

CASE STUDY : CAUSAL INFERENCE

Mohammed FELLAJI - Ahmed BEN AISSA

Supervisor : Frédéric PENNERATH †

†Associate professor at CentraleSupélec



CentraleSupélec

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. Determining the relationships between the inputs can be done on different levels : when recording the data or even once the data collection is done. Unlike correlation, causation cannot be extracted only from the distribution of the data. Thus, the aim of this case study is analyse the different literatures on causal inference for a better understanding of this notion.

Definitions

Here is some of the most important definitions [3, 5] :

- **Associational concept** : any relationship that can be defined in terms of a joint distribution of observed variables. For example: regression, correlation.
- **Causal concept** : 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.

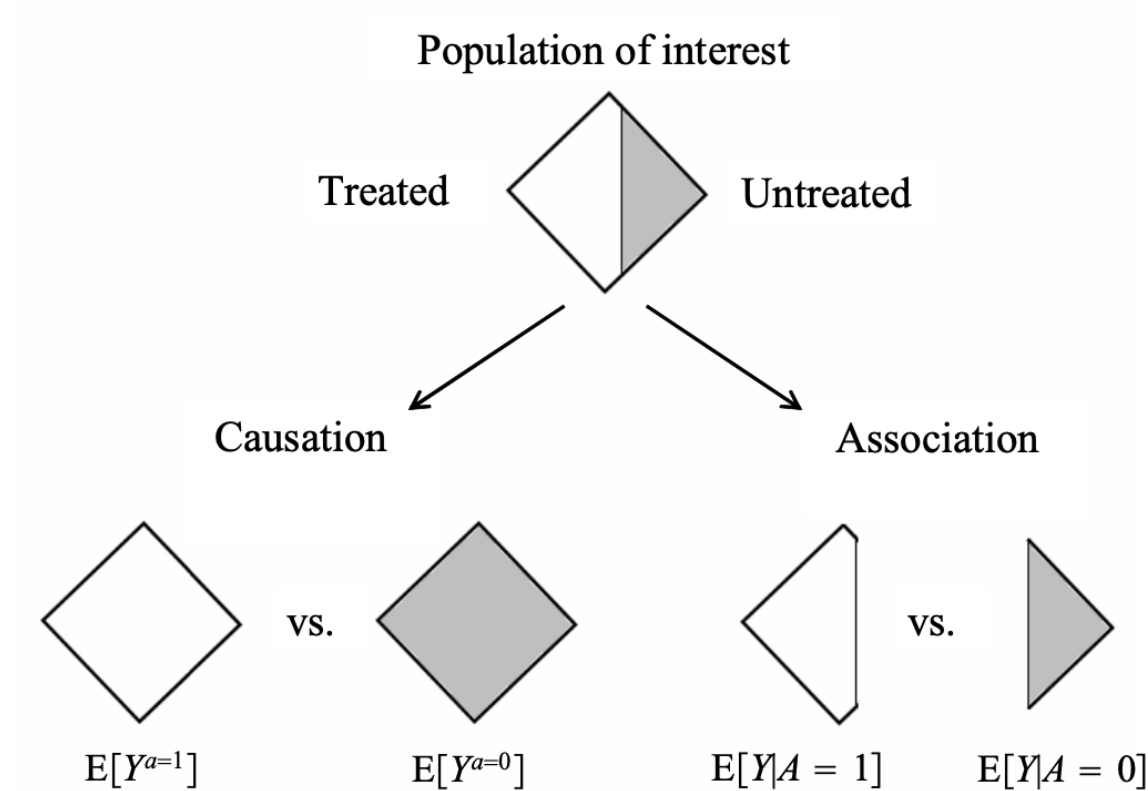


Fig. 1: a visualisation of the difference between causation and association

- **Counterfactuals** : the unobserved outcome. It is the answer to the question : what if?
- **Confounders** : an unmeasured variable that influences both the supposed cause and the supposed effect.

Notation

Here is a list of some useful notations :

- 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 [6]),
- 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 [6]),
- Y_i^a : controlled treatment

Metrics and Interpretations

In the case of a treatment (A) that has only 2 possible values $\{0,1\}$, we can define these metrics:

- **Individual level Causal Effect (ICE)** : the difference between an individual's two potential outcomes.

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

- **Average Causal Effect (ACE)** : it is a mesure of causation; it determines if the treatment has a causal effect on the outcome or not.

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

- **Standard estimator (S*)** : it is a mesure of association; it is computed from the treatment and control groups.

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

Fundamental Problem of Causal Inference

Because one can never observe both potentiel outcomes (the outcome under different treatments), it is impossible to mesure directly the causal effects (ACE and ICE). This is what Holland called as the "fundamental problem of causal inference" [2].

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 $E[Y^{a=0}|A=1]$. Thus, this problem can be considered as a missing data problem.

Randomised experimentation

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.

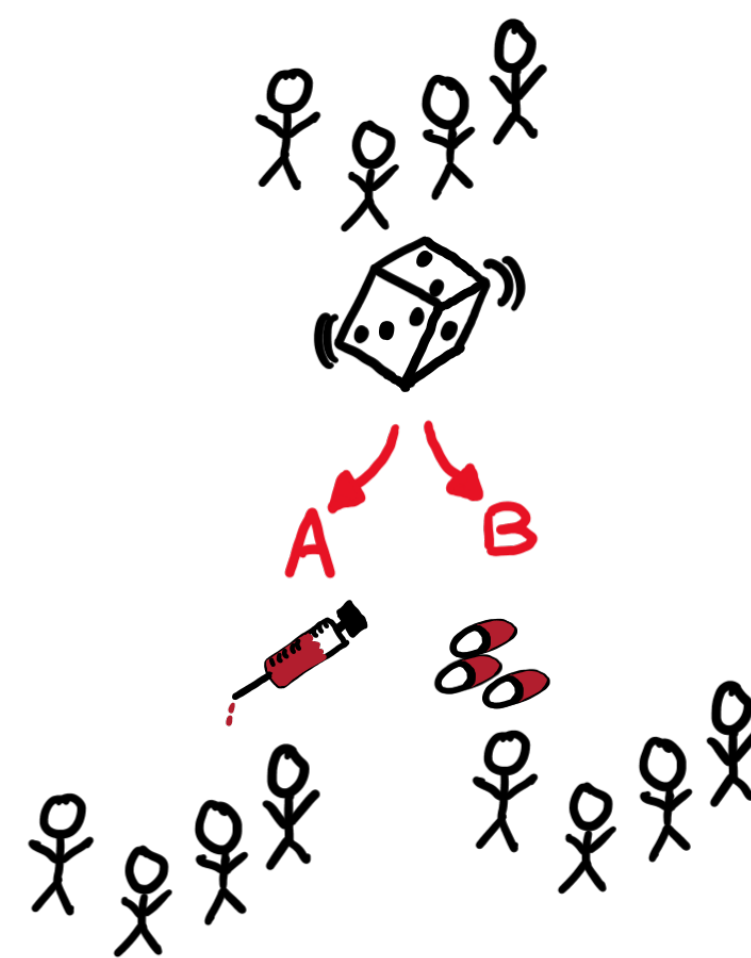


Fig. 2: an illustration of a randomised experiment

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

- $E[Y^{a=1}|A=1] = E[Y^{a=1}|A=0] = E[Y^{a=1}]$ (1)
- $E[Y^{a=0}|A=1] = E[Y^{a=0}|A=0] = E[Y^{a=0}]$ (2)
- from the equations (1) and (2) we can prove that : $ACE = S^*$

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.[1]

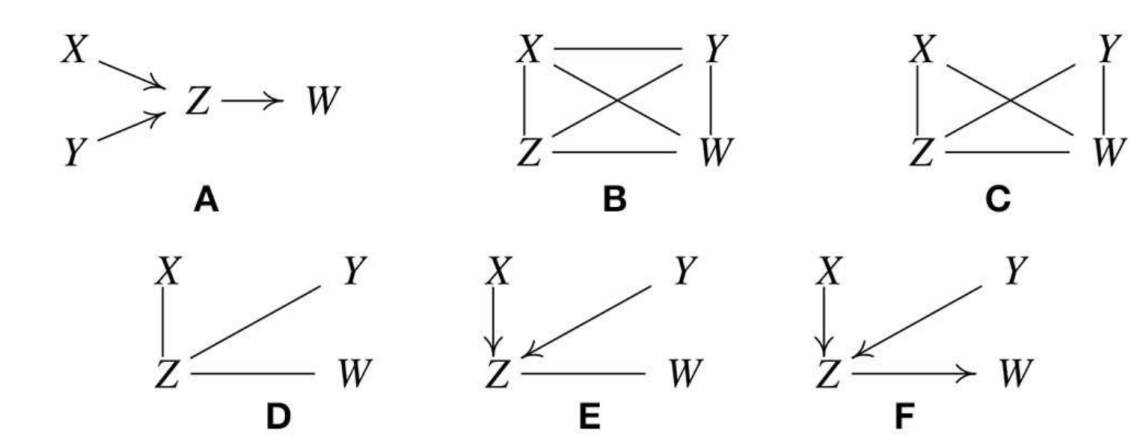


FIGURE 1 | Illustration of how the PC algorithm works. (A) Original true causal graph. (B) PC starts with a fully-connected undirected graph. (C) The $X - Y$ edge is removed because $X \perp Y$. (D) The $X - W$ and $Y - W$ edges are removed because $X \perp W|Z$ and $Y \perp W|Z$. (E) After finding v-structures. (F) After orientation propagation.

Fig. 3: an illustration of the PC algorithm

Inference engine - DoWhy Library

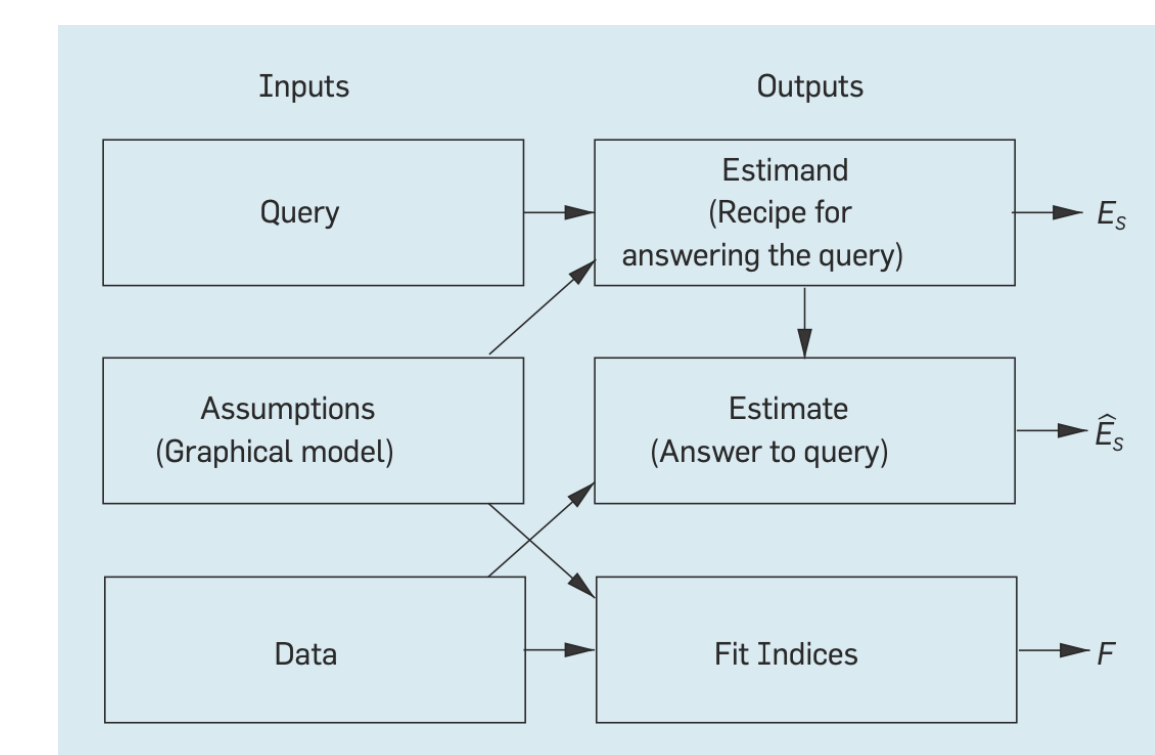


Fig. 4: The algorithm suggested by Judea Pearl [4]

References

- [1] Clark Glymour, Kun Zhang, and Peter Spirtes. "Review of Causal Discovery Methods Based on Graphical Models". In: *Frontiers in Genetics* 10 (2019), p. 524. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00524. URL: <https://www.frontiersin.org/article/10.3389/fgene.2019.00524>.
- [2] Paul W Holland. "Statistics and causal inference". In: *Journal of the American statistical Association* 81.396 (1986), pp. 945–960.
- [3] Judea Pearl. "The mathematics of causal relations". In: *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures* (P. Shrout, K. Keyes and K. Ornstein, eds.). Oxford University Press, Corvallis, OR (2010), pp. 47–65.
- [4] Judea Pearl. "The seven tools of causal inference, with reflections on machine learning". In: *Communications of the ACM* 62.3 (2019), pp. 54–60.
- [5] Judea Pearl et al. "Causal inference in statistics: An overview". In: *Statistics surveys* 3 (2009), pp. 96–146.