Proceedings of Machine Learning Research vol 291:1–2, 2025 38th Annual Conference on Learning Theory

Structure-agnostic Optimality of Doubly Robust Learning for Treatment Effect Estimation

Jikai Jin JKJIN@STANFORD.EDU
Stanford University
Vasilis Syrgkanis VSYRGK@STANFORD.EDU
Stanford University

Editors: Nika Haghtalab and Ankur Moitra

Abstract

Average treatment effect (ATE) estimation is one of the most central problems in causal inference, with applications across numerous disciplines. While many estimation strategies have been proposed in the literature, the statistical optimality of these methods remains an open area of investigation, especially in regimes where these methods do not achieve parametric rates. Given a covariate vector X, a binary treatment $D \in \{0, 1\}$, and an outcome $Y \in \mathbb{R}$ generated from

$$\begin{aligned} Y &= g_0(D, X) + U, \quad & \mathbb{E}[U \mid D, X] &= 0 \\ D &= m_0(X) + V, & \mathbb{E}[V \mid X] &= 0, \end{aligned} \tag{1}$$

we let $Y(1), Y(0)$ denote the potential outcomes and are interested in estimating the ATE $\theta^{\mathrm{ATE}} := \mathbb{E}[Y(1) - Y(0)]$. Under the standard conditional ignorability assumption $Y(1), Y(0) \perp D \mid X$, the ATE can be identified by $\theta^{\mathrm{ATE}} = \mathbb{E}[g_0(1, X) - g_0(0, X)]$.

Assuming access to nuisance estimates $\hat{g}, \hat{m}$ of g and m with mean-squared errors ϵ_g and ϵ_m, and n auxiliary observational data $\{(X_i, D_i, Y_i)\}_{i=1}^n$, the celebrated doubly robust learning (Robins et al., 1994; Chernozhukov et al., 2017) achieves $\mathcal{O}(\epsilon_g \epsilon_m + n^{-1/2})$ estimation error. While higher-order debiasing methods (Robins et al., 2008) can achieve better rates under certain structural assumptions such as Hölder smoothness, these assumptions can easily be violated or are hard to verify in practice, making these methods cumbersome to deploy. In this paper, we adopt the recently introduced structure-agnostic framework of statistical lower bounds (Balakrishnan et al., 2023), which imposes no structural properties on the nuisance functions other than access to black-box estimators that achieve some statistical estimation rate. This framework is particularly appealing when one is only willing to consider estimation strategies that use nonparametric regression and classification oracles as black-box subprocesses. We prove that within this framework, doubly robust learning is minimax optimal up to constants.

Specifically, given fixed nuisance estimates $(\hat{g}, \hat{m})$, we let $\mathcal{F}$ denote all possible ground-truth data distributions with nuisances (g, m) under the assumed mean-squared errors ϵ_g and ϵ_m, and define the minimax $(1-\gamma)$-quantile risk of estimating θ^{ate} over function space $\mathcal{F}$ as

$$\mathfrak{M}_{n,\gamma}^{\mathrm{ate}}(\mathcal{F}) = \inf_{\hat{\theta}:(\mathcal{X}\times\mathcal{D}\times\mathcal{Y})^n \mapsto \mathbb{R}} \sup_{P\in\mathcal{F}} \mathcal{Q}_{P,1-\gamma}\left(\left|\hat{\theta} - \theta^{\mathrm{ate}}\right|^2\right), \tag{2}$$

where $\mathcal{Q}_{P,\gamma}(X) = \inf\{x \in \mathbb{R} : P[X \leqslant x] \geqslant \gamma\}$ denotes the quantile function of random variable X. We show that under mild regularity conditions,

$$\mathfrak{M}_{n,\gamma}^{\mathrm{ate}}(\mathcal{F}) = \Omega\left(\epsilon_g \epsilon_m + n^{-1/2}\right). \tag{3}$$

We also establish the optimality of doubly robust learning for estimating the Average Treatment Effect on the Treated (ATT), as well as weighted variants of ATE that arise in policy evaluation (Athey and Wager, 2021). [1]

Keywords: Causal inference, semiparametric estimation, minimax lower bounds

1. Extended abstract. Full version appears as [arXiv 2402.14264, v3].

Acknowledgments

VS is supported by NSF Award IIS-2337916. JJ is partially supported by NSF Award IIS-2337916.

References

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1): 133–161, 2021.

Sivaraman Balakrishnan, Edward H Kennedy, and Larry Wasserman. The fundamental limits of structure-agnostic functional estimation. *arXiv preprint arXiv:2305.04116*, 2023.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5):261–265, 2017.

James Robins, Lingling Li, Eric Tchetgen, Aad van der Vaart, et al. Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and statistics: essays in honor of David A. Freedman*, volume 2, pages 335–422. Institute of Mathematical Statistics, 2008.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89 (427):846–866, 1994.