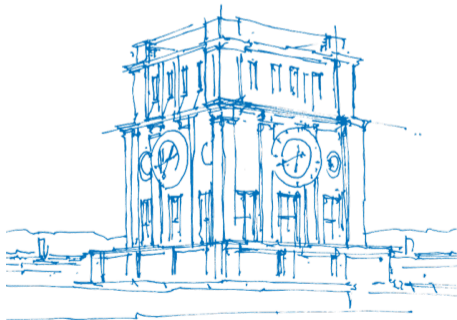


Dual Likelihood for Causal Inference under Structure Uncertainty

David Strieder, Mathias Drton

CLeaR (Causal Learning and Reasoning)
Los Angeles, CA, April 1-3, 2024



TUM Uhrenturm

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

- **Research question:** What is the total causal effect of X_i on X_j ? Confidence?
- **Given:** Observational data in form of n samples of (X_1, \dots, 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

Underlying Linear SCM with equal error variances

■ Example:

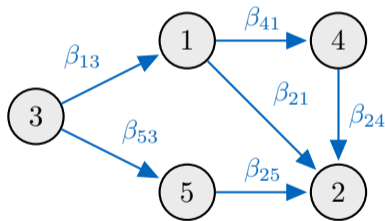
$$X_1 = \beta_{13}X_3 + \varepsilon_1$$

$$X_2 = \beta_{21}X_1 + \beta_{24}X_4 + \beta_{25}X_5 + \varepsilon_2$$

$$X_3 = \varepsilon_3$$

$$X_4 = \beta_{41}X_1 + \varepsilon_4$$

$$X_5 = \beta_{25}X_3 + \varepsilon_5$$



where $\varepsilon_j \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

Setup

Underlying Linear SCM with equal error variances

- **Example:** Causal effect $\mathcal{C}(1 \rightarrow 2) := \frac{d}{dx_1} \mathbb{E}[X_2 | \text{do}(X_1 = x_1)] = \Sigma_{12|p(1)} / \Sigma_{11|p(1)}$.

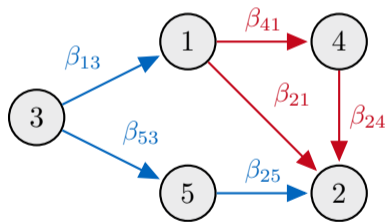
$$X_1 = \beta_{13}X_3 + \varepsilon_1$$

$$X_2 = \beta_{21}X_1 + \beta_{24}X_4 + \beta_{25}X_5 + \varepsilon_2$$

$$X_3 = \varepsilon_3$$

$$X_4 = \beta_{41}X_1 + \varepsilon_4$$

$$X_5 = \beta_{25}X_3 + \varepsilon_5$$



where $\varepsilon_j \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

- **Main Idea:** Use test inversion.



- **Goal:** Construct suitable **tests for all possible effects.**

- **Main Idea:** Use test inversion.



- **Goal:** Construct suitable **tests for all possible effects**.

- **Difficulty:** Each Hypothesis of fixed effect is **union of single hypotheses** over all DAGs on d nodes.

$$H_0^{(\psi)} := \bigcup_{G \in \mathcal{G}(d)} H_0^{(\psi)}(G)$$

- Linear SCM with equal error variances: $\mathcal{M} := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}(G)$, where

$$\mathcal{M}(G) = \left\{ \Sigma \in \text{PD}(d) : \exists \sigma^2 > 0 \text{ with } \sigma^2 = \Sigma_{k,k|p(k)} \quad \forall k = 1, \dots, d \right\}.$$

- Linear SCM with equal error variances: $\mathcal{M} := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}(G)$, where

$$\mathcal{M}(G) = \left\{ \Sigma \in \text{PD}(d) : \exists \sigma^2 > 0 \text{ with } \sigma^2 = \Sigma_{k,k|p(k)} \quad \forall k = 1, \dots, d \right\}.$$

- Hypothesis of fixed total causal effect: $\mathcal{M}_\psi := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}_\psi(G)$, where

$$\mathcal{M}_\psi(G) := \left\{ \Sigma \in \mathcal{M}(G) : \psi = \Sigma_{j,i|p(i)} / \Sigma_{i,i|p(i)} \right\}.$$

- Linear SCM with equal error variances: $\mathcal{M} := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}(G)$, where

$$\mathcal{M}(G) = \left\{ \Sigma \in \text{PD}(d) : \exists \sigma^2 > 0 \text{ with } \sigma^2 = \Sigma_{k,k|p(k)} \quad \forall k = 1, \dots, d \right\}.$$

- Hypothesis of fixed total causal effect: $\mathcal{M}_\psi := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}_\psi(G)$, where

$$\mathcal{M}_\psi(G) := \left\{ \Sigma \in \mathcal{M}(G) : \psi = \Sigma_{j,i|p(i)} / \Sigma_{i,i|p(i)} \right\}.$$

- **Task:** For all $\psi \in \mathbb{R}$ and $G \in \mathcal{G}(d)$ invert joint test of structure and effect size:

$$H_0^{(\psi)}(G) : \Sigma \in \mathcal{M}_\psi(G) \quad \text{against} \quad H_1 : \Sigma \in \mathcal{M} \setminus \mathcal{M}_\psi(G).$$

- **Problem:** Maximizing the Gaussian likelihood

$$\frac{2}{n} \ell_n(\Sigma) = -\log \det(2\pi\Sigma) - \text{tr}(\Sigma^{-1}\hat{\Sigma})$$

with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\text{tr}((I_d - B)^T(I_d - B)\hat{\Sigma})$$

where $B \in \mathbb{R}^G$ represents the direct causal effects between variables.

- **Problem:** Maximizing the Gaussian likelihood

$$\frac{2}{n} \ell_n(\Sigma) = -\log \det(2\pi\Sigma) - \text{tr}(\Sigma^{-1}\hat{\Sigma})$$

with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\text{tr}((I_d - B)^T(I_d - B)\hat{\Sigma})$$

where $B \in \mathbb{R}^G$ represents the direct causal effects between variables.

- Fixing total causal effects is complex **polynomial constraints** on direct effects, namely

$$(I_d - B)_{j,i}^{-1} = \psi.$$

- **Solution:** Maximizing the Dual likelihood

$$\frac{2}{n} \ell_n^{dual}(\Sigma) := -\log \det(2\pi\Sigma^{-1}) - \text{tr}(\Sigma\hat{\Sigma}^{-1})$$

with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\text{tr}((I_d - B)^{-1}(I_d - B)^{-T}\hat{\Sigma}) = \text{tr}((I_d - T)^T(I_d - T)\hat{\Sigma})$$

where $T \in \mathbb{R}^{-G}$ represents the negative total causal effects between variables.

- **Solution:** Maximizing the Dual likelihood

$$\frac{2}{n} \ell_n^{dual}(\Sigma) := -\log \det(2\pi\Sigma^{-1}) - \text{tr}(\Sigma\hat{\Sigma}^{-1})$$

with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\text{tr}((I_d - B)^{-1}(I_d - B)^{-T}\hat{\Sigma}) = \text{tr}((I_d - T)^T(I_d - T)\hat{\Sigma})$$

where $T \in \mathbb{R}^{-G}$ represents the negative total causal effects between variables.

- Fixed total effect constraint only pertains to one parameter, namely

$$T_{i,j} = -\psi$$

■ Main steps:

- (1) Intersection union test.
- (2) Stochastic upper bound by relaxing alternative.
- (3) Dual-LRT with conservative critical values from upper bound.

■ Main steps:

- (1) Intersection union test.
- (2) Stochastic upper bound by relaxing alternative.
- (3) Dual-LRT with conservative critical values from upper bound.

■ Result: Asymptotic $(1 - \alpha)$ -confidence set for causal effect $\mathcal{C}(i \rightarrow j)$ is

$$\{\psi \in \mathbb{R} : \min_{G \in \mathcal{G}(d) : i <_G j} \text{dual-}\lambda_n^{(\psi)}(G) \leq \chi_{d,1-\alpha}^2\} \cup \{0 : \min_{G \in \mathcal{G}(d) : j <_G i} \text{dual-}\lambda_n^{(0)}(G) \leq \chi_{d-1,1-\alpha}^2\}$$

- Closed-form solution for confidence regions for total causal effects capturing both types of uncertainty: numerical size of effects and causal structure.
- Branch and bound type search algorithm through causal orderings. Feasible up to 12 involved variables (already more than 10^{26} structures).
- Conceptual ideas of leveraging test inversions of joint tests for causal structure and effect size generalizable to other modeling assumptions.
- Matrix inversion interplay between direct and total effects that is behind the use of dual likelihood can be exploited in other causal inference tasks.

Thank you!

■ Our related papers:

- Strieder and Drton (2024). *Dual Likelihood for Causal Inference under Structure Uncertainty*. CLeaR 24.
- Strieder and Drton (2023). *Confidence in Causal Inference under Structure Uncertainty in Linear Causal Models with Equal Variances*. J. Causal Inference 11 (1).
- Strieder, Freidling, Haffner and Drton (2021). *Confidence in Causal Discovery with Linear Causal Models*. UAI 21.