The optmatch package: flexible, optimal matching for observational studies

Ben B. Hansen

Matching is used both for design and analysis of case-control studies, of quasiexperiments with a treatment and a control group, and of cohort studies, among others. In conjunction with propensity scores (Rosenbaum and Rubin 1983), it offers an attractive alternative to adjustment based on regression, stably adjusting for the high-dimensional covariates that are common in social and medical science research. Greedy algorithms for pair matching are simple to implement; but much better matchings, matchings that support estimation of treatment effects that's both more efficient and less biased, can be had using a flexible, optimal matching routine (Hansen, 2004). Coding such a routine is somewhat subtle, involving discrete optimization among other steps (Hansen and Klopfer, 2006), so it is perhaps unsurprising that until recently none was publicly available. The R package "optmatch" was developed to fill this gap.

The presentation will briefly discuss the internals of the package before presenting an overview of its architecture from the perspective of the user, and discussing an application. Among the issues covered will be how it combines with propensity scoring and other multivariate distances; its adaptations for large, memory-intensive problems; its use to produce optimal pair matches, matches with multiple controls, full matches, and full matches with restrictions; analogues of goodness-of-fit tests for matchings, including some from a related package, "ritools", that J. Bowers and I are developing; and matched analysis for treatment effects.

Rosenbaum, P.R. and Rubin, D.B. (1983), The central role of the propensity score in observational studies for causal effects, Biometrika, 70, 41–55.

Hansen, B.B. (2004), Full matching in an observational study of coaching for the SAT, Journal of the American Statistical Association, 99, 609–618.

Hansen, B.B. and Klopfer, S.O. (2006), Optimal full matching and related designs via network flows, Journal of Computational and Graphical Statistics, to appear.