Statistical Science
2007, Vol. 22, No. 4, 540–543
DOI: 10.1214/07-STS227C
Main article DOI: 10.1214/07-STS227

Comment: Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data

Greg Ridgeway and Daniel F. McCaffrey

This article is an excellent introduction to doubly robust methods and we congratulate the authors for their thoroughness in bringing together the wide array of methods from different traditions that all share the property of being doubly robust.

Statisticians at RAND have been making extensive use of propensity score weighting in education (McCaffrey and Hamilton, 2007), policing and criminal justice (Ridgeway, 2006), drug treatment evaluation (Morral et al., 2006), and military workforce issues (Harrell, Lim, Castaneda and Golinelli, 2004). More recently, we have been adopting doubly robust (DR) methods in these applications believing that we could achieve further bias and variance reduction. Initially, this article made us second-guess our decision. The apparently strong performance of OLS and the authors' finding that no method outperformed OLS ran counter to our intuition and experience with propensity score weighting and DR estimators. We posited two potential explanations for this. First, we suspected that the high variance reported by the authors when using propensity score weights could result from their use of standard logistic regression. Second, stronger interaction effects in the outcome regression model might favor the DR approach.

1. METHODS

We felt the authors were somewhat narrow in their discussion of weighting by focusing only on propensity scores estimated by logistic regression in their simulation. The high variability in the weights reported by the authors could result from using this method. The authors state that none of the various IPW methods could overcome the problems with estimated propensity scores near 0 and 1, yet we believed that this is indicative of a problem with the propensity score estimator rather than IPW methods. In our experience weights estimated using a generalized boosted model (GBM) following the methods of McCaffrey, Ridgeway and Morral (2004) as implemented in the Toolkit for Weighting and Analysis of Nonequivalent Groups, the `twang` package for R, tend not to show the extreme behavior that resulted from logistic regression (Ridgeway, McCaffrey and Morral, 2006).

GBM is a general, automated, data-adaptive algorithm that can be used with a large number of covariates to fit a nonlinear surface and estimate propensity scores. GBM uses a linear combination of a large collection of piecewise constant basis functions to construct a regression model for dichotomous outcomes. Shrunken coefficients prevent the model from overfitting. The use of piecewise constants has the effect of keeping the estimated propensity scores relatively flat at the edges of the range of the predictors, yet it still produces well-calibrated probability estimates. This reduces the risk of the spurious predicted probabilities near 0 and 1 that cause problems for propensity score weighting. Many variants of boosting have appeared in machine learning and statistics literature and Hastie, Tibshirani and Friedman (2001) provide an overview. We optimized the number of terms in the GBM model to provide the best "balance" between the weighted covariate distributions $f(\mathbf{x}|t=1)$ and $f(\mathbf{x}|t=0)$. This approach to fitting propensity scores is fully implemented in the `twang` package.

We tested our conjectures about the performance of IPW and DR estimators based on GBM and in the presence of omitted interactions terms through a simulation experiment using the same design that the authors used.

Greg Ridgeway is Senior Statistician and Associate Director of the Safety and Justice Program at the RAND Corporation, Santa Monica, California 90407-2138, USA (e-mail: gregr@rand.org). Daniel F. McCaffrey is Senior Statistician and Head of the Statistics Group at the RAND Corporation, Pittsburgh, Pennsylvania 15213, USA (e-mail: danielm@rand.org).