A Gentle Introduction into Structural Causal Models

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

June 2021

Abstract The interest in understanding relationships of variables beyond co-occurrence has increased the popularity of causal modelling. To provide a comprehensive understanding of causal modelling, I introduce two prominent causal model specifications namely (1) Bayesian Causal Networks (BCN) and (2) Structural Causal Models (SCM), focusing on the latter.Probabilistic specifications such as a BCN cast a model based on conditional probabilities. SCMs cast a model based on assignment functions and extend probabilistic models by specifying the data generating process rather than solely utilizing conditional probabilities. Another difference between these models is their ability to address different queries such as *predictions*, *interventions* and *counterfactuals*. These queries are part of Pearl's causal hierarchy (2009). Pearl matches these queries with their respective actions namely *observing*, *doing* and *imagining*. I compare the feasibility of addressing these queries and undertaking respective actions for both specifications. To contextualize SCMs within the field of causality, I also discuss the role of time in causality. This paper uses various directed acyclic graphs to highlight the differences in these modelling approaches. The insights of this paper can be used as a baseline for subsequent research on structural causal models.

Contents

1 Introduction

For many research problems, we want to understand the relationship between variables beyond co-occurrence.

Research dealing with these question moved from traditional statistics, to ,

The most popular causal model is the structural causal model or in short SCM. The SCM is an expressive simulator to estimate causal relationships, accounting latent factors. The basis of these SCMs is a set of equations, or more precisely assignments, providing functions to derive the conditional probabilities for our model [1]. These assignments describe our variables in our model. Compared to probabilistic models, where we only specify conditional probabilities, the SCM actually enables the combination of theoretical and observational data. Conditional probabilities cannot represent latent variables because there is no conditional probability in our observational data for unobserved variables [2].

The aim of this paper is to summarise SCMs and it's intersection with social sciences and physics.

Section 2 introduces the assumptions in causal modelling. Further, section 2 discusses different assumptions in causal modelling. Section 3 Section 4 addresses Pearls Causal Hierachy. Section 5 provides a brief introduction into graphical models. Section 6 focuses on the intersection of SCMs and the perception of time.

2 SCM Example

$S := f_S(U_S)$ $E := f_E(U_E)$ $C := f_C(S, E, U_C)$

$\{S\}$ - Smoking $\{C\}$ - Cancer $\{E\}$ - Exercise

A structural causal model contains an underlying graphical model [1]. The most popular graphical model is the directed acyclic graph or in short DAG. A DAG entails nodes and edges. Nodes represent our different variables. Edges depict the assignment equations. All edges are directed in the DAG. The graph is called acyclic if no variable causes (directly or indirectly) itself [3]. There is a hierachy within DAGs.

Probabilisitic Representation

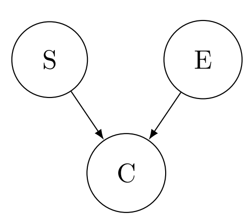

Figure 1: Probabilistic Model

3 Assumptions

This section discusses the independence assumption in SCMs.

Independence of Noise, Independence of Mechanism, Causal Sufficiency, and why time is ignored. The question should be really, what are we commiting to when we use SCMs.

Acyclicity

Assume acyclic because conditional probabilities with variables causing itself very hard .

Independence:

Noise indepedent so noise smoking and noise eating should be independent.

Independent Mechanisms:

Markov Condition

Truncated Factorization

$$P_{X_3=\mathrm{On}}\left(x_1, x_2, x_4, x_5\right)=P\left(x_1\right) P\left(x_2 \mid x_1\right) P\left(x_4 \mid x_2, X_3=\mathrm{On}\right) P\left(x_5 \mid x_4\right)$$

$$P_x(v)=\prod_{\{i \mid V_i \notin X\}} P\left(v_i \mid pa_i\right)$$
for all $\boldsymbol{v}$ consistent with x

4 Pearl's Causal Hierachy

Method	Action	
Association $P(a\|t)$	Co-occurrence	(Un-)Supervised ML, BN, Reg.
Intervention $P(a\|do(t),c)$	Do-manipulation	CBN,MDP,RL
Counterfactual $P(a_t\|a', t')$	Hypotheticals	SCM ,PO

Association:

-> example probabilistic terms

$P(C|s)$

Intervention:

$P(a|do(t),c)$

$do(t)$ -> replace function $T := f_T(\pi)$ with different conditional probability (or constant) [2].

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 $(Y = y)$

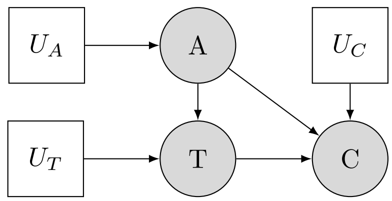

Figure 2: Structural Causal Model

5 SCMs and Time

largely ingore time

time in mechanical modeling crucial [4].

- 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:

$$\mathbf{x}\left(t_0\right) = \mathbf{x}_0$$

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:

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

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

Table: Source: [4]

6 Conclusion

7 References

[1] M. Hardt and B. Recht, *Patterns, predictions, and actions: A story about machine learning*. https://mlstory.org, 2021.

[2] J. Pearl, *Causality*. Cambridge university press, 2009.

[3] S. L. Morgan and C. Winship, *Counterfactuals and causal inference: Methods and principles for social research*, 2nd ed. Cambridge University Press, 2014.

[4] J. Peters, D. Janzing, and B. Schölkopf, *Elements of causal inference: Foundations and learning algorithms*. Cambridge, Massachuestts: The MIT Press, 2017.

[5] J. M. Mooij, D. Janzing, and B. Schölkopf, "From ordinary differential equations to structural causal models: The deterministic case," *arXiv preprint arXiv:1304.7920*, 2013.

[6] J. Pearl, "The causal foundations of structural equation modeling," CALIFORNIA UNIV LOS ANGELES DEPT OF COMPUTER SCIENCE, 2012.

[7] P. Tarka, "An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences," *Quality & quantity*, vol. 52, no. 1, pp. 313–354, 2018.

[8] E. Bareinboim, J. Correa, D. Ibeling, and T. Icard, "On pearl's hierarchy and the foundations of causal inference," *ACM Special Volume in Honor of Judea Pearl (provisional title)*, 2020.

[9] M. A. Hernán and S. L. Taubman, "Does obesity shorten life? The importance of well-defined interventions to answer causal questions," *International journal of obesity*, vol. 32, no. 3, pp. S8–S14, 2008.

[10] M. Oberst and D. Sontag, "Counterfactual off-policy evaluation with gumbel-max structural causal models," in *International conference on machine learning*, 2019, pp. 4881–4890.

[11] E. Creager, D. Madras, T. Pitassi, and R. Zemel, "Causal modeling for fairness in dynamical systems," in *Proceedings of the 37th international conference on machine learning*, 2020, vol. 119, pp. 2185–2195, [Online]. Available:http://proceedings.mlr.press/v119/creager20a.html.

[12] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and machine learning*. fairmlbook.org, 2019.

[13] S. Chiappa and W. S. Isaac, "A Causal Bayesian Networks Viewpoint on Fairness," *arXiv:1907.06430 [cs, stat]*, vol. 547, pp. 3–20, 2019, doi: 10.1007/978-3-030-16744-8_1.

[14] M. A. Hernán and J. M. Robins, "Causal Inference: What If," p. 311.

[15] E. C. Cortés and D. Ghosh, "A Simulation Based Dynamic Evaluation Framework for System-wide Algorithmic Fairness," *arXiv:1903.09209 [cs]*, Mar. 2019, Accessed: Apr. 12, 2021. [Online]. Available: http://arxiv.org/abs/1903.09209.

[16] M. J. Kusner, J. R. Loftus, C. Russell, and R. Silva, "Counterfactual Fairness," *arXiv:1703.06856 [cs, stat]*, Mar. 2018, Accessed: Apr. 12, 2021. [Online]. Available: http://arxiv.org/abs/1703.06856.

[17] N. Kilbertus, M. Rojas-Carulla, G. Parascandolo, M. Hardt, D. Janzing, and B. Schölkopf, "Avoiding Discrimination through Causal Reasoning," *arXiv:1706.02744 [cs, stat]*, Jan. 2018, Accessed: Apr. 12, 2021. [Online]. Available: http://arxiv.org/abs/1706.02744.

[18] J. von Kügelgen, A.-H. Karimi, U. Bhatt, I. Valera, A. Weller, and B. Schölkopf, "On the Fairness of Causal Algorithmic Recourse," *arXiv:2010.06529 [cs, stat]*, Feb. 2021, Accessed: Apr. 12, 2021. [Online]. Available: http://arxiv.org/abs/2010.06529.

[19] R. O. Ness, K. Paneri, and O. Vitek, "Integrating Markov processes with structural causal modeling enables counterfactual inference in complex systems," *arXiv:1911.02175 [cs, q-bio, stat]*, Nov. 2019, Accessed: Apr. 12, 2021. [Online]. Available: http://arxiv.org/abs/1911.02175.

[20] A.-H. Karimi, B. Schölkopf, and I. Valera, "Algorithmic Recourse: From Counterfactual Explanations to Interventions," *arXiv:2002.06278 [cs, stat]*, Oct. 2020, Accessed: Apr. 12, 2021. [Online]. Available: http://arxiv.org/abs/2002.06278.

[21] M. A. Hernan, "Invited Commentary: Agent-Based Models for Causal Inference–Reweighting Data and Theory in Epidemiology," *American Journal of Epidemiology*, vol. 181, no. 2, pp. 103–105, Jan. 2015, doi: 10.1093/aje/kwu272.