



Causal inference: the role of instrumental variables

Nataliya Sokolovska
in collaboration with Pierre-Henri Wuillemin

Laboratory of Computational, Quantitative and Synthetic Biology (CQSB),
Sorbonne University

ECLAIRE

March 24, 2026

Outline

Motivation

Independence of cause and mechanism

Instrumental variables

A symmetric causal inference algorithm

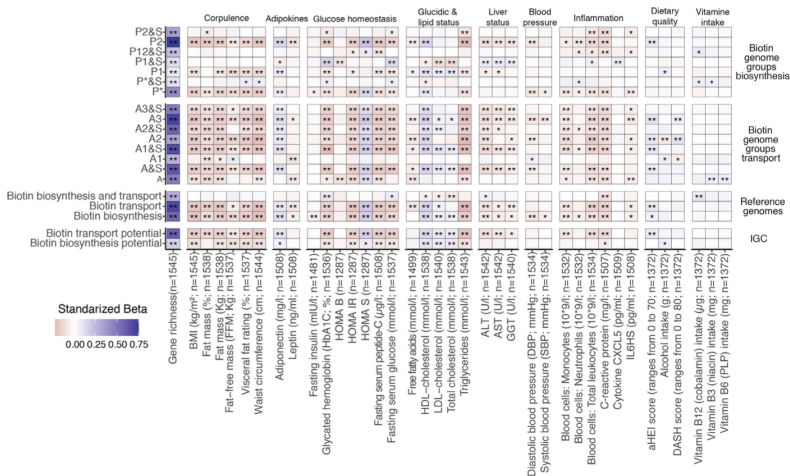
Practical issues, numerical results, and conclusions

Motivation

Causal inference from **observational non-temporal** data
interventions not possible

Motivation

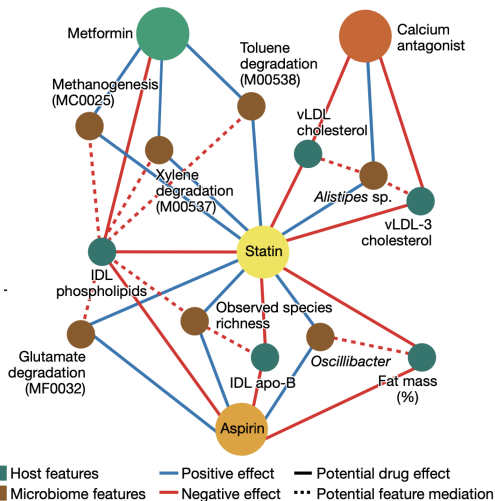
Bivariate relations



E. Belda et al., Impairment of gut microbial biotin metabolism and host biotin status in severe obesity: effect of biotin and prebiotic supplementation on improved metabolism, Gut microbiome, 2022

Motivation

Bivariate relations



S. K. Forslund et al., Combinatorial, additive and dose-dependent drug-microbiome associations. Nature, 2021.

Limits of the state-of-the-art

- ▶ Bivariate case ($X \rightarrow Y$ or $Y \rightarrow X$)
- ▶ Conditional independence-based causal discovery methods (J. Pearl, P. Spirtes): graphs are Markov equivalent; not applicable in the case of two variables, since $X \rightarrow Y$ and $Y \rightarrow X$ are Markov equivalent
- ▶ Methods based on the assumption of independence of cause and mechanism (J. Peters, D. Janzing, B. Schölkopf): if a simple function that fits data exists, then it is likely that it also describes a causal relation in the data
- ▶ Our main goal is to try to reconcile these two modern viewpoints on causal inference

Outline

Motivation

Independence of cause and mechanism

Instrumental variables

A symmetric causal inference algorithm

Practical issues, numerical results, and conclusions

Independent Mechanisms

The principle of independent mechanisms

If $X \rightarrow Y$, then the marginal distribution of the cause $P(X)$ and the conditional distribution of the effect given the cause $P(Y|X)$ are “independent” in the sense that $P(Y|X)$ contains no information about $P(X)$ and vice versa.

Probability distribution as random variables

Probability distributions are considered as random variables; $P(X)$ is a function of $X \in [0, 1]$, they are random variables distributed in $[0, 1]$.

Janzing and Schölkopf, 2010; Janzing et al., 2012, Sgouritsa et al., 2015.

Independent Mechanisms: an example

We have estimated the joint density $p(a, t)$ of the **altitude** A and the **temperature** T on a sample of cities in some country:

$$p(a, t) = p(a|t)p(t) \tag{1}$$

$$= p(t|a)p(a) \tag{2}$$

Independent Mechanisms: an example

We have estimated the joint density $p(a, t)$ of the **altitude** A and the **temperature** T on a sample of cities in some country:

$$p(a, t) = p(a|t)p(t) \tag{1}$$

$$= p(t|a)p(a) \tag{2}$$

$$p(a|t)p(t) \text{ (factorization according to } T \rightarrow A \text{)} \tag{3}$$

$$p(t|a)p(a) \text{ (factorization according to } A \rightarrow T \text{)} \tag{4}$$

from *J. Peters, D. Janzing, B. Schölkopf. Elements of causal inference, 2017*

Independent Mechanisms

$$p(a|t)p(t) \text{ (factorization according to } T \rightarrow A) \quad (5)$$

$$p(t|a)p(a) \text{ (factorization according to } A \rightarrow T) \quad (6)$$

Consider the **effect of interventions**:

- ▶ Intervening on A has changed T
 - ▶ We climb higher
 - ▶ The temperature is lower
- ▶ Intervening on T has not changed A
 - ▶ We do not change the altitude
 - ▶ We build a massive heating system around the city that raises the temperature

Physical Mechanism

If we change the altitude A , we assume that the **physical mechanism**

$$p(t|a) \tag{7}$$

is responsible for producing an average temperature, and changes T .

Physical Mechanism

If we change the altitude A , we assume that the **physical mechanism**

$$p(t|a) \tag{7}$$

is responsible for producing an average temperature, and changes T .

$$\underbrace{p(t|a) \text{ and } p(a)}_{\text{are independent}} \tag{8}$$

- ▶ $p(a)$: locations of cities can be different
- ▶ $p(t|a)$: would apply for different locations

Even a more detailed example

Let us consider the joint distributions of altitude and temperature in **Austria** $p^{\ddot{o}}(a, t)$ and **Switzerland** $p^s(a, t)$.

- ▶ Austrians and Swiss founded their cities in different places, and $p^{\ddot{o}}(a)$ and $p^s(a)$ are different
- ▶ However,
 - ▶ $p^{\ddot{o}}(a, t) = p(t|a)p^{\ddot{o}}(a)$
 - ▶ $p^s(a, t) = p(t|a)p^s(a)$
- ▶ $p(t|a)$ is the same
- ▶ no influence of $p^{\ddot{o}}(a)$ and $p^s(a)$ on $p(t|a)$

from *Peters, 2017*

Independence mechanism: a remark

- ▶ Cities locations depend on a country, since each country has its own urban policy, population density, etc.
 - ▶ Thus, $P(A)$ has at least one **latent variable** which is county C
- ▶ No matter what country is chosen, the physical mechanism $P(T|A)$ holds
 - ▶ The true underlying causal structure is $C \rightarrow A \rightarrow T$
- ▶ A country defines distribution of cities. Having two or more countries leads to a family of distributions.
 - ▶ This **mixture of probability distributions** is **independent** from $P(T|A)$

This remark also explains what is meant under the *independence between probability distributions*

Outline

Motivation

Independence of cause and mechanism

Instrumental variables

A symmetric causal inference algorithm

Practical issues, numerical results, and conclusions

Independence of probability distributions and instrumental variables

- ▶ $P(X, Y)$: the joint distribution of X, Y in a population $\mathcal{P}$
- ▶ $Q(X, Y)$: the joint distribution of X, Y in a population $\mathcal{Q}$
- ▶ If $X \rightarrow Y$:
 - ▶ $P(X, Y) = P(X) \cdot P(Y|X)$, $Q(X, Y) = Q(X) \cdot Q(Y|X)$,
 - ▶ $P(Y|X) = Q(Y|X)$, and on the contrary, $P(X|Y) \neq Q(X|Y)$

More generally, for all mixed populations between $\mathcal{P}$ and $\mathcal{Q}$, and for all mixtures $Q_\lambda = \lambda P + (1 - \lambda)Q$ with $\lambda \in [0, 1]$:

$$\forall \lambda \in [0, 1], Q_\lambda(X) \perp\!\!\!\perp Q_\lambda(Y|X) \quad (9)$$

$$\iff Q_\lambda(Y|X) = P(Y|X). \quad (10)$$

The example (temperature and altitude) revisited

- ▶ $P(X)$ and $Q(X)$: the **distributions of cities locations** in two different countries
- ▶ $P(Y|X)$ be a physical mechanism predicting weather given location
- ▶ λ is the hyper-parameter controlling the **proportion of observations** in each country
- ▶ λ , $P(X)$, and $Q(X)$ are independent from $P(Y|X)$

Instrumental variables

$$X = \alpha_0 + \alpha I_X + \delta U + \epsilon_X, \quad (11)$$

$$Y = \beta_0 + \beta X + \gamma U + \epsilon_Y, \quad (12)$$

- ▶ X , Y , and I_X are observed, and U is an unobserved confounder
- ▶ ϵ_X and ϵ_Y are noise terms, independent of each other
- ▶ U , ϵ_X , and ϵ_Y have mean zero
- ▶ The common cause U can be absent
- ▶ We do not assume that U exists when modelling dependencies between X and Y .

The instrumental variable is a source of variation for X , and it only influences Y through X . Studying how X and Y respond to perturbations of I_X can help to deduce how X influences Y .

Instrumental variables and the independence mechanism

Theorem 1

Let X and Y be two correlated random variables, and they do not have any common cause. We assume that either X causes Y or visa versa. If there exists a random variable I_X such that $I_X \rightarrow X$, and if $I_X \perp\!\!\!\perp Y|X$, then we are able to infer causality and decide that $X \rightarrow Y$.

Proof.

Several directed acyclic graphs (DAGs) may be Markov equivalent (Pearl, 2009). We assume that once an essential graph is found, the directed arcs of this graph are interpreted causally. Under the assumption that $I_X \rightarrow X$, and if $I_X \perp\!\!\!\perp Y|X$, the only possible directed graph is $I_X \rightarrow X \rightarrow Y$. In case where $I_X \not\perp\!\!\!\perp Y|X$, we get $I_X \rightarrow X \leftarrow Y$.



Instrumental variables and the independence mechanism

Theorem 2

If the true causal structure is $I_X \rightarrow X \rightarrow Y$, and X and Y do not have any common cause, then $P(Y|X)$ does not contain any information about $P(X)$ and vice versa, however, $P(X|Y)$ and $P(Y)$ are not independent.

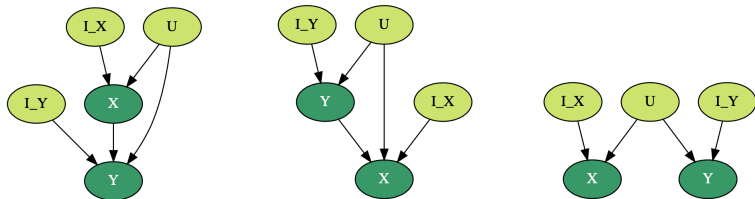
Proof.

Assume that $I_X \perp\!\!\!\perp Y|X$. Let us consider the relation between $P(Y|X)$ and $P(X)$. $P(Y|X)$, $P(X|Y)$, $P(X)$, and $P(Y)$ are random variables, and we can write: $P(Y|I_X, X) = P(Y|X)$. Note that we do not have $P(X)$ when we express $P(Y|X)$ for $I_X \rightarrow X \rightarrow Y$. Let us consider the relation between $P(X|Y)$ and $P(Y)$ for the same graphical structure. We get:

$$P(X|Y) = \frac{P(Y|X)P(X|I_X)}{P(Y)}, \quad (13)$$

where the form of the nominator is due to the fixed dependencies $I_X \perp\!\!\!\perp Y|X$. From eq. (13), we clearly see that $P(X|Y)$ is not independent from $P(Y)$ for this graphical structure. □

Structures with instrumental variables



The dark nodes: observed; the light coloured nodes: latent.

Assumption

In case of observational non-temporal data, if I_X exists such that $I_X \rightarrow X$, and if I_Y exists such that $I_Y \rightarrow Y$, and if the random variables X and Y are correlated, then we assume that it is impossible that both $I_X \perp\!\!\!\perp Y|X$ and $I_Y \perp\!\!\!\perp X|Y$ hold.

Outline

Motivation

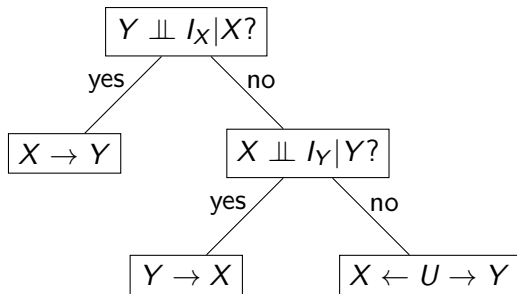
Independence of cause and mechanism

Instrumental variables

A symmetric causal inference algorithm

Practical issues, numerical results, and conclusions

Our Causal Inference Algorithm



If a conditional independence test tells that $Y \perp\!\!\!\perp I_X | X$ is true, then $X \rightarrow Y$ is inferred, otherwise we test whether $X \perp\!\!\!\perp I_Y | Y$, and if it is true, then Y causes X . The last case where X and Y are correlated but both $Y \perp\!\!\!\perp I_X | X$ and $X \perp\!\!\!\perp I_Y | Y$ are false, let us conclude that there is a common hidden cause U , and $Y \leftarrow U \rightarrow X$.

Outline

Motivation

Independence of cause and mechanism

Instrumental variables

A symmetric causal inference algorithm

Practical issues, numerical results, and conclusions

How to apply it in practice?

Instrumental variables approximation

- ▶ If $I_X \rightarrow X \rightarrow Y$, I_X does not contain information about Y , and I_X has to be constructed from X only. If $X \rightarrow Y \leftarrow I_Y$, I_Y is not independent from X , and I_Y has to contain information about both X and Y .
- ▶ Construct the IV from X , Y (I_{X_X} , I_{Y_Y}), and from both ($I_{X_{XY}}$, $I_{Y_{YX}}$); test what is more relevant.

Intuition

- ▶ If Y influences clustering of X less, than X impacts clustering of Y , then we apply I_X constructed from X only, and I_Y is constructed from X and Y . And visa versa. A criterion: test $if(\text{dist}(I_{X_X}, I_{X_{XY}}) < \text{dist}(I_{Y_Y}, I_{Y_{YX}}))$
- ▶ Similar to methods based on the *Kolmogorov complexity* and the *Minimum Description Length*: chose a simpler model.

Data: synthetic and real

► Simulated data

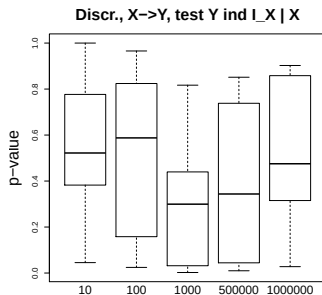
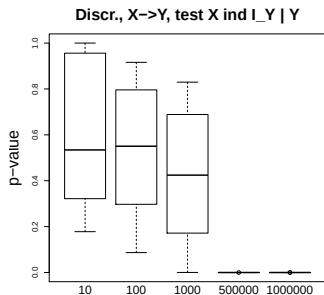
- I_X and I_Y , X and Y , and the hidden variable U are generated
- bnlearn R package
- the conditional independence tests
- discrete setting with binary variables, the asymptotic mutual information independence test `ci.test(test='mi')`
- continuous setting with Gaussian variables, the exact t-test for Pearson's correlation `ci.test(test='cor')`
- conditional independence tests from the bnlearn R package return “big” p-values if variables are conditionally independent, and the p-values are small (with an arbitrary threshold 0.05) for dependent variables

► Real data: Cause-effect pairs

<http://webdav.tuebingen.mpg.de/cause-effect>

- 100 pairs from different domains, ground truth is provided (some pairs excluded)

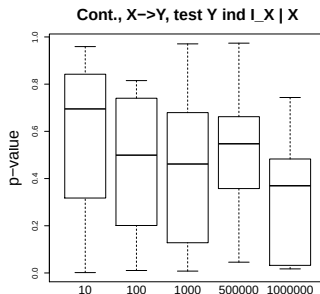
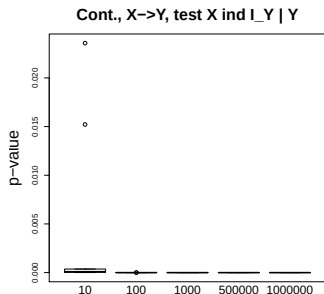
Simulated discrete data, ground truth $I_X \rightarrow X \rightarrow Y \leftarrow I_Y$



- ▶ The p-values of an asymptotic mutual information test (for the discrete case).
- ▶ Statistical hypothesis: H_0 : variables are independent, H_1 : dependent variables
- ▶ p-value < 0.05 : reject H_0
- ▶ $X \not\perp I_Y | Y$, $Y \perp I_X | X$

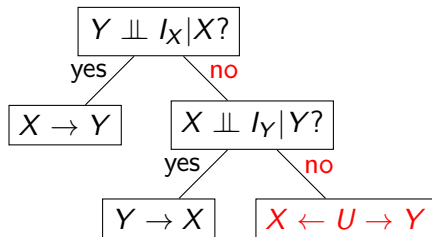
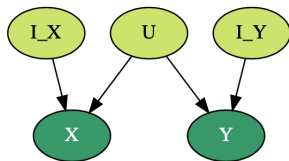
Simulated continuous data, ground truth

$$I_X \rightarrow X \rightarrow Y \leftarrow I_Y$$

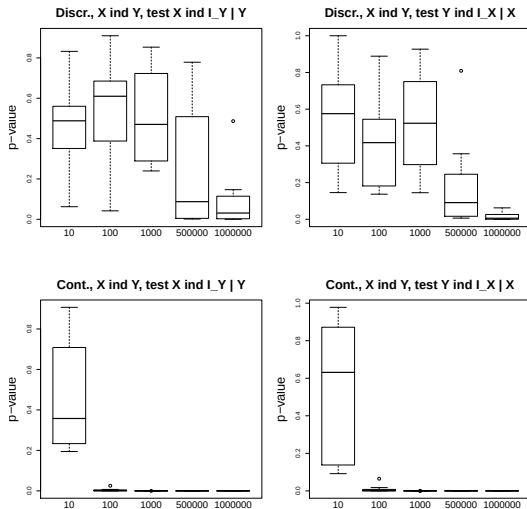


- ▶ The p-values of the exact t-test for Pearson's correlation
- ▶ Statistical hypothesis: H_0 : variables are independent
 H_1 : dependent variables
- ▶ $p\text{-value} < 0.05$: reject H_0
- ▶ $X \not\perp I_Y | Y$, $Y \perp I_X | X$

Ground truth: $I_X \rightarrow X \leftarrow U \rightarrow Y \leftarrow I_Y$

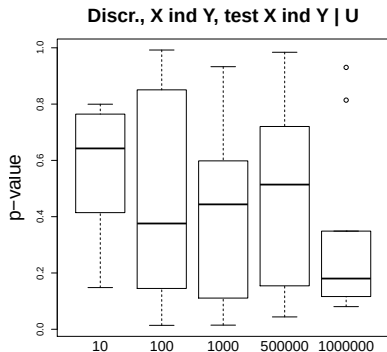
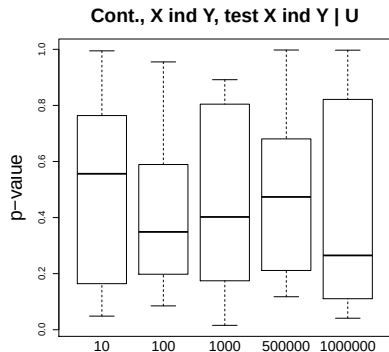


Simulated data, ground truth: $X \leftarrow U \rightarrow Y$



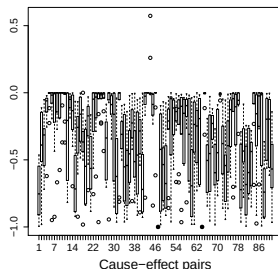
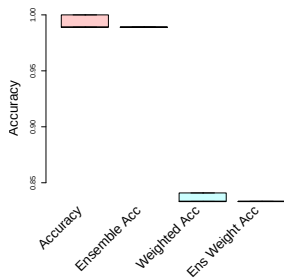
Small p-values, H_0 rejected, $X \not\perp I_Y | Y$ and $Y \not\perp I_X | X$.

Simulated data, ground truth: $X \leftarrow U \rightarrow Y$



The results of the conditional independence tests for $X \perp\!\!\!\perp Y|U$ for continuous (on the left) and discrete (on the right) data. On the x-axis: the number of observations.

Cause-effect pairs (Tübingen dataset)



- ▶ Accuracy on the Cause-Effect Benchmark.
- ▶ Difference between the test statistics $X \perp\!\!\!\perp I_Y|Y$ and $Y \perp\!\!\!\perp I_X|X$.
- ▶ $X \perp\!\!\!\perp I_Y|Y$ “small” p-values, $Y \perp\!\!\!\perp I_X|X$ “big” p-values
- ▶ $X \not\perp\!\!\!\perp I_Y|Y$, $Y \perp\!\!\!\perp I_X|X$, $X \rightarrow Y$.

Conclusions and Perspectives

► Conclusions

- Nataliya Sokolovska, Pierre-Henri Wuillemin. *The Role of Instrumental Variables in Causal Inference Based on Independence of Cause and Mechanism*. Entropy 23(8), 2021
- We challenged to bring together two principle research avenues in causal inference: causal inference using conditional independence and methods based on the postulate of independence of cause and mechanism
- Our method is simple, a number of its versions can be explored

► Perspectives

- Extensions to more complex graphs
- Huge applications (modelling gene interactions)
- Novel metrics for conditional independence