Causality & Algorithms Virtual Reading Group

June 19, 2020

Why are we meeting?

- Goal: understand current work in causal inference and figure out interesting questions from a TCS perspective.

- Make concrete connections to property testing? non-asymptotic bounds/sample complexity? robust statistics? approximation algorithms? hardness? Also, (re-)defining things in more CS-friendly language.

Administrivia

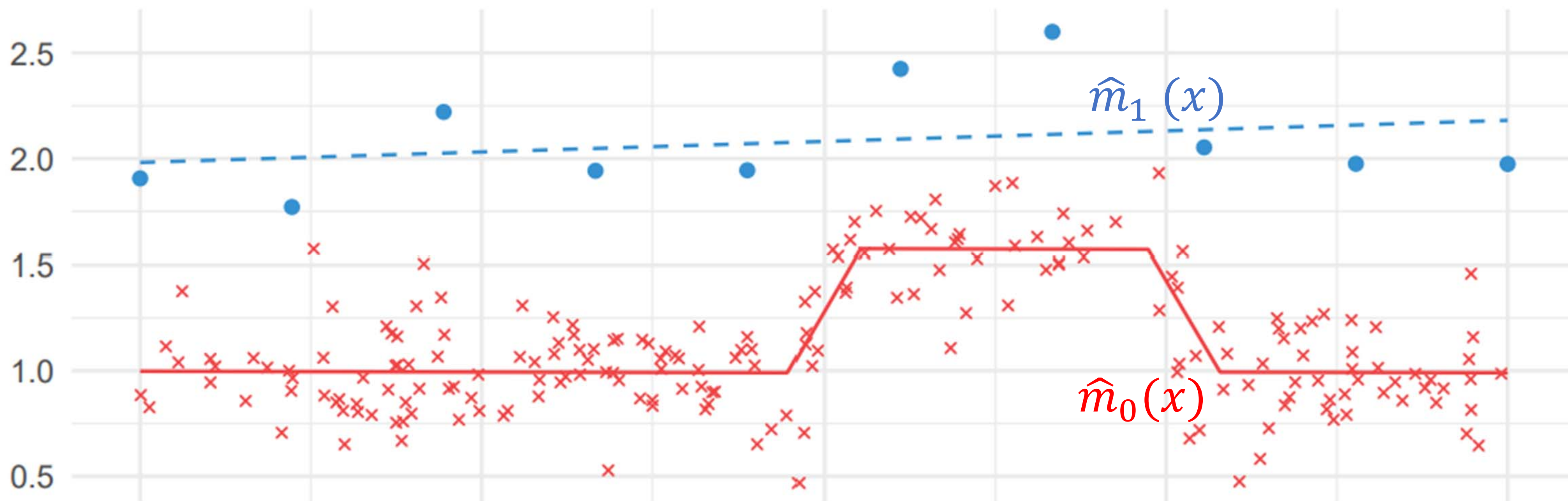
- Plan is to meet once every two weeks.
- Please volunteer! You don’t necessarily have to be an expert on the topic. The goal is to learn and discuss.
- I will post video recordings of the meetings.

Individual Treatment Effect Estimation & Causal Forests

Causality & Algorithms Virtual Reading Group

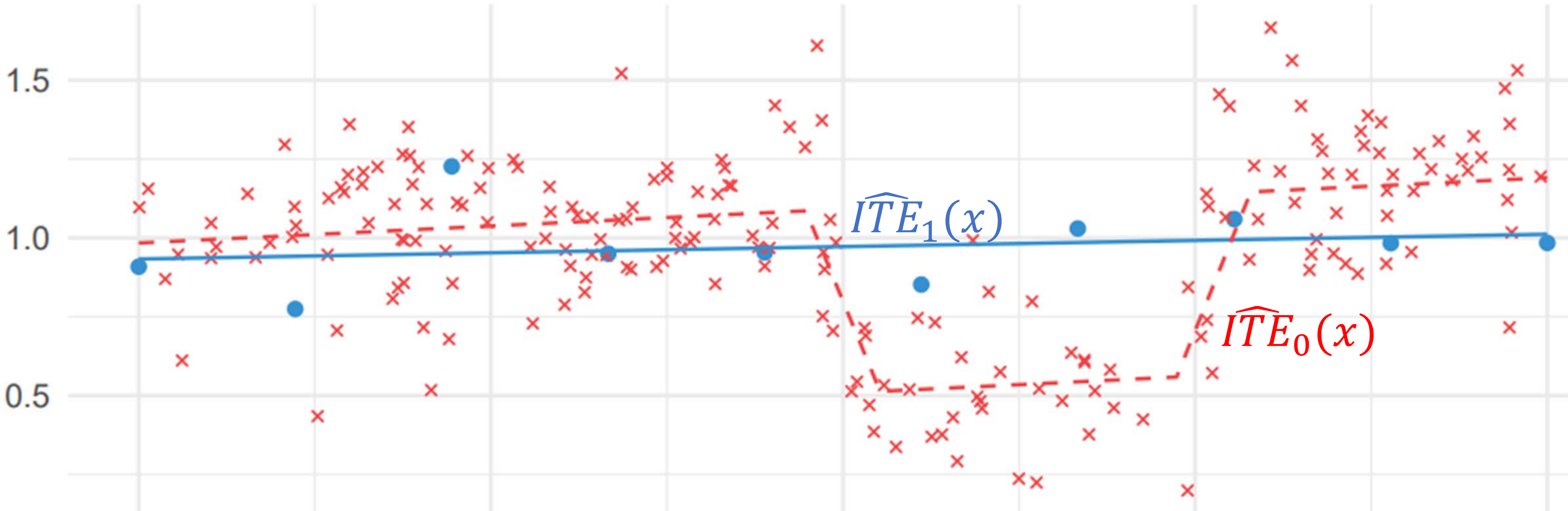
Arnab Bhattacharyya

June 19, 2020



Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning. Kunzel, Sekhon, Bickel, Yu. PNAS, 116 (10), pg. 4156—4165, 2019.

X-learner



Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning. Kunzel, Sekhon, Bickel, Yu. PNAS, 116 (10), pg. 4156—4165, 2019.

Procedure 1. DOUBLE-SAMPLE TREES

Double-sample trees split the available training data into two parts: one half for estimating the desired response inside each leaf, and another half for placing splits.

Input: n training examples of the form (X_i, Y_i) for regression trees or (X_i, Y_i, W_i) for causal trees, where X_i are features, Y_i is the response, and W_i is the treatment assignment. A minimum leaf size k.

1. Draw a random subsample of size s from $\{1, \ldots, n\}$ without replacement, and then divide it into two disjoint sets of size $|\mathcal{I}| = \lfloor s/2 \rfloor$ and $|\mathcal{J}| = \lceil s/2 \rceil$.
2. Grow a tree via recursive partitioning. The splits are chosen using any data from the $\mathcal{J}$ sample and X- or W-observations from the $\mathcal{I}$ sample, but without using Y-observations from the $\mathcal{I}$-sample.
3. Estimate leafwise responses using only the $\mathcal{I}$-sample observations.

Double-sample *regression* trees make predictions $\hat{\mu}(x)$ using (4) on the leaf containing x, only using the $\mathcal{I}$-sample observations. The splitting criteria is the standard for CART regression trees (minimizing mean-squared error of predictions). Splits are restricted so that each leaf of the tree must contain k or more $\mathcal{I}$-sample observations.

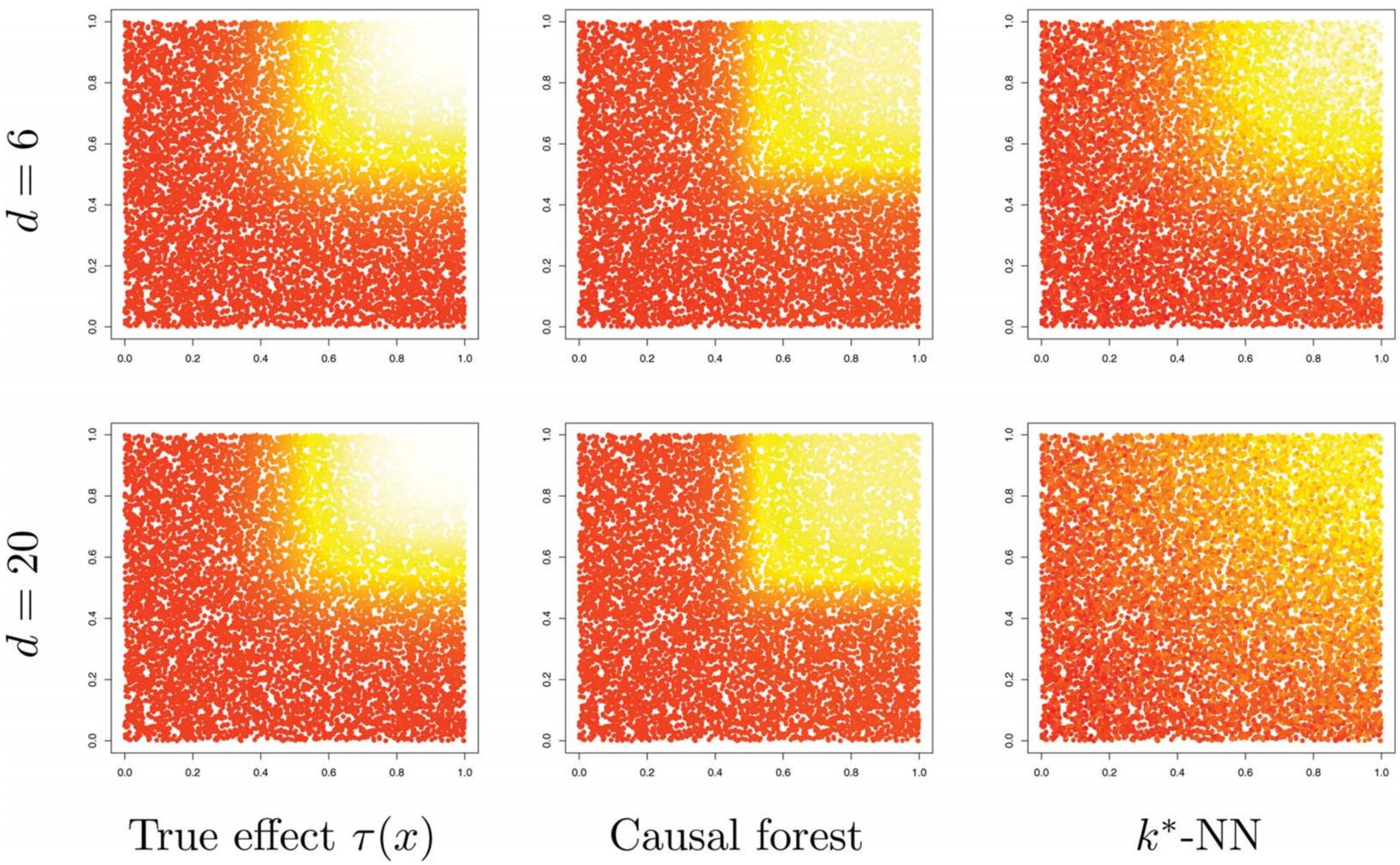
Double-sample *causal* trees are defined similarly, except that for prediction we estimate $\hat{\tau}(x)$ using (5) on the $\mathcal{I}$ sample. Following Athey and Imbens (2016), the splits of the tree are chosen by maximizing the variance of $\hat{\tau}(X_i)$ for $i \in \mathcal{J}$; see Remark 1 for details. In addition, each leaf of the tree must contain k or more $\mathcal{I}$-sample observations of *each* treatment class.

$$\widehat{ITE}(x) = \frac{1}{|\{i: T_i = 1, X_i \in L(x)\}|} \sum_{\substack{i \in I: T_i = 1, \\ X_i \in L(x)}} Y_i - \frac{1}{|\{i: T_i = 0, X_i \in L(x)\}|} \sum_{\substack{i \in I: T_i = 0, \\ X_i \in L(x)}} Y_i$$

Choose split so as to maximize:

$$\sum_{i \in J} \widehat{ITE}(X_i)^2$$

Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Wager & Athey. Journal of the American Statistical Association, 113:523, pg. 1228—1242, 2018.



Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Wager & Athey. Journal of the American Statistical Association, 113:523, pg. 1228—1242, 2018.