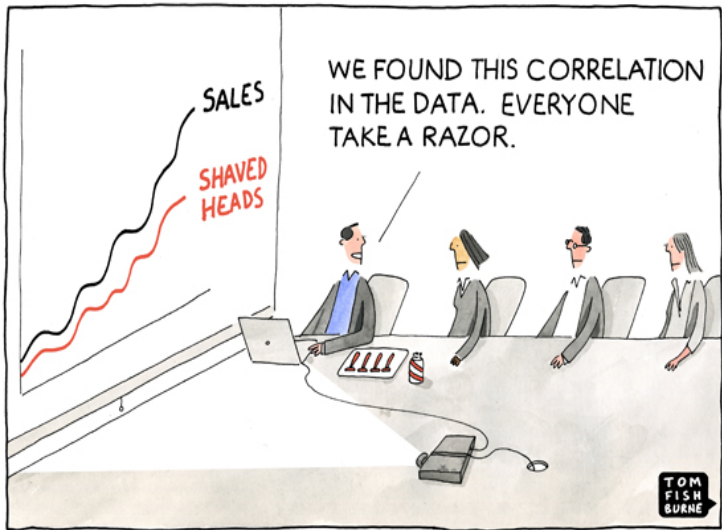


# Introduction to Structural Causal Models

Daniel Saggau

May 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 Hierarchy
  - Prediction
  - Intervention
  - Counterfactuals
- Graphical Models
- Causality and Time

- Nonparametric SEM
- System of equations with functions
- **Assignments**  $':='$  (non-symmetric)
- algebraic equation  $'='$  (symmetry)
- **Error in Regression:** Omissible 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 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 under which causal inference is possible (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 two hold for Counterfactuals in SCM
- third not testable but can check via backdoor criterion in SCM (Hardt and Recht 2021)

# SCM Applications:

- Flexible simulations for higher order problems (intervention, counterfactual) based on elementary noise variables
- Graphical visualization via directed acyclic graph
- Example: SCM for fairness in dynamical system (Creager et al. 2020)  
: credit loan approval, time allocation, college admission



# Fundamental Differences (1)

- Conflict whether to use **graphs** or not and to use **parameters** in SEM
- SEM is a parametric model used in applied sciences
- A Bayesian causal network (BCN) use conditional probabilities instead of functions
- PO is mostly equivalent but has not graphical framework
- SCM and PO are only frameworks that allow for counterfactuals
- Every SCM has the same information as respective probabilistic model + more (DGP)
- SCMs can derive potential outcomes but PO cannot derive SCMs

# Table of Contents

- Foundations of SCMs
  - Assumptions
  - Comparing Causal Tools
- **Pearl's Causal Hierarchy**
  - Prediction
  - Intervention
  - Counterfactuals
- Graphical Models
- Causality and Time

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
- Association-based methods ignore external changes outside of our data
- Intervention distribution needed for higher level information

- 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 Intervention (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 Intervention (2)

## Policy Intervention:

- replacing function with a different conditional probability

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

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

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

- 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 ( $Y = y$ )



# Table of Contents

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

- Nodes  $\rightarrow$  Variables (endogenous/exogenous)
- Edges  $\rightarrow$  relationship (equations)
- Parents/Ancestors/Descendants
- 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.)
- More information: (Morgan and Winship 2014)

# Graphical Illustration - Probabilistic Model

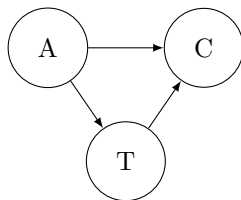


Figure 1: Probabilistic Model

# Graphical Illustration - Structural Causal Model

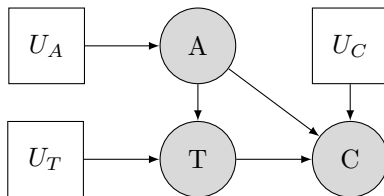


Figure 2: Structural Causal Model

# Table of Contents

- Foundations of SCMs
  - Assumptions
  - Comparing Causal Tools
- Pearl's Causal Hierarchy
  - Prediction
  - Intervention
  - Counterfactuals
- Graphical Models
- **Causality and Time**

# 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  $\mathbf{x}$  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)

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



# Concluding Remarks

- Association-based learning is easy to model because of lower information necessary
- 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)
- Note: *“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

# References

- Creager, Elliot, David Madras, Toniann Pitassi, and Richard Zemel. 2020. "Causal Modeling for Fairness in Dynamical Systems." In *Proceedings of the 37th International Conference on Machine Learning*, edited by Hal Daumé III and Aarti Singh, 119:2185–95. Proceedings of Machine Learning Research. PMLR. <http://proceedings.mlr.press/v119/creager20a.html>.
- 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, Massachusetstts: 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>