Causation in the Empirical Sciences: A Prefatory Note

Ashley I Naimi

June 2022

Contents

1 Causation in the Empirical Sciences

Before embarking on this short journey on how to estimate causal effects with machine learning algorithms, it would be important to take a step back and look at the current landscape of "causation" in the empirical sciences.

Generally speaking, one can divide the scientific interest displayed in causation in two areas:

1) Causal Discovery
2) Causal Inference

These two fields can be considered two distinct disciplines that deal with causation. While they do share some important relations between them, generally the areas of research in causal discovery and causal inference tend to be quite different.

Briefly, causal discovery concerns itself with uncovering causal structures using data. The objective of a causal discovery analysis is to generate a causal diagram (or DAG) using data, with minimal or no domain-specific knowledge (Glymour and Scheines, 1986, Spirtes et al. (1993),Spirtes and Zhang (2016),Glymour et al. (2019)). Causal discovery is a fast growing area in machine learning, and often synonymous with the "structural learning problem' ' in the Bayesian network context (Koller and Friedman, 2009).

In contrast, causal inference typically consists of collecting data based on domain-specific knowledge, assuming that a particular set of causal relations exist between the variables in this dataset (often articulated as a DAG Pearl (2000),Tennant et al. (2020)), and then proceeding with an analysis on the basis of these assumed relations. Investigators will often define a causal DAG using domain-specific knowledge to evaluate whether they can identify the effect of interest (with *d*-separation) with the data they possess (Tennant et al., 2020). This process is generally considered a part of causal inference, a process that requires translating domain-specific knowledge into assumed, but explicit, causal relations, which have implications for precisely how to quantify an exposure effect of interest (Pearl, 2000, Hernán and Robins (2020),Greenland and Brumback (2002)).

In this short course, we will be focusing on *causal inference*, and not causal discovery.

References

Clark Glymour and Richard Scheines. Causal modeling with the TETRAD program. page 27, 1986.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics*, 10, 2019. ISSN 1664-8021. DOI: 10.3389/fgene.2019.00524. URL https://www.frontiersin.org/articles/10.3389/fgene.2019.00524/full.

Sander Greenland and Babette Brumback. An overview of relations among causal modelling methods. *International Journal of Epidemiology*, 31(5): 1030–1037, October 2002. ISSN 0300-5771. DOI: 10.1093/ije/31.5.1030. URL https://academic.oup.com/ije/article/31/5/1030/745818. Number: 5 Publisher: Oxford Academic.

M. A. Hernán and JM Robins. *Causal Inference*. Chapman & Hall/CRC, Boca Raton, FL, 2020.

Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. Adaptive computation and machine learning. MIT Press, Cambridge, MA, 2009. ISBN 978-0-262-01319-2.

J. Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *UC Los Angeles: Department of Statistics, UCLA.*, page Retrieved from: http://escholarship.org/uc/item/8h99z8sp, 2000.

Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*, 3(1):3, February 2016. ISSN 2196-0089. DOI: 10.1186/s40535-016-0018-x. URL https://doi.org/10.1186/s40535-016-0018-x. Number: 1.

Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT Press, Cambridge, Mass., 1993.

Peter W G Tennant, Eleanor J Murray, Kellyn F Arnold, Laurie Berrie, Matthew P Fox, Sarah C Gadd, Wendy J Harrison, Claire Keeble, Lynsie R Ranker, Johannes Textor, Georgia D Tomova, Mark S Gilthorpe, and George T H Ellison. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *International Journal of Epidemiology*, (dyaa213), December 2020. ISSN 0300-5771. DOI: 10.1093/ije/dyaa213. URL https://doi.org/10.1093/ije/dyaa213.