Statistical Science
2020, Vol. 35, No. 3, 511–517
https://doi.org/10.1214/20-STS789
Main article: https://doi.org/10.1214/19-STS735

Rejoinder: A Nonparametric Superefficient Estimator of the Average Treatment Effect

David Benkeser, Weixin Cai and Mark J. van der Laan

1. INTRODUCTION

We thank each of the comment authors for their insights and perspectives on our work. The comments were wide-ranging in content and raised many interesting questions pertaining to our work and its place in the larger scope of research in the area. We address each commenter in turn.

2. LI

We thank Dr. Li for his interesting comment and novel proposal for stabilization in the context of estimating the average treatment effect. Li asks the question as to whether stabilization techniques that are common for inverse probability of treatment weighted (IPTW) estimators could stabilize doubly robust procedures in weakly identified settings. In essence, Li proposes to use a stabilized propensity score in combination with one-step estimation or TMLE. The stabilized propensity score is of the form $\bar{G}_0(w \mid h) = \bar{G}_0(w)/h(w)$, where $h : \mathcal{W} \to [0, 1]$ is some mapping that may depend on P_0. Several choices of h are discussed, such as

$$h(w) = \frac{\bar{G}_0(w)\{1 - \bar{G}_0(w)\}}{\int \bar{G}_0(w)\{1 - \bar{G}_0(w)\}\, dQ_{0,W}(w)}. \tag{1}$$

The author proposes a plug-in estimator h_n of h, based on an estimate of the propensity score, $\bar{G}_n$, and proceeds as usual with a one-step and TMLE procedure using the alternative propensity score estimator $\bar{G}_n(w \mid h_n) = \bar{G}_n(w)/h_n(w)$. The resultant estimators are found via simulation to have reasonable performance in the simulation settings considered in our paper.

Overall, Dr. Li's idea to bring in stabilization techniques from the IPTW literature to the doubly robust sphere is novel and interesting. However, we would like to highlight a potential difficulty when considering coupling this approach with machine learning or other nonparametric regression techniques. The potential problem is illustrated most directly by the analysis of Li's estimator in the case where $\bar{G}_0$ is known exactly, as in a stratified randomized trial. This setting is important, since it is one where asymptotically linear, doubly robust estimators can be generated under the weakest possible assumptions. We will argue that when the outcome regression is estimated nonparametrically Li's estimator may not achieve asymptotic linearity in even this "best-case" scenario.

Let $\psi^1_{n,*}$ be Li's TMLE of ψ^1_0, constructed based on the targeted outcome regression estimate $\bar{Q}^1_{n,*}$, the true stabilized propensity score $\bar{G}_0(\cdot \mid h)$ and the empirical distribution of W, $Q_{n,W}$. Below, we write Pf to denote $\int f(o)\, dP(o)$ for a given P-integrable function f and for each $P \in \mathcal{M}$. We also denote by P_n the empirical distribution function based on $O_1, \dots, O_n$, so $P_n f = n^{-1} \sum_{i=1}^n f(O_i)$. We define $R_{0n} = P_0\{D^1(\cdot \mid \bar{Q}^1_{n,*}, Q_{n,W}, \bar{G}_0(\cdot \mid h)) - D^1(\cdot \mid \bar{Q}^1_{n,*}, Q_{n,W}, \bar{G}_0)\}$. A linearization of Ψ^1 along with straightforward algebra gives

$$\begin{aligned}
&\Psi^1(Q^1_{n,*}) - \Psi^1(Q^1_0) \\
&\quad = -P_0 D^1(\cdot \mid \bar{Q}^1_{n,*}, Q_{n,W}, \bar{G}_0) \\
&\quad = -P_0 D^1(\cdot \mid \bar{Q}^1_{n,*}, Q_{n,W}, \bar{G}_0(\cdot \mid h)) + R_{0n} \\
&\quad = (P_n - P_0) D^1(\cdot \mid \bar{Q}^1_{n,*}, Q_{n,W}, \bar{G}_0(\cdot \mid h)) + R_{0n},
\end{aligned} \tag{2}$$

where the third line follows since, by construction, the targeted estimate $\bar{Q}^1_{n,*}$ is such that $P_n D^1(\cdot \mid \bar{Q}^1_{n,*}, Q_{n,W}, \bar{G}_0(\cdot \mid h)) = 0$. The first term in the final equality is an empirical process and standard conditions can be assumed to control its behavior (Appendix B of the web supplement accompanying the original paper). However,

$$\begin{aligned}
R_{0n} &= \mathrm{E}_{P_0}\left(\{h(W) - 1\}\left[\frac{A}{G_0(W)}\{Y - \bar{Q}^1_n(W)\}\right]\right) \\
&= \mathrm{E}_{P_0}\Bigg(\{h(W) - 1\} \\
&\qquad \times \left[\frac{A}{G_0(W)}\{\mathrm{E}_{P_0}(Y \mid A, W) - \bar{Q}^1_n(W)\}\right]\Bigg) \\
&= \mathrm{E}_{P_0}\left(\{h(W) - 1\}\left[\frac{A}{G_0(W)}\{\bar{Q}^1_0(W) - \bar{Q}^1_n(W)\}\right]\right) \\
&= \mathrm{E}_{P_0}[\{h(W) - 1\}\{\bar{Q}^1_0(W) - \bar{Q}^1_n(W)\}].
\end{aligned}$$

In order for Li's estimator to be asymptotically linear with the claimed influence function, we would need to establish that $R_{0n} = o_p(n^{-1/2})$. However, the form of R_{0n} is not second-order unless $h(w) = 1$ for all $w \in \mathcal{W}$ (in which

David Benkeser is an Assistant Professor of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia 30322, USA (e-mail: benkeser@emory.edu). Weixin Cai is a researcher at Citadel, Seattle, Washington USA. Mark J. van der Laan is Professor of Biostatistics and Statistics, University of California, Berkeley, California 94720, USA.