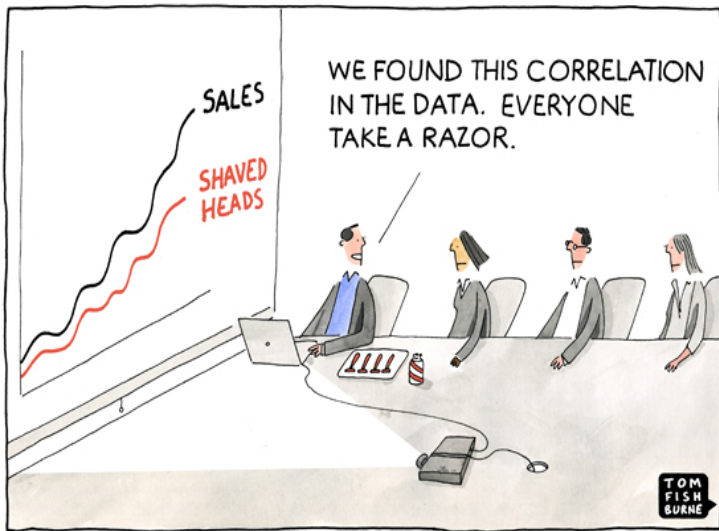


# Introduction to Structural Causal Models

Daniel Saggau

5/11/2021

# Motivation



© marketoonist.com

Table 1: (Pearl 2012)

Association-based Concepts	Causal Concepts
Correlation	Randomization
Regression	Confounding
Conditional Independence	Disturbance
Likelihood	Error Terms
Odds Ratio	Structural Coefficients
Propensity Score	Spurious Correlation

# Table of Contents

- Foundations of SCMs
  - Assumptions
  - Comparing Causal Tools
- Pearl's Causal Hierachy
  - Prediction
  - Intervention
  - Counterfactuals
  - Implications
- Graphical Models
  - Features
  - Implications
- Causality and Time

# Foundations of SCMs (1)

- System of equations with functions
- Assignment equation ' $:=$ ' rather than regular equation ' $=$ '
- Nonparametric SEM
- This algebraic equation would imply that  $B = A$  because the order has no concrete meaning in algebraic equations (symmetry)
- The initial ' $=$ ' sign was replaced with the ' $:=$ ' which is asymmetric (Pearl, 2009) and called an **assignment**.
- As mentioned, we define variables as functions e.g.  $A = f_A(B, U_A)$ .

**Error:\*** Regression: Omissible outside factor SCM/SEM; Latent influential factor that is pivotal for the model specification but not observable

$$C := N_C$$
$$E := f_E(C, N_E)$$

source: Peters, Janzing, and Schölkopf (2017)

# Assumptions in Causality

Independence:

- noise terms independent
- mechanisms independent (other variables invariant)(local changes)
- change in distribution stems from change in mechanism
- Causal Markov Condition allows to factorize
- Each Conditional is a causal mechanism

Source: (Peters, Janzing, and Schölkopf 2017)

# Assumptions in Causality (2)

## Established Conditions:

### Definition: **SUTVA**

'The treatment that one unit receives does not change the effect of treatment for any other unit.' (Hardt and Recht 2021)

### Definition: **Consistency**

The outcome  $Y$  agrees with the potential outcome corresponding to the treatment indicator.' (Hardt and Recht 2021)

### Definition: **Ignorability**

The potential outcomes are conditionally independent of treatment given some set of de-confounding variables. (Hardt and Recht 2021) (perfect RCT)

- First two hold for SCM counterfactuals
- third not testable but can check via backdoor criterion in SCM

# SCM Applications:

- Flexible simulations for higher order problems (intervention, counterfactual)
- Graphical visualization via directed acyclic graph
- Example: SCM for fairness in dynamical system : credit loan approval, time allocation, college admission



# Fundamental Differences (1)

- conflict whether to use graphs or not
- A SEM is a parametric specification used in applied sciences (parameters contested)
- A Bayesian causal network is another popular causal model using conditional probabilities and NO functions
- Differences in performance between BCN and SCM# Performance Evaluation

# Pearls Causal Hierachy

Table 2: Pearls Hierachy of Causation (2009)

Method	Action	Example	Usage
Association $P(a b)$	Co-occurrence	What happened. . .	(Un-)Supervised ML, BN, Reg.
Intervention $P(a do(b), c)$	Do- manipulation	What happens if . . .	CBN,MDP,RL
Counterfactual $P(a_b a', b')$	Hypotheticals	What would have happened if. . .	SCM ,PO

- ML, BN and regression are at the lowest level in the causal hierarchy
- Prediction methods demand the least information and depend on association alone
- Associational methods ignore external changes outside of our data.
- intervention distribution has information on these external changes.
- intervention distribution is only defined in high level causal methods.

- The second query deals with interventions
- Mathematical Tool: do-calculus
- The do-calculus enables us to study the manipulation of parent nodes
- There are various types of intervention
- One example is **atomic intervention**, where we set a variable to a constant
- In **policy intervention** we specify a different function for an equation
- Off-policy intervention models different intervention that is not in our historical data
- CBN , MDP and reinforcement learning model intervention.

# Example Intervention (1)

## Atomic Intervention:

- by replacing function with a constant

$$C_1 := f_{c_1}(p, N_1) \rightarrow C_1 := 600$$

$$C_2 := f_{c_2}(a, N_2)$$

$$E := f_E(C_1, C_2, N_E)$$

## Example Intervention (2)

### Policy Intervention:

- Intervention by replacing function with a different conditional probability

$$C_1 := f_{c_1}(p, N_1) \rightarrow C_1 := f(\pi)$$

$$C_2 := f_{c_2}(a, N_2)$$

$$E := f_E(C_1, C_2, N_E)$$

- missing data problem in PO framework

Process is described as follows:

- Abduction: Cast probability  $P(u)$  as conditional probability  $P(u|\epsilon)$
- Action: Exchange ( $X = x$ )
- Prediction: Compute ( $Y = y$ )

- Nodes  $\rightarrow$  Variables (endogenous/exogenous)
- Edges  $\rightarrow$  relationship (equations)
- Parents/Ancestors/Descendents
- No need to specify exact parametric shape
- highlight colliders
- estimation back door criterion



# Graphical Illustration - Probabilistic Model

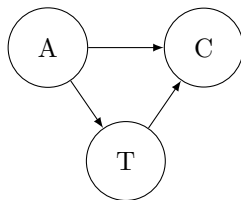


Figure 1: Probabilistic Model

# Graphical Illustration - Structural Causal Model

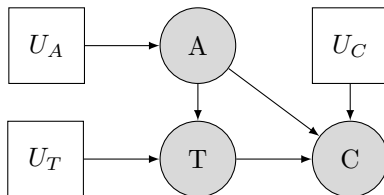


Figure 2: Structural Causal Model

- Time in Physical Sciences: Mechanical and exact
- Time in Social Sciences: Often Vague
- Regular Time Specification is also more vague
- To accommodate that issue, research on differential equation based SCMs started

Initial Value:

$$\mathbf{x}(t_0) = \mathbf{x}_0$$

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d$$

$$\mathbf{x}(t + dt) = \mathbf{x}(t) + dt \cdot f(\mathbf{x}(t))$$

# Graphical Overview

model	predict in IID setting	predict under changing distributions / interventions	answer counter-factual questions	obtain physical insight
mechanistic model	Y	Y	Y	Y
structural causal model	Y	Y	Y	N
causal graphical model	Y	Y	N	N
statistical model	Y	N	N	N

Source: (Peters, Janzing, and Schölkopf 2017)

- Mediation Analysis
- PO-Framework
- Causal Algorithms
- Causal Constraints Model

# Conclusions

- Hardt, Moritz, and Benjamin Recht. 2021. *Patterns, Predictions, and Actions: A Story about Machine Learning*. <https://mlstory.org>. <http://arxiv.org/abs/2102.05242>.
- Pearl, Judea. 2012. "The Causal Foundations of Structural Equation Modeling." CALIFORNIA UNIV LOS ANGELES DEPT OF COMPUTER SCIENCE.
- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.



# Appendix

Method	CBN	SCM
Prediction	<ul style="list-style-type: none"><li>• Unstable</li><li>• Volatile to parameter changes</li><li>• Re-Estimate entire model</li></ul>	<ul style="list-style-type: none"><li>• Stable</li><li>• More Natural Specification</li><li>• Only estimate <math>\Delta</math> CM</li></ul>
Intervention	<ul style="list-style-type: none"><li>• Costly for Non-Markovian Models</li><li>• Unstable(Nature CP)</li><li>• Only generic estimates(<math>\Delta</math> CP)</li></ul>	<ul style="list-style-type: none"><li>• Pot. Cyclic Representation</li><li>• Stable</li><li>• Context specific</li></ul>
Counterfactuals	<ul style="list-style-type: none"><li>• <b>Impossible</b></li><li>• no information on latent factors(<math>\epsilon</math>)</li></ul>	<ul style="list-style-type: none"><li>• Possible</li><li>• Inclusion of latent factors</li></ul>