

Main Figures

Rachael Caelie (Rocky) Aikens

5/3/2019

Set Up

We compare the performance of propensity score matching, Mahalanobis distance matching, and Buffalo Matching (described in the previous section) 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

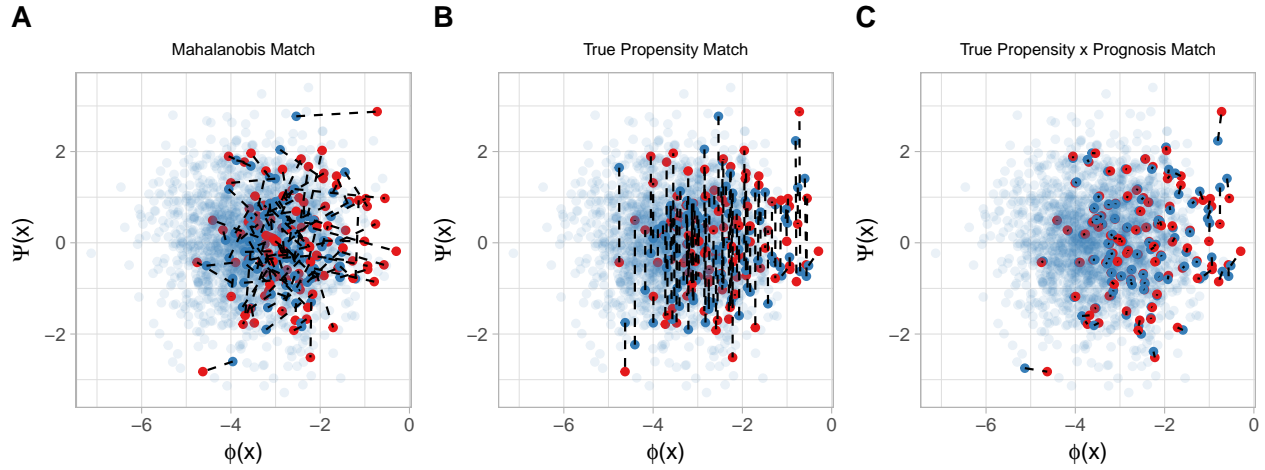
$$\begin{aligned}\phi(X_i) &= X_{i1} - 10/3, \\ \Psi(X_i) &= \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2},\end{aligned}$$

so that $\text{Cor}(\phi(X_i), \Psi(X_i)) \propto \rho$. The 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 $N = 2000$ times. We fix the treatment effect to be constant with $\tau = 1/2$ 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`

Motivation

The figure below gives a heuristic representation of what this algorithm is attempting to do. We imagine a scenario in which each individual in our data set, both treated and control, is represented in a reduced space of only two covariates: The variation determining the treatment assignment ($\phi(X_i)$), and the variation determining the outcome ($\Psi(X_i)$). These are the two features which are directly relevant to our matching: $\phi(X_i)$ balance (propensity balance) reduces bias, and $\Psi(X_i)$ balance (prognostic balance) reduces bias as well as variance and sensitivity to unobserved confounding.

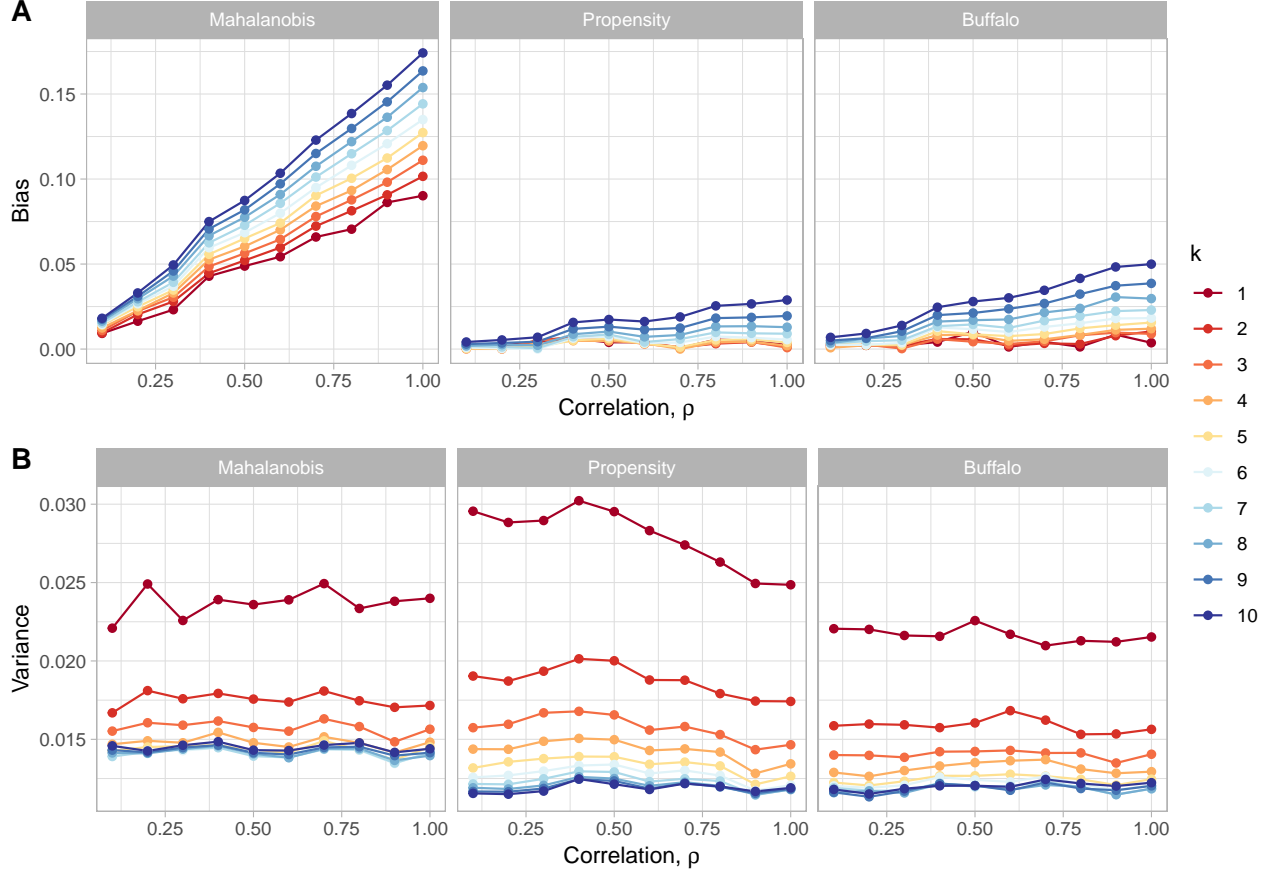
In our simulation, $\phi(X_i)$ and $\Psi(X_i)$ are known linear combinations of the covariates (see set up), so we can visualize them directly. For simplicity, we assume that the prognosis and treatment assignment are entirely uncorrelated ($\rho = 0$), although this need not always be the case (See supplementary figures 1 and 2). Optimal mahalanobis distance matching (Figure 1A), pairs individuals who are closest in the full covariate space. However, since only X_{i1} and X_{i2} are important for prognosis and treatment assignment, individuals who are close in the full covariate space may be very distant in the feature space of $\phi(X_i)$ and $\Psi(X_i)$. Propensity score matching (Figure 1B) pairs individuals who are close in the axis important for treatment assignment, $\phi(X_i)$, but not for prognosis, $\Psi(X_i)$. This matching will reduce bias compared to the unmatched dataset, but will lose the protection from variance and unobserved confounding conferred by prognostic balance. In contrast, if we match jointly on $\phi(X_i)$ and $\Psi(X_i)$, we obtain individuals who are close together in the feature space below. This optimizes for both desirable types of covariate balance: prognostic and propensity.



```
## pdf
## 2
```

Results

Below, we use the simulation parameters described in the set up to estimate the bias, variance, mse, and median gamma sensitivity of effect estimates produced from Mahalanobis distance matching, propensity score matching, and Buffalo matching.

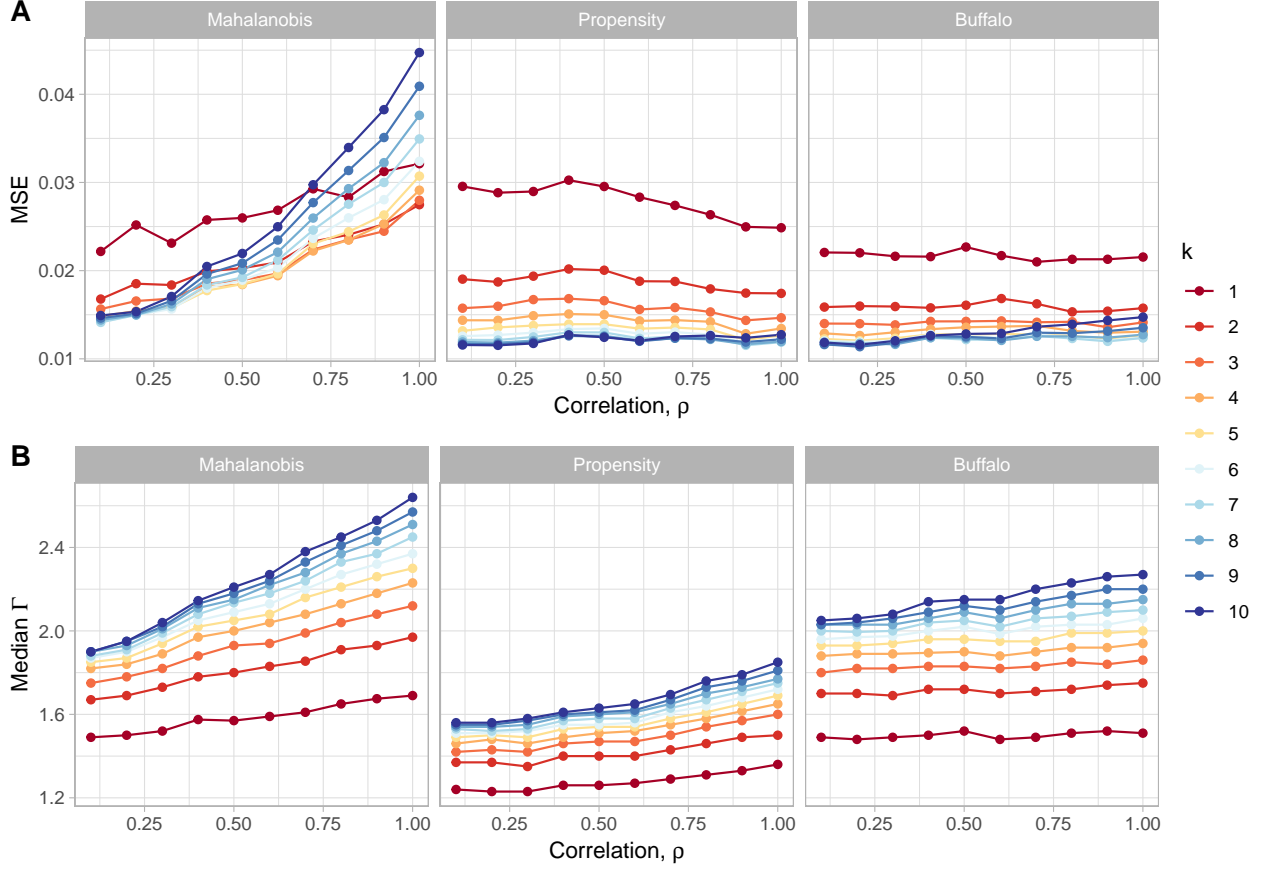


pdf
2

Figure 1 shows the bias and variance of 1 : k matching for each method on the same simulated data set as k increases (colored lines) and the correlation of propensity and prognosis, ρ increases from left to right. When $\rho = 0$, the treatment effect is completely unconfounded, since treatment is entirely determined (up to randomness) by variation in X_{i1} , and outcome (under the control assignment) is entirely determined by variation in X_{i2} . When $\rho = 1$, the data is highly confounded, since outcome and treatment assignment are both determined solely by variation in X_{i1} .

A first observation is that the bias from Mahalanobis distance matching is large, and it increases with the correlation between prognosis and treatment assignment (Figure 1A). This is probably because, as ρ approaches 1, only a single covariate, X_{i1} is important to match on, yet Mahalanobis distance considers 9 other covariates which are entirely random noise unimportant to the problem. It thus selects worse matches with respect to the true covariate of interest. This problem is exacerbated when the number of uninformative covariates is increased (Supplementary Figure 3).

Figure 1B shows the protective effect of Buffalo matching against the variance of the estimator. Propensity score matching has highest variance, and this is worse when ρ is close to zero. This is because, when ρ is small, the propensity score contains no information about prognosis under the control assignment. This causes the propensity score method to match observations which may be very different in terms of their potential outcome under the control assignment, giving poor prognostic balance and thus larger variance in the estimate of the causal effect. While Mahalanobis distance performs slightly better, it is again choosing lower quality matches because of the uninformative covariates in the data. The lowest variance is achieved by Buffalo matching, since this algorithm optimizes for prognostic balance as well as propensity score balance.



pdf
2

Figure 2 shows the mean squared error and median gamma sensitivity from the same set of simulations as above. Figure 2A demonstrates that Buffalo does well compared to the other two methods in terms of mean squared error, because of the protection against bias and variance shown in figure 1. Figure 2B shows the protective effect of prognostic balance against unobserved confounding. Propensity score matching gives lowest Γ values, especially when ρ is small. When this happens, the propensity score contains no prognostically relevant covariates, so the matches produced have poor prognostic balance. As ρ increases, the prognostic score and the propensity score become highly correlated, so matching on propensity score starts to give some prognostic balance as well as propensity balance. While the Gamma sensitivity of Mahalanobis distance matching seems promising when ρ is large, most of this is likely due to the large amount of bias in the effect estimate from this method (Figure 1A). Because Buffalo matching directly attempts to achieve prognostic balance in the matched sets, the Buffalo method has better protection against unobserved confounding, without giving a highly biased estimator.