A Gentle Introduction into Structural Causal Models

Daniel Saggau • daniel.saggau@campus.lmu.de
Department of Statistics, Ludwig Maximilian University Munich, Germany

April 5th, 2021

1 Introduction

Most students have heard the phrase 'correlation does not imply causation.' While correlation implies co-occurrence, for many problems correlation is not enough. Algorithmic decision making based on co-occurrence is insufficient in high stake settings (Bareinboim et al. 2020). For many problems we want to understand causal relationships between variables. There are different approaches on how to model causal relationships. This paper focuses on Structural Causal Models and Bayesian Causal Networks. Bayesian Causal Networks (BCN) cast a model based on **conditional probabilities.** In Figure 1 we can see an example of a probabilistic model depicted as a directed acyclic graph (DAG). The nodes are the white circles. The edges are the arrows, defined by the conditional probabilities. C is our **collider** variable because the affect of A on C and T on C collide (Pearl 2009).

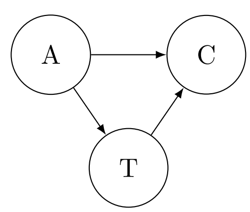

Figure 1: Probabilistic Model

Causal inference introduced tools that combine combine different sources of knowledge. We can e.g. include theoretical knowledge (Morgan and Winship 2014) and observational knowledge with structural causal models. Structural Causal Models (SCM) specify relationships based on **functional equations** (Pearl 2009). SCMs are non-parametric structural equations models with added features.

A structural equation model (SEM) is a set of autonous equations. This set represents the 'state of the world.' SEMs are popular in fields like economics, psychology and sociology (Pearl 2009). SCM are also defined by a (sub-)set of structural equations. Additionally, SCMs feature mathematical components from graphical models and the potential-outcome framework (Pearl 2012). There is a lot of controversy around

SEMs. Many scholars challenge the parameter-specification in a SEM (Pearl 2012).[1] SCMs specify an underlying data generating process without the computational effort of creating a parametric model. Note, that structural equations refer to assignment equations, used in computer science. Equations in causal models did not always have a concise notation (Pearl, 2009). First, there was no sign to express the assignment equation and people used the ‘=’ and one would e.g. write $A = B$. Treating an equation as a **algebraic equation** led to confusion because those have no causal information. This algebraic equation would imply that $B = A$ because the order has no concrete meaning in algebraic equations. The problem is that the equation is symmetric. The initial $'='$ sign was replaced with the ‘:=’ which is asymmetric (Pearl, 2009) and called an **assignment**. This misconception has caused a lot of challenges which I will address further on in this paper. As mentioned, we define variables as functions e.g. $A = f_A(B, U_A)$. B defines A and the latent factor U_A. To summarize, a SCM consists of a set of (autonomous) equations to generate (a) endogenous variables and (b) exogenous variables. In figure 2 we can see a vanilla structural causal model. The square nodes represent the latent variables. The circle nodes represent the observed variables. The arrows are our structural equations and depict the relationship between variables.[2]

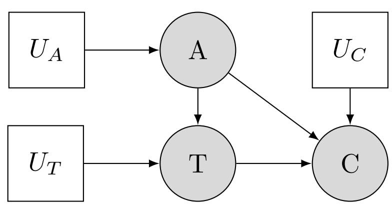

Figure 2: Structural Causal Model

To contextualise causal methods, Judea Pearl (2009) introduced the hierarchy of causation. Pearl focuses on three layers: association, intervention and counterfactuals. A higher level implies more detailed knowledge of the relationship between the variables. The first query is association, where we examine relationships based on observations.

Table 1: 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

Vanilla machine learning (ML), bayesian networks (BN) and regression models (Reg) are at the lowest level in the causal hierarchy (see table 1). These methods demand the least information and depend on association

[1]Inconsistency refers to inherent nature of creating parameters based of observational data. Observational data is seldom consistent. Henceforth respective estimates are often questioned. For further information see Hernán and Taubman (2008)

[2]Note that there are also cyclic structural causal models but no cyclic bayesian causal networks. For further information see Pearl (2009). Due to the confined scope of this paper, I will not explore cyclic structures.

alone. Associational methods ignore external changes outside of our data. The interventional distribution has information on these external changes. The interventional distribution is only defined in high level causal methods. The second query deals with interventions. Here we can use Pearls (2009) 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 (Oberst and Sontag 2019).

(tbc.) Hypothetical interventions:

Causal bayesian networks , Markov Decision Processes (MDP) and reinforcement learning model intervention.

The third query is counterfactual modelling. Here we deal with hypothetical settings. SCMs and potential outcome models allow for counterfactual modelling. These models can model counterfactuals because they include a interventional distribution (Oberst and Sontag 2019). BCNs only entail conditional probabilities. There is no information on relationships outside of the observations in the data. Henceforth, they cannot create counterfactuals (Pearl 2009).

Another issue in this paper is understanding the relationship of time and causality. SCMs make the assumption that causal relationships hold over time (Peters et al. 2017). These vague definitions of time is more prevalent in social sciences. Meanwhile hard sciences deal with time in a more concise manner. Differential equations model time on a more mechanic manner (Mooij et al. 2013).

Existing literature has provided an excellent introduction to SCMs and BCNs. Pearl (2009) provided a comprehensive introduction into this topic with his book 'Causality.' His work addressed misconceptions in social sciences, causality and statistics. Peters et al. (2017) address the relationship of causality, physical sciences, and causal inference. They also look at causal inference for observational data. E.g. (Bareinboim et al. 2020) worked on the application of causal methods in machine learning. Tarka (2018) discuss the history of SEMs and SCMs. This paper combines these contributions. The aim of this paper is to summarise SCMs and it's intersection with social sciences and physics. I structure the rest of the paper as follows:

2 Foundations of Structural Causal Models

Consists of graph and assignments: Baseline:

$$C := N_c$$

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

source: Peters et al. (2017)

3 Brief historical introduction:

Path analysis is the foundation of modern structural causal models. This path analysis is a structural equation model with one variable per indicator. Sewall Wright, a Statistican and Geneticist, introduced the topic in the 1920s. (Pearl, 2012) Various other disciplines such as econometrics, psychometrics and sociology adopted path analysis. Aside from path diagrams also introduced graph rules to formalise relationships.

(tbc. Cowles, Haavelmo,Aldrich,)

4 Assumptions in Causality

Reichenbachs common causal Principle Independence

If we specify the causal structure correctly:

(a) possible to undertake local intervention -> change f(x), regardless of f(y|x)
(b) these components are autonmous objects -> set of autonmous equations

Further: independence of noise ->

$$N_T, N_A$$

Example conditional probabilities: independence conditional probabilities/distribution and mechanism (ICM)

(peters et al., 2017)

- noise independent -> called **causal sufficiency** clause.
- Algorithmic independence

SCM fulfills via :

- dependecy only via vertice connection (noise)
- condition on noise, variables become independent

- Independence of mechanisms
- conditional independence:
- joint distribution

Density factorizes and their expectation

Causal assumptions differentiate causal models from association learning methods.

These causal concepts are not expressible based on distribution functions/statistical associations. (Pearl 2010)

Disturbance in SCM: Correlated and causal factor ; responsible for variation Disturbance in Regression: Uncorrelated

Causal assumption not testable (e.g. $Cov(U_a, U_b) = 0$). d-seperation to test assumptions in totality (cannot make assumptions in isolation).

exploit invariant characteristics of SEM whithout committing to shape. structural if function autonomous and invariant to change in form of other functions.

5 Pearl Causal Hierachy

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(Nature Eq.) · Context specific(Invariance of Eq.)
Counterfactuals	· **Impossible** · no information on latent factors(ϵ)	· Possible · Inclusion of latent factors

Counterfactuals

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 $(Y = y)$

5.1 Graphical Models

A graph in a SCM contains endogenous and exogenous variables. In a directed graph all edges have arrows. Directed arrows are direct (causal) effects. No edge between variables means that there is no causal effect. If we have some edges without arrows, we call that graph semi-directed. If we have all edges without arrows, we call that graph un-directed. The most famous graphical model is the acyclic graph. An acyclic graph has no roots that cause itself (directly and indirectly). Because we do not specify changes over time we typically only deal with acyclic models. This is partially because we assume the relationship to be constant over time.

There are various tools to derive graphical models.[3]

- d-seperation
- vertices variables and arrows are direct effects.

Any DAG can be written as SCM. given joint disrtribution can be discover grahp? Markov condition

5.2 Causal Inference and Time

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

Table: Source: Peters et al. (2017)

6 Conclusion

This paper provides a gentle introduction into structural causal models. SCMs entail many features, complementing research on association learning by providing depth. This in turn, is of particular benefit for

[3]The PC-Algorithm and the IC-Algorthm are two prominent examples.

high stake decision settings. SCMs differentiate from other methods through the specification of endogenous and exogenous variables, treating the exogenous factors as pivotal components of the actual model rather than assuming they are ommitable errors that are uncorrelated. As suggested by Pearls Hierachy, there are different levels to learning and each higher step can do anything the prior step can but with more detail and information. Henceforth, as machin

References

Bareinboim, Elias, JD Correa, Duligur Ibeling, and Thomas Icard. 2020. "On Pearl's Hierarchy and the Foundations of Causal Inference." *ACM Special Volume in Honor of Judea Pearl (Provisional Title).*

Hernán, Miguel A, and Sarah L Taubman. 2008. "Does Obesity Shorten Life? The Importance of Well-Defined Interventions to Answer Causal Questions." *International Journal of Obesity* 32 (3): S8–14.

Mooij, Joris M, Dominik Janzing, and Bernhard Schölkopf. 2013. "From Ordinary Differential Equations to Structural Causal Models: The Deterministic Case." *arXiv Preprint arXiv:1304.7920.*

Morgan, Stephen L., and Christopher Winship. 2014. "Causal Graphs." In *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd ed., 77–102. Analytical Methods for Social Research. Cambridge University Press. https://doi.org/10.1017/CBO9781107587991.004.

Oberst, Michael, and David Sontag. 2019. "Counterfactual Off-Policy Evaluation with Gumbel-Max Structural Causal Models." In *International Conference on Machine Learning*, 4881–90. PMLR.

Pearl, Judea. 2009. *Causality.* Cambridge university press.

———. 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.

Tarka, Piotr. 2018. "An Overview of Structural Equation Modeling: Its Beginnings, Historical Development, Usefulness and Controversies in the Social Sciences." *Quality & Quantity* 52 (1): 313–54.