HTE simulations

Rachael Caelie (Rocky) Aikens

2/25/2020

Set Up With HTE

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:

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

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

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

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 $p = 10$, $\rho = 0, 0.1, \ldots, 0.9, 1.0$, and $k = 1, \ldots, 10$. Each simulation consisted of a dataset of size $n = 2000$ and was repeated $N = 1000$ times. In this set-up, the treatment effect, τ always has mean 1, but it is allowed to be heterogeneous between individuals. The noise in the outcome is fixed as $\sigma = 1$. For a given matching, we estimate ATT and design sensitivity $\tilde{\Gamma}$ using the permutation t-statistic from the package `sensitivtymv`.

Tau random, independent of covariates

In this batch of simulations, we let τ_i be normally distributed with mean 1 and standard deviation 0.25. The value of τ is thus heterogeneous, but totally unrelated to the values of the covariates.

Tau random, dependent on X3

In this batch of simulations, $\tau_i = 1 + 0.25X_{i3}$. This way, since X_{i3} is normally distributed with mean 0 and standard deviation 1, τ_i is still normally distributed with mean 1 and standard deviation 0.25.