Greedy Propensity Score Matching for Binary Interventions in Quasi-Experimental Designs

Rodrigo Barrera
Doctorado en Estaística, Departmento de Estadística, Universidad de Valparaíso

Alejandra Christen
Departmento de Estadística, Universidad de Valparaíso

Abstract

This investigation presents a formulation of the 1:1 greedy matching procedure based on propensity scores for binary interventions in quasi-experimental studies. The aim is to describe the mechanism of the algorithm, the assumptions that allow its results to be interpreted in causal terms, and its computational complexity. In the framework of potential outcomes, the estimands ATE, ATT, and $CATE$ are defined, together with the assumptions of consistency, SUTVA, conditional ignorability, and positivity, which link the matching design to the parameters of interest.

The greedy algorithm selects, for each treated unit, the unused control with the smallest propensity score distance, generating matchings without replacement and dependent on the order in which the treated units are traversed. The 1-to-1 bipartite matching system is shown to not satisfy the exchange property of matroids, so the greedy optimality theorem does not apply in this context.

The proposed implementation in Stata's matrix language (Mata) uses a matrix of absolute distances between propensity scores and a record of controls already used. The computational complexity turns out to be $O(n_T n_C)$, where n_T and n_C are the sizes of the treatment and control groups, respectively, allowing its use in applications with large sample sizes. Future research lines are discussed, including sensitivity studies with limited support, systematic comparisons with other matching and weighting schemes, and integration with doubly robust estimators and flexible models for estimating propensity scores.

References

1. Austin, P. C. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34(28), 3661–3679.

2. Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521), 390–400.

3. Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21.

4. Rosenbaum, P. R. (2020). *Modern Statistical Methods for Causal Inference*. Springer.

5. Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, 173(7), 761–767.

6. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242.