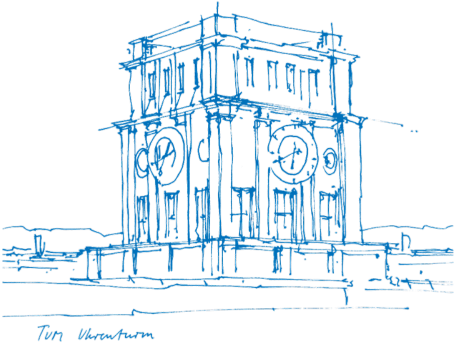
Research group Mathematical Statistics
TUM School of Computation, Information and Technology
Technical University of Munich

Confidence in Causal Discovery with Linear Causal Models

David Strieder

First results in:
Strieder, D., Freidling, T., Haffner, S., Drton, M. Confidence in causal discovery with linear causal models. PMLR 161:1217-1226 (2021).

ETH-UCPH-TUM Workshop on Graphical Models
Science and Study Center Raitenhaslach
October 10-14, 2022



Starting point

- **Research question:** What is the total causal effect of X_1 on X_2? Confidence?

Starting point

- **Research question:** What is the total causal effect of X_1 on X_2? Confidence?
- **Given:** Observational data in form of n samples of $(X_1, ..., X_d)$.
- **Problem:** Causal structure unknown.

Starting point

- **Research question:** What is the total causal effect of X_1 on X_2? Confidence?
- **Given:** Observational data in form of n samples of $(X_1, ..., X_d)$.
- **Problem:** Causal structure unknown.
- Naive two-step approach?
  (1) Learn causal structure.
  (2) Calculate confidence intervals for causal effects in inferred model.

Setup

Model assumptions that ensure identifiability

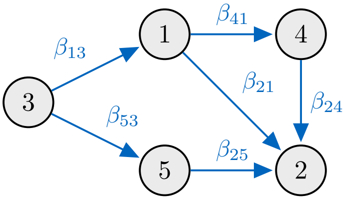
- **Linear** structural equation model with **Gaussian errors** with **equal variances**.

LSEM

$$X_j = \sum_{k \neq j} \beta_{jk} X_k + \epsilon_j, \qquad \epsilon_j = N(0, \sigma^2), \qquad j = 1, ..., d.$$

- Represented by directed acyclic graph $\mathcal{G}$.

Example:



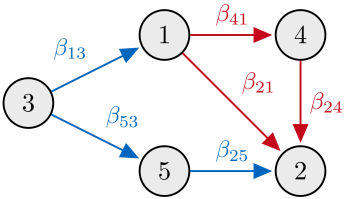
Setup

- **Target:** Total causal effect of an external intervention on variable X_1 onto variable X_2.

$$\mathcal{C}(1 \to 2) := \frac{\mathrm{d}}{\mathrm{d}x_1}\mathbb{E}[X_2 | \operatorname{do}(X_1 = x_1)] = (I_d - B)^{-1}_{21}$$

$$= \Sigma_{12|pa(1)}/\Sigma_{11|pa(1)}\mathbb{1}(1 <_{\mathcal{G}} 2)$$

Example:



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

Setup

- **Idea:** Use test inversion.

Tests for $\mathrm{H}_0^{(\psi)} : \mathcal{C}(1 \to 2) = \psi$ → Confidence interval for $\mathcal{C}(1 \to 2)$

- **Goal:** construct suitable **tests for all possible hypothesized causal effects!**

Setup

- **Idea:** Use test inversion.

  Tests for $\mathsf{H}_0^{(\psi)} : \mathcal{C}(1 \to 2) = \psi$ → Confidence interval for $\mathcal{C}(1 \to 2)$

- **Goal:** construct suitable **tests for all possible hypothesized causal effects!**

- **Difficulty:** Hypothesis of fixed effect ψ is a **union of single hypotheses** over all directed acyclic graphs across d nodes $\mathcal{G}(d)$, that is,

$$\mathsf{H}_0^{(\psi)} := \bigcup_{G \in \mathcal{G}(d)} \mathsf{H}_0^{(\psi)}(G)$$

Single Hypothesis $H_0^{(\psi)}(G)$

$$H_0^{(\psi)}(G) : \left\{ \Sigma \in \mathrm{PD(d)} : \exists \sigma^2 \text{ such that } \begin{cases} \psi & = \Sigma_{12|pa(1)}/\sigma^2 \, \mathbb{1}(1 <_{\mathcal{G}} 2) \\ \sigma^2 & = \Sigma_{jj|pa(j)} \ \forall \, j = 1, ..., d \end{cases} \right\}$$

- **Idea:** Use theory of intersection union test.
- Reject union if we reject each single hypothesis.

Constrained likelihood ratio test

- **Idea:** Relax alternative to entire cone of covariance matrices.

- Each single hypothesis for a given graph defines a smooth submanifold of different dimension depending on the causal order of the graph.

- The limit distribution is a chi-squared distribution.

- **Result:** Asymptotic $(1-\alpha)$ confidence set for causal effect $\mathcal{C}(1 \to 2)$ is

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

Split likelihood ratio test[1]

- **Idea:** Split data and use universal critical value.
- Calculate MLE of Σ under alternative based on Data set 1.
- Calculate MLE of Σ under hypothesis and likelihoods based on Data set 2.
- **Result:** $(1-\alpha)$ confidence set for causal effect $\mathcal{C}(1 \to 2)$ is

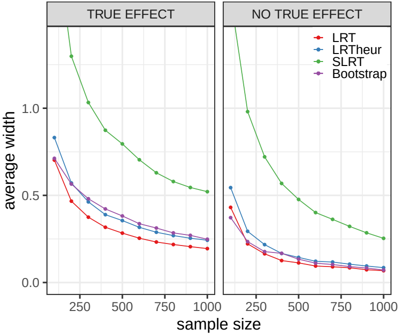
$$\{\psi \in \mathbb{R} : \min_{G \in \mathcal{G}(d) : 1 <_G 2} \tilde{\lambda}_n^{(\psi)}(G) \leq -2\log(\alpha)\} \cup \{0 : \min_{G \in \mathcal{G}(d) : 2 <_G 1} \tilde{\lambda}_n^{(0)}(G) \leq -2\log(\alpha)\}$$

[1] Wasserman L, Ramdas A, Balakrishnan S. Universal inference. *Proc. Natl. Acad. Sci. USA.* 2020.

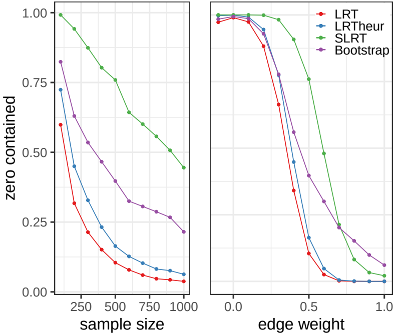
Simulations

method	$n \backslash \beta$	TRUE EFFECT 0.05	0.1	0.5
LRT	100	0.98	0.98	0.98
	500	0.99	0.99	0.98
	1000	0.97	0.98	0.98
LRTheur	100	1.00	1.00	1.00
	500	1.00	1.00	1.00
	1000	1.00	1.00	1.00
SLRT	100	1.00	1.00	1.00
	500	1.00	1.00	1.00
	1000	1.00	1.00	1.00
Bootstrap	100	0.66	0.75	0.97
	500	0.71	0.79	0.96
	1000	0.75	0.83	0.97

Empirical Coverage of 95%-CIs



Mean Width of 95%-CIs



Zero contained in 95%-CIs