

Model-based matching

Kellie Ottoboni

Draft January 30, 2016

Abstract

Drawing causal inferences from nonexperimental data is difficult due to the presence of confounders, variables that affect both the selection into treatment groups and the outcome. Matching methods can be used to subset the data to groups which are comparable with respect to important variables, but matching often fails to create sufficient balance between groups. Model-based matching is a nonparametric method for matching which groups observations that would be alike if none had received the treatment. We use model-based matching to conduct stratified permutation tests of association between the treatment and outcome, controlling for other variables. Under standard assumptions from the causal inference literature, model-based matching can be used to estimate average treatment effects.

1 Introduction

Observational studies in a range of fields including social sciences, epidemiology, and ecology are used to make inferences about cause and effect. Causal inference can be viewed as a missing data problem: one is only able to see each individual’s outcome after treatment or no treatment, but not both (Holland, 1984). To estimate the effect of treatment, one must use a control group as the counterfactual. The treatment effect is obscured by confounders, variables that are entangled with both the treatment and outcome. In an ideal situation, to adjust for the effect of confounders one would estimate the difference in outcomes between cases and controls who are identical with respect to all confounders, then average over the pairs.

In practice, the curse of dimensionality makes this impossible: studies typically account for a large number of covariates, so groups of individuals matched exactly on all pretreatment covariates can be too small to provide adequate statistical power to detect a significant treatment effect. For example, it is difficult to study the effect of a job-training program on income levels when the participants differ from non-participants on a wide range of socioeconomic variables, as well as potential unmeasured confounders (Dehejia & Wahba, 1999). Post hoc methods to construct a group of controls whose covariates balance the covariates in the treatment group include exact matching on covariates, matching and weighting by propensity score (Rosenbaum & Rubin, 1983), genetic matching (Diamond & Sekhon, 2013), and maximum-entropy weighting (Hainmueller, 2012). These balanced control groups are then used to estimate average treatment effects using standard methods such as unadjusted differences of means or linear regression controlling for other covariates. When matching and weighting methods fail to achieve balance in all the pretreatment covariates, estimates of the average treatment effect can be severely biased (Freedman & Berk, 2008).

The proposed method improves upon existing methods by eliminating the need for parametric assumptions such as Gaussian and homoscedastic errors, linearity, and adequate support. Observational studies often violate these key assumptions, so model-based matching may better accommodate real-world data from observational studies than traditional methods. Additionally, it is more flexible than traditional methods that assume that the treatment is binary and outcome is continuous; by modifying the statistic used to measure the strength of association, one may use categorical or continuous treatment and outcome variables.

The method has been applied before in a study of the effect of packstock use on the amphibian population in Yosemite National Park (Matchett, Stark, et. al 2015). In this paper, we develop the theory behind the testing method and discuss estimation strategies.

1.1 Notation

Let $Y_i(0)$ and $Y_i(1)$ be individual i 's potential outcomes under control and to treatment, respectively. T_i is the treatment assigned to individual i . Then the observed outcome is $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$.

A standard assumption is exogeneity, or conditional independence: $(Y(0), Y(1)) \perp\!\!\!\perp T \mid X$.

2 Matching

The great thing about randomized control trials is that the potential outcomes are balanced between treatment and control groups by construction: in other words, $(Y(0), Y(1)) \perp\!\!\!\perp T$. In observational studies, this is no longer guaranteed. Rosenbaum and Rubin (1983) achieve a weaker form of this balance by appealing to the propensity score, $p(x) = \mathbb{P}(T = 1 \mid X = x)$. The propensity score is a balancing score, in the sense that $X \perp\!\!\!\perp T \mid p(X)$. Under the assumptions of conditional independence given X and overlap in the distribution of propensity scores (together, called “strong ignorability”), they show that strong ignorability given X implies strong ignorability given $p(X)$, and that by the law of iterated expectations, we can recover the overall average treatment effect.

One issue is that in observational studies, the propensity score is unknown. One typically estimates the propensity score using logistic or probit regression models using a set of observed covariates. When the propensity score estimates are wrong, then estimates of the average treatment effect can be biased.

We’d like to try to achieve balance in the potential outcomes some other way: $(Y(0), Y(1)) \perp\!\!\!\perp T \mid \hat{Y}$, where \hat{Y} is the model-based prediction of Y , absent knowledge of the treatment, for all units.

3 Tests of Residuals

Tests of residuals after covariance adjustment appear in various forms in the literature. Rosenbaum (2002) uses residuals after fitting prediction models to stabilize estimates of treatment effects for more powerful randomization tests. Rosenbaum’s framework is limited to the case of binary

treatment, where individuals are either assigned to treatment or receive the control.

Shah and Bhlmann (2015) use residuals to test for the goodness of fit of high-dimensional linear models by testing for nonlinear signals in the residuals.

4 Theory

4.1 MM for Testing

4.2 MM for Estimation

4.2.1 MM Estimate when T is Binary

Recall that

$$ATE = \mathbb{E}(Y_1 - Y_0)$$

Let $\hat{Y} = f(X_1, \dots, X_p)$ be a prediction of the response using no information on the treatment.

$$\begin{aligned} \tau &= \mathbb{E}(\hat{Y} - Y_1 \mid T = 1) - \mathbb{E}(\hat{Y} - Y_0 \mid T = 0) \\ &= \mathbb{E} \left[\mathbb{E}(\hat{Y} \mid T = 1, X) - \mathbb{E}(Y_1 \mid T = 1, X) - \mathbb{E}(\hat{Y} \mid T = 0, X) + \mathbb{E}(Y_0 \mid T = 0, X) \right] \\ &= \mathbb{E} \left[\mathbb{E}(\hat{Y} \mid X) - \mathbb{E}(Y_1 \mid T = 1, X) - \mathbb{E}(\hat{Y} \mid X) + \mathbb{E}(Y_0 \mid T = 0, X) \right] \\ &\quad \text{by conditional independence, } \hat{Y} \perp\!\!\!\perp T \mid X \\ &= \mathbb{E} [-\mathbb{E}(Y_1 \mid T = 1, X) + \mathbb{E}(Y_0 \mid T = 0, X)] \end{aligned}$$

If selection on observables holds, i.e. you are so lucky to have accounted for all confounders X so that $(Y_0, Y_1) \perp\!\!\!\perp T \mid X$, then

$$\begin{aligned} \tau &= \mathbb{E} [-\mathbb{E}(Y_1 \mid T = 1, X) + \mathbb{E}(Y_0 \mid T = 0, X)] \\ &= \mathbb{E} [-\mathbb{E}(Y_1 \mid X) + \mathbb{E}(Y_0 \mid X)] \\ &= -ATE \end{aligned}$$

4.2.2 When T is discrete

We can generalize the previous result when there are more than two treatments.

4.2.3 When T is continuous

Suppose we estimate $\hat{Y} = f(X)$ and bin observations according to their predicted \hat{Y} . Then within bins, we can estimate $\mathbb{E}(Y \mid T)$. Suppose that $Y = f(X) + g(X, T) + \varepsilon(X)$, where $\varepsilon(X) \perp\!\!\!\perp T$, $\mathbb{E}(\varepsilon(X) \mid X) = 0$, and the $\varepsilon(X)$ are uncorrelated. If we assume this linear relationship, we obtain a generalized additive model:

$$\mathbb{E}(Y \mid T, X) = \mathbb{E}(f(X) + g(X, T) + \varepsilon(X) \mid T, X) = f(X) + \mathbb{E}(g(X, T) \mid X, T)$$

Within bins, all the \hat{Y} within bins should be close to each other simply because we've used it as the score by which to stratify. We can estimate $g_X(T)$ for each bin, too, using any method desired.

$$\mathbb{E}(Y \mid T) = \mathbb{E}_X [\mathbb{E}(Y \mid T, X)] = \mathbb{E}_X (f(X) + g(X, T) \mid T) = \mathbb{E}_X(\hat{Y}) + g_{\hat{Y}}(T)$$

where we define $g_{\hat{Y}}(T) = \mathbb{E}_X(g(X, T) \mid T)$, where the expectation is taken only over X with observations falling in the bin defined by \hat{Y} . The equation above implies that within small enough bins of \hat{Y} , the expected response is the mean predicted response plus the treatment effect, which is just a function of T and not of X .

Estimating $g_{\hat{Y}}(T)$ separately within bins allows for $g(X, T)$ to behave nonlinearly in the covariates.

Examples:

1. $f(X) = a + bX$ and the treatment effect is a multiple of τ according to the treatment dose, independent of X : then $Y = a + bX + \tau T + \varepsilon$
2. $f(X) = a + bX$ and the treatment effect is a multiple of τ according to the treatment dose interacted with X : then $Y = a + (b + \tau T)X + \varepsilon$
3. For $X < 5$, we are in case 1), and for $X \geq 5$ we are in case 2). Then for bins of \hat{Y} less than $a + 5b$, we estimate $g_{\hat{Y}}(T) = \tau T$ and for bins of \hat{Y} greater than $a + 5b$, we estimate $g_{\hat{Y}}(T) = \tau XT$.

5 Empirical results