

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).

large propensity score weights. However, if the OR is estimated nonparametrically or inconsistently, the OR model would not be expected to perform well on its own, so C-TMLE may produce finite-sample bias. Thus C-TMLE allows for bias-stability tradeoffs, though the extent of the tradeoff may be difficult or impossible to establish in an application.

3. Large-sample properties conditional on large-sample behavior of the collaborative tuning [14], A.18. Simulation studies investigating weak identifiability (near positivity violations) have revealed that the variance of C-TMLE may be underestimated when using methods based on the influence curve [7, 12].

In the current paper, the newly proposed C-TMLE estimator formally relies on the consistency of the OR estimator, and so is a departure from the typical double robustness. However, it still allows for a subparametric $n^{-1/4}$ rate of convergence of the OR estimate, which is comparable to the requirements of the nuisance estimators in the vanilla TMLE. They also require at least a $n^{-1/4}$ rate of convergence of their adaptive PS which is conditional on a single-dimensional covariate \bar{Q}_n^1 , and thus more easily achieved [3]. They can also incorporate cross-validation in the TMLE update step to avoid the Donsker condition on the nuisance estimators. Importantly, their algorithm is noniterative and does not depend on tuning parameters in a machine learning algorithm for the PS (as in [8, 10]) and so overcomes limitation #1. They explicitly require consistency of the OR estimate and so perhaps sidestep #2. The asymptotic results for this estimator and the proposed variance estimation resolve limitation #3. My only concern is that despite these theoretical results, the coverage of the confidence intervals (as displayed in Figure 4) decreases as sample size increases, which is not what would be expected.

3. THE GOAL OF WIDER ADOPTION IN APPLICATION

The authors' work on weak identifiability is important as this challenge is common in application, even with a well-defined causal question. I can see this as a useful method to add to an existing or future TMLE R package. The development of statistical packages and software is a massive and multidisciplinary undertaking, and so I applaud all previous and ongoing development by the Berkeley group and all those who share their algorithms and code. As both a methodology researcher in this area and a frequent user of existing R packages in application, I have noted that there may be particular challenges to wider usage beyond the usual technological inertia. More extensive uptake will require

- Software with model fitting diagnostics for users with various degrees of savvy. I am not of the opinion that these methods should be applied as a black box. For

example, the diagnosis of a weakly identifiable target parameter is relatively straightforward for users well-versed in estimators that involve inverse probability weighting. Large weights suggest potential problems, and could result in large finite sample bias in particular, though a specific size threshold for the weights is not a priori available [7]. A histogram of weights or the option to evaluate stability through bootstrap [1, 9] would give users access to important information. The user could then choose whether to use a C-TMLE rather than TMLE. Another example is that these types of estimators have many moving pieces (each nuisance estimator which may use multiple machine learners inside the ensemble learner, the TMLE update step, additional step(s) for the C-TMLE, nested cross-validation that can lead to data sparsity, depreciating numbers of observations over time in longitudinal problems) and any step may fail to numerically converge due to, for example, sparsity or extreme values. The user must be able to identify what failed, why, and would then choose how to proceed. Notably, the current proposal has far fewer steps than previous C-TMLE algorithms and so has fewer steps at which the algorithm may experience numerical problems.

- Software for diverse data structures and types that are common across various fields of application. There is already a vast range of TMLE methods that have been developed for a broad scope of settings, for example, case-control studies, two-stage sampling, longitudinal cohorts with fixed follow-up times, survival analysis, etc. Robust implementations of these methods, such as the correction to allow for doubly robust inference [2] and/or a C-TMLE implementation, would benefit applied research. The groundwork for such an implementation is currently being laid in the exciting TLverse project [4].

- Reasonably fast computational times and options for scalable algorithms. When an applied researcher is less concerned with or does not understand the nuances of nonparametric estimation, trading off an immediate result with hours of computation is a hard sell. Consider also that in a single study, models may be fit multiple times for various logistical reasons such as corrections to the dataset and multiple subgroup and secondary analyses.

- More translational papers, including papers directed at applied statisticians, explaining and demonstrating the utility and benefits of adaptive semiparametric estimators in various domains and statistical contexts.

To conclude, this is a strong proposal that avoids some of the most important drawbacks of previous C-TMLE implementations. Stable estimation in sparse settings leading to near practical positivity violations is important and the ongoing evaluation of such estimators in simulation and with real data is essential.

REFERENCES

- [1] BAHAMYIROU, A., BLAIS, L., FORGET, A. and SCHNITZER, M. E. (2019). Understanding and diagnosing the potential for bias when using machine learning methods with doubly robust causal estimators. *Stat. Methods Med. Res.* **28** 1637–1650. MR3961955 <https://doi.org/10.1177/0962280218772065>
- [2] BENKESER, D., CARONE, M., VAN DER LAAN, M. J. and GILBERT, P. B. (2017). Doubly robust nonparametric inference on the average treatment effect. *Biometrika* **104** 863–880. MR3737309 <https://doi.org/10.1093/biomet/asx053>
- [3] BENKESER, D. and VAN DER LAAN, M. (2016). The highly adaptive lasso estimator. In *Proceedings of the International Conference on Data Science and Advanced Analytics, IEEE International Conference on Data Science and Advanced Analytics*, 689–696.
- [4] COYLE, J., HEJAZI, N., MALENICA, I., PHILLIPS, R., HUBBARD, A. and VAN DER LAAN, M. J. The hitchhiker’s guide to the tverse or a targeted learning practitioner’s handbook.
- [5] GRUBER, S. and VAN DER LAAN, M. J. (2010). An application of collaborative targeted maximum likelihood estimation in causal inference and genomics. *Int. J. Biostat.* **6** Art. ID 18. MR2653847 <https://doi.org/10.2202/1557-4679.1182>
- [6] JU, C., GRUBER, S., LENDLE, S. D., CHAMBAZ, A., FRANKLIN, J. M., WYSS, R., SCHNEEWEISS, S. and VAN DER LAAN, M. J. (2019). Scalable collaborative targeted learning for high-dimensional data. *Stat. Methods Med. Res.* **28** 532–554. MR3903757 <https://doi.org/10.1177/0962280217729845>
- [7] JU, C., SCHWAB, J. and VAN DER LAAN, M. J. (2019). On adaptive propensity score truncation in causal inference. *Stat. Methods Med. Res.* **28** 1741–1760. MR3961963 <https://doi.org/10.1177/0962280218774817>
- [8] JU, C., WYSS, R., FRANKLIN, J. M., SCHNEEWEISS, S., HÄGGSTRÖM, J. and VAN DER LAAN, M. J. (2019). Collaborative-controlled LASSO for constructing propensity score-based estimators in high-dimensional data. *Stat. Methods Med. Res.* **28** 1044–1063. MR3934634 <https://doi.org/10.1177/0962280217744588>
- [9] PETERSEN, M. L., PORTER, K. E., GRUBER, S., WANG, Y. and VAN DER LAAN, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Stat. Methods Med. Res.* **21** 31–54. MR2867537 <https://doi.org/10.1177/0962280210386207>
- [10] SCHNITZER, M. E. and CEFALU, M. (2018). Collaborative targeted learning using regression shrinkage. *Stat. Med.* **37** 530–543. MR3748759 <https://doi.org/10.1002/sim.7527>
- [11] SCHNITZER, M. E., LOK, J. J. and GRUBER, S. (2016). Variable selection for confounder control, flexible modeling and collaborative targeted minimum loss-based estimation in causal inference. *Int. J. Biostat.* **12** 97–115. MR3505689 <https://doi.org/10.1515/ijb-2015-0017>
- [12] SCHNITZER, M. E., SANGO, J., FERREIRA GUERRA, S. and VAN DER LAAN, M. J. (2020). Data-adaptive longitudinal model selection in causal inference with collaborative targeted minimum loss-based estimation. *Biometrics*. To appear. Available at <https://onlinelibrary.wiley.com/doi/full/10.1111/biom.13135>.
- [13] VAN DER LAAN, M. J. (2014). Targeted estimation of nuisance parameters to obtain valid statistical inference. *Int. J. Biostat.* **10** 29–57. MR3208072 <https://doi.org/10.1515/ijb-2012-0038>
- [14] VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Series in Statistics. Springer, New York. MR2867111 <https://doi.org/10.1007/978-1-4419-9782-1>
- [15] ZHENG, W. and VAN DER LAAN, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning: Causal Inference for Observational and Experimental Data* (M. J. van der Laan and S. Rose, eds.). Springer Series in Statistics 459–474. Springer, New York. MR2867139 https://doi.org/10.1007/978-1-4419-9782-1_27