



Joint sufficient dimension reduction for estimating continuous treatment effect functions

Ming-Yueh Huang^{a,*}, Kwun Chuen Gary Chan^b

^a Institute of Statistical Science, Academia Sinica, Taiwan

^b Department of Biostatistics, University of Washington, USA



ARTICLE INFO

Article history:

Received 28 November 2017

Available online 10 July 2018

Keywords:

Central subspace

Cross-validation

Dose–response

Infinitesimal jackknife

Optimal bandwidth

AMS subject classifications:

62B05

ABSTRACT

The estimation of continuous treatment effect functions using observational data often requires parametric specification of the effect curves, the conditional distributions of outcomes and treatment assignments given multi-dimensional covariates. While nonparametric extensions are possible, they typically suffer from the curse of dimensionality. Dimension reduction is often inevitable and we propose a sufficient dimension reduction framework to balance parsimony and flexibility. The joint central subspace can be estimated at a $n^{1/2}$ -rate without fixing its dimension in advance, and the treatment effect function is estimated by averaging local estimates of a reduced dimension. Asymptotic properties are studied. Unlike binary treatments, continuous treatments require multiple smoothing parameters of different asymptotic orders to borrow different facets of information, and their joint estimation is proposed by a non-standard version of the infinitesimal jackknife.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

It is common in medical and social studies to collect data on treatments or exposures and a primary goal in these studies is to investigate the causal effects of the treatments on an outcome. In the case of continuous treatments, the effects are naturally described by functions, and model-free estimation typically involves smoothing to borrow information from adjacent treatment levels. While observational data are often the only data available in practice, direct smoothing of observed responses across different treatment levels usually results in biased estimates of causal treatment effects due to confounding.

In the literature, there are two major approaches for estimating treatment effects under confoundedness; they are based on outcome regression models and generalized propensity score models. Outcome regression models describe how the response relates to the treatment and covariates [9,14], while generalized propensity score models extend the classical propensity score models for binary treatments to account for the conditional density of treatment given covariates [4,10,13]. Doubly robust estimation was recently proposed in [17] to unify these two approaches. It combines outcome regression and generalized propensity score models through estimation. Model-based applications require at least one correctly specified parametric or semiparametric model for the outcome regression functions or generalized propensity scores. However, in practice, it is often hard to check if these models are adequate without prior knowledge. In contrast, one can use fully nonparametric smoothing techniques to replace the modeling procedures as in [4,17]. When the dimension of covariates is large, the model-free approaches suffer from the curse of dimensionality and can lead to unstable estimates for the treatment effects. Some middle ground between current model-based and model-free approaches would thus be desirable.

* Corresponding author.

E-mail address: myh0728@stat.sinica.edu.tw (M.-Y. Huang).

In this work, we study a nonparametric method that is more parsimonious and suffers less from the curse of dimensionality than the fully nonparametric smoothing estimator, while retaining robustness to potential model misspecification. Sufficient dimension reduction for treatment effect estimation has recently been studied in [11,19] for binary treatment variables. To extend these ideas to continuous treatments, we consider a joint sufficient dimension reduction model [11] to capture all the information of the outcome regression and generalized propensity score. In contrast, Luo et al. [19] considered separate dimension reduction for different subgroups, which is more difficult to generalize to continuous treatments.

An advantage of sufficient dimension reduction is that, unlike generalized propensity scores which cannot be estimated nonparametrically at a $n^{1/2}$ convergence rate, the central subspace can be estimated with a $n^{1/2}$ -rate, which turns out to be crucial in guaranteeing a desirable theoretical and practical performance of the treatment effects estimator. The estimation is complicated by the need to employ multiple smoothing parameters, one for borrowing information across different treatment levels, and another for undersmoothing with respect to marginalizing the observed covariate distribution. The smoothing parameters are chosen in a data-adaptive manner by minimizing an estimator of the asymptotic mean squared error via a non-standard application of the infinitesimal jackknife.

The rest of this article is organized as follows. Section 2 introduces the considered dimension reduction model and proposes an estimation procedure to estimate the joint central subspace and the treatment effects of interest. The results of a series of simulation studies are reported in Section 3 and an application on food patterns data analysis is described in Section 4. Some concluding remarks are given in Section 5.

2. The proposed methodology

2.1. Joint sufficient dimension reduction model

Let $Y(t)$ be the potential outcome associated with each treatment level t , where $t \in \mathcal{T}$, and \mathcal{T} is a connected subset in \mathbb{R} , the set of all real numbers. Also, let T be the continuous treatment variable. Then the observed outcome is $Y = Y(T)$. In addition, a vector $\mathbf{X} = (X_1, \dots, X_p)$ of covariates is observed for each subject. The main goal of this work is to estimate the continuous treatment effect $\mu(t) = E\{Y(t)\}$, $t \in \mathcal{T}$, based on a random sample $(Y_1, T_1, \mathbf{X}_1), \dots, (Y_n, T_n, \mathbf{X}_n)$. Since this parameter is defined in terms of potential outcomes which are not observed directly, some assumptions are needed to identify the quantity of interest by the observed data.

Assumption 1 (Positivity). $f_T(t | \mathbf{x}) > 0$ for all $t \in \mathcal{T}$ and $\mathbf{x} \in \mathcal{X}$, where $f_T(t | \mathbf{x})$ is the conditional density of T given $\mathbf{X} = \mathbf{x}$, and \mathcal{X} is the support of \mathbf{X} .

Assumption 2 (Ignorability). $Y(t) \perp\!\!\!\perp T | \mathbf{X}$ for all $t \in \mathcal{T}$, where $\perp\!\!\!\perp$ denotes independence.

The positivity assumption states that every subject has a non-zero chance of receiving treatment level t . The ignorability assumption says that the potential outcomes and the treatment are unconfounded when the covariates are given. That is, all the confoundedness is captured by the vector of covariates. According to these assumptions, a simple bias removal strategy is to use the property

$$E\{Y(t)\} = E[E\{Y(t) | \mathbf{X}\}] = E[E\{Y(t) | T = t, \mathbf{X}\}] = E\{E\{Y | T = t, \mathbf{X}\}\}. \tag{1}$$

Thus, an estimator for $\mu(t)$ can be obtained by averaging an estimator of the conditional effects $E\{Y | T = t, \mathbf{X}\}$. In practice, the number of covariates is usually large and the nonparametric estimators for the conditional effects typically suffer from the curse of dimensionality. In order to find a lower dimensional score to attain dimension reduction while retaining a property similar to (1), we first note that the likelihood of the observed variables is

$$f_{Y,T,\mathbf{X}}(y, t, \mathbf{x}) = f_Y(y | t, \mathbf{x})f_T(t | \mathbf{x})f_{\mathbf{X}}(\mathbf{x}) = f_{Y(t)}(y | \mathbf{x})f_T(t | \mathbf{x})f_{\mathbf{X}}(\mathbf{x})$$

based on Assumptions 1 and 2. One can see that all the information is conveyed through the conditional distributions of $Y(t)$ and T given \mathbf{X} . Thus, we propose a joint sufficient dimension reduction model to summarize both potential outcomes and treatment assignment, viz.

$$\forall_{t \in \mathcal{T}} Y(t) \perp\!\!\!\perp \mathbf{X} | B^T \mathbf{X}, \quad T \perp\!\!\!\perp \mathbf{X} | B^T \mathbf{X} \tag{2}$$

for some $p \times d$ full-rank parameter matrix B with $d \leq p$. Since model (2) has a nested structure in d , we can focus on searching the joint sufficient dimension reduction subspace $\text{span}(B)$ with the smallest dimension, which is called the joint central subspace and the corresponding basis matrix is denoted by B_0 with dimension d_0 . Related discussion on the model structure and identification is given in [11] for binary treatments. Model (2) leads to the following result on the ignorability of treatment assignment.

Proposition 1. Under Assumption 2 and model (2), the joint central subspace $\text{span}(B_0)$ satisfies $Y(t) \perp\!\!\!\perp T | B_0^T \mathbf{X}$ for all $t \in \mathcal{T}$.

Download English Version:

<https://daneshyari.com/en/article/7546347>

Download Persian Version:

<https://daneshyari.com/article/7546347>

[Daneshyari.com](https://daneshyari.com)