

Main Figures

Rachael Caelie (Rocky) Aikens

5/3/2019

Set Up

We compare the performance of propensity score matching, Mahalanobis distance matching, and pilot matching (described in the manuscript) on simulated data, varying the dimensionality of the problem, the fixed treatment to control ratio during matching, and the correlation between the true propensity and prognostic score. The generative model for all of our simulations is the following:

$$\begin{aligned}X_i &\sim_{iid} \text{Normal}(0, I_p), \\T_i &\sim_{iid} \text{Bernoulli}\left(\frac{1}{1 + \exp(-\phi(X_i))}\right), \\Y_i &= \tau T_i + \Psi(X_i) + \epsilon_i, \\ \epsilon_i &\sim_{iid} N(0, \sigma^2),\end{aligned}$$

where the true propensity and prognostic scores are given by the linear combinations

$$\begin{aligned}\phi(X_i) &= X_{i1}/3 - c, \\ \Psi(X_i) &= \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2},\end{aligned}$$

so that $\text{Cor}(\phi(X_i), \Psi(X_i)) \propto \rho$. The constant, c , in the propensity score formula was chosen such that there were approximately 100 treated observations in each dataset. We consider $p = 10$, $\rho = 0, 0.1, \dots, 0.9, 1.0$, and $k = 1, \dots, 10$. Each simulation consisted of a dataset of size $n = 2000$ and was repeated $N = 1000$ times. We fix the treatment effect to be constant with $\tau = 1$ and the noise to be $\sigma = 1$. For a given matching, we estimate ATT and design sensitivity $\tilde{\Gamma}$ using the permutation t -statistic from the package `sensitivtymv`.

Design sensitivity plots

Below, I tried to use our empirical simulation results (with the setup above) to make a version of the plots shown in figure 14.3 in Rosenbaum “Design of Observational Studies.” Long story short... you don’t see much. I think part of this is because our sample size is too small to get an idea of the asymptotic nature of the power of the sensitivity analysis, which you need in order to get a sense of where the design sensitivity is.

The issue is that at the end of the day, the sample size is determined primarily by the number of treated individuals in the sample. I.e. even though each simulated data set has 2,000 observations, there will only be as many matched pairs as treated observations (about 100). In order to keep all the simulation parameters stationary while getting a final sample size of, say 1000, I would have to simulate a sample size of 20,000. At that point, matching is going to start becoming really computationally burdensome.

