Chair of Mathematical Statistics
Department of Mathematics
Technical University of Munich

Confidence in Causal Discovery with Linear Causal Models

UAI 2021

David Strieder[1], Tobias Freidling[1,2], Stefan Haffner[1], and Mathias Drton[1]

Motivation

Structure learning achieves accuracy of $\approx$ 70% in a prominent benchmark study by Mooij et al. [2016] but:

What about confidence in causal effect estimation?

- Confidence statements are needed to reliably draw conclusions from estimated causal effects.

Naive two-step approach:

1. Apply causal structure learning algorithm.
2. Use standard methods to calculate confidence intervals for causal effects in the inferred model.

$\rightarrow$ **Fails** to account for uncertainty wrto. structure.

Example: If we incorrectly select model $X_1 \leftarrow X_2$, we are "certain" the effect of X_1 on X_2 is zero.

What are the **difficulties**?

- Cannot restrict to one fixed causal ordering, while respecting uncertainty in causal structure.
- Different causal structures allow for the same numerical size of the causal effect.
- Classical resampling/bootstrapping techniques and standard asymptotic MLE-theory do not work.

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

Setup

Start with simplest setting: Recursive linear structural equation model with homoscedastic Gaussian errors.

Bivariate case with two possible models:

$$\begin{array}{lll} (\text{M1: } 1 \rightarrow 2) & X_1 = \epsilon_1, & X_2 = \beta_{21} X_1 + \epsilon_2, \\ (\text{M2: } 1 \leftarrow 2) & X_1 = \beta_{12} X_2 + \epsilon_1, & X_2 = \epsilon_2, \end{array}$$

with parameters $\beta_{12}, \beta_{21} \in \mathbb{R}$ and $\epsilon_1, \epsilon_2 \overset{\text{ind}}{\sim} N(0, \sigma^2)$.

Write Σ for the covariance matrix of (X_1, X_2).

Method

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

$$\mathcal{C}(1 \rightarrow 2) := \frac{\mathrm{d}}{\mathrm{d}x_1} \mathbb{E}[X_2 | \, \mathrm{do}(X_1 = x_1)] = \beta_{21} \mathbb{1}\{(\text{M1})\} = \frac{\Sigma_{12}}{\Sigma_{11}} \mathbb{1}\{\Sigma_{11} \leq \Sigma_{22}\}.$$

Key Idea: Use test inversion

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

$$\mathrm{H}_0: \quad \mathcal{C}(1 \rightarrow 2) = \psi.$$

Three concrete tests based on likelihood ratio tests of order constraints [Silvapulle and Sen, 2005], and recent theory of universal inference [Wasserman et al., 2020].

1. Testing Causal Ordering (LRT1)

- Assumption of **homoscedasticity** implies that causal order is implied by a set of inequalities for variances.
- Testing these constraints leads to hypothesis

$$\mathrm{H}_0: \begin{cases} \Sigma_{12} = \psi \Sigma_{11} \text{ and } \Sigma_{11} \leq \Sigma_{22}, & \text{if } 0 \leq |\psi| \leq 1, \\ \Sigma_{12} = \psi \Sigma_{11}, & \text{if } 1 < |\psi|, \\ \Sigma_{11} \geq \Sigma_{22}, & \text{if } \psi = 0. \end{cases}$$

- General alternative of entire positive definite cone.
- Stochastically largest asymptotic distribution of the **likelihood ratio**

$$\lambda_n \xrightarrow{\mathcal{D}} \begin{cases} 0.5\chi_1^2 + 0.5\chi_2^2, & \text{if } 0 \leq |\psi| \leq 1, \\ \chi_1^2, & \text{if } 1 < |\psi|, \\ 0.5\chi_0^2 + 0.5\chi_1^2, & \text{if } \psi = 0, \end{cases} \quad \text{as } n \rightarrow \infty.$$

- Best performing method in experiments with real data.

2. Testing Structure Assumptions (LRT2)

- Assumption of underlying **LSEM imposes structure** on the covariance matrix.
- Testing those polynomial constraints representing different possible models leads to hypothesis

$$\mathrm{H}_0: \begin{cases} \Sigma_{12} = \psi \Sigma_{11} \text{ and } \Sigma_{11}^2 = \det(\Sigma), & \text{if } \psi \neq 0, \\ \Sigma_{22}^2 = \det(\Sigma), & \text{if } \psi = 0. \end{cases}$$

- General alternative of entire positive definite cone.
- Asymptotic distribution of the **likelihood ratio** λ_n

$$\lambda_n \xrightarrow{\mathcal{D}} \begin{cases} \chi_2^2, & \text{if } \psi \neq 0, \\ \chi_1^2, & \text{if } \psi = 0, \end{cases} \quad \text{as } n \rightarrow \infty.$$

- Explicit calculation of confidence interval possible.
- Best performing method for simulated data.

3. Split Likelihood Ratio Tests (SLRT)

- Employ theory of universal inference by Wasserman et al. [2020], a general framework to construct hypothesis test.
- Uses a modification of the classical likelihood-ratio statistic, termed **split likelihood ratio**.
- Based on **data splitting approach**.
- Appealing for irregular composite hypotheses where asymptotic distributions are intractable.
- Type-I error control via **Markov's inequality**.
- Explicit calculation of confidence interval possible.
- Conservative method but **finite sample guarantee**.

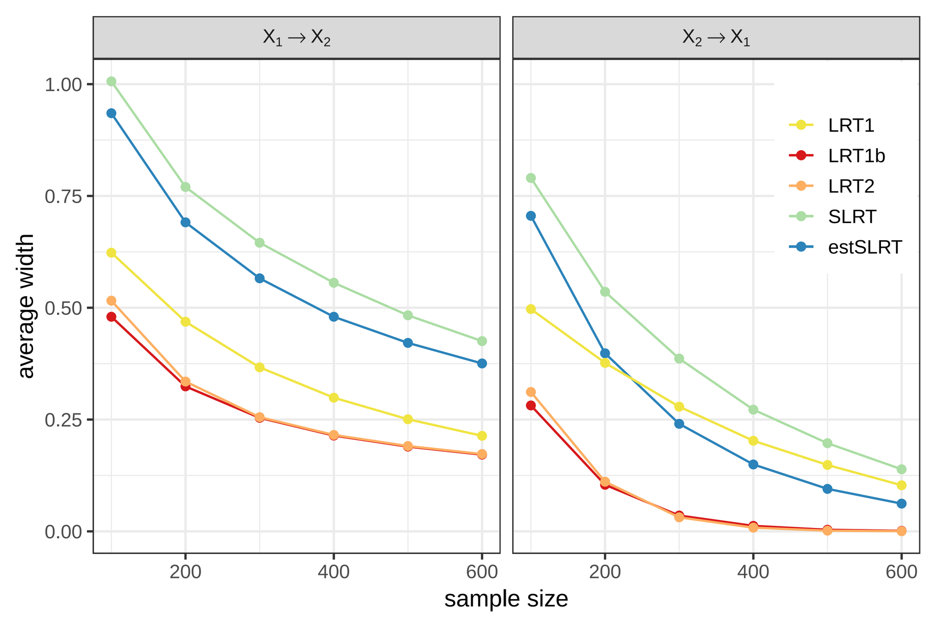
Note that confidence regions might be disconnected, reflecting the larger null hypothesis for a zero effect.

Simulations

- Synthetic data based on (M1) or (M2).
- All **proposed methods achieve** the desired empirical **coverage probability**.
- Bootstrap method with established GDS [Peters and Bühlmann, 2014] algorithm does not work in practice.
- Proposed methods account for the high uncertainty in the causal structure for small true causal effects.

method	$n \backslash \beta$	$X_1 \rightarrow X_2$ 0	0.05	0.5	$X_2 \rightarrow X_1$ 0	0.05	0.5
LRT1	500	1.00	0.95	0.97	1.00	0.99	1.00
	1000	1.00	0.96	0.98	1.00	0.98	1.00
LRT2	500	0.97	0.97	0.97	0.97	0.97	1.00
	1000	0.97	0.96	0.96	0.96	0.96	1.00
SLRT	500	1.00	1.00	1.00	1.00	1.00	1.00
	1000	1.00	1.00	1.00	1.00	1.00	1.00
Bootstrap	500	1.00	0.63	0.95	1.00	1.00	1.00
	1000	1.00	0.67	0.96	1.00	1.00	1.00

Empirical coverage of 95%-confidence intervals.



Average maximum width of 95%-confidence intervals.

- True causal effect of size 0.5 (in different directions).
- For no true effect confidence intervals converge to zero for reasonably large sample sizes.
- Provide correct confidence sets that successfully **help decide whether there is an effect or not**.

Outlook

- **Generalize proposed framework** in future work.
- **First promising results** with SLRT method
  - for higher dimensions,
  - for different model assumptions (linear non gaussian additive noise models) via empirical likelihood.

References

Joris M. Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. *J. Mach. Learn. Res.*, 17:Paper No. 32, 102, 2016.

Jonas Peters and Peter Bühlmann. Identifiability of Gaussian structural equation models with equal error variances. *Biometrika*, 101(1):219–228, 2014.

Mervyn J. Silvapulle and Pranab K. Sen. *Constrained statistical inference.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2005.

Larry Wasserman, Aaditya Ramdas, and Sivaraman Balakrishnan. Universal inference. *Proceedings of the National Academy of Sciences*, 117(29): 16880–16890, 2020.

[1]Chair of Mathematical Statistics, Technical University of Munich, Germany
[2]Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, UK

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).