Liang, Wei; Yan, Ying
Empirical likelihood-based estimation and inference in randomized controlled trials with high-dimensional covariates. (English) Zbl 07493257

Summary: In this paper, we propose a data-adaptive empirical likelihood-based approach for treatment effect estimation and inference, which overcomes the obstacle of the traditional empirical likelihood-based approaches in the high-dimensional setting by adopting penalized regression and machine learning methods to model the covariate-outcome relationship. In particular, we show that our procedure successfully recovers the true variance of Zhang’s treatment effect estimator [B. Zhang, “Empirical likelihood inference in randomized clinical trials”, Stat. Methods Med. Res. 27, No. 12, 3770–3784 (2018; doi:10.1177/0962280217711205)] by utilizing a data-splitting technique. Our proposed estimator is proved to be asymptotically normal and semiparametric efficient under mild regularity conditions. Simulation studies indicate that our estimator is more efficient than the estimator proposed by S. Wager et al. [Proc. Natl. Acad. Sci. USA 113, No. 45, 12673–12678 (2016; Zbl 1407.62264)] when random forest is employed to model the covariate-outcome relationship. Moreover, when multiple machine learning models are imposed, our estimator is at least as efficient as any regular estimator with a single machine learning model. We compare our method to existing ones using the ACTG175 data and the GSE118657 data, and confirm the outstanding performance of our approach.

MSC:
62-XX Statistics

Keywords:
average treatment effect; datasplitting; machine learning; multiple robustness; semiparametric efficiency bound

Full Text: DOI arXiv