Seminar: Causal Inference

Winter Semester 2025/26

`AG – Velten`

UNIVERSITÄT HEIDELBERG

Introduction

Important Information

Time and location:

1400 hrs c.t., 0/200, Mathematikon

Dates:

kw50, kw51 (Dec2025)

kw3, kw5 (Jan 2026)

contact: `nikolai.koehler`[1], `purusharth.saxena`[2], `jan.sprengel`[1]

[1]`@cos.uni-heidelberg.de`

[2]`@iwr.uni-heidelberg.de`

Important Information / Participation

- Each appointment will have a maximum of 2 presentations.
- Groups of up to two people can be formed for longer or more complex papers.
- Please send your slides for review at least one week before your presentation
- A seminar report[1] is expected within three weeks after your presentation (Note: pro-seminar doesn't require a seminar report)

[1] maximum 5 pages excluding references

Important Information / Participation

Circular Feedback

Important Information / Participation

Circular Feedback

- Each group is expected to provide feedback to one other group.
- The quality of this feedback will account for *10% of the grade* of the group giving the feedback.
- The grade of the *group receiving* the feedback *will not be affected*
- Feedback form will be provided.

Deliverables

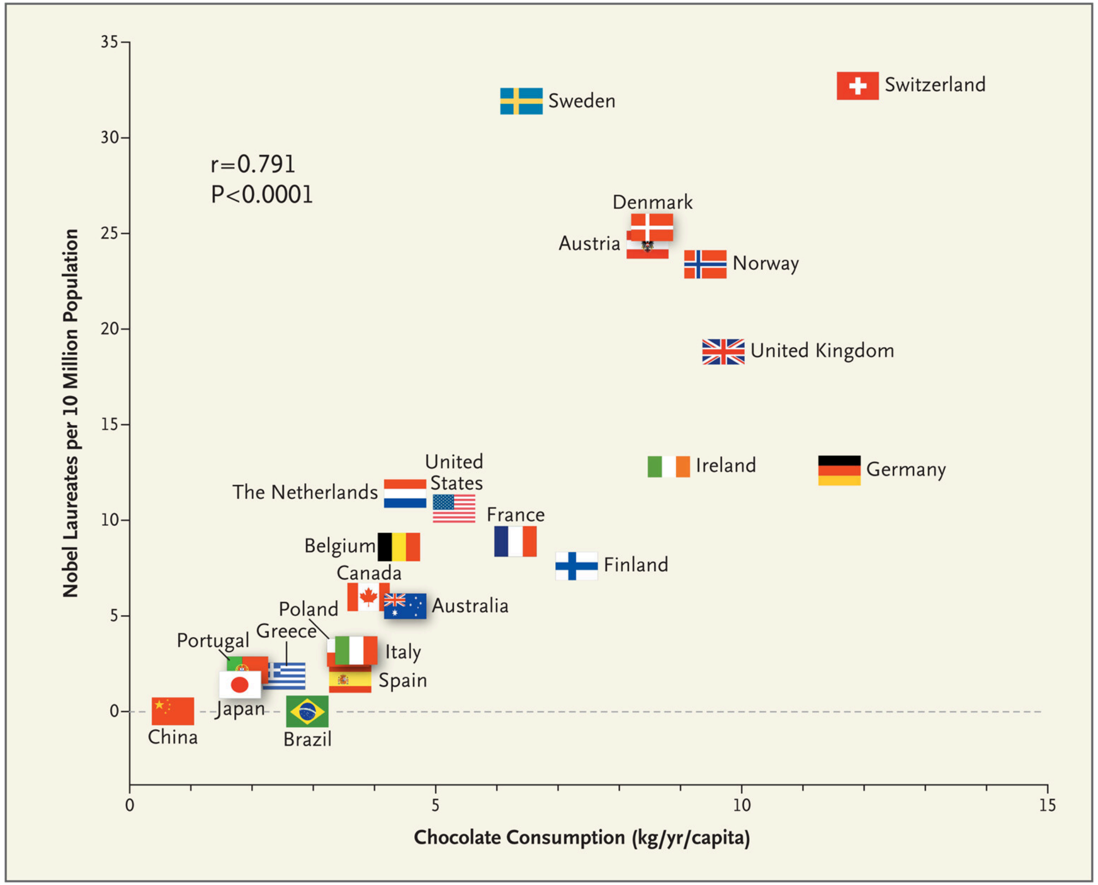
- A cohesive presentation on the selected paper or chapter, lasting 45 minutes, followed by 10 minutes for discussion.
- Seminar report (if applicable)
- *Active participation is mandatory since there are only 4 dates for presentations. i.e don't disappear before/after your presentation ;)*

Motivation

Causality / Correlation $\neq$ Causation

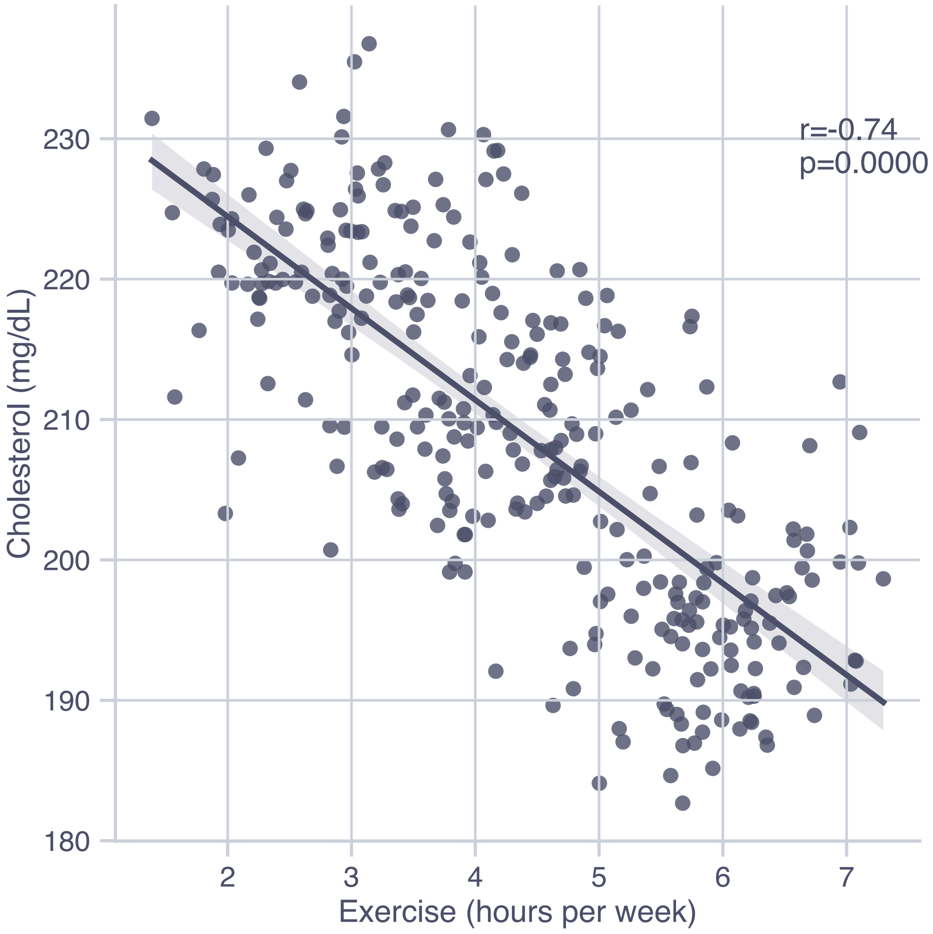
https://x.com/simonmaechling/status/1866423324861206615

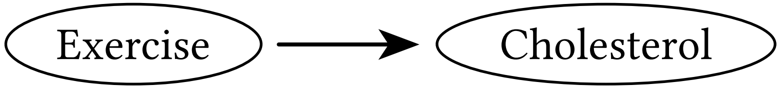
Causality / Correlation ≠ Causation



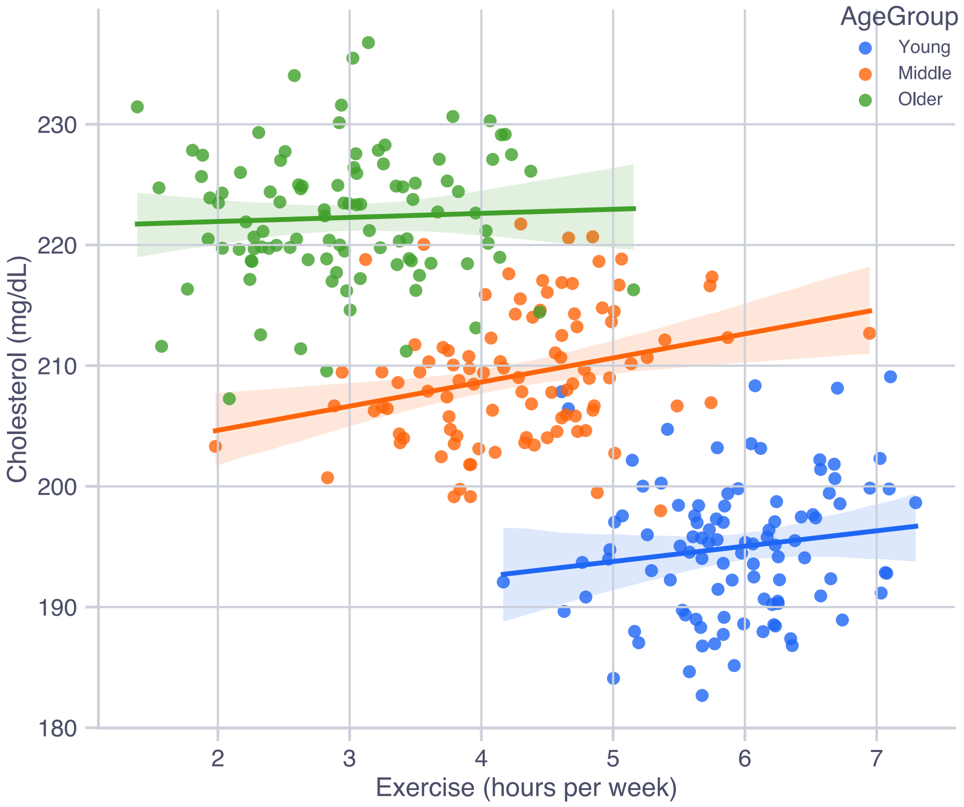
Messerli, 2012 [1]

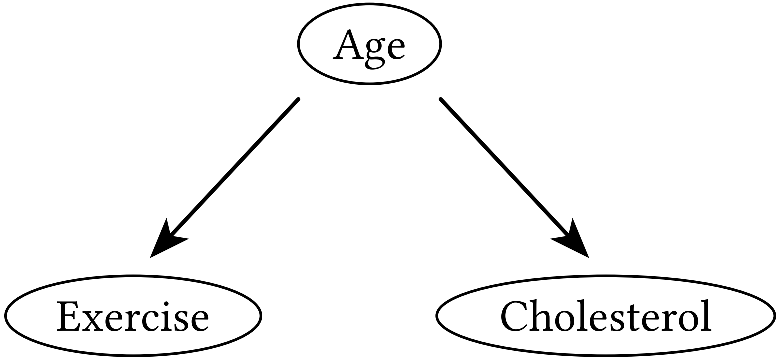
Causality / Synthetic Example I



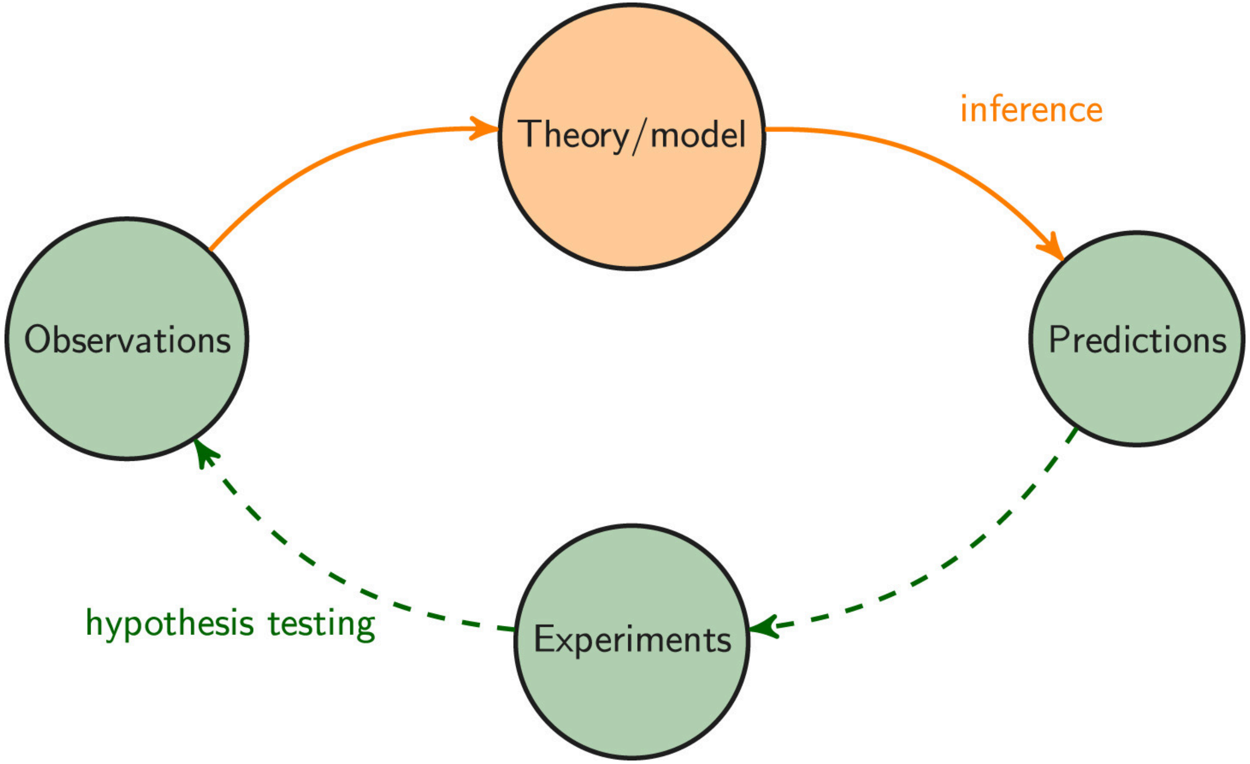


Causality / Synthetic Example II

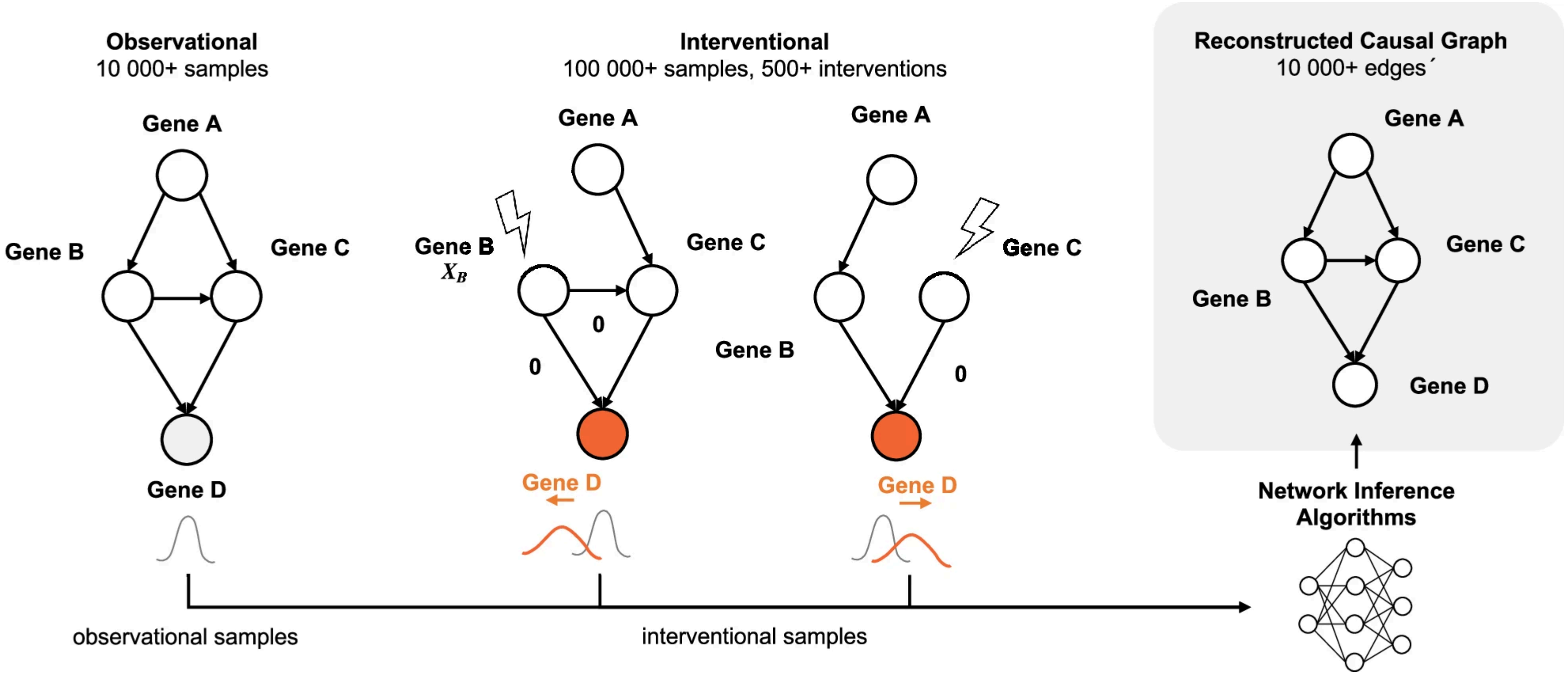




Causality / Causality from Data

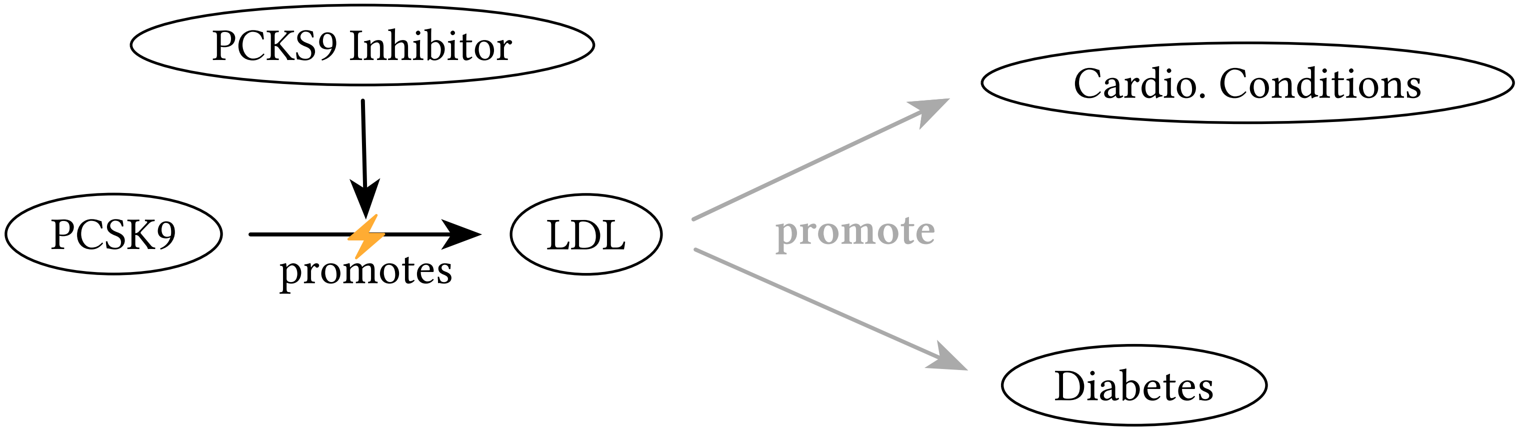


Camps-Valls et al., 2023 [2]



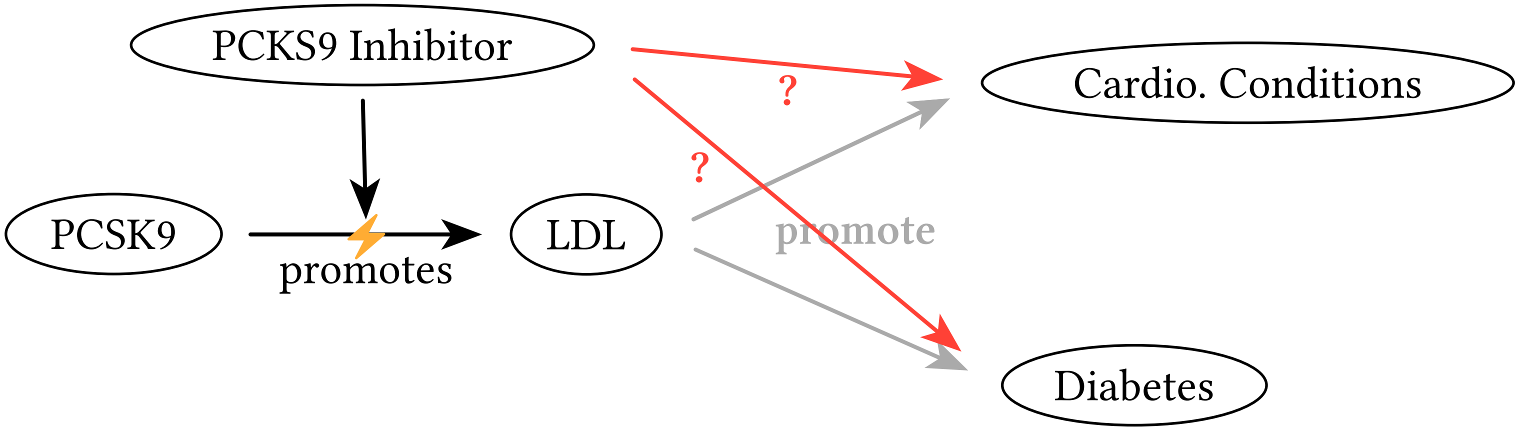
Chevalley et al., 2025 [3]

Causality / Medical Example - Treatment Effects



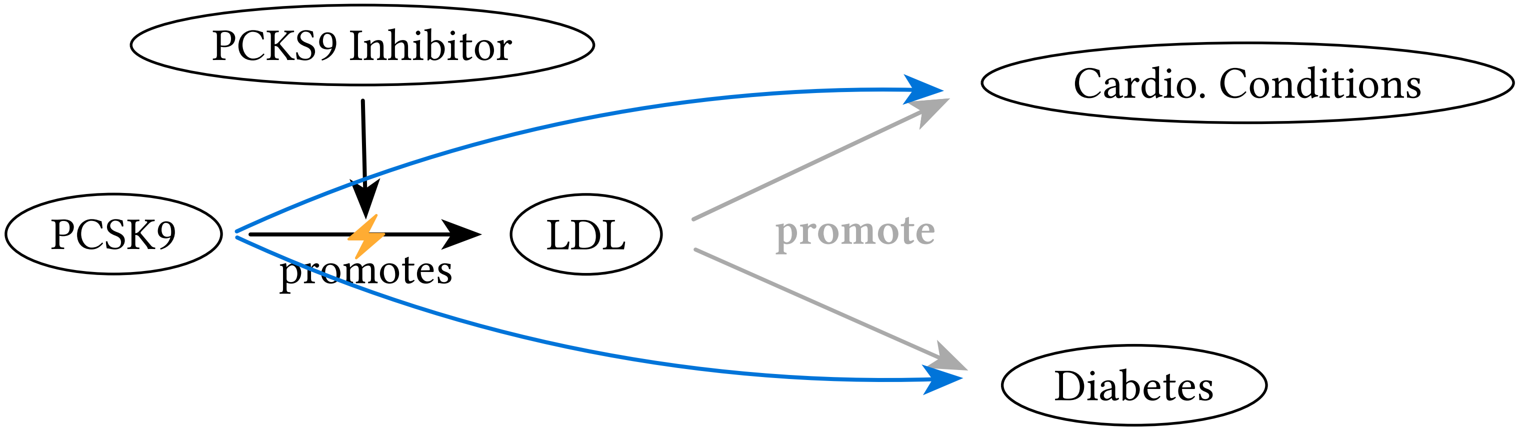
Roughly from Ference et al., 2016 [4]

Causality / Medical Example - Treatment Effects



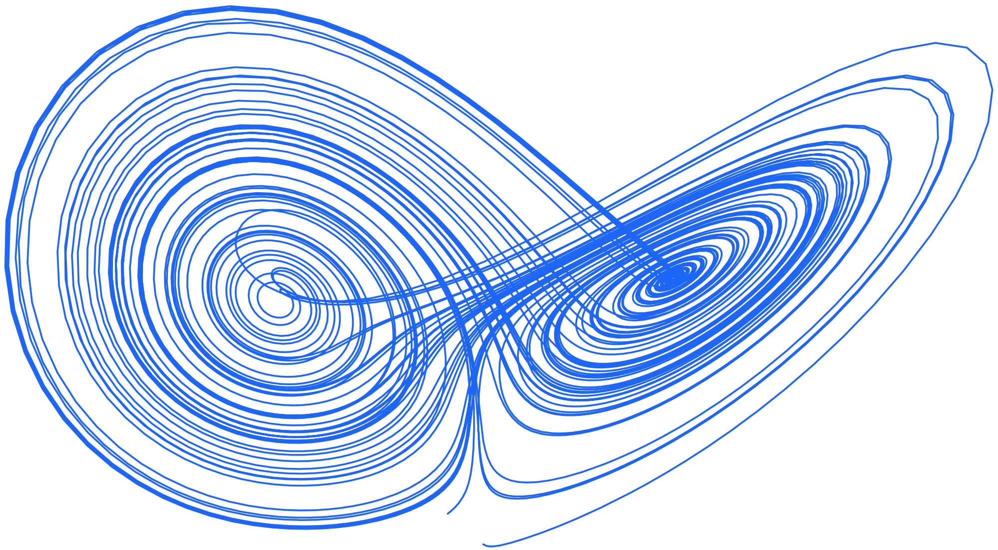
Roughly from Ference et al., 2016 [4]

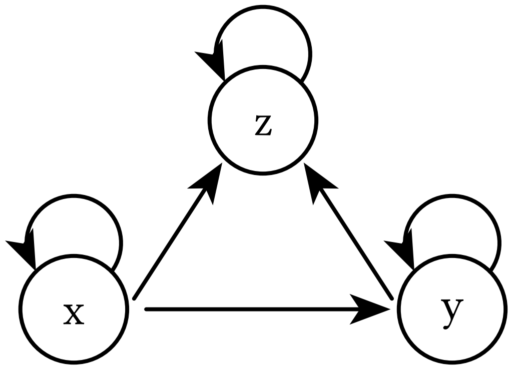
Causality / Medical Example - Treatment Effects



Roughly from Ference et al., 2016 [4]

Causality / Physics Example - Lorenz System





$$\frac{d}{dt}x = \sigma(y - x)$$
$$\frac{d}{dt}y = x(\rho - z) - y$$
$$\frac{d}{dt}z = xy - \beta z$$

Typical assumption: i.i.d.

Example case: recommender systems

1. I buy a laptop on amazon - what do I get recommended?

Typical assumption: i.i.d.

Example case: recommender systems

1. I buy a laptop on amazon - what do I get recommended?
2. A case, a bag, ...

Causality / Causal Machine Learning

Typical assumption: i.i.d.

Example case: recommender systems

1. I buy a laptop on amazon - what do I get recommended?
2. A case, a bag, ...
3. I buy said case, bag, ...

Is this buy still i.i.d?

Typical assumption: i.i.d.

Example case: recommender systems

1. I buy a laptop on amazon - what do I get recommended?
2. A case, a bag, ...
3. I buy said case, bag, ...

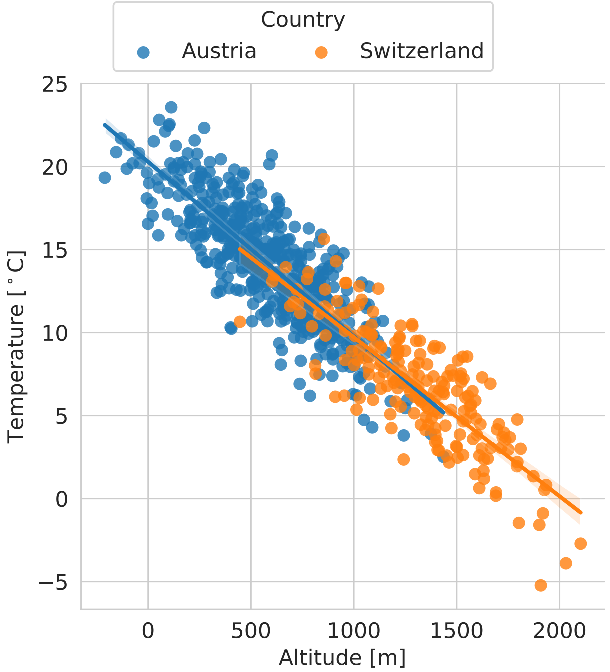
Is this still i.i.d?
$\Rightarrow$ no, recommendation interferes with the decision ("intervention")

Additionally, knowing whether laptops or accessories are bought reduces uncertainty symmetrically $\Rightarrow$ cause-effect-relationship lost

Formalising Causality

Independent Mechanisms

> **Example:** Altitude a and temperature t of cities in the Alps



A joint distribution of cities with a and t can be observed:

$$(a, t) \sim p(a, t)$$

We can factorize this in two ways:

$$p(a, t) = p(t|a)p(a) \qquad \text{temperature as function of altitude}$$

$$= p(a|t)p(t) \qquad a \text{ as function of } t$$

Independent Mechanisms

> **Independent Causal Mechanisms Principle:** For a generative process governing the variables of a system:
> 1. It is composed of individual autonomous mechanisms
> 2. The conditional distribution of each variable given it's causes does not influence those of others
>
> *Also called independence of cause and mechanism*

Independent Mechanisms

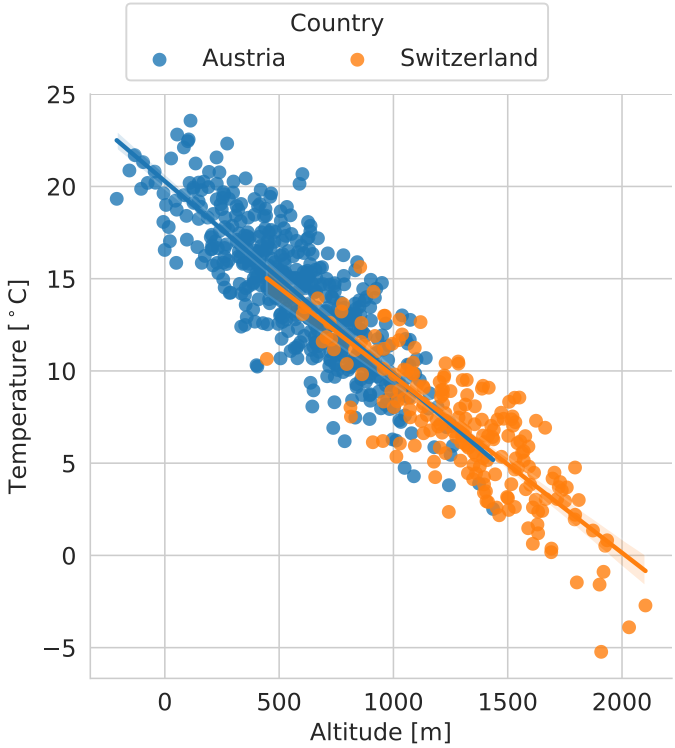
> **Example:** Altitude a and temperature t of cities in the Alps

In this example there is an (physical) mechanism $a \to t$ that indicates meaningful factorization:

$$p(a,t) = p(t|a)p(a)$$

Effect Mechanism Cause

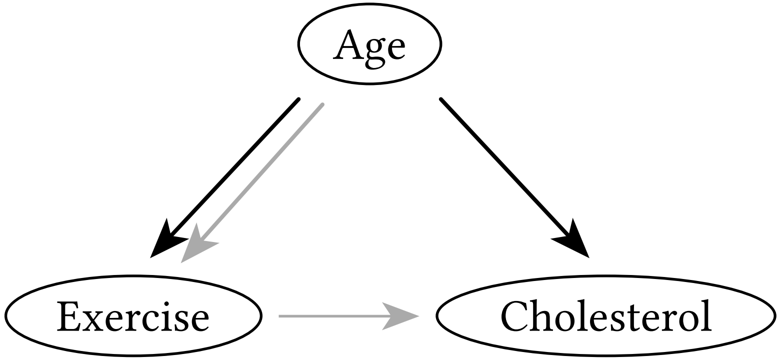
$$p(a,t) = p(a|t)p(t)$$



Independent Mechanisms

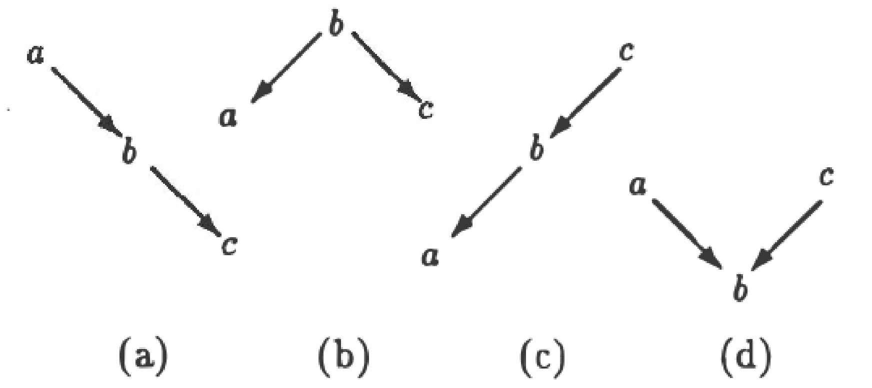
Causal factorization:
effect as conditional of cause

Here:
Exercise: $\text{ex} \sim p_{\text{ex}}(\text{ex} \mid \text{age})$
Cholesterol: $\text{ch} \sim p_{\text{ch}}(\text{ch} \mid \text{age})$



$$p(\text{age}, \text{ex}, \text{ch}) = p(\text{ex}|\text{age})p(\text{ch}|\text{age})p(\text{age})$$
$$= p(\text{ch}|\text{ex})p(\text{ex}|\text{age})p(\text{age})$$

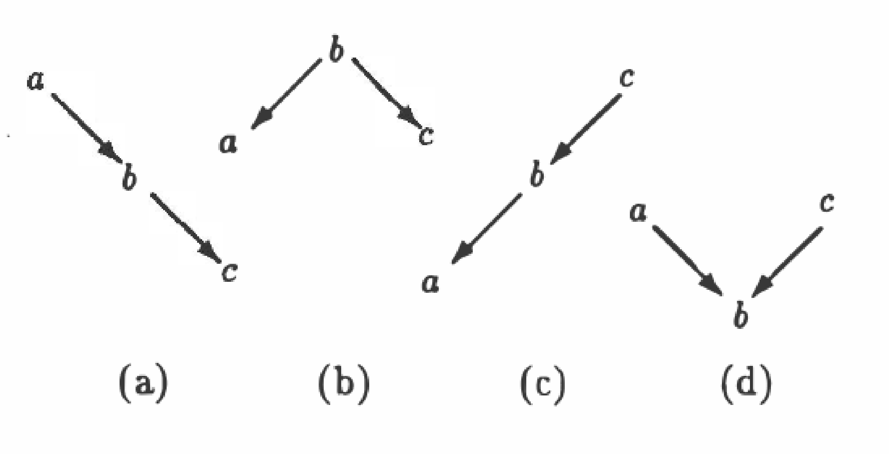
Causal Models



Adapted from Verma and Pearl, 1990 [5]

Q: Is there any observable difference?

Causal Models



Adapted from Verma and Pearl, 1990 [5]

Q: Is there any observable difference?

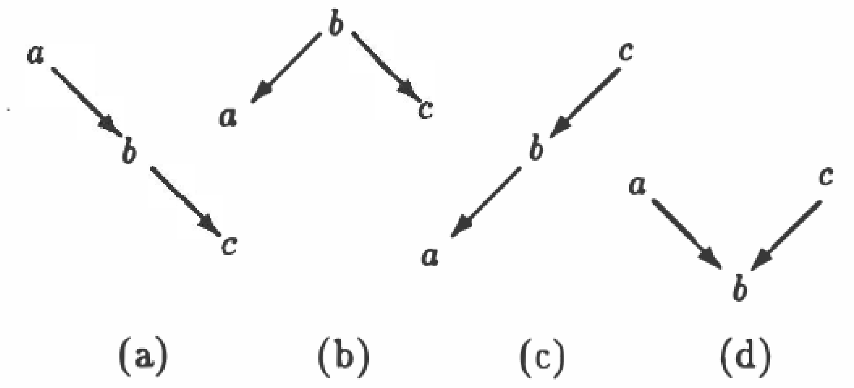
(a) $P(c|b), P(b|a), P(a)$

(b) $P(a|b), P(b|c), P(c)$

(c) $P(a|b), P(c|b), P(b)$

(d) $P(b|a,c), P(a), P(c)$

Causal Models



Adapted from Verma and Pearl, 1990 [5]

Q: Is there any observable difference?

(a) $P(c|b), P(b|a), P(a)$

(b) $P(a|b), P(b|c), P(c)$

(c) $P(a|b), P(c|b), P(b)$

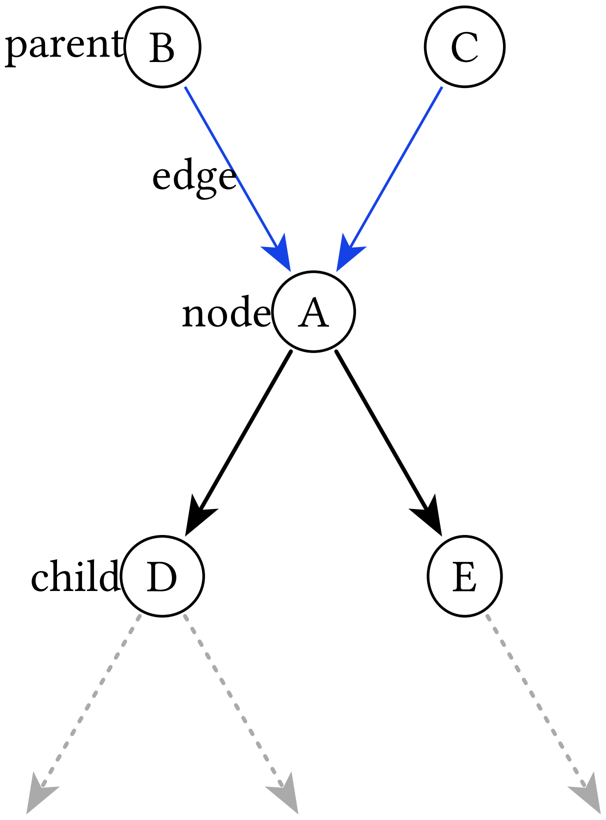
(d) $P(b|a, c), P(a), P(c)$

Causal Models / Graph Terminology

Directed Graphs: Represent relationships between variables in a causal model.

Useful terms:

- directed acyclic graph (DAG)
- node, edge
- parent, child
- ancestors, descendants
- **v-structure/collider**



Collect a set of variables and the mechanisms that generate them in a *Causal Model*:

Structural Causal Model (SCM). A SCM $\mathfrak{C}$ consists of a causal graph $\mathcal{G}$ and a set of (structural) assignments

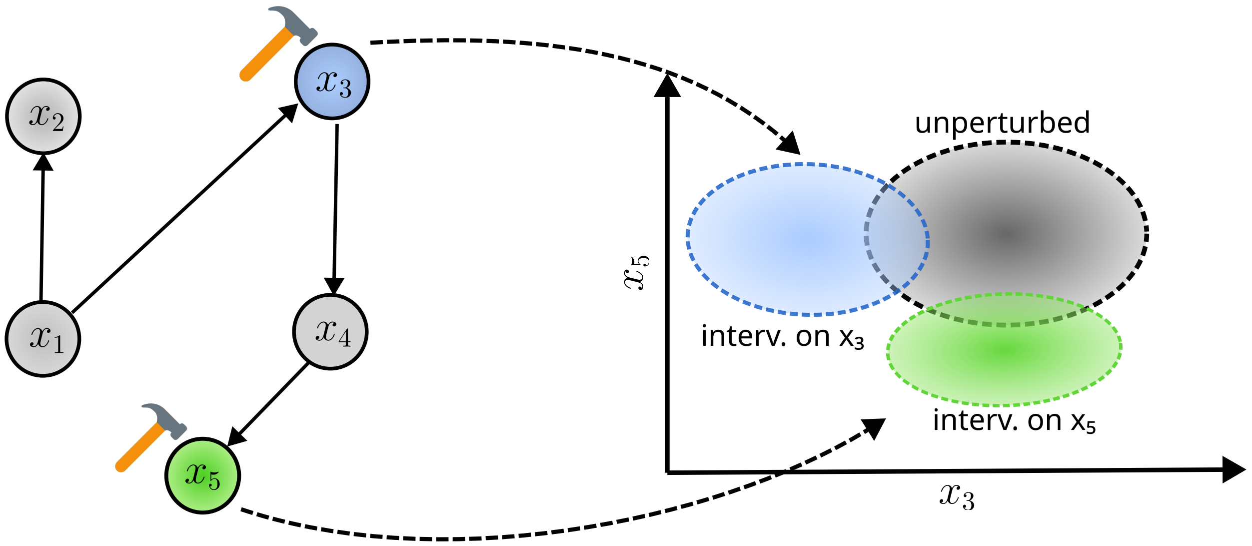
$$x_i := f_i\left(X_{\mathrm{PA}_i}, N_i\right), \qquad i = 1, \ldots, d$$

where X_{Pa_i} are the parents of variable x_i in $\mathcal{G}$, N_i are independent noise variables.

Typically the SCM then entails a joint probability distribution $p(x)$, which can be factorized by each variable x_i.

Causal Models / Interventions

Intervention: What happens if one variable is changed?



Interacting Variables **Induced Distributions - changed by Interventions**

> **Counterfactual:** What would have happened *in a given observation* if one variable had been different?

For a SCM $\mathfrak{C}$ with variables x_i, $i = 1, \ldots, d$, if we have a **given observation** X:

$$\text{Set } X_i := 2 \text{ and see how the other variables } X_{j \neq i} \text{ change.}$$

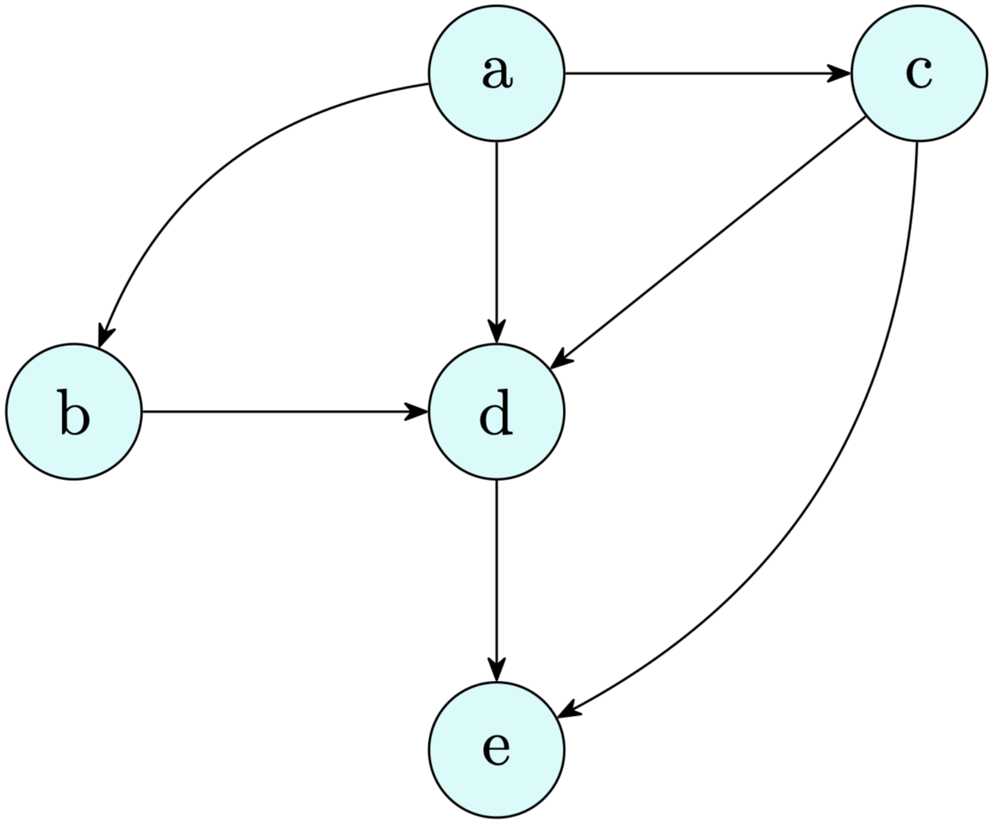
Causal Models / Outlook: Identifiability

Goal: Identify the correct causal model for a system.
Obstacle: Different SCMs can entail identical distributions.

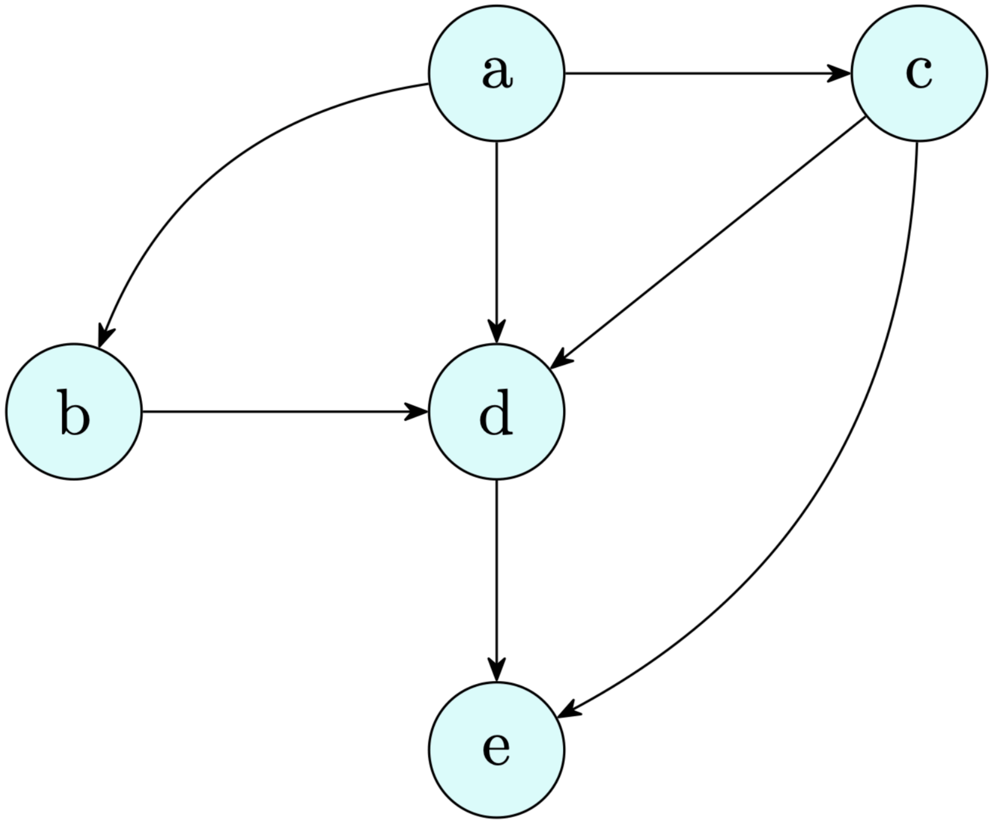
Approaches that can be taken:

- restrictions on allowed models:
  - functional dependency: e.g. linear functions
  - noise type: e.g. additive Gaussian noise,…
  - …
- leverage data from different experimental contexts (i.e. interventions)

Revisited: Structural Causal Model



Revisited: Structural Causal Model



$$\begin{aligned}
a &:= \varepsilon_a \\
b &:= f_b(a) &&+ \varepsilon_b \\
c &:= f_c(a) &&+ \varepsilon_c \\
d &:= f_d(a, b, c) &&+ \varepsilon_d \\
e &:= f_e(c, d) &&+ \varepsilon_e
\end{aligned}$$

Do - Calculus

Interventional Distribution: $\mathbb{P}(b \mid \mathrm{do}(a))$

For the previous graph we have:

$$P(e \mid \mathrm{do}(a)) = \sum_{b,c,d} P(e \mid d, c) \cdot P(d \mid a, b, c) \cdot P(b \mid a) \cdot P(c \mid a)$$

say, given intervention $\mathrm{do}(a = 1)$, what is the probability that $e = 1$?

ODE / SDE

Let $X \to Y$ be the changes in X directly influencing the changes in Y.

[1] describes evolution not causation

ODE / SDE

Let $X \rightarrow Y$ be the changes in X directly influencing the changes in Y.
Let a general ODE with n-variables be

$$\dot{x} = f_i(x_1, x_2, ..., x_n)^1$$

[1] describes evolution not causation

ODE / SDE

Let $X \to Y$ be the changes in X directly influencing the changes in Y.
Let a general ODE with n-variables be

$$\dot{x} = f_i(x_1, x_2, ..., x_n)^1$$

Specify parent set: $\mathrm{Pa}(i) \subset \{1, ..., n\}$

And restrict the structural equation to :

$$\dot{x} = f_i\left(x_{\mathrm{Pa}(i)}\right)$$

Variables on RHS are direct caues of x_i

[1] describes evolution not causation

ODE / SDE

A causal intervention replaces only the structural equation of the intervened variable.

$$do(x_j = c) : \dot{x}_j = 0$$

Removes incoming edges to x_j

Tl;dr: An ODE / SDE sufficiently shows causal relationships between variables [6], [7]

Markov Equivalence Class

Equivalence Class: A set objects that are treated as the same because they satisfy some equivalence relation[1]

[1] eg: reflexive, symmetric, or transitive.

[2] X, Y are conditionally independent of Z if $P(X, Y|Z) = P(X|Z)P(Y|Z)$

Markov Equivalence Class

Equivalence Class: A set objects that are treated as the same because they satisfy some equivalence relation[1]

Markov Equivalence: The set of all DAGs that encode the same conditional independences[2] and therefore look identical from observational data.

[1] eg: reflexive, symmetric, or transitive.

[2] X, Y are conditionally independent of Z if $P(X, Y|Z) = P(X|Z)P(Y|Z)$

Markov Equivalence Class

Equivalence Class: A set objects that are treated as the same because they satisfy some equivalence relation[1]

Markov Equivalence: The set of all DAGs that encode the same conditional independences[2] and therefore look identical from observational data.

A *Markov equivalence class* is the collection of all causal DAGs that imply exactly the same conditional independence relations

[1] eg: reflexive, symmetric, or transitive.

[2] X, Y are conditionally independent of Z if $P(X, Y|Z) = P(X|Z)P(Y|Z)$

Time based measurement

- Dynamic Causal Modeling
  - Mostly from Neuroscience as in [8]
    - activity of n brain regions with fMRI
  - Similar in setting to ODE model
- Granger Causality
  - Connecting two time series

Seminar Literature

Papers / Basics | Introductory

- Equivalence and Synesis of Causal Models

Papers / Basics | Introductory

- Equivalence and Synesis of Causal Models
- Causal models for dynamical systems

Papers / Machine Learning

- Causality for Machine Learning

[1]proseminar

Papers / Machine Learning

- Causality for Machine Learning
- Towards Causal Representation Learning
- Causal Machine Learning: A Survey and Open Problems[1]

[1] proseminar

Papers / Causal Discovery + Inference

- Differentiable Causal Discovery from Interventional Data

Papers / Causal Discovery + Inference

- Differentiable Causal Discovery from Interventional Data
- Amortized Inference for Causal Structure Learning

Papers / Causal Discovery + Inference

- Differentiable Causal Discovery from Interventional Data
- Amortized Inference for Causal Structure Learning
- Causal inference using invariant prediction: identification and confidence intervals

- Optimal Transport and Wasserstein Distances for Causal Models

Papers / Causal Optimal Transport

- Optimal Transport and Wasserstein Distances for Causal Models
- A primer on optimal transport for causal inference with observational data

Papers / Theory

- Stationary Diffusions
- Quantum Causal Modelling
- A Measure-theoretic axiomatization of Causality

Appendix

References

[1] F. H. Messerli, “Chocolate Consumption, Cognitive Function, and Nobel Laureates,” *New England Journal of Medicine*, vol. 367, no. 16, pp. 1562–1564, Oct. 2012, doi: 10.1056/NEJMon1211064.

[2] G. Camps-Valls *et al.*, “Discovering Causal Relations and Equations from Data,” no. arXiv:2305.13341. arXiv, May 2023. doi: 10.48550/arXiv.2305.13341.

[3] M. Chevalley, Y. H. Roohani, A. Mehrjou, J. Leskovec, and P. Schwab, “A Large-Scale Benchmark for Network Inference from Single-Cell Perturbation Data,” *Communications Biology*, vol. 8, no. 1, p. 412, Mar. 2025, doi: 10.1038/s42003-025-07764-y.

[4] B. A. Ference *et al.*, “Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes,” *New England Journal of Medicine*, vol. 375, no. 22, pp. 2144–2153, Dec. 2016, doi: 10.1056/NEJMoa1604304.

[5] T. Verma and J. Pearl, “Equivalence and Synthesis of Causal Models,” in *Probabilistic and Causal Inference: The Works of Judea Pearl*, 1st ed., New York, NY, USA: Association for Computing Machinery, 2022, pp. 221–236. doi: 10.1145/3501714.3501732.

[6] J. M. Mooij, D. Janzing, and B. Schölkopf, “From Ordinary Differential Equations to Structural Causal Models: the deterministic case.” 2013.

[7] A. Sokol and N. R. Hansen, “Causal interpretation of stochastic differential equations,” Mar. 2013.

[8] K. Friston, L. Harrison, and W. Penny, “Dynamic causal modelling,” *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, Aug. 2003, doi: 10.1016/s1053-8119(03)00202-7.