



CentraleSupélec

Case Study : Causal Inference

*Toward Ethical Algorithms ?*

MOHAMMED FELLAJI, AHMED BEN AISSA

SUPERVISOR : FRÉDÉRIC PENNERATH

September, 2020

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Fundamental notions</b>	<b>3</b>
2.1	Causation vs Association . . . . .	3
2.1.1	Example . . . . .	3
2.1.2	Definitions . . . . .	3
2.1.3	The difference between the 2 notions . . . . .	4
2.2	Counterfactuals . . . . .	4
2.3	Confounders . . . . .	5
2.4	Simpson's paradox . . . . .	5
<b>3</b>	<b>Causal inference</b>	<b>6</b>
3.1	Individual level Causal Effect : ICE . . . . .	6
3.2	Average Causal Effect : ACE . . . . .	6
3.3	Standard estimator : $S^*$ . . . . .	6
3.4	Fundamental Problem of Causal Inference . . . . .	6
3.5	Randomised experimentation . . . . .	7
3.6	The Causal Hierarchy . . . . .	7
3.7	Graph representation . . . . .	8
<b>4</b>	<b>Paper : Judea Pearl – remove this</b>	<b>9</b>
<b>5</b>	<b>Notations – remove this</b>	<b>9</b>
<b>6</b>	<b>Definition of the causality – remove this</b>	<b>9</b>
<b>7</b>	<b>Causal discovery</b>	<b>10</b>
<b>8</b>	<b>Python Library : DoWhy</b>	<b>10</b>

# 1 Introduction

One of the most challenging questions in every problem is the one related to understanding the reason(s) why an action happened and whether or not we can explain it with the information we have at our disposal. Another interesting question one might ask is what would be the outcome if the conditions 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.<sup>1</sup>

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.

---

<sup>1</sup>After 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 Fundamental notions

### 2.1 Causation vs Association

#### 2.1.1 Example

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 :

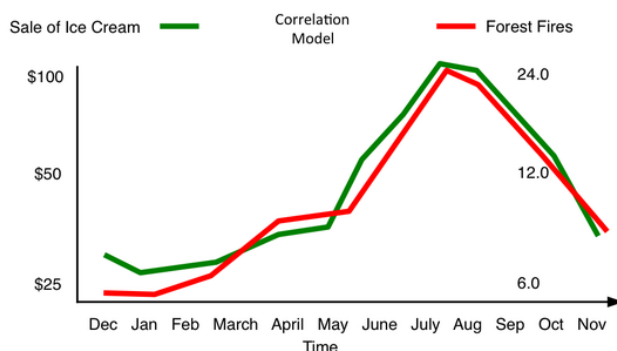


Figure 1: an example of correlation (source)

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.[1]

#### 2.1.2 Definitions

The previous example shows a clear difference between causation and correlation. More generally, correlation is a special example of an association concept. The definition of these two concepts is given by Judea Pearl[1, 2] as follows :

- An **associational concept** is any relationship that can be defined in terms of a joint distribution of observed variables. For example: regression, correlation.
- A **causal concept** is any relationship that cannot be defined from the distribution alone. One should also rely on causal assumption that explains the different relationships. For example: randomisation, confounding (it will be detailed later).

### 2.1.3 The difference between the 2 notions

To have a better understanding of how causation and association differ, the following example is considered. The goal of a study is to test the effectiveness of a vaccine on a sick population. Let's denote by  $A$  the action of injecting the vaccine and by  $Y$  the effect of the vaccine on the population. In an association approach, the population will be divided into 2 groups : only one group will be injected with the vaccine. The effectiveness of the treatment is thus measured on the treated population (  $E[Y|A = 1]$  ) and it is also possible to see if the untreated population will be treated without the vaccine (  $E[Y|A = 0]$  ). However, in a causation approach, we suppose that is possible to inject the whole population and study the effect of the vaccine (  $E[Y^{a=1}]$  ) and at the same time, have the same population **not** injected (  $E[Y^{a=0}]$  ).

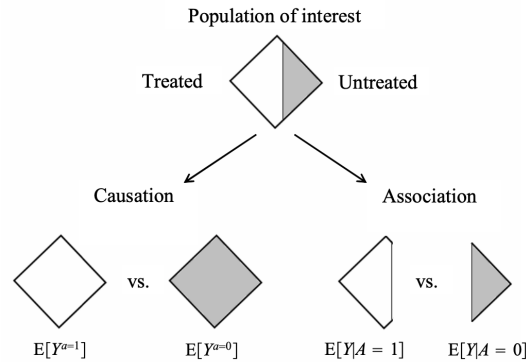


Figure 2: a visualisation of the difference between causation and association

One can see clearly the impossibility of applying in the causation approach : it is impossible to go back in time and inject an individual who was not injected; we can only observe one of the different potential outcomes. Meanwhile, there is some techniques so that such an experiment can be possible: randomised experimentation is one example that will be detailed later.

## 2.2 Counterfactuals

This is an important concept for causation. Counterfactuals could be defined as the answer to the question "what if ?" or simply as the unobserved outcome. If we consider the previous example in the case of an association, the unobserved outcomes will be the observations of "what if we have injected the proportion of the population that was not injected ?" and of "what if we have not injected the proportion of the population that was injected ?".

## 2.3 Confounders

A confounding variable is an unmeasured variable that influences both the supposed cause and the supposed effect. Ignoring the confounder may lead to conclude an association between 2 variables when actually it is not the case. Confounders can also increase the variance or introduce bias.

Different technics exist to reduce the effect of the confounders (random samples, control variable ...)<sup>2</sup>. However, knowing in advance all the the confounders in the data/model is still very important.

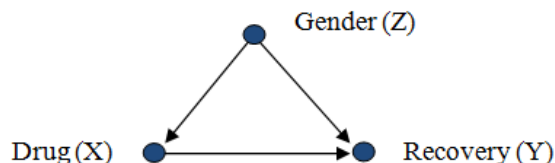


Figure 3: An example of a confounding variable : in this example, the Gender is not observed and it is affecting the variables Drug and Recovery

## 2.4 Simpson's paradox

Simpson's paradox, also called Yule-Simpson effect, in statistics, an effect that occurs when the marginal association between two categorical variables is qualitatively different from the partial association between the same two variables after controlling for one or more other variables.

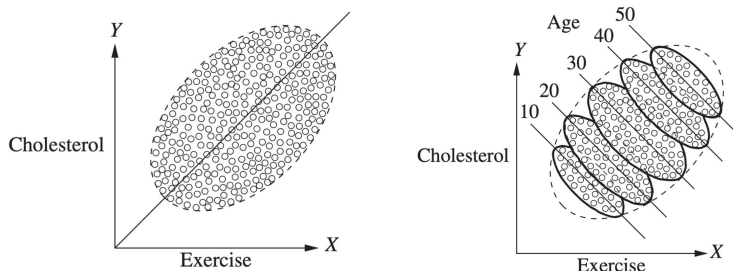


Figure 4: Illustration of the Simpson's Paradox[3]

The above example is one of the best illustration of the Simpson's paradox : when analysing the whole dataset, the first conclusion we can draw is that the amount of exercise is positively correlated with the level of cholesterol. Although, when including the information of age, we can see clearly, for every age category, a negative correlation between the two variables. Thus, domain knowledge is something very important when investigating the relationships in the data and sometimes the expected results are achieved only after increasing/decreasing the studied set of data - adding/excluding some variables.

<sup>2</sup>An interesting article about confounders could be found at [www.statisticshowto.com](http://www.statisticshowto.com)

### 3 Causal inference

In the this section, we suppose the following :

- $Y_i$  : random variable of the outcome for the "unit"  $i$  (a unit is a physical object, for example, a person, at a particular point in time [4]),
- $N$  : the number of units,
- $Y$  :  $(Y_1, Y_2, \dots, Y_n)$ ,
- $A$  : treatment (a treatment is an action that can be applied or withheld from the unit [4]),
- $Y_i^a$  : controlled treatment

For simplicity, we assume that the treatment ( $A$ ) can take only 2 values :  $\{0,1\}$ .

#### 3.1 Individual level Causal Effect : ICE

The ICE is defined as the difference between an individual's two potential outcomes :

$$ICE = \delta = Y_i^{a=1} - Y_i^{a=0}$$

#### 3.2 Average Causal Effect : ACE

The ACE is the population average of the individual level causal effects :

$$ACE = E[\delta] = E[Y^{a=1}] - E[Y^{a=0}]$$

ACE is computed over the entire population. If the treatment has a causal effect on the outcome, the value of the ACE will be different than 0 and it will be equal to zero in the other case. This is a mesure of the causation.

#### 3.3 Standard estimator : $S^*$

$$S^* = E[Y^{a=1}|A = 1] - E[Y^{a=0}|A = 0]$$

$S^*$  is computed from the treatment and control groups, and not from the entire population as for the ACE. This is a mesure of the association.

#### 3.4 Fundamental Problem of Causal Inference

Because one can never observe both potentiel outcomes ( the outcome under different treatments ), it is impossible to measure directly the causal effects ( ACE and ICE ). This is what Holland called as the "fundamental problem of causal inference" [5].

This can be seen as missing data problem : we have  $E[Y^{a=1}|A = 1]$  and  $E[Y^{a=0}|A = 0]$  which are equal respectively to  $E[Y|A = 1]$  and  $E[Y|A = 0]$  but we don't know  $E[Y^{a=1}|A = 0]$  and  $E[Y^{a=0}|A = 1]$ .

Since  $S^*$  is a mesure of association and ACE is a mesure of the causation, these 2 values are not necessarily equal.

### 3.5 Randomised experimentation

One of the solution to tackle the fundamental problem of causal inference is to use a randomised experience.

A randomised experiment is when the investigator carried out the action of interest and it was randomised because the decision to act on any study subject was made by a random device. Randomisation results in convincing causal inferences, the downsides being mostly financial and ethical.

A scientific study in which the investigator observes and records the relevant data is referred to as an observational study. And while observational studies are not considered robust causal inference tools, much human knowledge is derived from them.

Statistical data, derived from studies and experimentation, can represent many forms of bias. The two most relevant to causal inference are Measurement Bias and Selection Bias.

In the case of randomised experiment, we have the following :

- $E[Y^{a=1}|A = 1] = E[Y^{a=1}|A = 0] = E[Y^{a=1}]$
- $E[Y^{a=0}|A = 1] = E[Y^{a=0}|A = 0] = E[Y^{a=0}]$

Based on the assertions above, it is easy to prove that a randomised experiment leads to :

$$ACE = S^*$$

which means that it is possible to determine causation from the existing data. This also suppose comparability conditions so that the individuals in both groups are identical.

### 3.6 The Causal Hierarchy

**Figure 1. The causal hierarchy. Questions at level 1 can be answered only if information from level i or higher is available.**

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y x)$	Seeing	What is? How would seeing $X$ change my belief in $Y$ ?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y do(x), z)$	Doing, Intervening	What if? What if I do $X$ ?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y, x', y')$	Imagining, Retrospection	Why? Was it $X$ that caused $Y$ ? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past two years?

Figure 5: The causal hierarchy[6]

One of the very interesting papers that explains causal inference is : "The Seven Tools of Causal



Inference, with Reflections on Machine Learning" [6]. This section is thus inspired mostly by this paper.

[6]

### **3.7 Graph representation**

## **4 Paper : Judea Pearl – remove this**

see [1]

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 – remove this**

see [7] and [8]

## **6 Definition of the causality – remove this**

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.[1]

[4]

## 7 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.[9]

## 8 Python Library : DoWhy

library developed by Microsoft [10], [blog article available link](#)

## References

- [1] 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.
- [2] J. Pearl *et al.*, “Causal inference in statistics: An overview,” *Statistics surveys*, vol. 3, pp. 96–146, 2009.
- [3] J. Pearl, M. Glymour, and N. P. Jewell, *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- [4] 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.
- [5] P. W. Holland, “Statistics and causal inference,” *Journal of the American statistical Association*, vol. 81, no. 396, pp. 945–960, 1986.
- [6] 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.
- [7] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, and A. Zhang, “A survey on causal inference,” *arXiv preprint arXiv:2002.02770*, 2020.
- [8] M. A. Hernán and J. M. Robins, “Causal inference: what if,” *Boca Raton: Chapman & Hill/CRC*, vol. 2020, 2020.
- [9] C. Glymour, K. Zhang, and P. Spirtes, “Review of causal discovery methods based on graphical models,” *Frontiers in Genetics*, vol. 10, p. 524, 2019.
- [10] “DoWhy: A Python package for causal inference.” <https://github.com/microsoft/dowhy>, 2019.
- [11] A. Marx and J. Vreeken, “Testing conditional independence on discrete data using stochastic complexity,” *arXiv preprint arXiv:1903.04829*, 2019.
- [12] B. Schölkopf, “Causality for machine learning,” *arXiv preprint arXiv:1911.10500*, 2019.
- [13] P. Judea, “Causality: models, reasoning, and inference,” *Cambridge University Press. ISBN 0*, vol. 521, no. 77362, p. 8, 2000.
- [14] J. Pearl, “The science and ethics of causal modeling,” *Handbook of ethics in quantitative methodology*, pp. 383–416, 2011.
- [15] 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.
- [16] P. K. Lam, *Estimating individual causal effects*. PhD thesis, 2013.