Type-II errors of independence tests can lead to arbitrarily large errors in estimated causal effects: an illustrative example

Workshop UAI 2014

Nicholas Cornia & Joris M. Mooij

University of Amsterdam

27/07/2014

Introduction

- **Task:** Inferring causation from observational data
- **Challenge:** Presence of hidden confounders.
- **Approach:** Causal discovery algorithms based on conditional independence (CIs) tests .
- **Simplest case:** Three random variables, a single CI test (LCD-Trigger setting).
- **Contribution:** Causal predictions are extremely unstable when type II errors arise.

LCD-Trigger Algorithm

Cooper (1997) and Chen et al. (2007).
The following causal model

$$X_1 \overset{\leftrightarrow}{\rightarrow} X_2 \rightarrow X_3$$

is implied by

Prior assumptions

- No Selection Bias
- Acyclicity
- Faithfulness
- X_2, X_3 do not cause X_1

Statistical tests

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- $X_1 \perp\!\!\!\perp X_3 | X_2$

Application of the LCD in biology

Example

Gene expression

Example

Disease Treatment

Linear Gaussian model

For simplicity: linear-Gaussian case.
Structural equations:

$$X_i = \sum_{i \neq j} \alpha_{ij} X_j + E_i \qquad X = AX + E$$

where

$$E \sim \mathcal{N}(0, \Delta) \quad \Delta = diag(\delta_i^2)$$

and $A = \{\alpha_{ij}\}$ is the weighted adjacency matrix of the causal graph ($\alpha_{ij} \neq 0 \iff X_i \to X_j$).

Example

$$X_1 \overset{\alpha_{12}}{\to} X_2 \overset{\alpha_{23}}{\to} X_3$$

$$\begin{cases} X_1 = E_1 \\ X_2 = \alpha_{12} X_1 + E_2 \\ X_3 = \alpha_{23} X_2 + E_3 \end{cases}$$

Then:

$$X \sim \mathcal{N}(0, \Sigma) \qquad \Sigma = \Sigma(A, \Delta)$$

Causal effect estimator

Causal effect of X_2 on X_3:

$$A \ni \alpha_{23} = \frac{\partial}{\partial x_2} \mathbb{E}(X_3 | do(X_2 = x_2))$$

Under the LCD assumptions

$$\mathbb{E}(X_3 | X_2) = \frac{\Sigma_{32}}{\Sigma_{22}}$$

is a valid estimator for the causal effect of X_2 on X_3.

Example

Structural equations (observed)

$$\begin{cases} X_1 = E_1 \\ X_2 = \alpha_{12} X_1 + E_2 \\ X_3 = \alpha_{23} X_2 + E_3 \end{cases}$$

Structural equations after an intervention

$$\begin{cases} X_1 = E_1 \\ X_2 = x_2 \\ X_3 = \alpha_{23} x_2 + E_3 \end{cases}$$

Fundamental question

- What happens to the error in the causal effect estimator if in reality there is a weak dependence $X_1 \not\perp\!\!\!\perp X_3 | X_2$, but we do not have enough data to detect it?

- **Type II error:** Erroneously accepting the null hypotesis of independence in the statistical test $X_1 \perp\!\!\!\perp X_3 | X_2$. Can we still guarantee some kind of bound for the distance

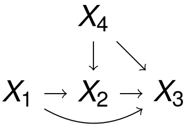
$$|\mathbb{E}(X_3|X_2) - \mathbb{E}(X_3|do(X_2))|$$

From LCD to our model

Starting from the chain

$$X_1 \rightarrow X_2 \rightarrow X_3 \qquad\qquad X_1 \perp\!\!\!\perp X_3 | X_2$$

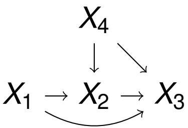
If we consider a possible weak dependence not detected by our test suddenly the causal graph gains complexity



$$X_1 \not\!\perp\!\!\!\perp X_3 | X_2$$

where X_4 is a confounding variable between X_2 and X_3.

True model



Prior assumptions

- No Selection Bias
- Acyclicity
- Faithfulness
- X_2, X_3 do not cause X_1
- No confounders between X_1 and X_2, or X_3, or both (for simplicity)

Statistical tests

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- **A weak conditional dependence** $X_1 \not\perp\!\!\!\perp X_3 | X_2$

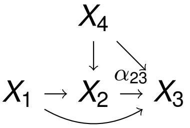
Causal effect estimation error function

Belief

$$X_1 \rightarrow X_2 \overset{\alpha_{23}}{\rightarrow} X_3$$

$$\alpha_{23} = \frac{\Sigma_{32}}{\Sigma_{22}}$$

True model



$$\alpha_{23} \neq \frac{\Sigma_{32}}{\Sigma_{22}}$$

Error in the causal effect estimation function

$$g(A, \Sigma) = \frac{\Sigma_{32}}{\Sigma_{22}} - \alpha_{23}$$

Constraint equations

Proposition

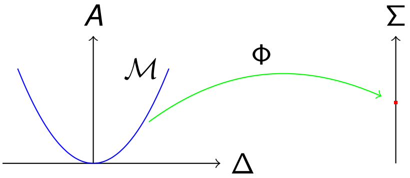
There exists a map

$$\Phi : (A, \Delta) \to \Sigma$$

from the model parameters to the observed covariance matrix that defines a set of polynomial equations.

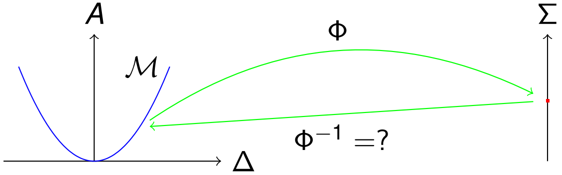
From a geometrical point of view, given Σ

$$(A, \Delta) \in \mathcal{M} \subset \mathbb{R}^9$$



Non-identification of the model parameters

- In our model the map Φ is not injective. Thus, the manifold $\mathcal{M}$ does not reduce to a single point.



- **Nevertheless it is still an interesting question whether the function g is a bounded function on $\mathcal{M}$ or not.**

Main result

Theorem

There exists a map

$$\Psi(\Sigma, \delta_2^2, \delta_3^2, s_1, s_2) = A$$

where s_1, s_2 *are two signs and the* δ_2^2, δ_3^2 *are the variance of the noise sources of* X_2 *and* X_3 *respectively.*

Corollary

It is possible to express the error in the causal effect estimation function g as

$$g(\Sigma, \Psi(\Sigma, \delta_2^2, \delta_3^2, s_1, s_2)) = \underbrace{\frac{\vartheta\Sigma_{12}}{m\Sigma_{22}}}_{\text{small for weak dep.}} \underbrace{+ s_1 s_2 \frac{\sqrt{\det\Sigma - m\delta_3^2}\sqrt{m - \Sigma_{11}\delta_2^2}}{m\sqrt{\delta_2^2}}}_{\text{arbitrarily large}}$$

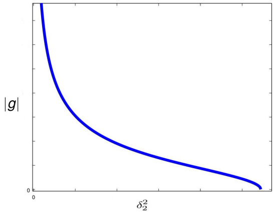
where $\vartheta = \Sigma_{13}\Sigma_{22} - \Sigma_{12}\Sigma_{23}$ *and* $m = \Sigma_{11}\Sigma_{22} - \Sigma_{12}^2$.

Approaching the singularity

Proposition

$$\lim_{\delta_2^2 \to 0} |g| = +\infty$$

$$\forall\ \delta_3^2 \in [0, \det \Sigma / m] \quad (s_1, s_2) \in \{-1, 1\}^2$$



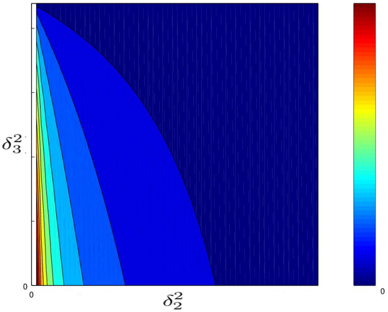
Probabilistic estimation of the error

$$(\delta_2^2, \delta_3^2) \in D(\Sigma) \subset \mathbb{R}^2$$

$$\mathcal{M}_M = \{(\delta_2^2, \delta_3^2) \ : \ |g| \leq M\}$$

If we put a uniform prior on the noise variances

$$Pr(|g| \leq M) = \frac{||\mathcal{M}_M||}{||D(\Sigma)||}$$



- What would be a reasonable prior distribution for δ_2^2, δ_3^2?

Looking for an approximate bound

The causal effect error function g can be optimized over the δ_3^2 parameters, giving a confidence interval for the causal weight α_{23}

$$\alpha_{23} \in [b_-, b_+] \subset \mathbb{R}$$

where

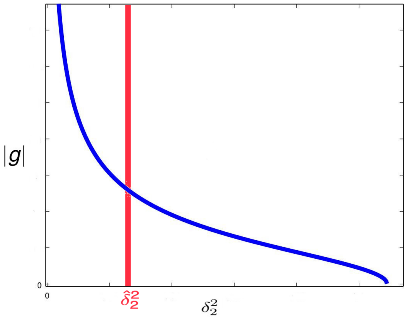
$$b_\pm(\delta_2^2) = \frac{\gamma}{m} \pm \frac{\sqrt{\det \Sigma}\sqrt{m - \Sigma_{11}\delta_2^2}}{m\sqrt{\delta_2^2}}$$

Looking for an approximate bound

Suppose we would have a lower bound

$$\delta_2^2 \geq \hat{\delta}_2^2$$

then this implies an upper bound on $|g|$.



- What would be a practical example where we can assume such a lower bound for the variance δ_2^2?

Conclusions

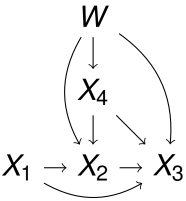
- The causal effect estimation error is sensible to erroneous conclusions in conditional independence tests.
- The result is in accord with Robins et al. (2003), on the lack of uniform consistency of causal discovery algorithms, but through this paper we wish to emphasize this issue on the more practical matter of type II errors.
- In our case it was not possible to identify the model parameters explicitly.

Proposal for future work

- **Bayesian model selection**: What would be a reasonable prior distribution for the model parameters?
- **Bayesian Information Criterion**: Will the BIC still give reasonable results even though the model parameters are not identifiable? Could it deal with irregular or even singular models?

Proposal for future work

- **Adding an "environment" variable**: Might it be reasonable to assume that a part, or most, of the external variability is carried by the covariance between the environment variable W and the other measured ones, including possible confounders?



Thanks for your attention!