Internship proposal

Applicability of root cause analysis in epidemiology and development of strategies for non-identifiable root causes in summary causal graphs

Context: The Pierre Louis Institute of Epidemiology and Public Health (IPLESP), co-accredited by Inserm and Sorbonne University, brings together research strengths 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, Hernan and Robins, 2023] emerges as an indispensable tool. Therefore, at IPLESP, we are establishing a new team dedicated to developing advanced methodologies rooted in causal inference. One objective of this team is to address root cause analysis challenges in epidemiology, requiring robust methodologies amid increasing data volumes and the intricate interplay of various factors.

A novel root cause analysis framework [Assaad et al., 2023] has recently emerged, designed to leverage observational temporal data and a summary causal graph derived from a Markovian model [Assaad et al., 2022] to identify the root causes that explain all anomalies within a system. While initially developed for IT systems, this framework shows promising potential for adaptation to epidemiological data, potentially advancing root cause analysis in public health research. However, root causes within this framework are not always identifiable. The framework's approach depends on estimating and comparing direct effects across normal and anomalous regimes, meaning that a root cause can only be identified if its corresponding direct effect can be identified solely using the summary causal graph. For the framework to be effectively applied in public health, it would need to offer solutions for cases where root causes are not identifiable.

Proposal: The primary objective of this internship is to review and compare different root cause analysis frameworks and methods, including the one based on summary causal graphs, and evaluate their applicability to epidemiological data. The secondary objective is to develop strategies for handling cases where root causes cannot be identified using only summary causal graphs.

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. The candidate will have the opportunity to persue a PhD in causal inference.

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, early 2025, for a duration of 5-6 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.

C. K. Assaad, I. Ez-Zejjari, and L. Zan. Root cause identification for collective anomalies in time series given an acyclic summary causal graph with loops. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 8395–8404. PMLR, 25–27 Apr 2023.

M. Hernan and J. Robins. *Causal Inference: What If*. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2023. ISBN 9781420076165.

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