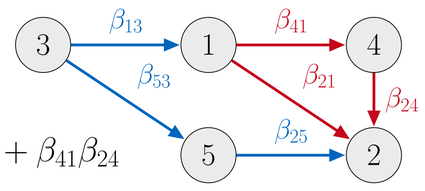
- and **Gaussian errors** with **equal variances**

$$\epsilon_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \quad j = 1, ..., d.$$

Target quantity is the **total causal effect** that an external intervention on variable X_1 has onto variable X_2, that is,

$$\begin{aligned} \mathcal{C}(1 \to 2) &:= \tfrac{\mathrm{d}}{\mathrm{d}x_1} \mathbb{E}[X_2 | \, \mathrm{do}(X_1 = x_1)] \\ &= \Sigma_{12|\mathrm{pa}(1)} / \Sigma_{11|\mathrm{pa}(1)} \; \mathbb{1}(1 <_{\mathcal{G}} 2) \end{aligned}$$

Example:



$\mathcal{C}(1 \to 2) = \beta_{21} + \beta_{41}\beta_{24}$

Method

Key Idea: Use test inversion

- Leverage duality between statistical hypothesis tests and confidence regions.
- Shifts task to construction of **tests for all possible values of the total causal effect**, i.e., for all $\psi \in \mathbb{R}$ we have to test the hypothesis

$$\mathsf{H}_0^{(\psi)} : \mathcal{C}(1 \to 2) = \psi.$$

Difficulty: Hypothesis of fixed effect ψ is **union of single hypotheses** over all possible directed acyclic graphs $\mathcal{G}(d)$ that allow ψ.

Insight: Each single hypothesis for a given graph defines a **smooth submanifold** of the cone of covariance matrices, namely all $\Sigma \in$ PD(d) such that there exists $\sigma \in \mathbb{R}$ with

$$\begin{cases} \psi &= \Sigma_{12|pa(1)} / \sigma^2 \; \mathbb{1}(1 <_{\mathcal{G}} 2) \\ \sigma^2 &= \Sigma_{jj|pa(j)}, \qquad j = 1, ..., d. \end{cases}$$

Idea: **Intersection union tests**, that is, reject the union if we reject each single hypothesis.

We provide **three concrete solutions** based on likelihood ratio tests of order constraints (Silvapulle & Sen. *Constrained Statistical Inference,* 2005) and recent theory of universal inference (Wassermann et al. *Universal Inference*, 2020).

Constrained likelihood ratio test (LRT)

- Idea: **Relax alternative** and test against the entire cone of covariance matrices.
- Under each single hypothesis the asymptotic distribution of the likelihood ratio statistic is a chi-square distribution.
- **Result:** An asymptotic $(1 - \alpha)$-confidence set for the causal effect $\mathcal{C}(1 \to 2)$ is the union of

$$\{\psi \in \mathbb{R} : \min_{G \in \mathcal{G}(d) \,:\, 1 <_G 2} \lambda_n^{(\psi)}(G) \leq \chi^2_{d,1-\alpha}\}$$

and $\{0 : \min_{G \in \mathcal{G}(d) \,:\, 2 <_G 1} \lambda_n^{(0)}(G) \leq \chi^2_{d-1,1-\alpha}\}$.

Split likelihood ratio test (SLRT)

- Idea: Employ **data splitting approach**.
- Uses modification of classical likelihood ratio statistic with a universal critical value.
- Type-I error control via Markov's inequality.
- Conservative but finite sample guarantee.

Note that confidence regions might be disconnected, reflecting the larger null hypothesis for a zero effect.

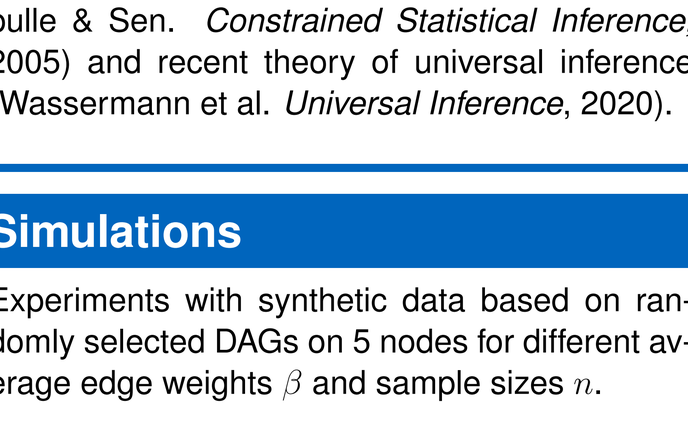
Simulations

Experiments with synthetic data based on randomly selected DAGs on 5 nodes for different average edge weights β and sample sizes n.

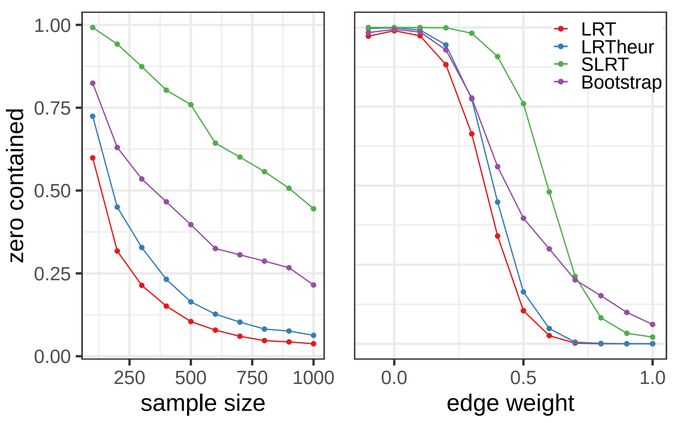
- All **proposed methods achieve** the desired empirical **coverage probability**.

method	LRT		LRTheur		SLRT		Bootstrap	
$\beta \backslash n$	100	1000	100	1000	100	1000	100	1000
0.05	0.98	0.97	1.00	1.00	1.00	1.00	0.66	0.75
0.1	0.98	0.98	1.00	1.00	1.00	1.00	0.75	0.83
0.5	0.98	0.98	1.00	1.00	1.00	1.00	0.97	0.97

Empirical Coverage of 95%-CIs.



Mean width of 95%-CIs.



Percentage of times zero contained in 95%-CIs.

- Bootstrap method does not work in practice (with established GDS algorithm (Peters & Bühlmann. *Identifiability of Gaussian Structural Equation Models with Equal Variances,* 2014)).
- Proposed methods correctly account for the uncertainty in the causal structure and successfully **help to draw conclusions about the existence and size of causal effects.**

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 883818).

[1]Munich Center for Machine Learning (MCML)
Research group Mathematical Statistics
Technical University of Munich, Germany

IMS International Conference on Statistics and Data Science (ICSDS)
December 13-16, 2022, Florence, Italy