Causal inference and mechanism investigation under time-varying system

Sheng-Hsuan Lin

Institute of Statistics, National Chiao-Tung University

Abstract

Causal inference has become a popular approach for investigating causality and mechanism in complex data format. In the first part, I will provide a general picture of the causal inference framework: defining a scientific question as a causal parameter based on counterfactual model, identifying this causal parameter as a statistical parameter, and then estimating (or testing) this statistical parameter by statistics inference. The mechanism investigation for the effect of religious behavior on survival rate will be used as the motivating example. In the second part, I will show how to define and identify the longitudinal total and mediation effects (two effects of interest in mechanism investigation) as analytic formula, and obtains parametric estimates by Monte-Carlo Simulation and bootstrapping method. In the Framingham Heart Study data, we apply this method to estimate the mediation effect of smoking behaviors sustained over a 10-year period on blood pressure when considering weight change as a time-varying mediator. Compared with non-smoking, smoking 20 cigarettes per day for 10 years was estimated to increase blood pressure by 1.18 (95 % CI: -0.68, 2.69) mm-Hg. The mediation effect was -0.34 (95 % CI: -0.52, -0.13) mm-Hg, which is negative because smoking leads to lower weight which leads to lower blood pressure. These results provide evidence that weight change in fact partially conceals the detrimental effects of cigarette smoking on blood pressure. In addition, I will also introduce an approach to conduct mediation analysis for survival data with time-varying exposures, mediators, and confounders. We identify certain interventional direct and indirect effects through a survival mediational g-formula and describe the required assumptions. We also provide a feasible parametric approach along with an algorithm and software to estimate these effects. We apply this method to analyze the Framingham Heart Study data to investigate the causal mechanism of smoking on mortality through coronary artery disease. The risk ratio of smoking 30 cigarettes per day for ten years compared with no smoking on mortality is 2.34 (95 % CI = (1.44, 3.70)). Of the overall effect, 7.91% (95% CI: = 1.36%, 19.32%) is mediated by coronary artery disease. Our work represents the first application of the parametric mediational g-formula in an epidemiologic cohort study.

Keyword: mediation analysis, causal inference, time-varying system, g-computation algorithm