Chapter 1

Introduction

1.1 From causation to causality

The most effective vessel for bringing our thoughts from one state to the desired other, anchored in the essence of reasoning, is logic. It gives us insight into most peoples' thinking mechanisms and provides a robust tool in the search for knowledge and truth to anyone who is not afraid to find them.

We use argumentation known as *inference* (Ross, 1924), to rationalize something within our mind. This process increases or decreases our proclivity to believe the thought we wanted to rationalize or helps us see why it must be true. The question of why something must be true or even why something *is* does not simply make us think, but makes us think causally.

To gain a perspective on the theory of causality, we will embark on a brief guided tour through its history. In the middle of Fig. 1.1, we see a statue of Aristotle, who was the first to introduce a theory of causality (Falcon, 2023) in his efforts to understand how humans experience the physical world around them. He built upon and formalized the thoughts of his contemporaries, conceiving four types of explanations

Figure 1.1: ***The Hall of Causation.*** *Illustration created by the author.*

that comprise causation: material, formal, efficient, and final cause. To determine those causes and learn why something is in its current state or form, one needs to answer the following questions (Ross, 1924): Out of which material is something composed? What shape or form does it have? What is its primary source of the change or rest? What purpose does it fulfill?

Aristotle's thoughts on causality were so profound that they influenced philosophers centuries and even millennia later. In the middle ages, the most prominent thinker was Thomas Aquinas (the third portrait on the left in Fig. 1.1). He mainly relied on early causation ideas to rationalize the cause of existence. Moreover, he expanded the theory of causality by stating that the final cause comes first and manifests itself through the efficient causes, which can be regarded as instrumental, but subservient to it (Aquinas, 1265-1274).

The idea of causation developed further in the next few centuries and was slowly becoming more scientific. However, it was still orbiting around proving the existence of a divine creator. René Descartes, a great thinker, philosopher, and mathematician, whose portrait is the second on the left in Fig. 1.1, supported the idea that the cause of anything needs to contain at least as much reality as the effect, either formally or eminently (Williams, 1978). For instance, a cup of tea is hot because of the tea being hot, making the tea temperature a formal cause of the state of the cup. However, possessing a water boiler (and a tea bag) is an eminent cause of the cup heating up.

In order to grow and improve any idea, no matter how great it seemed at a time, it should be analyzed, challenged, and open for discussion. Gottfried Wilhelm Leibniz, portrayed the first on the left in Fig. 1.1, was the one to do precisely this regarding his predecessors' views on causation. He reasoned their arguments away and proposed a different theory of cause and effect. Namely, he stood firmly behind the thought that each monad, a simple substance reflecting the order of the world, had inherent properties according to which it would be the only source of its modifications (Hulswitt, 2002). In this sense, all cause-effect relationships are completely predestined, and necessarily, the effects logically follow their causes.

Now we turn to the first portrait on the right in Fig. 1.1, depicting Leibniz's contemporary, English scientist Isaac Newton. Newton was one of the prominent individuals contributing to the philosophical revolution called *the Enlightenment*, and equally revolutionary was his stance on causality. He explicitly denies that every event must have a cause, implying that there is no law of universal causation (Collingwood, 1938). This further indicates that a fundamental distinction must be made between causation and a law-like behavior, such as motion according to his first law (Hulswitt, 2002).

Another mathematician and philosopher to disagree with his predecessors and conceive a different concept of the cause was Immanuel Kant, portrayed next to Newton in Fig. 1.1. He was provoked to thought after reading the work of Hume that causation cannot be rationally justified, yielding that one cannot rationally account for scientific knowledge (Hulswitt, 2002). A new idea of Kant that would mitigate this issue was to declare causality an a priori concept. He further differentiated between effects following their causes as a matter of fact and following them necessarily. Kant stood by the latter principle and claimed that it was governed by a universal rule known to us a priori and not from experience (Hulswitt, 2002).

The last portrait in Fig. 1.1 on the right depicts an English philosopher, and a political economist, John Stuart Mill. He criticized the contemporary notion of a cause as being selected from a broader set of causes just because it occurred the last or was superficially the most apparent (Mill, 1874). However, he did not define cause as just the set of all involved conditions necessary for the effect to materialize, but it had to do so unconditionally (Hulswitt, 2002).

The balance between these differing views on causality throughout history would be recognizing that there are more causes to a specific event but that not all of them are relevant. Therefore, the correct course of further action is prioritizing causes that could have the highest impact on future events. In the more modern and formal sense, however, the concept of causality still takes multiple shapes.

In 1956, a US-American mathematician Norbert Wiener proposed that one variable could be called 'causal' to another if the prediction of the second one is improved by including information about the first (Wiener, 1956). This idea was practically implemented by the economics Nobel Prize winner, Granger (1969), and the concept is known as the Wiener-Granger causality ever since.

Forty years later, a computer scientist, Judea Pearl, proposed graphically analyzing causal effects using a tool known as the calculus of interventions, or *do*-calculus (Pearl, 2009). We introduce this causality analysis tool in detail in Chapter 2 and rely on it throughout this thesis.

We now end the guided tour through the history of causation, as we proceed to explore the ways of using expert knowledge, which can aid in directing automated causality analysis and decision-making on specific domains.

1.2 Domain knowledge integration

The purpose of defining causal reasoning throughout history was to uncover the truth and gain ever more knowledge about the world.

In our time, when vast amounts of data are becoming available each day, one could assume that, in turn, at least as much knowledge is gained daily. That is still unfathomably more than even a few years ago. However, it is one thing to collect data, and another, usually much more complex, to extract meaningful information from it and draw conclusions resulting in new knowledge.

In the realm of machine learning research, we are not only interested in obtaining useful information from data but also in having a computer learn everything it possibly can about it. We then either want it to tell us if it can recognize a similar data example, generate a new one, or describe the underlying process according to which the data was created in the first place. In some cases, no matter how much data we use to train the machine, obtaining the underlying data-governing process is impossible. This obstacle is illustrated by the model of population growth, known as the *logistic map* (Verhulst, 1838):

$$x_{t+1} = r \cdot x_t \cdot (1 - x_t), \tag{1.1}$$

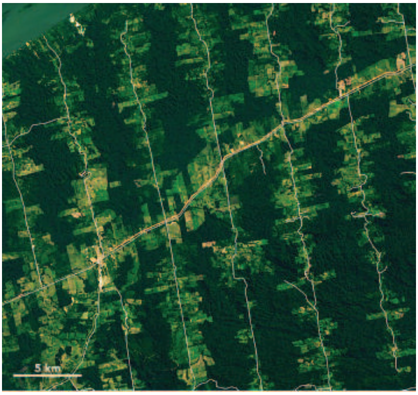


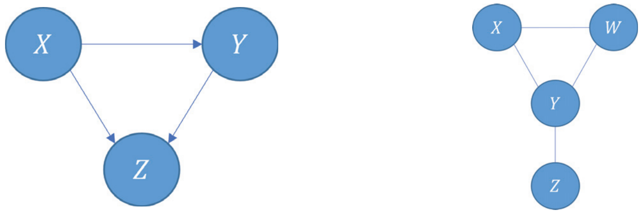
Figure 1.2: *Deforested region of Amazon rainforest.*
Source: earthobservatory.nasa.gov

for $x_t \in [0, 1], t \in \mathbb{N}, r \in [0, 4]$. It is shown to be non-learnable when $r = 4$, as its behavior becomes chaotic (Wang, 2017).

Climate science is another great example of an entire research domain where comparably few advances have been made despite big data. Due to data complexity and the ever-changing Earth system, Faghmous and Kumar (2014) made a case for theory-guided data science, *i.e.*, for domain knowledge integration.

Expert knowledge that gets integrated to improve machine learning architectures' efficacy considerably varies in format and type, depending on the domain. It can take a form of an equation describing certain physical phenomena, textual information gathered from reliable online sources, an insight about a pattern in remote sensing images denoting deforestation to experts (see Fig. 1.2), and many others.

We will now introduce a concept that, along with causality analysis, allows for domain knowledge integration in its more general form, in an attempt to make certain forms of reasoning more accessible to computers.



(a) *Bayesian network* (b) *Markov random field*

Figure 1.3: *Examples of probabilistic graphical models.*

1.3 Probabilistic graphical models

Artificial neural networks' lack of interpretability is the main obstacle currently preventing the wider use of artificial intelligence in fields where the rationale behind critical decision-making is of utmost importance, such as medicine and finance. One step towards mitigating this problem is using Probabilistic Graphical Models (PGMs) (Koller and Friedman, 2009) alongside purely data-driven methods.

PGMs are a framework combining uncertainty and logical structure through the use of probability, graph theory, and independence constraints. More precisely, they are graphical representations of complex multivariate joint distributions between random variables shown as nodes and connected with edges. The most well-known types of PGMs are directed acyclic graphs (DAGs), also known as Bayesian networks (BNs) (Howard and Matheson, 2005; Pearl, 1988), and undirected graphical models or Markov Random Fields (MRFs) (Kindermann and Snell, 1980), respectively shown in Fig. 1.3a, and Fig. 1.3b for random variables W, X, Y, and Z. Particularly, in addition to being directed, the underlying graph of a BN must also be acyclic.

Inspired by the argumentation process happening in our minds when trying to arrive at a logical conclusion with high certainty, PGMs allow reasoning about random variables of interest by using their distribution for inference. Namely, certain inference algorithms can be applied to estimate the posterior probability of desired variables given information about specific others. For instance, to discover how likely it is for a hurricane to form over the Atlantic, let us assume that it is summer and that the winds are stronger than usual. An inference algorithm can then compute the probability $P(\text{Hurricane} = \mathit{true} \mid \text{Season} = \mathit{summer}, \text{WindSpeedIncrease} = \mathit{true})$ and give us an estimated answer.

These probabilistic models can be effectively constructed using expert knowledge within a certain domain or by learning the model from data. Nevertheless, they are the most effective when these two approaches are employed together. That is because experts can provide important attributes and general guidelines the model should contain and follow. In contrast, obtaining details is more potent when done automatically by fitting the model to data (Koller and Friedman, 2009).

The success of this synthesis culminated through the use of graphical models in combination with deep neural networks (Kingma and Welling, 2014; Chung et al., 2015; Krishnan et al., 2017).

1.4 Variational inference

Uniting deep learning and PGMs would not have been possible without *variational inference* (Blei et al., 2017). Complicated probability distributions of PGMs used to be approximated by drawing samples via Markov chain Monte Carlo (MCMC) methods described by Robert and Richardson (1998). However, with more data flowing through deep learning pipelines, this approach needed to be replaced by a more efficient one. That is why variational inference was introduced to find a similar distribution that minimizes the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951).

Variational inference has been applied in computer vision, computational neuroscience, and large-scale document analysis (Blei et al., 2017).

The first method emerging as a union between deep learning and graphical models was the deep Boltzmann machine (Salakhutdinov and Hinton, 2009). Furthermore, a deep graphical model with the most traction at present is a Variational Autoencoder (VAE) by Kingma and Welling (2014). We will discuss the VAE in more detail in Section 2.5.1, and in Chapter 7 we explain how its variants can be expanded for multivariate time series causality analysis in difficult real-world scenarios.

1.5 Challenges

We will now examine some of the most prevalent challenges of time series causality analysis and justify using deep graphical models to tackle them. We note that some methods solve a few of the following challenges simultaneously (see Section 7.2), but they are not as effective at tackling a larger multitude of such obstacles at once. Moreover, they might be limited to a specific domain.

One of the first challenges of causal link identification for classical time series causality analysis models, *i.e.*, models not using deep learning, arises when **causal links are nonlinear**. The nonlinearity occurs, however, for many interactions in domains such as neuroscience, finance, and climate science (Weber and Oehrn, 2022). For this reason, using neural networks emerged as a natural choice due to the Universal Approximation Theorem (see Theorem 2.4.1), *i.e.*, the neural networks' ability to model arbitrarily complex functions.

Another issue specific to time series data is the occurrence of **delayed causal effects**. In environmental science, for instance, variables can affect one another with a delay, known as *time lag*. One notable method that tackles this problem was introduced by Sugihara et al. (2012), called Convergent Cross Mappings (CCMs). It aims to measure how well the cause variable X can be reconstructed from the history of the effect variable Y, assuming that both variables stem from the same dynamical system. Using a Neural ODE latent process modeling, CCMs were further extended by Brouwer et al. (2021a) to incomplete observations at irregular intervals. Their method is an example of how using neural networks can improve a successful causality analysis method and make it more data-driven even when data is imperfect.

Causal relationships of real-world data that originates from dynamical mechanisms over time, also known as *dynamical systems*, change according to those mechanisms. This often makes that data **nonstationary** (Box and Jenkins, 1970; Hargreaves, 1994), which promotes misleading results if an unsuitable method that assumes stationarity is applied for causal discovery.

Furthermore, in real-world scenarios, two variables we are causally interested in are often influenced by a third, unobserved one. This is the problem of **hidden confounding**, and the unobserved variable is referred to as the *hidden confounder*. When it occurs, the causal link between the variables of interest may be falsely detected due to the influence of the hidden confounder. One of the approaches to determine if the causal link is actually present is to use deconfounding methods (Bica et al., 2019; Hatt and Feuerriegel, 2021) and generate proxies of the hidden confounders. Another strategy would be to intervene on the potential cause variable in the sense of *do*-calculus (Pearl, 2009) to remove the influence of the confounding variable.

However, it is often the case that controlled experiments using **direct interventions are infeasible, costly, or unethical**. For example, one cannot intervene on the air temperature over a particular geographical region, and should not manipulate experimental results to support a certain hypothesis. Deep learning is particularly beneficial in such cases as it enables data-driven approaches and, in combination with counterfactual reasoning, it can provide useful answers without tangibly altering the examined system.