

A Gentle Introduction into Structural Causal Models

Daniel Saggau • `daniel.saggau@campus.lmu.de`

Department of Statistics, Ludwig Maximilian University Munich, Germany

April 2021

Abstract

This paper provides a gentle introduction into causal inference focusing on structural causal models.

1 Introduction

1.1 Time in Causality

In causal research, we are interested in the relationship of our variables outside of a fixed time frame. We want to capture relationships that hold outside of an available dataset and results are intended to be generalizable to other time periods¹. The causal model entails a probabilistic specification but also entails additional information. This additional information is a specification for the model when intervention, outcome or feature distributions are modified (Peters et al., 2017). Probability theory deals with outcomes given a specific underlying data generating process.

1.2 SEM and Structural Causal Model

A lot of research ignores the distinction between structural equation models (SEM) and structural causal models (SCM). Essentially, the SCM builds on the SEM but there are some differences. Firstly, the structural equation model specification stems from disciplines like economics and psychology and is inherently a parametric specification. The SCM on the other hand is a non-parametric specification. Secondly, the SCM allows one to create causal graphs.²

Nevertheless, both models also have a lot in common.

Structural Causal Models specify the relationship in functional forms. This is very different to working with conditional probabilities. There are two popular schools of thought on how to specify causal mechanisms.

¹other datasets

²Note that we can also create causal graphs with other models like a bayesian causal network model

There are a number of reasons why people are interested in structural causal models. We can specify exogenous and endogenous variables in our model specification. When all our variables are fully specified within our model specification (endogenous variables only) we have a deterministic SCM.

Note that while this is very unreasonable because we assume we can explain all exogenous error terms and further we assume that there are no unobserved variables. Henceforth, when applying these models, they are seldom fully deterministic. This paper focuses on structural causal models. A lot of research ignores the distinction between structural equation models (SEM) and structural causal models (SCM). Essentially, the SCM builds on the SEM but there are some important differences. Firstly, the structural equation model specification stems from disciplines like economics and psychology and is inherently a parametric specification. The SCM is a non-parametric specification with additional mechanisms to cast causal graphs.³

Nevertheless, both models also have a lot in common. Structural equation models entail a fully specified set of equations. Henceforth the model looks at a system perspective of different variables. These different equations are autonomous which allows us to manipulate equations without modifying the rest of the system.

Structural Causal Models specify the relationship in functional forms, including latent/unobserved variables. This is very different to working with conditional probabilities. There are two popular schools of thought on how to specify causal mechanisms.

The first section provides introduces probabilities and will focus on clarifying the difference between observations and causal concepts. The second section discusses the hierarchy of causation. The third section discusses causal graphs. The fourth

1.3 Bayesian Causal Models

Meanwhile structural causal models approach causality via functions. There are some differences inherent to these two approaches. Conditional probabilities are less stable for causal methods (prediction and intervention).

SCMs are able to create counterfactuals, hypothetical knowledge. CBNs are unable to produce counterfactuals because we fail to specify exogenous variables in a probabilistic model. CBNs only specify observable characteristics. SCMs on the other hand specify latent factors too and are built on various distributional assumptions. In a SCM we may specify the counterfactual distribution from which we can draw counterfactual observations.

Method	CBN	SCM
Prediction	<ul style="list-style-type: none"> • Unstable • Volatile to parameter changes 	<ul style="list-style-type: none"> • Stable • More Natural Specification

³Note that we can also create causal graphs with other models like a Bayesian Causal Network Model

Method	CBN	SCM
	<ul style="list-style-type: none"> • Re-Estimate entire model 	<ul style="list-style-type: none"> • Only estimate Δ CM
Intervention	<ul style="list-style-type: none"> • Costly for Non-Markovian Models • Unstable(Nature CP) • Only generic estimates(Δ CP) 	<ul style="list-style-type: none"> • Pot. Cyclic Representation • Stable(Nature Eq.) • Context specific(Invariance of Eq.)
Counterfactuals	<ul style="list-style-type: none"> • Impossible • no information on latent factors(ϵ) 	<ul style="list-style-type: none"> • Possible • Inclusion of latent factors

There are a number of reasons why people are interested in structural causal models. We can specify exogenous and endogenous variables in our model specification for a SCM. When all our variables are fully specified within our model specification (endogenous variables only) we have a deterministic SCM.

Note that while this is very unreasonable because we assume we can explain all exogenous error terms and further we assume that there are no unobserved variables. Henceforth, when applying these models, they are seldom fully deterministic.

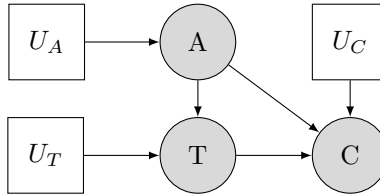


Figure 1: Structural Causal Model

2 Common Queries in Causal Models

We can apply a number of different queries in these different causal models. Note that one central advantage of SCMs is that we can cast counterfactuals. This is unique to structural causal models due to the specification of our distribution of unobserved variables.

Table 2: Pearl - Hierachy of Causation

Method	Action	Example
Prediction	Observation/Co-occurrence	What happened...
Intervention	Do-manipulation	What happens if ...
Counterfactual	Hypothetical Realities	What would have happened if...

2.1 Prediction

Note that prediction in a causal model and prediction in a machine learning context are not the same thing. In prediction in causal models we can only modify root nodes. We are unable to modify the child nodes in causal prediction. Meanwhile, in a machine learning context we can modify any variable. We do not have an underlying causal specification, henceforth we are able to modify any variable and undertake prediction. One should note that clearly the prediction in causal models has more explanatory power than the prediction in a vanilla-machine learning model.

2.2 Intervention

This paper focuses on atomic, policy and off-policy intervention. Atomic intervention means that we replace a variable with a constant so we replace variable T with constant t . Policy intervention means we replace a variable with a new distribution π . Off-policy intervention is an intervention when we only have historical data and no data on the specific intervention. Off-policy has gained traction in disciplines like reinforced learning.

3 Common Queries in Causal Models

Table 3: Pearl - Hierarchy of Causation

Method	Action	Example
Prediction	Observation/Co-occurrence	What happened...
Intervention	Do-manipulation	What happens if ...
Counterfactual	Hypothetical Realities	What would have happened if...

3.1 Prediction

3.2 Intervention

3.3 Counterfactuals

Process is described as follows:

- (a) Abduction: Cast probability $P(u)$ as conditional probability $P(u|\epsilon)$
- (b) Action: Exchange ($X = x$)

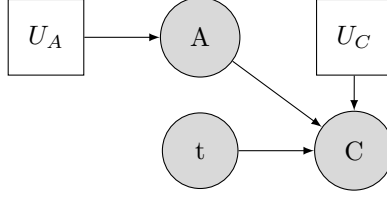


Figure 2: Atomic Intervention

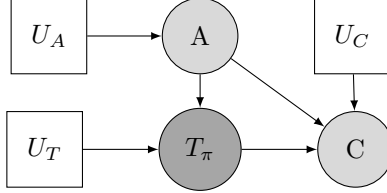


Figure 3: Policy Intervention

(c) Prediction: Compute ($Y = y$)

3.4 Causal Graphs

Causal graphs are graphical models to visualize causal models. Causal graphs are based on various different causal models. A graphical models contains nodes and edges (vertices). If we are dealing with a causal graph based on a structural causal model, we usually see latent variables. These latent variables are typically differentiated by either using a different color and/or shape than the nodes in probabilistic models where we only special conditional probabilities. The absence of a edge between nodes can many different things depending on the underlying model. Absence in a SCM means that there is no causal relationship. Absence in a BCN means that there is no conditional probability.

The most common causal graph is the acyclic directed graph. A directed graph is a graph where every edge has a direction. An un-directed graph has edges but there is no arrow which means that the relationship is unclear. We can depict confounding variables via a edge with two arrows. An acyclic graph is a graph without cycles. That means that a variable does not cause itself, directly or indirectly. Given the way we look at time where the relationship between variables is indifferent to time, this representation provides a foundation for graphical models.

3.5 Criteria

3.6 Reichenbach common cause principle

If two [random] variables X and Y are dependent when ignoring z and are independent when accounting for z is the true underlying causal influence. We can observe Z by looking at the model independent of Z ($X \setminus \setminus Y$)

and once dependent on Z ($X \perp\!\!\!\perp Y|Z$). (Peters et al., 2017) Note that there may be a different explanation for the dependence (selection bias)

3.7 Independence Assumption

3.8 D-Seperation

3.8.1 Markovian Condition

3.8.1.1 In Graphical Model

3.8.1.2 in SCM

3.8.2 Backdoor criteria

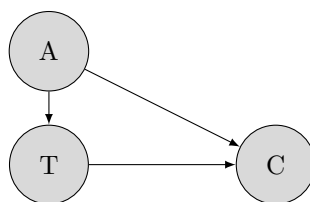


Figure 4: Backdoor Criterion

3.9 Causal Mechanisms via Differential Equations

Differential equations can describe a dynamic system accounting with time considerations.

4 Conclusion

This paper illustrates the versatile nature of structural causal models. Casting causal mechanisms via a functional specification entails many advantages. SCMs provide a stable foundation for causal graphs.

5 References