

IMPROVE EFFICIENCY OF DOUBLY ROBUST ESTIMATOR WHEN PROPENSITY SCORE IS MISSPECIFIED

Liangbo Lyu and Molei Liu*

University of Michigan and Peking University

Abstract: Doubly robust (DR) estimation is a crucial technique in causal inference and missing data problems. We propose a novel **P**ropensity score **A**ugmented **D**oubly robust (PAD) estimator to enhance the commonly used DR estimator for average treatment effect on the treated (ATT), or equivalently, the mean of the outcome under covariate shift. Our proposed estimator attains a lower asymptotic variance than the conventional DR estimator when the propensity score (PS) model is misspecified and the outcome regression (OR) model is correct while maintaining the double robustness property that it is valid when either the PS or OR model is correct. These are realized by introducing some properly calibrated adjustment covariates to linearly augment the PS model and solving a restricted weighted least square (RWLS) problem to minimize the variance of the augmented estimator. Both the asymptotic analysis and simulation studies demonstrate that PAD can significantly reduce the estimation variance compared to the standard DR estimator when the PS model is wrong and the OR is correct, and maintain close performance to DR when the PS model is correct. We further applied our method to study the effects of eligibility for 401(k) plan on the improvement of net total financial assets using data from the Survey of Income and Program Participation of 1991.

Key words and phrases: Causal inference, covariate shift correction, double robustness, intrinsic efficiency, outcome regression, propensity score.

1. Introduction

1.1. Background

Doubly robust (DR) estimation has attracted extensive interest in the literature on semiparametric theory and causal inference and is frequently used in biomedical science, economics, and policy science studies. It incorporates two nuisance models, a propensity score (PS) model, and an outcome regression (OR) model to characterize distributions of the exposure and outcome against the adjustment covariates respectively, and draws valid inferences when either one of them is correctly specified. It has been well-established that when both the PS and OR models are correct, the DR estimator is semiparametric efficient and its asymptotic variance does not really depend on the estimating equations

*Corresponding author. E-mail: moleiliu95@gmail.com