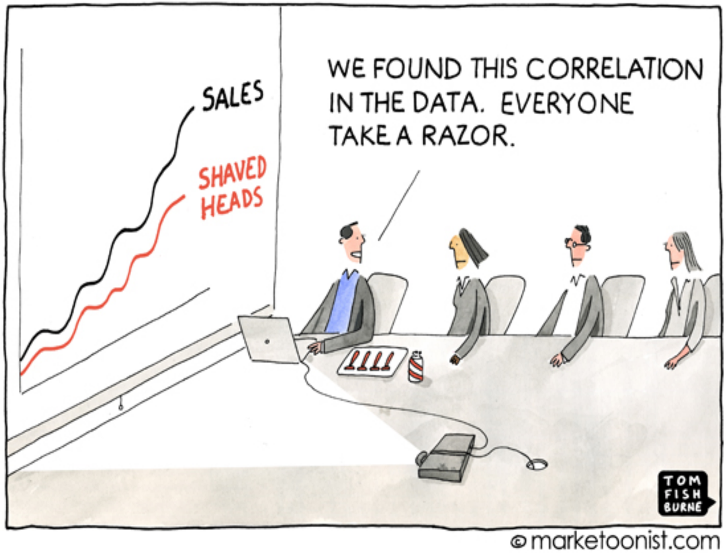
# Introduction to Structural Causal Models

Daniel Saggau

5/11/2021

# Motivation


SALES
SHAVED HEADS
WE FOUND THIS CORRELATION IN THE DATA. EVERYONE TAKE A RAZOR.
TOM FISH BURNE
© marketoonist.com

# Introduction

Table 1: [1]

| 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

- Nonparametric SEM
- System of equations with functions
- Assignments ':=' (non-symmetric) instead of algebraic equation '=' (symmetry)
- As mentioned, we define variables as functions e.g. $A = f_A(B, U_A)$. **Error in Regression:** Omittable outside factor **Error in SCM/SEM:** Latent (influential) factor that is pivotal for the model

$$C := N_c$$

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

source: [2]

# Assumptions in Causality (1)

Independence:

- Noise terms independent $(N_c, N_e)$
- Mechanisms independent (other variables invariant)(local changes)
- Change in distribution stems from change in mechanism
- Causal Markov Condition

# Assumptions in Causality (2)

**SUTVA**; 'The treatment that one unit receives does not change the effect of treatment for any other unit.'

**Consistency**: The outcome Y agrees with the potential outcome corresponding to the treatment indicator.'

**Ignorability**: The potential outcomes are conditionally independent of treatment given some set of de-confounding variables. (perfect RCT)

- First tow hold for SCM counterfactuals
- third not testable but can check via backdoor criterion in SCM [3]

# SCM Applications:

- Flexible simulations for higher order problems (intervention, counterfactual)
- Graphical visualization via directed acyclic graph
- Example: SCM for fairness in dynamical system [4] : 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

# Table of Contents

# 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 |

# Prediction

- ML, BN and regression are at the lowest level in the causal hierarchy
- Prediction methods demand the least information and depend on association alone
- Association-based methods ignore external changes outside of our data
- Intervention distribution needed for higher level information

# Intervention

- Mathematical Tool: do-calculus
- The do-calculus enables us to study the manipulation of parent nodes
- **Atomic intervention:** where we set a variable to a constant
- **Policy intervention:** we specify a different function for an equation
- CBN , MDP and reinforcement learning model intervention.

# Example Intevention (1)

**Atomic Intervention**:

- replacing function with a constant

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

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

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

# Example Intevention (2)

**Policy Intervention:**

- replacing function with a different conditional probability

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

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

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

# Counterfactuals

- missing data problem in PO framework

Process is described as follows:

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

# Table of Contents

# Graphical Tools

- Nodes -> Variables (endogenous/exogenous)
- Edges -> relationship (equations)
- Parents/Ancestors/Descendents
- No need to specify exact parametric shape
- Highlight colliders
- Estimation back door criterion
- Test theoretical model structure via causal algorithms to detect structure in data (IC/PC Algo.)

# Graphical Illustration - Probabilisitic Model

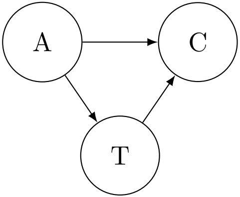



Figure 1: Probabilistic Model

# Graphical Illustration - Structural Causal Model

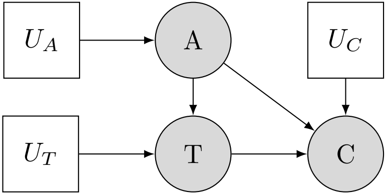



Figure 2: Structural Causal Model

# Table of Contents

# Causality and Time

- 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

# Causal Modelling with Differential Equations

Intial Value:

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

Derivative of function x with respect to time t:

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

Value of Function at time t + dt:

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

- dependence on prior time point and change in time contribute to the value at time point

# 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: [2]

# Extensions

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

# Concluding Remarks

- Assosicational learning is easy to model because of lower information neccessary
- but not always appropiate in high stake settings
- SCM as simulator for causal modelling, entailing a lot of information (DGP, intervention distribution)
- Note: *"Garbage in, Garbage out"*
- computational advantage casting causal model as system of assignments
- Enables modelling of higher order concepts like counterfactuals
- Extensions through differential equations for concise modelling of time

# References

[1]
J. Pearl, "The causal foundations of structural equation modeling," CALIFORNIA UNIV LOS ANGELES DEPT OF COMPUTER SCIENCE, 2012.

[2]
J. Peters, D. Janzing, and B. Schölkopf, *Elements of causal inference: Foundations and learning algorithms*. Cambridge, Massachuestts: The MIT Press, 2017.

[3]
M. Hardt and B. Recht, *Patterns, predictions, and actions: A story about machine learning*. https://mlstory.org, 2021.

[4]
E. Creager, D. Madras, T. Pitassi, and R. Zemel, "Causal modeling for fairness in dynamical systems," in *Proceedings of the 37th international conference on machine learning*, 2020, vol. 119, pp. 2185–2195, [Online]. Available: http://proceedings.mlr.press/v119/creager20a.html.

# Appendix

| Method | CBN | SCM |
|---|---|---|
| Prediction | · Unstable<br>· Volatile to parameter changes<br>· Re-Estimate entire model | · Stable<br>· More Natural Specification<br>· Only estimate $\Delta$ CM |
| Intervention | · Costly for Non-Markovian Models<br>· Unstable(Nature CP)<br>· Only generic estimates($\Delta$ CP) | · Pot. Cyclic Representation<br>· Stable<br>· Context specific |
| Counterfactuals | · **Impossible**<br>· no information on latent factors($\epsilon$) | · Possible<br>· Inclusion of latent factors |