Internship proposal

Investigating the applicability of causal representation in epidemiology

Context: The Pierre Louis Institute of Epidemiology and Public Health (IPLESP), co-accredited by Inserm and Sorbonne University, unites expertise in epidemiology and public health within Sorbonne University. IPLESP's main objective is to produce original knowledge on pressing public health issues and related intervention effectiveness, focusing on emerging infectious diseases, chronic diseases, environmental health, and mental health. To tackle these challenges, causal inference [Pearl, 2000, Hernán and Robins, 2016] emerges as an indispensable tool. Therefore, at IPLESP, we are establishing a new team dedicated to developing advanced methodologies rooted in causal inference.

Access to causal graphs is essential for estimating causal effects [Greenland et al., 1999, Savitz and Wellenius, 2016]. However, in many applications, it is challenging for a practitioner to provide such a graph. In certain cases, it is feasible to discover the causal graph from data under specific assumptions [Spirtes et al., 2000, Assaad et al., 2022]. However, the field of epidemiology, particularly with the increasing use of medico-administrative databases such as the EDS of AP-HP or the National Health Data System (SNDS), presents unique challenges. These include the high dimensionality of data and the reliance on variables that often serve only as proxies for the variables of real interest. These factors increase the risk of failure in accurately discovering causal graphs from data. In this context, causal representation emerges as a promising research approach [Schölkopf et al., 2021]. This approach aims to identify confounding factors among observed variables and to detect causal relationships between these hidden factors. However, the assumptions required for causal representation are generally more stringent than those for causal graph discovery from data [Yao et al., 2022, Sturma et al., 2023], raising questions about their applicability in the healthcare sector.

Proposal: The primary objective of the internship is to explore the applicability of causal representation methods for uncovering causal relationships between latent variables that give rise to observed proxy variables in medico-administrative databases. In addition, the internship involves writing a concise survey of existing causal representation methods, with a particular emphasis on clearly outlining their underlying assumptions.

Required skills: Highly motivated candidate with an M2 degree and strong background in probability, machine learning, and causal inference, along with a keen interest in epidemiology. Proficiency in programming is also required.

Location: The intern will work at IPLESP (`https://iplesp.fr/`), located in Paris. She/he will be supervised by Charles Assaad.

Dates:: Starting date: To be discussed, end of 2025 or early 2026, for a duration of 4-6 months.

Contact:: To apply, please send a CV and a cover letter to Charles Assaad (charles.assaad@inserm.fr)

References

C. K. Assaad, E. Devijver, and E. Gaussier. Survey and evaluation of causal discovery methods for time series. *J. Artif. Int. Res.*, 73, apr 2022. doi: 10.1613/jair.1.13428.

S. Greenland, J. Pearl, and J. Robins. Causal diagrams for epidemiologic research. *Epidemiology (Cambridge, Mass.)*, 10(1):37—48, January 1999. ISSN 1044-3983. doi: 10.1097/00001648-199901000-00005.

M. A. Hernán and J. M. Robins. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology*, 183(8):758–764, 03 2016. ISSN 0002-9262. doi: 10.1093/aje/kwv254.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000. ISBN 0-521-77362-8.

D. A. Savitz and G. A. Wellenius. Causal Diagrams for Epidemiologic Inference. In *Interpreting Epidemiologic Evidence: Connecting Research to Applications*. Oxford University Press, 08 2016. ISBN 9780190243777. doi: 10.1093/acprof:oso/9780190243777.003.0003.

B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, May 2021. ISSN 1558-2256. doi: 10.1109/JPROC.2021.3058954.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT press, 2nd edition, 2000.

N. Sturma, C. Squires, M. Drton, and C. Uhler. Unpaired multi-domain causal representation learning. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 34465–34492. Curran Associates, Inc., 2023.

W. Yao, G. Chen, and K. Zhang. Temporally disentangled representation learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 26492–26503. Curran Associates, Inc., 2022.