



CentraleSupélec

Case Study : Causal Inference

Toward Ethical Algorithms ?

MOHAMMED FELLAJI, AHMED BEN AISSA

SUPERVISOR : FRÉDÉRIC PENNERATH

September, 2020

Contents

1	Introduction	2
2	Causation vs Association	3
2.1	Example	3
2.2	Definitions	3
3	Association - Intervention - Counterfactuals	3
4	Paper : Judea Pearl	4
5	Notations	4
6	definition of counterfactual	4
7	Simpson's paradoxe	4
8	Average Causal Effect	4
9	Definition of the causality	4
10	Python Library : DoWhy	5
11	Causal discovery	5

1 Introduction

One of the most challenging questions in every problem is the one related to understanding the reason why an action happened and whether or not we can explain it with the information we have in our disposal. Another interesting one might ask is what would be the outcome if the condition of the experiment were different ? What would happen if we have more/less information ? What will we get if we change completely the set of inputs ?

When thinking about these questions, having a time machine seems as the perfect solution : we can then repeat the same experiment with different initial conditions and record the outcome in every scenario. A more realistic a possible solution would be to use Causal inference, which aims to estimate the likelihood of an event under static conditions and also under dynamic changing conditions.

When searching about causal inference, one will definitely come across some of the work of Judea Pearl who is credited for developing a theory of counterfactuals and causal inference based on structured models.¹

In this document, we will try to study the different literatures about causal inference and collect the results in a simple and yet detailed way. Many papers and book are mentioned in the references section, some were used in writing these document, some are not. Those interested more in the subject may take a look at them.

¹After watching dozens of Judea Pearl lectures and reading many of his papers, we can only recommend doing the same. One of his main ideas is that even if machine learning is shaping millions of industries around the globe, this is done without attention to fundamental theoretical impediments. This might lead machine learning algorithms, very soon, to reach the barriers of impossibility. According to Judea Pearl, the goal is to use the knowledge acquired from causal inference and combine it with the success of machine learning in order to achieve more general models.

2 Causation vs Association

2.1 Example

[1] [2] One of the most common phrases in statistics is "correlation does not imply causation". To understand it well, let's take a look at the following example :

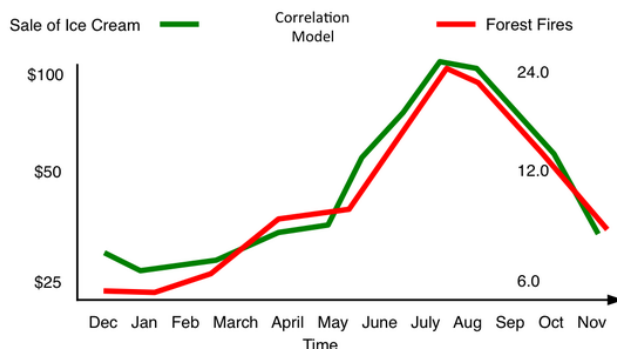


Figure 1: an example of correlation : source

Without having any information about what we are trying to model, we might conclude that there is a cause-effect relationship between these two measures. The graph also shows a correlation close to 1 for the two curves. When we start to analyse the data in details, the first thing that comes to our mind is that there is no logical relationship (and thus no cause-effect relationship) between the sale of ice cream and the forest fires. However, we can see clearly in the graph that the values are higher between May and November (with a peak in July) compared to the rest of the year. In fact, the heat is the reason behind forest fires and the sale of ice cream. We can then conclude the following :

- a causation between the heat and the sale of ice cream;
- a causation between the heat and the forest fire;
- a correlation between the sale of ice cream and the sale of ice cream.

This simple example shows us that, in general and without having a good knowledge about the problem, we tend to assume simple correlations between 2 variables when in fact there is a third variable that causes both of them. This is a core idea in causal inference : unlike for correlation, we can not rely only on the distribution of the data, even at the population level, but we also should rely on causal assumption that is always not testable in observational studies.[3]

2.2 Definitions

3 Association - Intervention - Counterfactuals

One of the very interesting papers that explains causal inference is : "The Seven Tools of Causal Inference, with Reflections on Machine Learning" [1]. This section is thus inspired mostly by this

paper.

[1]

4 Paper : Judea Pearl

see [3]

One of the best paper so far. It explains clearly the difference between the causal concept and the associational concept (for example : correlation). The first concept ...

5 Notations

see [4] and [2]

6 definition of counterfactual

7 Simpson's paradox

- give definition :
- give example : Cholesterol

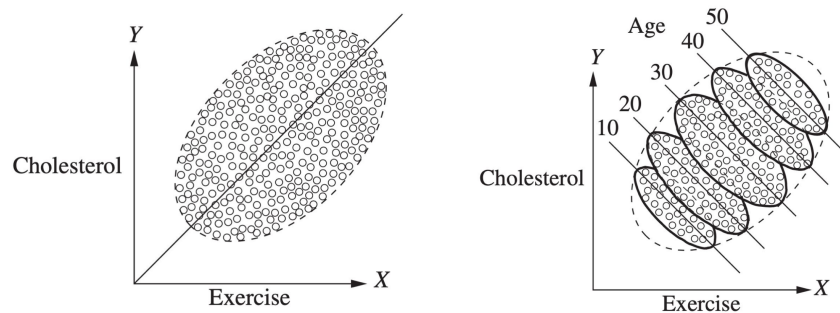


Figure 2: Illustration of the Simpson's Paradox[12]

8 Average Causal Effect

see [2]

9 Definition of the causality

Perhaps the most important message of the discussion and methods presented in this paper would be a widespread awareness that (1) all studies concerning causal relations must begin with causal

assumptions of some sort and (2) that a friendly and formal language is currently available for articulating such assumptions.[3]

[5]

10 Python Library : DoWhy

library developed by Microsoft [6], blog article available link

11 Causal discovery

A traditional way to discover causal relations is to use interventions or randomized experiments, which is, however, in many cases of interest too expensive, too time- consuming, unethical, or even impossible. Therefore, inferring the underlying causal structure from purely observational data, or from combinations of observational and experimental data, has drawn much attention in various disciplines. With the rapid accumulation of huge volumes of data, it is necessary to develop automatic causal search algorithms that scale well.[7]

References

- [1] J. Pearl, “The seven tools of causal inference, with reflections on machine learning,” *Communications of the ACM*, vol. 62, no. 3, pp. 54–60, 2019.
- [2] M. A. Hernán and J. M. Robins, “Causal inference: what if,” *Boca Raton: Chapman & Hill/CRC*, vol. 2020, 2020.
- [3] J. Pearl, “The mathematics of causal relations,” *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures (P. Shrout, K. Keyes and K. Ornstein, eds.)*. Oxford University Press, Corvallis, OR, pp. 47–65, 2010.
- [4] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, and A. Zhang, “A survey on causal inference,” *arXiv preprint arXiv:2002.02770*, 2020.
- [5] D. B. Rubin, “Causal inference using potential outcomes: Design, modeling, decisions,” *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 322–331, 2005.
- [6] “DoWhy: A Python package for causal inference.” <https://github.com/microsoft/dowhy>, 2019.
- [7] C. Glymour, K. Zhang, and P. Spirtes, “Review of causal discovery methods based on graphical models,” *Frontiers in Genetics*, vol. 10, p. 524, 2019.
- [8] A. Marx and J. Vreeken, “Testing conditional independence on discrete data using stochastic complexity,” *arXiv preprint arXiv:1903.04829*, 2019.
- [9] B. Schölkopf, “Causality for machine learning,” *arXiv preprint arXiv:1911.10500*, 2019.
- [10] J. Pearl *et al.*, “Causal inference in statistics: An overview,” *Statistics surveys*, vol. 3, pp. 96–146, 2009.
- [11] P. Judea, “Causality: models, reasoning, and inference,” *Cambridge University Press. ISBN 0*, vol. 521, no. 77362, p. 8, 2000.
- [12] J. Pearl, M. Glymour, and N. P. Jewell, *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- [13] J. Pearl, “The science and ethics of causal modeling,” *Handbook of ethics in quantitative methodology*, pp. 383–416, 2011.
- [14] H. Matute, F. Blanco, I. Yarritu, M. Díaz-Lago, M. A. Vadillo, and I. Barberia, “Illusions of causality: how they bias our everyday thinking and how they could be reduced,” *Frontiers in Psychology*, vol. 6, p. 888, 2015.