Regression and Matching

Susanna Makela

March 1, 2017

My response this week is a (partial) attempt to connect the Shalizi reading to matching. The matching estimator for the ATE in Chapter 24 of Shalizi (2015)

$$ATE = \mathbb{E}[\mu(1, S) - \mu(0, S)],$$

where $\mu(x, s) = \mathbb{E}[Y | X = x, S = s]$, X is the causal variable, and S is a set of variables satisfying the back-door criterion. If we could observe μ, the law of large numbers would lead to the approximation

$$ATE \approx \frac{1}{n} \sum_{i=1}^{n} (\mu(1, s_i) - \mu(0, s_i)). \tag{1}$$

We of course can't actually observe μ. Let's assume $Y_i = \mu(x_i, s_i) + \epsilon_i$, where ϵ_i is mean-zero noise. Suppose we have $n = 2m$ observations and can exactly match each treated observation i (those observations for which $x_i = 1$) with an untreated observation i^* ($x_{i^*} = 0$) such that $s_i = s_{i^*}$. Then since

$$Y_i - Y_{i^*} = \mu(1, s_i) + \epsilon_i - \mu(0, s_i) - \epsilon_{i^*},$$

we can take an average over all m pairs

$$
\begin{aligned}
\frac{1}{m} \sum_{i=1}^{m} (Y_i - Y_{i^*}) & \\
&= \frac{1}{m} \sum_{i=1}^{m} (\mu(1, s_i) - \mu(0, s_i)) + \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \\
&\approx ATE,
\end{aligned}
\tag{2}
$$

since the noise should converge to 0 as $n \to \infty$.

This works fine if we are able to match exactly, but in practice exact matching is often infeasible. The set S may consist of many variables or continuous variables that have to be (arbitrarily) discretized. Forcing exact matches may leave us with only a small subset of our observations, making the above approximation less exact. As noted in Ho et al. (2007), inexact matching followed by the difference-in-means estimator (2) assumes that any remaining imbalance in the data is strictly unrelated to the treatment, which we know to be false (since we are making the assumption that S is the set of variables satisfying the back-door criterion). It also assumes that the imbalance has no effect on the outcome, which we have no evidence about before consulting the outcome, and if our matching is to be valid, we obviously cannot base it on the outcome variable.

I guess one way to connect this to Pearl's framework is that matching provides some assurance that our estimates may still be valid even if our causal graph is wrong. If the graph is wrong and there is some back door path that remains open (in other words, S is missing some variable), then simply conditioning on S will give us biased estimates. But if we are able to make the distributions of variables in S as similar as possible between treatment and control, and if any lurking confounder is highly correlated with S (a possibly strong assumption) then the fact that we are not able to observe and control for this lurking confounder will be less problematic.

Propensity score matching is an attractive option, in that it is the coarsest balancing score as described in Chapter 12 of Imbens and Rubin (2015). One issue is that it requires us to estimate the propensity score, since the true propensity scores are not found. However, if we can match on estimated propensity scores in such a way that the distributions of covariates are balanced between treatment and control groups, then we can be confident that our estimated propensity scores are consistent for the true propensity scores; this is referred to as the "propensity score tautology" by Ho et al. (2007).

The matching-followed-by-regression estimator is of the form

$$\widehat{ATE} = \frac{1}{m}\sum_{i=1}^{m}(\widehat{\mu}(1, s_i) - \widehat{\mu}(0, s_i)),$$

where $\widehat{\mu}(x, s_i)$ is a regression estimate of $\mathbb{E}[Y|X = x, S = s]$ of the form $g(\alpha + \beta X_i + \gamma S_i)$. In the case of linear regression, $g(\alpha + \beta X_i + \gamma S_i) = \alpha + \beta X_i + \gamma S_i$ and the estimator simplifies to

$$\widehat{ATE} = \widehat{\beta}.$$

One attractive property of matching followed by a parametric regression of the form $\mathbb{E}[Y|X = x, S = s] = g(\alpha + \beta X_i + \gamma S_i)$ is that this procedure is doubly robust: under weak conditions, "if either the matching or the parametric model is correct, but not necessarily both, causal estimates will still be consistent" (Ho et al. 2007).

References

Ho, D.E. et al. (2007). "Matching as Nonparameteric Preprocessing for Reducing Model Dependence in Parameteric Causal Inference". In: *Political Analysis* 15, pp. 199–236. DOI: 10.1093/pan/mpl013. URL: http://dx.doi.org/10.1093/pan/mpl013.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York, NY, USA: Cambridge University Press. ISBN: 0521885884, 9780521885881.

Shalizi, Cosma (2015). "Advanced Data Analysis from an Elementary Point of View". In: URL: http://stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf.