NORMALIZING FLOWS FOR PROPENSITY SCORE ADJUSTMENT

Assume that a certain drug was approved for treating a disease. We now want to determine if the drug reduces the 5-year mortality due to another disease. We are then interested in estimating the causal effect of the drug on mortality. More specifically, we are interested in the decrease in average mortality when every individual in the population of interest takes the drug and does not take the drug. Since computing this quantity is impossible (an individual cannot take and not take the drug and/or the population may be too large), we may instead consider a group of voluntary individuals, each of which randomly receives the drug or a placebo (i.e., a pill with no therapeutic effect). This approach is called randomized controlled trial and, unfortunately, it may still be unfeasible due to ethical concerns and/or time and cost constraints.

An alternative to the experimental data produced by a randomized controlled trial is to use observational data. These data are sometimes called register data, and they are routinely collected to record the habits of the individuals in terms of, for instance, medications they take and diseases they suffer. Estimating the causal effect from observational data requires adjusting for confounders (i.e., factors that affect both the drug intake and the mortality such as gender or age) to avoid producing a biased estimate. For instance, if more men than women take the drug in the observational data whereas the population contains roughly as many men and women, then the estimate obtained from the data may be a bias estimate of the causal effect in the population.

A widely used method for confounder adjustment is inverse probability weighting (IPW). This method weights each individual in the data with his/her propensity score (i.e., the probability of the individual taking the drug given the confounders' values) and, then, it uses the weighted data to learn a regression of mortality on drug intake. Logistic regression is typically used to compute the propensity score. However, this implies that the log odds of drug intake is assumed to be linear in the confounders. We would like to drop this assumption. In particular, we would like to use our previous work on normalizing flows (NFs) for this purpose. Briefly, a NF is a deep neural network that model the causal mechanism behind the data and, thus, the join distribution over drug, mortality and confounders. Therefore, it may be used to estimate the causal effect directly but, also, to produce the propensity scores needed by IPW. This project aims to compare these two approaches on synthetic data. If time permits, real data may be used. We will provide have access to our implementation of NFs. The articles below provide further information on our current work with NFs. The first article is particularly relevant for this project.

- Balgi, S., Peña, J. M., and Daoud, A. (2022). Personalized Public Policy Analysis In Social Sciences Using Causal-Graphical Normalizing Flows. In Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI 2022), 11810–11818.
- Balgi, S., Peña, J. M., and Daoud, A. (2022). Counterfactual Analysis of the Impact of the IMF Program on Child Poverty in the Global-South Region using Causal-Graphical Normalizing Flows. arXiv:2202.09391 [cs.AI].
- Balgi, S., Peña, J. M., and Daoud, A. (2022). ρ-GNF : A Novel Sensitivity Analysis Approach Under Unobserved Confounders. arXiv:2209.07111 [stat.ME].

Prerequisites

Knowledge of deep learning and Python. No previous knowledge of causality is required.

Contact

Jose M. Peña, jose.m.pena@liu.se
Sourabh Balgi, sourabh.balgi@liu.se