World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:16, No:06, 2022

A Targeted Maximum Likelihood Estimation for a Non-Binary Causal Variable: An Application

Authors : Mohamed Raouf Benmakrelouf, Joseph Rynkiewicz

Abstract : Targeted maximum likelihood estimation (TMLE) is well-established method for causal effect estimation with desirable statistical properties. TMLE is a doubly robust maximum likelihood based approach that includes a secondary targeting step that optimizes the target statistical parameter. A causal interpretation of the statistical parameter requires assumptions of the Rubin causal framework. The causal effect of binary variable, E, on outcomes, Y, is defined in terms of comparisons between two potential outcomes as E[YE=1 − YE=0]. Our aim in this paper is to present an adaptation of TMLE methodology to estimate the causal effect of a non-binary categorical variable, providing a large application. We propose coding on the initial data in order to operate a binarization of the interest variable. For each category, we get a transformation of the non-binary interest variable into a binary variable, taking value 1 to indicate the presence of category (or group of categories) for an individual, 0 otherwise. Such a dummy variable makes it possible to have a pair of potential outcomes and oppose a category (or a group of categories) to another category (or a group of categories). Let E be a non-binary interest variable. We propose a complete disjunctive coding of our variable E. We transform the initial variable to obtain a set of binary vectors (dummy variables), E = (Ee : e ∈ {1, ..., |E|}), where each vector (variable), Ee, takes the value of 0 when its category is not present, and the value of 1 when its category is present, which allows to compute a pairwise-TMLE comparing difference in the outcome between one category and all remaining categories. In order to illustrate the application of our strategy, first, we present the implementation of TMLE to estimate the causal effect of non-binary variable on outcome using simulated data. Secondly, we apply our TMLE adaptation to survey data from the French Political Barometer (CEVIPOF), to estimate the causal effect of education level (A five-level variable) on a potential vote in favor of the French extreme right candidate Jean-Marie Le Pen. Counterfactual reasoning requires us to consider some causal questions (additional causal assumptions). Leading to different coding of E, as a set of binary vectors, E = (Ee : e ∈ {2, ..., |E|}), where each vector (variable), Ee, takes the value of 0 when the first category (reference category) is present, and the value of 1 when its category is present, which allows to apply a pairwise-TMLE comparing difference in the outcome between the first level (fixed) and each remaining level. We confirmed that the increase in the level of education decreases the voting rate for the extreme right party.

Keywords : statistical inference, causal inference, super learning, targeted maximum likelihood estimation
Conference Title : ICAS 2022 : International Conference on Applied Statistics
Conference Location : Tokyo, Japan
Conference Dates : June 09-10, 2022