Introduction To Causality: A Modern Approach

Florian Brezina

Contents

Welcome

This is the HTML version of **“Introduction to Causality: A Moden Approach”**, a gentle but rigorous introduction into the art and science of causal inference. This book covers the basics of causal inference: you will learn

- how **causal inference** differs from **statistical inference** or **prediction**;
- how to express these differences in **unambiguous mathematical notation** and **causal graphs**;
- a variety of techniques to probe causal questions, from **randomized controlled experiments** to **structural equation models**;
- the current scientific edge on causal analysis, including reinforcement learning.

We approach this topic by closely examining most simple scenarios first and build upon those chapter by chapter. Throughout the book, we will use modern notation and language, primarily following Pearl (2000).

The book further contains an extensive appendix containing code snippets in the statistical programming language R as well as auxilliary material on statistics. We hope that this will allow the book to be a good standalone source for all those interested in causality, whether or not they have a solid foundation in statistics.

Chapter 1

Introduction

> Correlation doesn't imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there.'
> — Randall Munroe

Causal analysis is a fascinating field. It deals with the fundamental relation between cause and effect in complex environments. Being able to infer what the effect is going to be after doing A versus B is of utmost importance in a wide variety of applications, from policy analysis, drug prescription to marketing. Despite its ubiquity in all disciplines concerned with complex phenomena, the concept of **causality** has eluded a mathematically rigorous treatment for a long time, resulting in puzzling paradoxes and ambiguous statements. Only in recent decades a new formalism has emerged to solve these problems, with main contributors from the computer science and economics departments. The **causal revolution** (Pearl and Mackenzie 2018) swept away decades of experts and students arguing about the correct interpretation of phenomena such as **Simpson's paradox**, the nature and properties of the **error term in regression equations** and the interpretation of **structural parameters** in SEMs. The revolution established a new regime, which introduced new notation and a unified language, where causal and statistical concepts are finally separated. At last, **correlation** is never again to be confused with **causation**.

"Introduction To Causality" is a gentle introduction into this modern understanding of causality as it unfolded after the revolution. It will help you learn the fundamentals of this art and science of causal analysis. After reading this book, you'll have the tools to understand and communicate causal concepts and you will know how to tackle the common questions. The code in the appendix will help you to apply these methods using the R programming language.

1.1 What you will learn

First we will discuss causality in a trivial lab environment where we are able to control every important aspect of the environment. This will help us get familiar with the vocabulary and notation and will provide some insights in how to think about causality.

Once we have become familiar with the simple setting, we will loosen the assumption that we are able to fully control the environment. At this point, we will introduce causality as a probabilistic concept. Our inability to fully control and understand our environment forces us to settle for a less precise inference on the effects. We can't say what will happen, but we can still provide robust inference on what will happen **on average**. Most importantly, we will see why classical statistical concepts and notation are not sufficient for a rigorous and unambiguous treatment of causality and we will get a sense that there is a fundamental difference between what can be learned from passive observation ("correlation") and active intervention ("causation").

Once we mastered the probabilistic nature of causality we will discuss on how we can **measure** the effect of **actions** in a variety of settings. We will start with the easiest scenario, the randomized controlled trial. It is often considered the gold standard for clinical trials and applied across scientific disciplines. It will serve as a benchmark in our further discussion, where we will look at scenarios that violate the assumption behind the randomized controlled trial: we will discuss observational studies, synthetic cohorts and time series analysis.

After this tour de force, we shift gears and have a closer look at a couple of applications. We will discuss how to measure and interpret the placebo effect in clinical trials, how to optimize marketing using A/B tests and multi-armed bandits, and how to evaluate government interventions.

Finally, I will wrap up this book by providing some parting thoughts on epistemology and the importance of causality in the evolution of artificial intelligence and machine learning.

These chapters will hopefully provide you with a solid foundation and will allow you to find the right solution for your causal problem. But for most of your problems they will not be enough. Throughout this book we'll point you to resources where you can learn more. The appendix provides some additional material on statistics and programming.

1.2 What you won't learn

There are some important topics that this book doesn't cover. We hope that this book will leave you wanting more and that you will continue in your journey to master causality by going deeper into this topic or by exploring closly related fields and applications that we did not cover in sufficient length.

1.2.1 Statistics

The book focuses on causal inference rather than statistics. Some basic statistical concepts are discussed in the appendix, but they primarily serve as a refresher. We assume that the reader is (or has been) familiar with statistics as it is taught in most Statistics 101 classes. Details on estimation methods and properties of estimators (e.g. consistency) are not discussed. We will provide references that provide more details. We will, however, extensively discuss the differences between these two types of inferences and how they relate. Our discussion on causal inference will, except for the basic introduction, be probabilistic in nature and statistical notation will be used throughout the book. We have summarized information on notation and terminology in CHAPTER XX.

1.2.2 Machine Learning

We will address issues of machine learning where we see a connection to causal concepts. We do not go deep on any causal and non-causal ML algorithms. The discussion will focus on the discussion of supervised ML versus reinforcement learning.

1.2.3 Proofs

The book does not contain any proof or any heavy mathematical derivations. We will link to reference material. Despite that, we do intend to be rigorous in argumentation and notation and some discussions might seem overly verbose at first. We believe however that this is necessary, especially to avoid confusion between statistical and causal concepts.

1.2.4 Type Causality vs Actual Causality

When referring to causality, we will always mean what philosophers typically call *type causality* rather than *actual causality*. The former takes a forward-looking approach by inferring the *effects of causes*. This allows to predict outcome for interventions, e.g. it allows to answer questions like "what will be the outcome if we prescribe this new drug X to patients with heart disease". *Actual causality* instead mostly takes a backward look at a given instance and tries to infer *causes of effects*. This allows to answer questions such as "what caused person Y to die from heart disease". This type of inference is important if the goal is to assign responsibility, e.g. in a legal case. For a thorough introduction I recommend to take a look at (Halpern 2016).

1.3 How this book is organised

1.4 Prerequisites

To get the most out of this book, you should be familiar with basic concepts of statistical analysis, nomeclature and notation. If “expected value”, “conditional probability” or “hypothesis test” are only vaguely familiar to you, please review the appendix before digging into the main text.

The code snippets at the end of the book are purely optional. If you want to follow along on these, you need to have R on your computer. To download the software, go to CRAN, the **c**omprehensive **R** **a**rchive **n** etwork. CRAN is composed of a set of mirror servers distributed around the world and is used to distribute R and R packages. Don’t try and pick a mirror that’s close to you: instead use the cloud mirror, https://cloud.r-project.org, which automatically figures it out for you. RStudio is an integrated development environment, or IDE, for R programming. Download and install it from http://www.rstudio.com/download.

1.5 Acknowledgements

The book has been compiled from markdown documents using R package bookdown. This package has allowed me to adopt a very flexible workflow where the compilation and publication of an HTML version only takes seconds.

1.6 Links

A free HTML version of this book is available at https://flobrez.github.io/itc/. The markdown sources and supplementary material is available at https://github.com/flobrez/itc.

Chapter 2

Causal Models

2.1 Causality, Asymmetry and Entropy

Causality is strongly linked to the concept of *time*. Cause precedes effect, never the other way round. Symptoms occur after infection This asymmetry is mirrored in the physical notion of entropy and (as an emergent property) time. While the past is determined we feel that we are able to act on the future, that we are able to choose one among many possible futures. This is due to the low entropy the universe had in the past. All fundamental physical laws are perfectly symmetrical and therefore reversable. Asymmetry is only introduced by a coarse-grained look at the world, and therefore an *emergent* property.

2.2 Basic Definitions

We assume the world can be modelled by *variables*. Variables can take various values. The variables themselves are denoted by upper-case latin letters, e.g. X, whereas we use lower-case letters for their values, e.g. x. In case X is *categorical*, different values will be denoted by a subscript x_j. Where X has two values only, we will encode them with 0 and 1.

2.2.1 Causal Graphs

Definition 2.1 (Causal Graph)**.** A graph is a mathematical structure. It consists of a set of nodes and and a set of edges, where edges connect ordered pairs of nodes. In *causal graphs*, nodes represent variables; edges represent the causal relation from cause to effect. Note that in a causal graph, an edge is an *ordered* pair of nodes, the edge therefore directed. In most graphs in this book, we will consider causal systems that can be represeted as directed acyclical graphs (DAGs)[1]. These DAGs have no feedback loops.

The causal graphs convey the qualitative pattern of causal relations. They do not quantify that relation, i.e. specify how two variables are related. A graph with relation $A \to B$ It The quantitative aspects are better represented in a set of structural equations.

Definition 2.2 (Exogeneous and Endogeneous Variables)**.** An exogeneous variable in a graph G has no edges pointing into itself. An endogeneous variable in a graph G has at least one edge point into itself.

2.2.2 Structural Equations

Structural equations represent the causal relations between *variables*. The *absence* of a variable from the model assumes that it is not relevant for the causal description of the system. We will focus exposition on *categorical variables* which can assume a

$X \to Y$ means that X causes Y. Manipulating X determines the value of Y, but not the other way round. We call X the *cause* and Y the *outcome*. Others call Y the "*effect*", but we will use *effect* to denote changes in the outcome due to manipulations of the cause. This is in line with conventions in statistical literature (e.g. "average treatment effect") and its usage in everyday language (e.g. "tipping on that button had no effect on the brightness of the screen").

2.3 Simple Environments

2.3.1 An electric circuit with one switch

Let's first take a look at a most simple environment, shown in figure xxx. It represents a circuit diagram with a voltage source, a switch (X) and a lamp (Y). Both, X and Y, can assume one of two states. We will encode these as 0 and 1:

- switch is open (0) or closed (1)
- light bulb is off (0) or on (1).

This system is very easy to reason about. Assuming that the power source has enough capacity, the light bulb will be on if and only if the switch is closed. The system can be in one of only two states:

[1] Readers familiar with DAGs data processing pipelines will recognize that these too describe causal mechanisms. Datasets are manipulated in an ordered sequence of steps to produce a final outcome where the result of each step is determined by the outcome of its parents steps (the input datasets) and the mechanism itself (the transformation of the datasets).

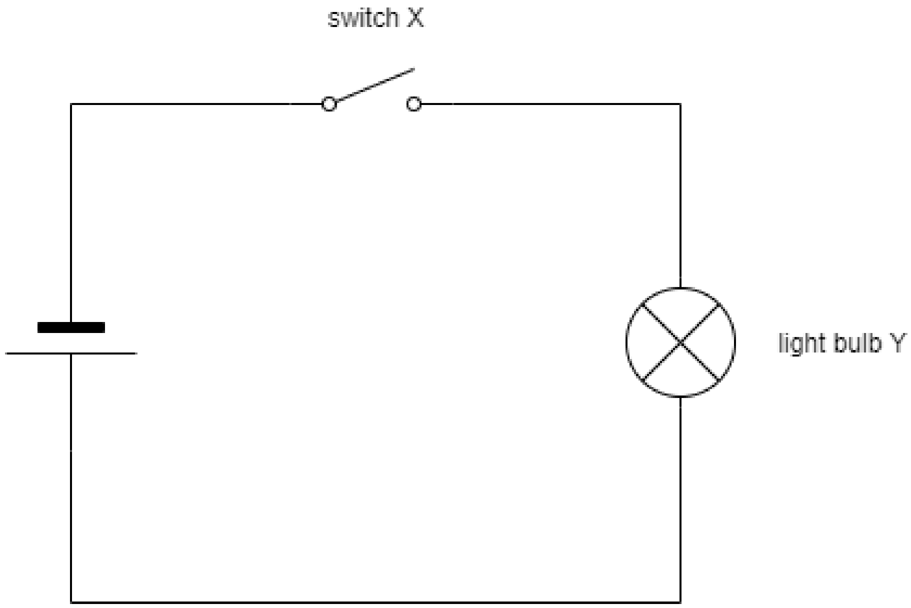

Figure 2.1: A simple circuit diagram with a single power source, a switch (X) and a light bulb (Y)

switch X	light bulb Y
0	0
1	1

This table, however, does not contain any information on the causal relationship between X and Y and will therefore not be sufficient to correctly reason about the system. Hence, let's take an extra second to translate this circuit diagram into a *causal graph*. This will become much more useful in more complex systems, but it's probably a good idea to get used to this representation from the start.

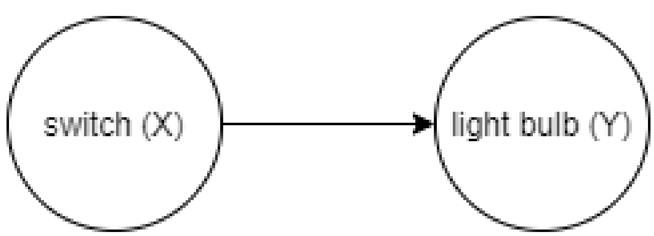

Figure 2.2: The simple circuit diagram converted to a causal graph, where X is the (only) cause of Y

A *causal graph* represents variables as *nodes* and causal relationships as *directed edges* between nodes. $X \to Y$ means that X causes Y. This in turn means that manipulating X determines the state of Y, but not the other way round. Here, the graph consists of two nodes, the switch X and the light bulb Y, connected by a directed edge from X to Y.

While the graphical representation of the causal structure makes it easy to *qualitatively* reason about the causal structure of the system, an algebraic representation will be needed for quantitative analysis. The algebraic representation of a causal relationship is called a *structural equation*. In any system we have one infer $f()$ right away:

$$Y := f(X) = X \tag{2.2}$$

Now that we managed to represent this system in various forms, we can start to reason about *interventions*. An *intervention* is an operation in the system that fixes a variable to a certain value. Here, two interventions are of interest: we can open the switch, i.e. set $X := 0$, or close it, i.e. set $X := 1$. The *structural equation* is then

$$Y := f(X) = 0 \tag{2.3}$$

and

$$Y := f(X) = 1 \tag{2.4}$$

respectively.

This simple system has a nice property: for the light bulb to be on, the closed switch is a *necessary* and *sufficient* condition. This property is “nice” as it allows us to falsify the causal model from observation: a *single* observation where the light bulb is on but the switch is open (or the light bulb is off but the switch is closed) allows us to refute the causal model, e.g. the power source might not be strong enough or the circuit might have flaws.

2.3.2 An electric circuit with two switches

Let us now slightly increase the complexity of the system by adding a second switch. Both switches are connected in series as shown in the circuit diagram in figure XXX

Again, our understanding of the physical nature of this system allow us to derive the corresponding causal graph. Both switches cause the state of the light bulb whereas the light bulb does not cause any of the switches and switch X_1 does not cause switch X_2, nor vice versa. Hence our causal graph has three nodes and directed edges $X_1 \to Y$ and $X_2 \to Y$:

The structural equation for this system is also very simple:

$$Y := f(X_1, X_2) \tag{2.5}$$

but the functional form of $f()$ might not be obvious. The two switches allow the environment to be in four different states. Only if *both* switches are on, will the light bulb be on, in the other three states it will be off:

switch X_1	switch X_2	light bulb Y
0	0	0
0	1	0
1	0	0
1	1	1

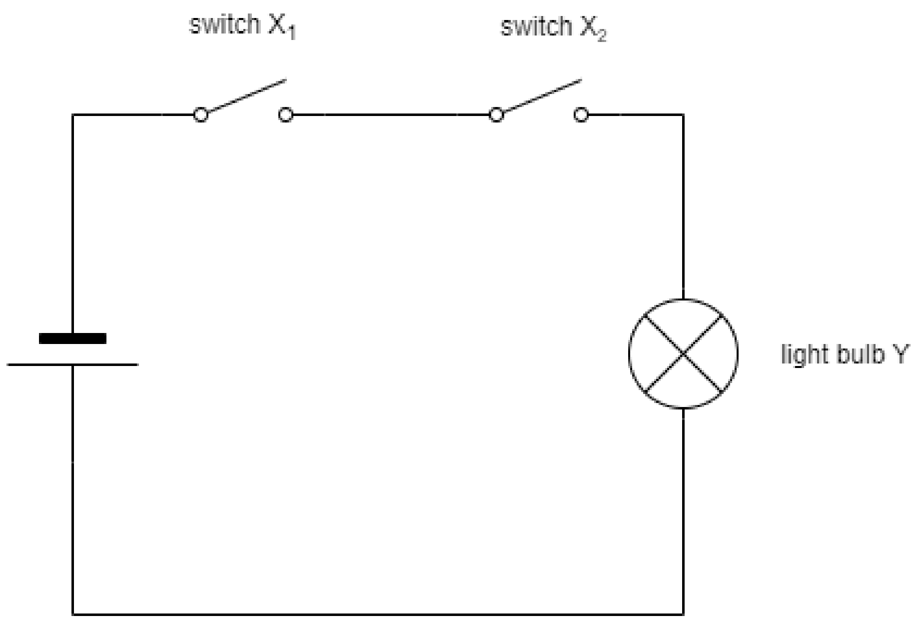

Figure 2.3: A circuit diagram with a single power source, two switches (X1 and X2) and a light bulb (Y)

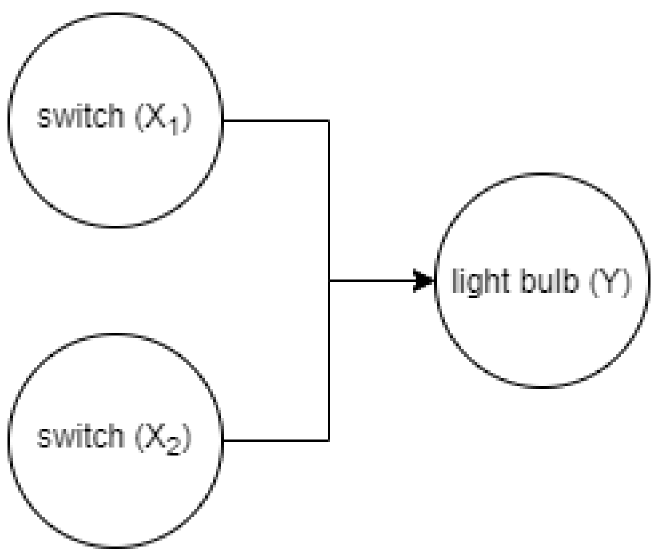

Figure 2.4: The simple circuit diagram converted to a causal graph, where X is the (only) cause of Y

Hence, the structural equation has to be

$$Y := f(X_1, X_2) = X_1 \cdot X_2 \tag{2.6}$$

For the light bulb to be on, $X_1 = 1$ and $X_2 = 1$ are necessary conditions but neither is (on its own) sufficient.

Let us now turn to interventions in this system. While in the system with just one switch, every intervention on X caused a *change* in the state of Y, this is not true in the case of interventions on a single switch now. In a state where $X_1 = 0$, we are unable to change the state of Y by intervening on X_2:

$$Y := f(0, X_2) = 0 \cdot X_2 = 0 \tag{2.7}$$

Conversely, if the first switch is closed, $X_1 = 1$, we're basically back in a system with just one switch, which solely determines by the state of the light bulb.

$$Y := f(1, X_2) = 1 \cdot X_2 = X_2 \tag{2.8}$$

2.3.3 Unobservability

So far, we were able to reason about the effectiveness of interventions due to our ability to fully specify the structural equations (i.e. we knew $f()$) *and* were able to *observe* the state of all causes. This allowed us to reason that closing switch 2 will *not* have an effect on the light bulb when switch 1 is open, but will change its state if switch 1 is closed.

Problems arise, when one of these conditions fails. Suppose we are still dealing with the system with two switches connected in series, but the state of switch 1 cannot be observed as the switch is hidden in a box, see figure XXX.

Imagine now that we observe that switch 2 is open (and the light bulb is off). Will closing switch 2 turn on the light?

Unfortunately, we are not any more able to answer this question. Our best answer is "It depends.". It depends on the state of switch 1, which is not observable. What were left is to resort to a different kind of reasoning, *probabilistic* reasoning. While we're not able to make any statements about any single system of that kind, we can still make statments on the likelihood of X_2 effectiveness based on probability distributions for X_1. If we knew that the likelihood that switch 1 is closed is 0.8 in all cases where we encounter switch 2 to be open and and the light to be off, then closing switch 2 will turn on the light in 80% of the cases.

2.3.4 Probabilistic Models of Causality

This example has shown that even in very simple causal systems, not being able to observe (or accurately measure) a single variable requires us to revert