Statistical Science
2020, Vol. 35, No. 3, 496–498
https://doi.org/10.1214/20-STS770
Main article: https://doi.org/10.1214/20-STS735

Comment: Increasing Real World Usage of Targeted Minimum Loss-Based Estimators

Mireille E. Schnitzer

1. INTRODUCTION

I would like to congratulate the authors David Benkeser, Weixin Cai and Mark van der Laan on their important contributions and wonderfully written article. Like the authors, I am interested in the development of high performing and theoretically principled estimation approaches that are robust, interpretable and diagnosable in messy real-world applications. Over the past 5–10 years, there has been much progress made in adaptive semiparametric estimation for causal and statistical parameters. The current paper represents a notable contribution as it trades off robustness for both stability and computational efficiency in contexts where the parameter of interest is weakly identifiable. In this discussion, I provide my perspective on the proposed estimator in the context of the current literature with a goal of real-world usage.

2. ROBUSTNESS TRADEOFFS AND EXTRAPOLATION

For what is often described as a "vanilla" targeted minimum loss-based estimator (TMLE) of the average treatment effect [14], Chapter 4, the user must specify nuisance estimators for the outcome regression (OR) and the propensity score (PS). The consistency of the point estimate is guaranteed if either nuisance estimator is consistent. However, the convergence of the TMLE to a mean-zero Gaussian variable relies on the nuisance estimators being Donsker and also the product of errors of the two nuisance estimators converging to zero at a $n^{-1/2}$ rate (with respect to the $L^2(P_0)$ norm). The usage of cross-validated TMLE [15] relaxes the Donsker condition, which allows the analyst to choose from a wider selection of machine learning methods. If both nuisance estimators are consistent, convergence speeds above $n^{-1/4}$ will do and nonparametric estimators compatible with this requirement are available [3]. However, if only one is consistent then it must have parametric rates of convergence. The general philosophy is that in ignorance of the exact data-generating functions, nonparametric nuisance estimators allow for consistency while parametric methods, with conventional but often unjustifiable restrictions on the relationships between variables, will not. The authors have previously proposed estimators with even stronger robustness properties [2, 13] that converge to a normally distributed random variable with known variance even under the nonconvergence of one of the two nuisance estimators. Note that the incorporation of slower-converging nonparametric methods for the PS in inverse probability of treatment weighting and for the OR in G-Computation has no theoretical guarantees and may perform poorly. Asymptotic linearity under nuisance estimators that converge at slower-than-parametric-rates is one of the major benefits of doubly robust approaches in general as it allows for more flexible nuisance modeling.

Weak identifiability may arise when certain patient covariates are strongly related to the treatment taken. If these covariates are not confounding, then they can be excluded from the estimation procedure and a TMLE will behave well. If they are confounders, then they must be adjusted for in some way, typically in both the OR and PS models. Practical problems arise when the covariates in the PS model allow for good discrimination between treatments. Previous work has shown that types of collaborative-TMLE (C-TMLE), that allow for either thresholding, the precision of the treatment predictions [7] or limiting the complexity of the PS model [1, 5, 8, 11] will outperform TMLE in such settings. As in the current proposal, C-TMLE was introduced as a way of stabilizing estimation in weakly identifiable, finite sample settings.

I have noted three limitations of previous implementations of C-TMLE in my experience. These are:

1. Slow computational times as many C-TMLE algorithms involve nested loops of model fitting and are thus computationally expensive (with an exception in the work by [6] though the authors noted some potential tradeoffs). In longitudinal settings, computational time may increase quickly in the number of time points (multiplicatively or exponentially depending on restrictions) [12].
2. Extrapolation in finite samples where the OR model is relied upon to model associations between the outcome and confounders where there is little to no data support for a given treatment. Extrapolation is essentially the goal of C-TMLE and is both a blessing and a curse. In the extreme case, C-TMLE can entirely avoid adjusting for confounding in the PS model, so that the resulting estimate fully relies on the model for the OR. This avoids the instability of

Mireille E. Schnitzer is Associate Professor of Biostatistics, Faculty of Pharmacy and School of Public Health, Université de Montréal, 2940, chemin de Polytechnique, Montréal, Québec, Canada, H3T 1J4 (e-mail: mireille.schnitzer@umontreal.ca).