Advances in Cyclic Structural Causal Models

Joris Mooij

`j.m.mooij@uva.nl`

Amsterdam Machine Learning Lab

University of Amsterdam

June 1st, 2018

Part I

Introduction to Causality

Causation $\neq$ Correlation

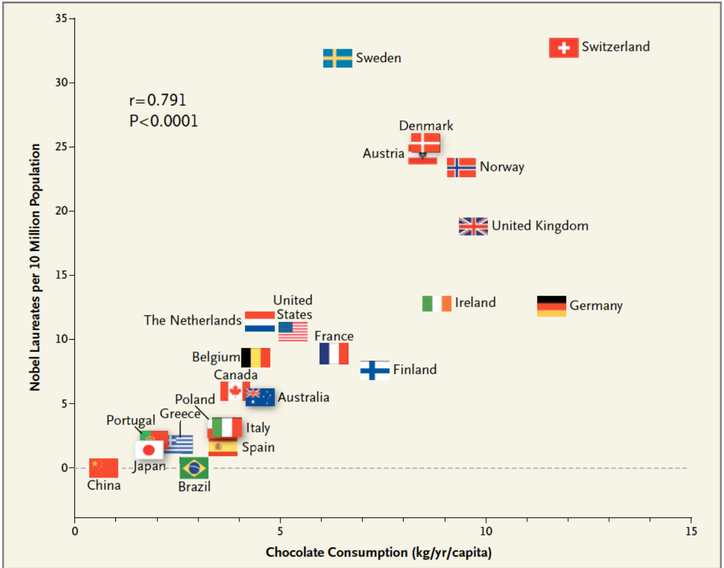
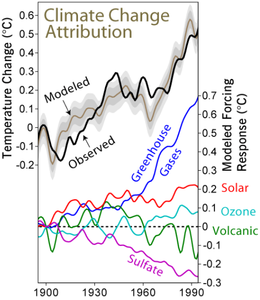


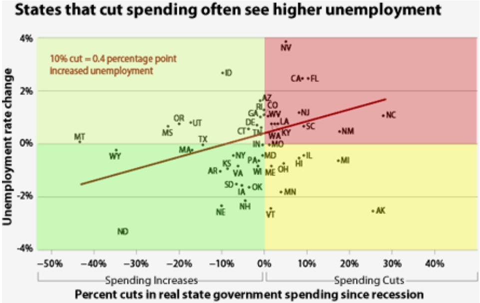
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Many questions in science are *causal*

Climatology:

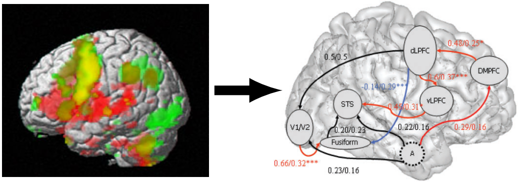


Economy:



Medicine:

Neuroscience:



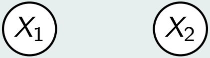
Causal Relations

Definition (Informal)

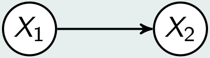
Let A and B be two distinct variables of system. A causes B ($A \dashrightarrow B$) if changing A (*intervening on* A) leads to a change of B.

Causal graph represents causal relationships between variables graphically.

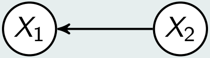
Example



X_1 and X_2 are causally unrelated

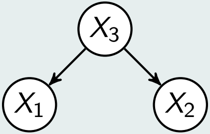


X_1 causes X_2

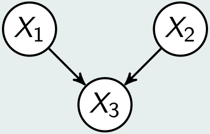


X_2 causes X_1

X_1 and X_2 cause each other



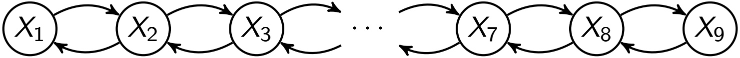
X_1 and X_2 have a common cause X_3



X_1 and X_2 have a common effect X_3

Direct vs. indirect causation: example

- Each stone causes *all* subsequent stones to topple.
- Each stone only directly causes the next neighboring stone to topple.
- Causal graph:



Direct causation

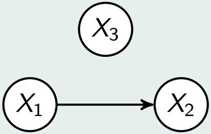
Let $\boldsymbol{V} = \{X_1, \dots, X_N\}$ be a set of variables.

Definition

If X_i causes X_j even if all other variables $\boldsymbol{V} \setminus \{X_i, X_j\}$ are hold fixed at arbitrary values, then

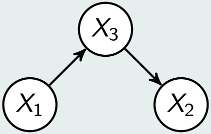
- we say that X_i causes X_j directly with respect to $\boldsymbol{V}$
- we indicate this in the causal graph on $\boldsymbol{V}$ by a directed edge $X_i \to X_j$

Example



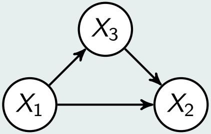
X_1 causes X_2;

X_1 causes X_2 directly w.r.t. $\{X_1, X_2, X_3\}$



X_1 causes X_2;

X_1 does not cause X_2 directly w.r.t. $\{X_1, X_2, X_3\}$



X_1 causes X_2;

X_1 causes X_2 directly w.r.t. $\{X_1, X_2, X_3\}$

Confounders: Example

A (latent) common cause of two variables is called a confounder.

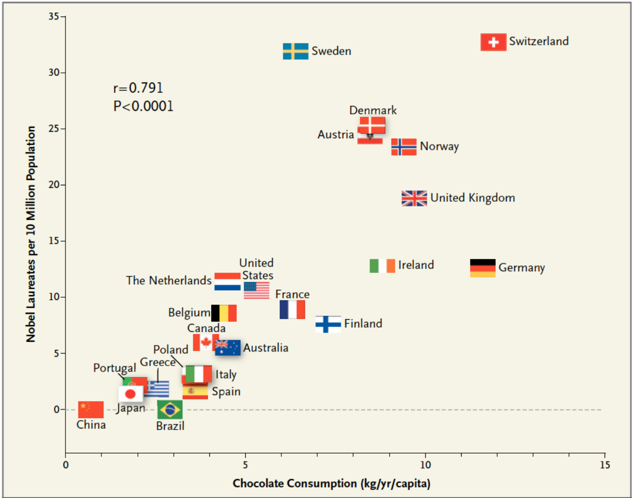
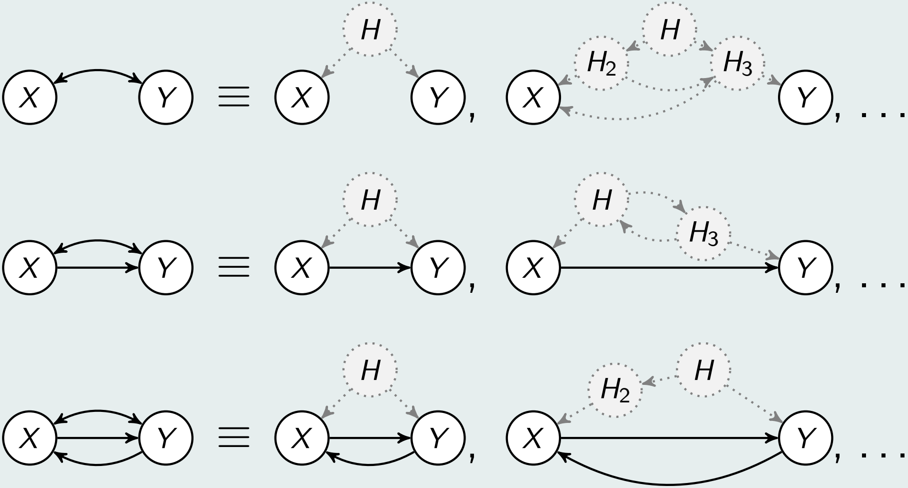


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Confounders: Graphical notation

We denote latent confounders by bidirected edges in the causal graph:

Example



Cycles: Definitions

Let A, B be two variables in a system.

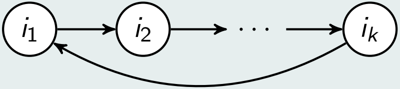
Definition

If A causes B and B causes A, then we say that A and B are involved in a causal cycle ("feedback loop").

Let $\mathcal{G}$ be a Directed Mixed Graph with nodes $\{1, \ldots, n\}$ (with directed and bidirected edges).

Definition

$\mathcal{G}$ is cyclic if it contains a directed cycle

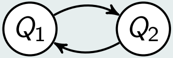


If $\mathcal{G}$ does not contain such a directed cycle, it is called acyclic, and known as an Acyclic Directed Mixed Graph (ADMG). If in addition, $\mathcal{G}$ does not contain any bidirected edges, it is called a Directed Acyclic Graph (DAG).

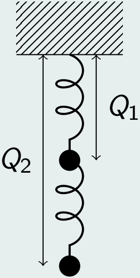
Feedback loops: Toy example

Example (Damped Coupled Harmonic Oscillators)

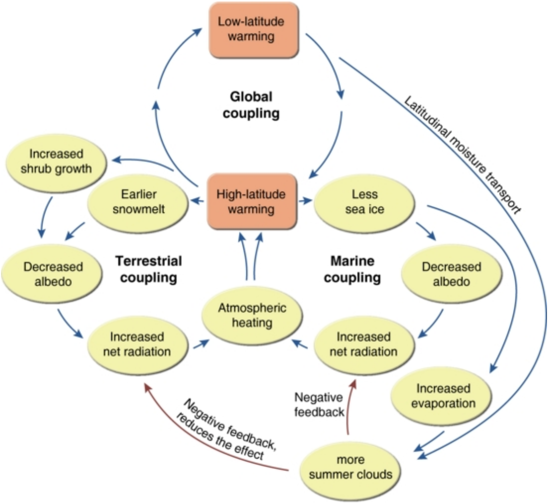
- Two masses, connected by a spring, suspended from the ceiling by another spring.
- Variables: vertical **equilibrium** positions Q_1 and Q_2.
- Q_1 causes Q_2.
- Q_2 causes Q_1.
- Causal graph:



- Cannot be modeled with acyclic causal model.

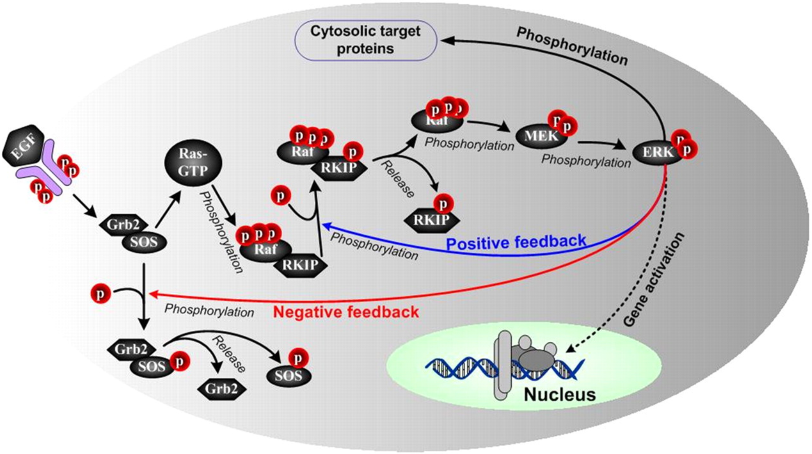


Feedback loops: Climatology



"Part of the uncertainty around future climates relates to important feedbacks between different parts of the climate system: air temperatures, ice and snow albedo (reflection of the sun's rays), and clouds." [Ahlenius, 2007]

Feedback loops: Biology



“Feedback mechanisms may be critical to allow cells to achieve the fine balance between dysregulated signaling and uncontrolled cell proliferation (a hallmark of cancer) as well as the capacity to switch pathways on or off when needed for physiologic purposes.” [McArthur, 2014]

Modeling causality

Question: How to model quantitatively the causal semantics of equilibrium states of systems, taking into account possible confounders and feedback loops. . . ?

Modeling causality

Question: How to model quantitatively the causal semantics of equilibrium states of systems, taking into account possible confounders and feedback loops. . . ?

Here we use Structural Causal Models (SCMs), a.k.a. Structural Equation Models (SEMs).

We present recent theoretical advances regarding cyclic SCMs:

- SCMs are causal models of fixed points of ODEs [Mooij et al., 2013]
- Marginalization: summarizing a subsystem [Bongers et al., 2016]
- Markov property: generalizing d-separation [Forré and Mooij, 2017]

We use this to develop an algorithm for causal discovery.

Part II

Structural Causal Models

Definition

Definition (Wright 1921, Pearl, 2000; [Bongers et al., 2016])

A Structural Causal Model (SCM), also known as Structural Equation Model (SEM), is a tuple $\mathcal{M} = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ with:

1. a product of standard measurable spaces $\boldsymbol{\mathcal{X}} = \prod_{i \in \mathcal{I}} \mathcal{X}_i$ (domains of the endogenous variables)
2. a product of standard measurable spaces $\boldsymbol{\mathcal{E}} = \prod_{j \in \mathcal{J}} \mathcal{E}_j$ (domains of the exogenous variables)
3. a measurable mapping $\boldsymbol{f} : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ (the causal mechanism)
4. a probability measure $\mathbb{P}_{\boldsymbol{\mathcal{E}}} = \prod_{j \in \mathcal{J}} \mathbb{P}_{\mathcal{E}_j}$ on $\boldsymbol{\mathcal{E}}$ (the exogenous distribution)

Definition

A pair of random variables $(\boldsymbol{X}, \boldsymbol{E})$ is a solution of SCM $\mathcal{M}$ if $\mathbb{P}^{\boldsymbol{E}} = \mathbb{P}_{\boldsymbol{\mathcal{E}}}$ and the structural equations $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{E})$ hold a.s..

Example

Example

Structural Causal Model $\mathcal{M}$:

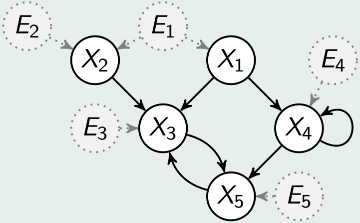
Formally:

$$(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}}) = (\prod_{i=1}^{5} \mathbb{R}, \prod_{j=1}^{5} \mathbb{R}, (f_1, \dots, f_5), \prod_{j=1}^{5} \mathbb{P}_{\mathcal{E}_j})$$

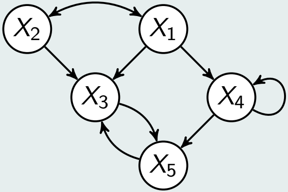
Informally:

$$\begin{aligned}
X_1 &= f_1(E_1) & \mathbb{P}^{E_1} &= \dots \\
X_2 &= f_2(E_1, E_2) & \mathbb{P}^{E_2} &= \dots \\
X_3 &= f_3(X_1, X_2, X_5, E_3) & \mathbb{P}^{E_3} &= \dots \\
X_4 &= f_4(X_1, X_4, E_4) & \mathbb{P}^{E_4} &= \dots \\
X_5 &= f_5(X_3, X_4, E_5) & \mathbb{P}^{E_5} &= \dots
\end{aligned}$$

Augmented functional graph $\mathcal{G}^a(\mathcal{M})$:



Functional graph $\mathcal{G}(\mathcal{M})$:



(Augmented) Functional Graphs

Definition

The components of the causal mechanism usually do not depend on *all* variables: for $i \in \mathcal{I}$,

$$X_i = f_i(\boldsymbol{X}_{\mathrm{pa}_i^{\mathcal{I}}}, \boldsymbol{E}_{\mathrm{pa}_i^{\mathcal{J}}})$$

where f_i only depends on $\mathrm{pa}_i^{\mathcal{I}} \subseteq \mathcal{I}$ (the endogenous parents of i) and $\mathrm{pa}_i^{\mathcal{J}} \subseteq \mathcal{J}$ (the exogenous parents of i).

Definition

The augmented functional graph $\mathcal{G}^a(\mathcal{M})$ of an SCM $\mathcal{M}$ is a directed graph with nodes $\mathcal{I} \dot\cup \mathcal{J}$ and an edge $k \to i$ iff $k \in \mathrm{pa}_i^{\mathcal{I}} \dot\cup \mathrm{pa}_i^{\mathcal{J}}$ is a parent of $i \in \mathcal{I}$.

Definition

The functional graph $\mathcal{G}(\mathcal{M})$ of an SCM $\mathcal{M}$ is a directed mixed graph with nodes $\mathcal{I}$, directed edges $k \to i$ iff $k \in \mathrm{pa}_i^{\mathcal{I}}$, and bidirected edges $k \leftrightarrow i$ iff $\mathrm{pa}_i^{\mathcal{J}} \cap \mathrm{pa}_k^{\mathcal{J}} \neq \emptyset$.

Interventions

To interpret an SCM as a *causal* model, we also need to define its semantics under interventions.

Definition (Perfect Interventions, [Pearl 2000])

- The perfect intervention $\mathrm{do}(\boldsymbol{X}_I = \boldsymbol{\xi}_I)$ enforces $\boldsymbol{X}_I$ to attain value $\boldsymbol{\xi}_I$.
- This changes the SCM $\mathcal{M} = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ into the intervened SCM $\mathcal{M}_{\mathrm{do}(\boldsymbol{X}_I = \boldsymbol{\xi}_I)} = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ where

$$\tilde{f}_i = \begin{cases} \xi_i & i \in I \\ f_i(\boldsymbol{X}_{\mathrm{pa}_i^{\mathcal{I}}}, \boldsymbol{E}_{\mathrm{pa}_i^{\mathcal{J}}}) & i \notin I. \end{cases}$$

- Interpretation: overriding default causal mechanisms that normally would determine the values of the intervened variables.

Interventions (Example)

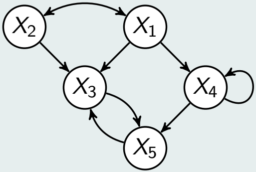
Example

Observational (no intervention):

Structural Causal Model $\mathcal{M}$:

$$\begin{aligned} X_1 &= f_1(E_1) & \mathbb{P}^{E_1} &= \dots \\ X_2 &= f_2(E_1, E_2) & \mathbb{P}^{E_2} &= \dots \\ X_3 &= f_3(X_1, X_2, X_5, E_3) & \mathbb{P}^{E_3} &= \dots \\ X_4 &= f_4(X_1, X_4, E_4) & \mathbb{P}^{E_4} &= \dots \\ X_5 &= f_5(X_3, X_4, E_5) & \mathbb{P}^{E_5} &= \dots \end{aligned}$$

Functional graph $\mathcal{G}(\mathcal{M})$:

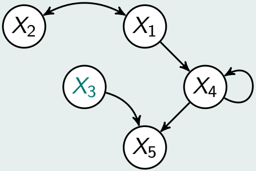


Intervention $\text{do}(X_3 = \xi_3)$:

Structural Causal Model $\mathcal{M}_{\text{do}(X_3=\xi_3)}$:

$$\begin{aligned} X_1 &= f_1(E_1) & \mathbb{P}^{E_1} &= \dots \\ X_2 &= f_2(E_1, E_2) & \mathbb{P}^{E_2} &= \dots \\ X_3 &= \xi_3 & \mathbb{P}^{E_3} &= \dots \\ X_4 &= f_4(X_1, X_4, E_4) & \mathbb{P}^{E_4} &= \dots \\ X_5 &= f_5(X_3, X_4, E_5) & \mathbb{P}^{E_5} &= \dots \end{aligned}$$

Functional graph $\mathcal{G}(\mathcal{M}_{\text{do}(X_3=\xi_3)})$:



Distributions

Definition

A pair of random variables $(\boldsymbol{X}, \boldsymbol{E})$ is a solution of SCM $\mathcal{M}$ if $\mathbb{P}^{\boldsymbol{E}} = \mathbb{P}_{\boldsymbol{\mathcal{E}}}$ and the structural equations $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{E})$ hold a.s..

Definition

We call the set of probability distributions of the solutions $\boldsymbol{X}$ of an SCM $\mathcal{M}$ the observational distributions of $\mathcal{M}$.

A perfect intervention on $\mathcal{M}$ may change the distributions.

Definition

We call the family of sets of probability distributions of the solutions of $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ (for $I \subseteq \mathcal{I}$, $\boldsymbol{\xi}_I \subseteq \boldsymbol{\mathcal{X}}_I$) the interventional distributions of $\mathcal{M}$.

Crucial difference with more usual statistical models: SCMs simultaneously model the distributions under all perfect interventions on a system.

Causal Graph

Proposition

If $\mathcal{M}$ has no self-loops, the causal graph of $\mathcal{M}$ is a subgraph of the functional graph $\mathcal{G}(\mathcal{M})$.

In that case, generically:

- The directed edges in $\mathcal{G}(\mathcal{M})$ represent direct causal effects w.r.t. $\mathcal{I}$;
- The bidirected edges in $\mathcal{G}(\mathcal{M})$ represent the existence of confounders w.r.t. $\mathcal{I}$;
- A direct causal relation $X_i \to X_j$ w.r.t. $\mathcal{I}$ can be detected experimentally by intervening on all variables $\boldsymbol{X}_{\mathcal{I}\setminus\{j\}}$ except X_j, and testing if the marginal distributions of the solutions on X_j depend on the value to which X_i is set.

Part III

SCMs model fixed points of ODEs

Modeling (Random) ODE fixed points with an SCM

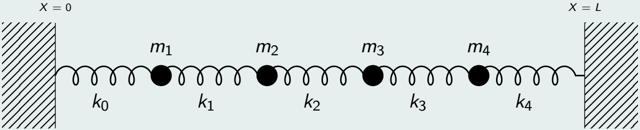
Theorem ([Mooij et al., 2013, Bongers and Mooij, 2018])

An ODE describing a dynamical system induces an SCM that models its fixed points, and how these change under perfect interventions.

$$\begin{tikzcd}
\boxed{\begin{array}{c} \mathcal{D}: \\ \begin{cases} \dot{X}_i(t) = f_i(\boldsymbol{X}_{\mathrm{pa}\,i}), \\ X_i(0) = (\boldsymbol{X}_0)_i \end{cases} \qquad i \in \mathcal{I} \end{array}} \arrow[r, "\textit{fixed points}"] \arrow[d, "\textit{intervention}"', "\mathrm{do}(I, \boldsymbol{\xi}_I)"] & \boxed{\begin{array}{c} \mathcal{M}_{\mathcal{D}}: \\ X_i = X_i + f_i(\boldsymbol{X}_{\mathrm{pa}(i)}) \qquad i \in \mathcal{I} \end{array}} \arrow[d, "\textit{intervention}"', "\mathrm{do}(I, \boldsymbol{\xi}_I)"] \\
\boxed{\begin{array}{c} \mathcal{D}_{\mathrm{do}(\boldsymbol{X}_I = \boldsymbol{\xi}_I)}: \\ \begin{cases} \dot{X}_i(t) = 0, \\ X_i(0) = \xi_i \end{cases} \qquad i \in I \\ \begin{cases} \dot{X}_i(t) = f_i(\boldsymbol{X}_{\mathrm{pa}\,i}), \\ X_i(0) = (\boldsymbol{X}_0)_i \end{cases} \qquad i \notin I \end{array}} \arrow[r, "\textit{fixed points}"] & \boxed{\begin{array}{c} \mathcal{M}_{\mathcal{D}_{\mathrm{do}(\boldsymbol{X}_I = \boldsymbol{\xi}_I)}}: \\ X_i = \xi_i \qquad i \in I \\ X_i = X_i + f_i(\boldsymbol{X}_{\mathrm{pa}(i)}) \qquad i \notin I \end{array}}
\end{tikzcd}$$

From ODE to SCM: Example

Example (Damped coupled harmonic oscillators)



- ODE $\mathcal{D}$:

$$\ddot{X}_i = \frac{k_i}{m_i}(X_{i+1} - X_i - l_i) - \frac{k_{i-1}}{m_i}(X_i - X_{i-1} - l_{i-1}) - b_i \dot{X}_i$$

- Induced SCM $\mathcal{M}_{\mathcal{D}}$:

$$X_i = \frac{k_i(X_{i+1} - l_i) + k_{i-1}(X_{i-1} + l_{i-1})}{k_i + k_{i+1}}$$

- Causal graph of induced SCM $\mathcal{G}(\mathcal{M}_{\mathcal{D}})$:

Part IV

Marginalization of SCMs

Marginalization (Example)

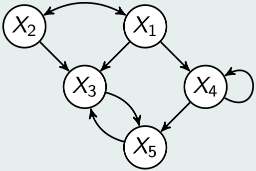
Example

SCM for complete system:

Structural Causal Model $\mathcal{M}$:

$$\begin{aligned} X_1 &= f_1(E_1) & \mathbb{P}^{E_1} &= \dots \\ X_2 &= f_2(E_1, E_2) & \mathbb{P}^{E_2} &= \dots \\ X_3 &= f_3(X_1, X_2, X_5, E_3) & \mathbb{P}^{E_3} &= \dots \\ X_4 &= f_4(X_1, X_4, E_4) & \mathbb{P}^{E_4} &= \dots \\ X_5 &= f_5(X_3, X_4, E_5) & \mathbb{P}^{E_5} &= \dots \end{aligned}$$

Functional graph $\mathcal{G}(\mathcal{M})$:

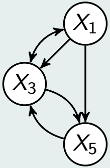


Marginalizing out X_2, X_4:

Marginalization $\mathcal{M}^{\setminus\{2,4\}}$:

$$\begin{aligned} X_1 &= f_1(E_1) & \mathbb{P}^{E_1} &= \dots \\ & & \mathbb{P}^{E_2} &= \dots \\ X_3 &= f_3(X_1, g_2(E_1, E_2), X_5, E_3) & \mathbb{P}^{E_3} &= \dots \\ & & \mathbb{P}^{E_4} &= \dots \\ X_5 &= f_5(X_3, g_4(X_1, E_4), E_5) & \mathbb{P}^{E_5} &= \dots \end{aligned}$$

Functional graph $\mathcal{G}(\mathcal{M}^{\setminus\{2,4\}})$:



Structural Causal Models: Unique Solvability

Definition

An SCM $\mathcal{M}$ is uniquely solvable w.r.t. a subset $\mathcal{C} \subseteq \mathcal{I}$ if there exists a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost surely unique, measurable mapping $\boldsymbol{g}_{\mathcal{C}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}^{\mathcal{I}}_{\mathcal{C}} \setminus \mathcal{C}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}^{\mathcal{J}}_{\mathcal{C}}} \to \boldsymbol{\mathcal{X}}_{\mathcal{C}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$, for all $\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}} \in \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}}$:

$$\boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L})\setminus\mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}) = \boldsymbol{f}_{\mathcal{L}}(\boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L})\setminus\mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}), \boldsymbol{x}_{\mathrm{pa}(\mathcal{L})\setminus\mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}) \,.$$

Informally: the set of structural equations for $\mathcal{C}$ has a unique solution for (almost) any input.

Example

An SCM with structural equations $X_1 = X_1, \quad X_2 = E_1, \quad X_3 = X_3 + X_1$ is only uniquely solvable w.r.t. $\{X_2\}$.

Example

Acyclic SCMs are uniquely solvable w.r.t. *any* set of endogenous variables.

Marginalization

Definition ([Bongers et al., 2016])

If $\mathcal{M} = \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ is uniquely solvable w.r.t. $\mathcal{L} \subseteq \mathcal{I}$, then it has a marginalization $\mathcal{M}^{\setminus \mathcal{L}} = \langle \boldsymbol{\mathcal{X}}_{\mathcal{I} \setminus \mathcal{L}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}^{\setminus \mathcal{L}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$, where the marginal causal mechanism $\boldsymbol{f}^{\setminus \mathcal{L}}$ is obtained by substituting the solution function $\boldsymbol{g}_{\mathcal{L}}$ for $\boldsymbol{X}_{\mathcal{L}}$ in terms of $\boldsymbol{X}_{\mathcal{O}}$ (with $\mathcal{O} := \mathcal{I} \setminus \mathcal{L}$) and $\boldsymbol{E}$ into the causal mechanism $\boldsymbol{f}$:

$$\boldsymbol{f}^{\setminus \mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) := \boldsymbol{f}_{\mathcal{O}}\big(\boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}), \boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}\big).$$

The marginalization preserves the causal semantics (restricted to the remaining part of the system, $\mathcal{I} \setminus \mathcal{L}$):

Theorem ([Bongers et al., 2016])

The marginalization $\mathcal{M}^{\setminus \mathcal{L}}$ is interventionally equivalent to $\mathcal{M}$ w.r.t. $\mathcal{I} \setminus \mathcal{L}$. In other words, for any perfect intervention on a subset of $\mathcal{I} \setminus \mathcal{L}$, $\mathcal{M}^{\setminus \mathcal{L}}$ and $\mathcal{M}$ admit the same solutions (marginalized onto $\boldsymbol{\mathcal{X}}_{\mathcal{I} \setminus \mathcal{L}}$).

Marginalization: Latent Projection of Functional Graph

The functional graph $\mathcal{G}(\mathcal{M}^{\setminus\mathcal{L}})$ of the marginalization of $\mathcal{M}$ on $\mathcal{I} \setminus \mathcal{L}$ is always a subgraph of the latent projection of $\mathcal{G}(\mathcal{M})$ on $\mathcal{I} \setminus \mathcal{L}$:

Definition

For a DMG $\mathcal{G}$ and a subset $\mathcal{L} \subseteq \mathcal{I}$ of nodes, the latent projection $\mathcal{G}^{\setminus\mathcal{L}}$ is defined as the DMG with nodes $\mathcal{I} \setminus \mathcal{L}$ and edges

- $i \to j$ iff there is a directed path $i \to \ell_1 \to \cdots \to \ell_k \to j$ in $\mathcal{G}$ with $\ell_1, \dots, \ell_k \in \mathcal{L}$
- $i \leftrightarrow j$ iff there is a path $i \leftarrow \ell_1 \leftarrow \cdots \leftarrow \ell_{k_1} \leftrightarrow \ell_{k_1+1} \to \cdots \to \ell_{k_2} \to j$ in $\mathcal{G}$ with $\ell_1, \dots, \ell_{k_1}, \dots, \ell_{k_2} \in \mathcal{L}$

Part V

Markov Properties of SCMs

The generalized directed global Markov property

We introduce a notion σ-separation that generalizes d-separation:

- σ-separation implies d-separation.
- For acyclic graph, σ-separation is equivalent to d-separation.

Inspired by ideas by [Spirtes, 1996], we show:

Theorem ([Forré and Mooij, 2017])

If an SCM $\mathcal{M}$ is uniquely solvable w.r.t. every strongly connected component in $\mathcal{G}(\mathcal{M})$, then the generalized directed global Markov property holds for any solution $\boldsymbol{X}$ of $\mathcal{M}$ with respect to the functional graph $\mathcal{G}(\mathcal{M})$:

$$A \underset{\mathcal{G}(\mathcal{M})}{\overset{\sigma}{\perp}} B \mid Z \implies \boldsymbol{X}_A \underset{\mathbb{P}^{\boldsymbol{X}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \mid \boldsymbol{X}_Z \qquad A, B, Z \subseteq \mathcal{I}.$$

Markov properties: σ-separation

Definition (σ-separation, [Forré and Mooij, 2017])

In a DMG $\mathcal{G}$, a path

$$i_1 \;\substack{\leftarrow\\ \rightarrow\\ \leftrightarrow}\; \cdots \;\substack{\leftarrow\\ \rightarrow\\ \leftrightarrow}\; i_n$$

is called σ-blocked by a set of nodes Z iff

- one or both end nodes i_1, i_n are in Z, or
- it contains a collider $i_{k-1} \;\substack{\rightarrow\\ \leftrightarrow}\; i_k \;\substack{\leftarrow\\ \leftrightarrow}\; i_{k+1}$ with $i_k \notin \mathrm{an}_{\mathcal{G}}(Z)$, or
- it contains a non-collider with $i_k \in Z$:

  $$i_{k-1} \;\substack{\rightarrow\\ \leftarrow\\ \leftrightarrow}\; i_k \rightarrow i_{k+1}, \quad i_{k-1} \leftarrow i_k \;\substack{\rightarrow\\ \leftarrow\\ \leftrightarrow}\; i_{k+1},$$

  where the child i_{k+1} (resp. i_{k-1}) is not in $\mathrm{sc}_{\mathcal{G}}(i_k)$.

We say that A is σ-separated from B by Z, denoted $A \perp^{\sigma} B \mid Z$, if every path with one end node in A and one end node in B is σ-blocked by Z.

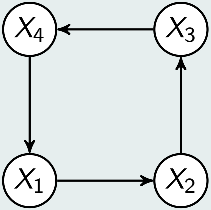
Markov properties: Example

Example

SCM $\mathcal{M}$:

$$\begin{aligned} X_1 &= f_1(X_4, E_1) = X_4 + E_1 \\ X_2 &= f_2(X_1, E_2) = X_1 \cdot E_2 \\ X_3 &= f_3(X_2, E_3) = X_2 + E_3 \\ X_4 &= f_4(X_3, E_4) = X_3 \cdot E_4 \end{aligned}$$

Functional graph $\mathcal{G}(\mathcal{M})$:



$$X_1 \perp^d X_3 \mid X_2, X_4$$

but

$$X_1 \not\perp^\sigma X_3 \mid X_2, X_4$$

So for any solution $\boldsymbol{X}$ of the SCM $\mathcal{M}$, in general we do not have that $X_1 \perp\!\!\!\perp X_3 \mid X_2, X_4$.

In general: No σ-separations between nodes within the same strongly connected component.

Directed global Markov property

Stronger statements can be derived for special cases:

Theorem ([Forré and Mooij, 2017])

If an SCM $\mathcal{M}$ satisfies at least one of the following three conditions:

1. *$\mathcal{M}$ is linear, its exogenous variables have a density with respect to Lebesgue measure, and $\mathcal{M}$ is solvable w.r.t. $\mathcal{I}$;*
2. *all endogenous variables are discrete-valued, $\mathcal{M}$ is uniquely solvable w.r.t. each ancestral subgraph of $\mathcal{G}(\mathcal{M})$;*
3. *$\mathcal{M}$ is acyclic;*

then the directed global Markov property holds for any solution $\boldsymbol{X}$ of $\mathcal{M}$ with respect to the functional graph $\mathcal{G}(\mathcal{M})$:

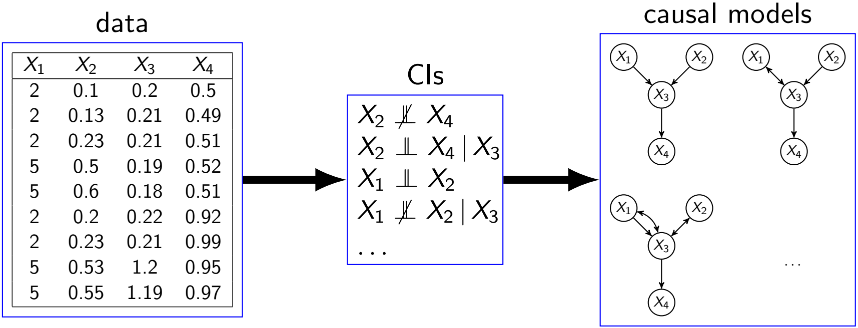
$$A \underset{\mathcal{G}(\mathcal{M})}{\overset{d}{\perp}} B \mid Z \implies \boldsymbol{X}_A \underset{\mathbb{P}^{\boldsymbol{X}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \mid \boldsymbol{X}_Z \qquad A, B, Z \subseteq \mathcal{I}.$$

Part VI

Causal Discovery from Data

Constraint-based Causal Discovery

From the pattern of conditional independences in the data we can reconstruct a set of possible causal graphs describing the data generating mechanism.



data

X_1	X_2	X_3	X_4
2	0.1	0.2	0.5
2	0.13	0.21	0.49
2	0.23	0.21	0.51
5	0.5	0.19	0.52
5	0.6	0.18	0.51
2	0.2	0.22	0.92
2	0.23	0.21	0.99
5	0.53	1.2	0.95
5	0.55	1.19	0.97

CIs

$X_2 \not\perp\!\!\!\perp X_4$

$X_2 \perp\!\!\!\perp X_4 \mid X_3$

$X_1 \perp\!\!\!\perp X_2$

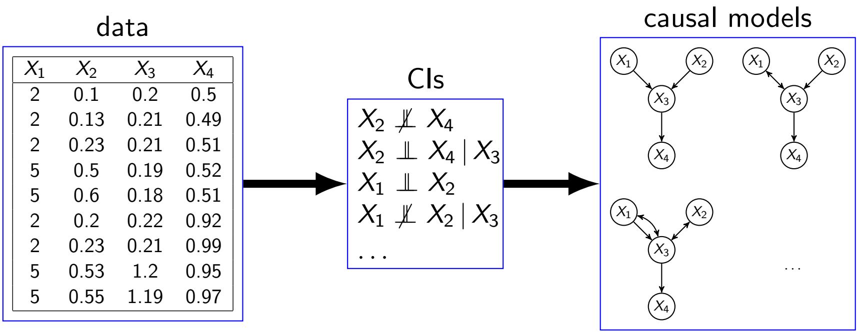
$X_1 \not\perp\!\!\!\perp X_2 \mid X_3$

$\ldots$

causal models

Constraint-based Causal Discovery

From the pattern of conditional independences in the data we can reconstruct a set of possible causal graphs describing the data generating mechanism.



data

X_1	X_2	X_3	X_4
2	0.1	0.2	0.5
2	0.13	0.21	0.49
2	0.23	0.21	0.51
5	0.5	0.19	0.52
5	0.6	0.18	0.51
2	0.2	0.22	0.92
2	0.23	0.21	0.99
5	0.53	1.2	0.95
5	0.55	1.19	0.97

CIs

$X_2 \not\perp\!\!\!\perp X_4$

$X_2 \perp\!\!\!\perp X_4 \mid X_3$

$X_1 \perp\!\!\!\perp X_2$

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

$\dots$

causal models

[Forré and Mooij, 2018]: the first causal discovery algorithm that can handle cycles, nonlinear relationships, latent (confounding) variables and data from different (interventional) contexts.

First Results on Synthetic Data [Forré and Mooij, 2018]

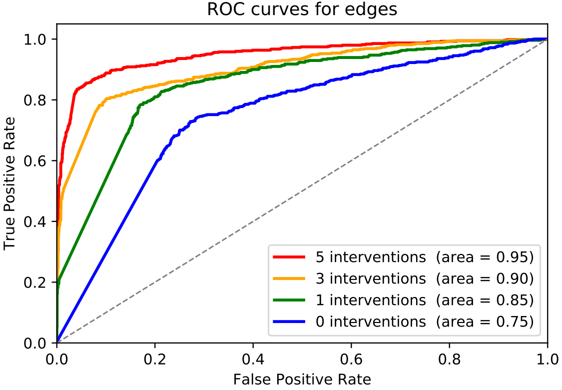


Figure: ROC curves for detecting direct causal relations from observational and interventional data.

Conclusion and Outlook

Motivated by:

- SCMs are a popular framework for causal modeling,
- SCMs can model confounders and causal feedback,

we developed theory for cyclic SCMs regarding:

- SCMs for modeling fixed points of ODEs [Mooij et al., 2013, Bongers and Mooij, 2018],
- Marginalization [Bongers et al., 2016],
- Markov property (σ-separation) [Forré and Mooij, 2017].

Based on this theory, we developed an algorithm for causal discovery from data [Forré and Mooij, 2018], that can handle:

- Cycles
- Nonlinear relationships
- Latent (confounding) variables
- Data from different (interventional) contexts

Acknowledgments


Stephan Bongers

Patrick Forré



Dominik Janzing


Jonas Peters


Bernhard Schölkopf

This work was supported by NWO, the Netherlands Organization for Scientific Research (VIDI grant 639.072.410), and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 639466).

References

Bongers, S. and Mooij, J. M. (2018).
From random differential equations to structural causal models: the stochastic case.
arXiv.org preprint, arXiv:1803.08784 [cs.AI].

Bongers, S., Peters, J., Schölkopf, B., and Mooij, J. M. (2016).
Structural causal models: Cycles, marginalizations, exogenous reparametrizations and reductions.
arXiv.org preprint, arXiv:1611.06221 [stat.ME].

Forré, P. and Mooij, J. M. (2017).
Markov properties for graphical models with cycles and latent variables.
arXiv.org preprint, arXiv:1710.08775 [math.ST].

Forré, P. and Mooij, J. M. (2018).
Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders.
In *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

Mooij, J. M., Janzing, D., and Schölkopf, B. (2013).
From ordinary differential equations to structural causal models: the deterministic case.
In Nicholson, A. and Smyth, P., editors, *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, pages 440–448. AUAI Press.