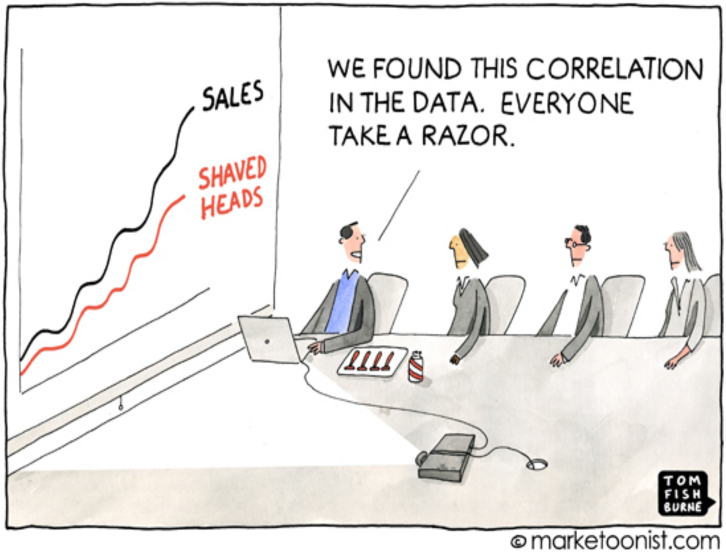
Introduction to Structural Causal Models

Daniel Saggau

May 2021

Motivation


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

Introduction

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

Seminar Topics

- Topic 1: **Structural Causal Models**
- Topic 2: Linking d-separation with (conditional) independence & Causal Structure Learning
- Topic 3: do-calculus: estimating the effect of interventions & counterfactuals
- Topic 4: Rubin's Potential Outcome Framework & Causal Trees
- Topic 5: Algorithmic Recourse
- Topic 6: Causal Interpretation of Black-Box Models

Foundations of SCMs (1)

- Flexible simulations for higher order problems (intervention, counterfactual)
- based on (elementary) ‘noise’/latent variables
- Graphical visualization via directed acyclic graph
- Nonparametric SEM
- System of equations with functions

Foundations of SCMs (2)

- **Assignments** ‘$:=$’ (non-symmetric)
- algebraic equation ‘$=$’ (symmetry)
- **Error in Regression:** Omittable outside factor
- **Error in SCM:** Latent (influential) factor that is pivotal for the model

$$C := N_c$$

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

Source: Peters, Janzing, and Schölkopf (2017)

Assumptions - 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

Differences

SEM:

- Conflict whether to use **graphs** or not and to use **parameters** in SEM
- SEM is a parametric model used in applied sciences

BCN:

- A Bayesian causal network (BCN) use conditional probabilities instead of functions
- Every SCM has the same information as respective probabilistic model + more (DGP)
- Computationally less efficient and cannot cast counterfactuals

Potential Outcome Framework:

- PO is mostly equivalent but has not graphical framework
- SCMs can derive potential outcomes but PO cannot derive SCMs

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

Intervention

- Mathematical Tool: do-calculus
- The do-calculus enables us to study the manipulation of parent nodes (direct effect)
- **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
- SCM 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 $(Y = y)$

Assumptions

- **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 Counterfactuals in SCM
- third not testable but can check via backdoor criterion in SCM
- Source: (Hardt and Recht 2021)

Table of Contents

Graphical Tools

- Nodes -> Variables (endogenous/exogenous)
- Edges -> relationship (equations)
- Parents/Ancestors/Descendants
- No need to specify exact parametric shape
- Focus on acyclical structures (DAG)
- Estimation back door criterion
- Test theoretical model structure via causal algorithms to detect structure in data (IC/PC Algo.)
- More information: (Morgan and Winship 2014)

Undirected/Semidirected Graph

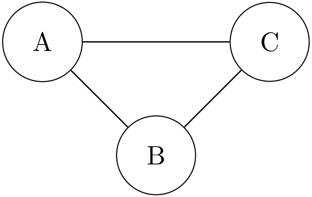


Figure 1: Undirected graph

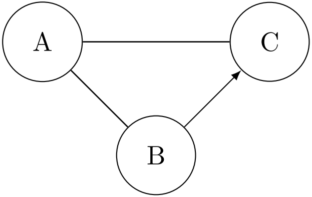


Figure 2: Semi-directed graph

Graphical Illustration - Probabilisitic Model

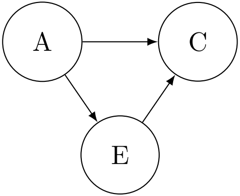


Figure 3: Probabilistic Model

Graphical Illustration - Structural Causal Model

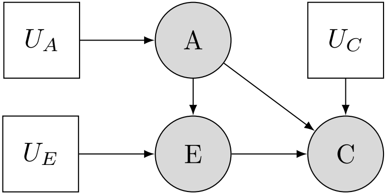


Figure 4: Structural Causal Model

Table of Contents

Causal Modelling with Differential Equations

- Time in Social Sciences: Often Vague
- Time in Physical Sciences: Mechanical via **Differential equations**
- dependence on prior time point and change in time contribute to the value at time point

Initial 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:

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

Graphical Overview

model	IID setting	changing distributions	counter-factual questions	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)

Concluding Remarks

- Association-based learning is easy to model because of lower information neccessary
- but not always appropriate in high stake settings
- SCM as simulator for causal modelling based on noise variables
- SCMs entail a lot of information (DGP, intervention distribution)
- BUT: **“Garbage in, Garbage out”**
- Computational advantage casting causal model as assignments
- Enables modelling of higher order concepts like counterfactuals
- Extensions through differential equations for concise modelling of time

Dos and Do nots

If you. . .

- ..only have conditional probabilities
- ..are only interested in association
- ..have very limited knowledge of your actual underlying structure

$\rightarrow$ you should probably go for a probabilistic model, association based model or a model based on conditional probabilities

But if you

- ..have an understanding of underlying relationships in the data
- ..are dealing with high stake settings
- ..want to simulate different interventions

$\rightarrow$ you could probably go for a SCM

References

Hardt, Moritz, and Benjamin Recht. 2021. *Patterns, Predictions, and Actions: A Story about Machine Learning*. https://mlstory.org. http://arxiv.org/abs/2102.05242.

Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Analytical Methods for Social Research. Cambridge University Press. https://doi.org/10.1017/CBO9781107587991.

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, Massachuestts: The MIT Press.

Appendix

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

Questions:

Thank you so much for your time!

If you have any questions contact me via:

- Mattermost: daniel.saggau
- Email: daniel.saggau@campus.lmu.de