

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

Interest in a deeper understand of relationships between variables has increased the popularity of causal models. This paper provides a gentle introduction into causality focusing on structural causal models. For illustrative purposes, I will explore two models namely (1) Bayesian causal networks (BCN) and (2) structural causal models (SCM). Probabilistic specifications such as a BCN cast a model based on conditional probabilities. SCMs cast a model based on functions. SCMs extend probabilistic models by specifying the data generating process rather than solely focusing on conditional probabilities. These SCMs can be applied to number of different queries. Using Pearl's hierarchy of causation (2009), I focus on prediction, intervention and counterfactuals. To further clarify central notions of causal modelling, this paper briefly discusses the role of time in causality and will illustrate an general overview of the history of causal modelling to contextualize the SCM modelling approach. I will also introduce causal graphs, specifically focusing on the popular directed acyclical graphical (DAG) models, to highlight differences in various queries and modelling assumptions.

1 Introduction

“Correlation does not imply Causation” is one of the most pivotal lessons many students learn during introductory undergraduate statistics courses.

In causal research, 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).

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.

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.

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.

¹other datasets

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.

Firstly, the structural equation model specification stems from disciplines like economics and psychology and is inherently a parametric specification with algebraic equations. The SCM is a non-parametric specification (with assignment equations stemming from CS) with additional mechanisms to cast causal graphs.²

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. This 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.

1.1 Criteria

1.2 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 ($X \not\perp Y$) and once dependent on Z ($X \perp Y|Z$). (Peters et al., 2017) Note that there may be a different explanation for the dependence (selection bias)

1.3 Independence Assumption

The first principle is independence of our mechanisms. As a reminder, each equation in a SCM is a separate independent mechanism. In a two-variable case of a causal model one can differentiate between the cause distribution and the effect distribution (Peters et al., 2017).

Implications:

1. Change mechanism without changing the rest of the model
2. Change of one mechanism does not contain information on other changes
3. Systems can never be isolated from environment, henceforth weak dependence of noise possible without violating assumption about independent mechanisms

²Note that we can also create causal graphs with other models like a Bayesian Causal Network Model

1.3.1 Markovian Condition

1.3.2 Backdoor criteria

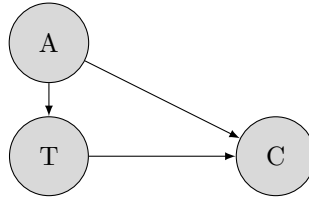


Figure 1: Backdoor Criterion

1.3.3 (Causal) Minimality

1.3.4 Interventioanl equivalence

1.3.5 Cyclic SCMs

1.3.6 d-seperation

A baseline SCM consists of two assingments.

$$A = N_A \tag{1}$$

$$B = f_B(A, N_B) \tag{2}$$

N_A and N_B are independent. A is a direct cause of B. $A \rightarrow B$ is causal graph (Peters et al., 2017). The intervention distribution is unequal to the observational distribution.

Canonical Representation

1.4 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.

1.5 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.

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.

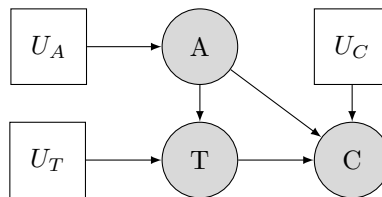


Figure 2: Structural Causal Model

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 1: 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 a 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

3 Common Queries in Causal Models

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. . .

³Another type of intervention is soft intervention when we only change the noise distribution and keep the prior causal effect

3.1 Prediction

3.2 Intervention

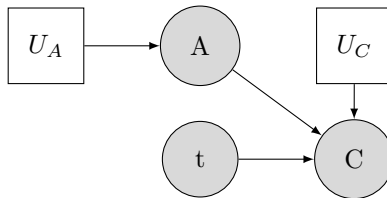


Figure 3: Atomic Intervention

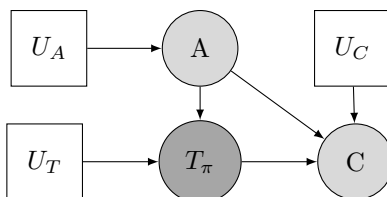


Figure 4: Policy Intervention

3.3 Counterfactuals

Process is described as follows:

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

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

Given the way we look at time where the relationship between variables is indifferent to time, this representation provides a foundation for graphical models.

3.4 Causal Mechanisms via Differential Equations

SCMs are abstractions of physical processes. Differential equations can describe a dynamic cyclic system accounting with more sophisticated time considerations. Henceforth SCMs are predominately used for acyclic causal graphs. Note that acyclic causal graphs are still dynamical models but they do not have loops where variables reinforce themselves.

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