HW 1: Exploring Causal Inference in Experimental and Observational Studies

See the course syllabus for more instructions about working in teams. Students should turn in individual write-ups but may collaborate on code.

Getting Data and HW Overview

For this assignment, you should start with data from a randomized experiment. You are welcome to work with the dataset provided for the warm-up exercise, but several additional datasets are provided on Github. You may also bring your own dataset from another source. You can change datasets from assignment to assignment; you will be asked to complete this first assignment about average treatment effects, and then a second assignment about heterogeneous treatment effects, and it may be easiest to use the same dataset for both.

For the provided datasets, you should first take a quick look at the associated papers to understand the data. Also take a look at the key results. You will end up facing different types of issues depending on data set size, number of covariates, and strength of the basic treatment effect results. For example, in the paper on Charitable Giving, using ML methods to look for treatment effect heterogeneity in my experience often lead to spurious results, as the baseline results are not that strong. So if you pick this dataset, you may be emphasizing lack of findings or false findings rather than positive findings. On the other hand, some of the other datasets have richer or stronger results and heterogeneity, and/or more covariates. A few have some additional complications. (E.g. the paper on social voting has multiple observations per household; I suggested for simplicity analyzing only one household member so you don't have to worry about independence of observations.) Many have multiple outcomes and multiple treatments, so you'll want to pick one of each to focus on. It is fine to use linear models for binary outcomes for the purposes of this class, or you can also use logistic versions of procedures if you like.

You should not assume that because a dataset is included, it has a particular type of results associated with it. There aren't that many large-scale, publicly available randomized experiments out there. (If you find more that look interesting, please send me an email with a link or source, as I'd love to build up my collection. I intentionally didn't include the Lalonde data here because it is over-studied and the experiment is pretty small, but I don't object to using it if you have an interest in that area.)

Your assignment is to test out some different methods for estimating average treatment effects. To do so, the first part of the assignment is to turn the randomized experiment into an observational study. I'd like you to systematically delete some observations as a function of X's and treatment status (but not as a function of observed outcomes), and give a little thought to how you would like to do so in order to make things interesting. In general, X's that are correlated with both treatment assignment and outcomes create challenges for causal inference. This part of the exercise is a good warm-up to thinking about causal inference, as it helps you think about how different data-generating processes lead to observational datasets and associated biases in estimation.

In addition to the code in the warm-up exercise, you can also find sample code in run.all.R in http://github.com/swager/balanceHD/test -- this has code to run several different methods using simulated data --and also in https://github.com/swager/balanceHD/blob/master/R/baselines.R

Specific Assignment

For your assignment:

1. Before beginning, estimate the average treatment effect and the confidence interval in the randomized experiment.
2. Describe your method(s) for systematically deleting some observations as a function of X's and treatment status (but not as a function of observed outcomes), provide summary statistics, and show the old and new (simple average estimated) treatment effect.
    - Try to drop enough observations (and with an aggressive enough rule) that your new point estimate of the treatment effect is significantly different than the point estimate in the full randomized experiment. This is not a hard and fast rule, just a guideline.
    - Also make sure your rule is complex enough that it is not trivial to recover the propensity score; for example include some nonlinearities and multiple variables.
    - If all the methods below give the same, correct answer, try a more aggressive or complex dropping rule.
    - Be sure to preserve *overlap*, so that the estimated propensity score is not equal to 0 or 1 for any observations.
    - Plot the histogram of the bias function as in Athey, Imbens, Pham and Wager (2017) to summarize how challenging your problem is after dropping observations
3. In the modified dataset (note that most of the code required to do the exercises below can be written by combining bits of code from the warm-up R exercise, and see also the sample code in swager/balanceHD as mentioned above):
    - Test out the following traditional methods for estimating the ATE: (i) propensity score weighting; (ii) direct regression analysis; (iii) traditional double robust analysis weighting using inverse propensity score weighting; the lm command in R has a weights option.
    - Then try using lasso or regularized logistic regression (optionally try CART or random forest --classification trees, or classification forests), to estimate the propensity score and re-estimate the ATE using the methods above.
    - Next try using a single-equation lasso of Y on X and W to estimate the ATE.
        - Note that there is an option to not penalize the treatment indicator. You may wish to use that option anyway so that the treatment effect estimate is not shrunk. See http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html for the syntax for setting penalties for some coefficients to 0.
    - Next try using the Belloni-Chernozhukov-Hansen method, where you use the lasso to select variables with non-zero coefficients from the propensity equation and the outcome equation, take the union of the two sets, and finally run OLS. You can either use cross-validation to choose lambda in each case, or you can follow the approaches suggested by BCH (those are more complicated but probably doesn't make a difference).
    - Look at how your ATE coefficient changes with regularization.
        - Consider the single-equation LASSO of Y on X and W. Similar to the plot included in the handout, plot how the coefficient on the treatment indicator changes with lambda. Interpret your findings. (See

http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html for some sample code on plotting.)

- Optional: force the LASSO to include the variables from the propensity equation by manually setting the penalties for those variables equal to zero. Then repeat the previous exercise of plotting how the coefficient on the average treatment effect changes with lambda.
- Optional: also plot the coefficients on some other covariates as lambda changes.

4. Also in the modified dataset
    - Use double machine learning to estimate the ATE. Specifically, use a LASSO or random forest to estimate regressions of Y on X and separately Y on W. Then, run a residual on residual regression.
    - Use approximate residual balancing (package: http://github.com/swager/balanceHD) to estimate ATE. (Please email athey@stanford.edu if you find any bugs.)
    - Compare and interpret your results.

Your write-up should include code with output (if it works, generated by "knitting" as per the R instructions, but this sometimes fails, it is finnicky), as well as an electronic document (submitted individually) that discusses the results. Try to make your document self-contained by pasting in figures and referring to specific numbers/standard errors in the text where relevant. If you worked with group members on your code, indicate the group members on the assignment, but your write-ups should be done individually, and each member should submit the code/knit file.